СВЕЩЕНИЯТ СМЕТАЧ
ТОДОР АРНАУДОВ – ТОШ

# ЛИСТОВЕ

## ПО ВСИЧКО

# ПРОРОЦИТЕ НА МИСЛЕЩИТЕ МАШИНИ
## ИЗКУСТВЕН РАЗУМ И РАЗВИТИЕ НА ЧОВЕКА
### ИСТОРИЯ ТЕОРИЯ И ПИОНЕРИ
### МИНАЛО НАСТОЯЩЕ И БЪДЕЩЕ

**от автора на първия в света**
**университетски курс по**
**Универсален изкуствен разум и**
**Теория на разума и вселената**

THE PROPHETS OF THE THINKING MACHINES
**ARTIFICIAL GENERAL INTELLIGENCE & TRANSHUMANISM**
HISTORY THEORY AND PIONEERS; PAST PRESENT AND FUTURE

## THE SACRED COMPUTER

**TODOR ARNAUDOV – TOSH**

# THE PROPHETS OF THE THINKING MACHINES

## ARTIFICIAL GENERAL INTELLIGENCE & TRANSHUMANISM

### HISTORY THEORY AND PIONEERS
### PAST PRESENT AND FUTURE

# LISTOVE

## REFLECTIONS ON EVERYTHING

by the author of the world's first university course in
Artificial General Intelligence and the
Theory of Universe and Mind

## ПРОРОЦИТЕ НА МИСЛЕЩИТЕ МАШИНИ

**ИЗКУСТВЕН РАЗУМ И РАЗВИТИЕ НА ЧОВЕКА**

ИСТОРИЯ ТЕОРИЯ И ПИОНЕРИ; МИНАЛО НАСТОЯЩЕ И БЪДЕЩЕ

© **Тодор Арнаудов** и всички цитирани автори и източници.
© **Todor Arnaudov** and all cited authors and sources.




**Последни корекции и добавки до:** 13.4.2026
**Редакция от:** 29.11.2025
**Дребни корекции:** 15.1.2026
**Публикуван на:** 5.11.2025
http://twenkid.com/agi
https://github.com/twenkid/sigi-2025
http://artificial-mind.blogspot.com
http://research.twenkid.com/

# ПРОРОЦИТЕ НА МИСЛЕЩИТЕ МАШИНИ

Изкуствен разум и развитие на човека:

История, теория и пионери

Минало настояще и бъдеще

Тодор Арнаудов – Тош*

## ПРИЛОЖЕНИЕ

# ЛИСТОВЕ

## по всичко

*предварително насочващо описание*

***** Отговор на възможни критики, че творбите в „Свещеният сметач“, „Изкуствен разум“ (Artificial Mind), „Разумир“, „Пророците на мислещите машини“ и пр. „не са научни публикации“, „нямат научна стойност“ и не се броят за приноси, защото не са публикувани на конференции, в „научни списания“ – по-точно в „признати“, „рецензирани“, „индексирани“ и пр. 48
* Обобщение на някои от съвпаденията:

* Теория на Разума и Вселената от 2001-2004 предвиди принципите и насоката на развитието на изкуствения интелект и универсалния изкуствен разум, и в интердисциплинарен контекст нейни основни заключения се преоткриват и потвърждават от школата на принципа на свободната енергия и извод чрез действие на Карл Фристън и др. (~2006-2009+), Анди Кларк; „осъществяването на уместност“ на Джон Фервеке и др. (2012-; 2024-); няколко теореми на Дейвид Уолпърт във връзка с предсказуемостта (2007;2018); теорията за УИР на Майкъл Бенет (~2023-2025); теорията за създаването на смисъл в универсалния изкуствен разум на Торисон и Талави (2024); множество интердисциплинарни теории обединяващи живото, неживото и ума на Майкъл Левин и др. (2020-те); подобни кибернетични идеи на Йоша Бах (~2019?+); идея на Йошуа Бенджио „consciousness prior” (2017-2018) в мета-обучението и ученето на причините; теорията на Ян Льокан за пътя към автономен ИИ (2022); „пет основни принципа на роботиката на развитието“ на Александър Стойчев (2006-); диалог между мислещата машина и човека от Томас Мецингер през 2009; мярката на Франсоа Шоле за машинна интелигентност (2019) и много други ~ 52

* Дейвид Уолпърт и няколко заключения от ТРИВ, преоткрити в негови по-късни публикации, но с объркани или неизпълними предпоставки
***** Осъществяване на уместност/Relevance Realization и преоткриването на принципи от ТРИВ – школата около John Vaerveke; накратко за преоткриването от Торисен и Талави и др.

***** Биофизика, квантови теории за съзнанието – на квантовата информация, на микротубулите; биология, психофизика, връзка и единство между материя и съзнание, „ ум-тяло“ (mind-body problem) – М.Левин, К.Фристън; Пенроуз-Хамероф, Д.Георгиев и мн.др.

***** Алгоритмична сложност и нейната философия, Теория на сглобяването, компресия, симулационна интелигентност – обзор, бележки, връзка с Теория на Разума и Вселената (ТРИВ); връзки с квантовите теории на съзнанието, изчислимостта, какво е изчисление, изчислимост, нетюрингови модели на изчислителни машини и пр.

***** Невронауки, невроморфни системи, съчетание на невронауки и машинно обучение и учене с подкрепление (Reinforcement Learning), непрекъснато обучение (Continual Learning); теория на системите и динамични системи и изчислителни невронауки – Михаил Рабинович и др. Български учени: Момчил Томов и отговор „от миналото и настоящето“ към неговата дисертация от 2019 г. от работата на Тош от ТРИВ от 2004 г. „Анализ на смисъла на изречение...“; Парашкев Начев, Ралица Димитрова и др.

***** Принцип на свободната енергия и извод чрез действие (Free Energy Principle and Active Inference) – Карл Фристън и др.

***** Класическа, съвременна и епигенетична роботика и мултиагентни системи, класическо и съвременно възстановяване на обема и динамично построяване на карти (SLAM), навигация, планиране.

***** Съзнание и панпсихизъм: откъси, тълкуване и бележки към множество автори; подобни или изпреварилите ги публикации от ТРИВ и с какво се различават от тях: Федерико Фаджин, Бернардо Каструп, Томас Мецингер, Томас Кембъл, Доналд Хофман, Тонони, Филип Гоф; Данко Георгиев; Люк Ройлофс (Luke Roelofs); феноменология и др. обзори.

***** Философия на живота и парадокси на живото и неживото, мета-биология, автопоеза, самоорганизация, теория на системите.

***** Философия на ума и на разума. Информационни и кибернетични теории

за развитието на вселената, разума и живота. Хипотези за симулираната вселена.
***** Философия на ума и на разума, изкуствения интелект и философия на изследванията по изкуствен интелект, мета-изкуствен интелект, кибернетична теория за Вселената, теории на всичко, Вселената Сметач (Вселената като компютър; цифрова вселена; Universe as Computer): pancomputationalism; Йоша Бах, Фристън, М.Левин и пр. Математическата Вселена на Тегмарк.

***** Схема за болката и философски анализ на Т.А. – въпросите за съзнанието, които възникват от някои нейни особености.

***** Големи езикови модели и машинно обучение: преглед на литература; изкуствени невронни мрежи. Други видове методи за машинно обучение, насочвано физични модели; други мрежи за обучение като на Колмогоров-Арнолд, „течни“ невронни мрежи и др.

***** Обзори на голям брой големи езикови модели и научните статии за тях; мултиагентни системи; обзори на обзори на основни модели с общо предназначение за роботи и за архитектури на агенти и бъдещи направления за развитие и разработка; езикови модели за действия.

***** Преобразители (трансформатори): механизъм на внимание; преглед на публикации за множество архитектури на преобразители с памет и други видове трансформатори като “Titans” и др.

***** Когнитивна лингвистика и когнитивна наука и мислене по аналогия.
– Съветски изкуствен интелект от 1960-те и 1970-те: Бонгард, Лосев, Максимов и др.; „пророчеството“ на Г.Херц от 1959. Проект „Животно“. Немонотонната логика на Александър Зиновиев и др.

***** Компании за УИР и преглед на събития свързани с тях., Преглед на поток от събития за публикувани невронни модели, продукти с ИИ, създаване на компании и др. Бележки за някои изследователски институти, свързани с разглежданата материя.

***** Проекти за символен универсален изкуствен разум
***** Дискусии и коментари в групи за изследователи по ИИ за паметта и др.; целенасоченият ИИ и този с учене с подкрепление са съотносими; връзка с ТРИВ
***** Преглед на голям брой беседи по ИИ с разнообразие от водещи учени и инженери, публикувани в Ютюб канала MLST*
***** Бележки и обобщения на най-важното, особено и пр. от всичко.
***** За отживели и неотговаряи на времето особености във високите степени

на академичното образование: срокове на докторантура, специализация, ограничено измисляне на ново и др.; разговор с Г.Чаитин.
***** За творчеството и „радикалната“ новост
***** (...) и др.

Виж и другите приложения, които продължават и разширяват темите тук, както Основния том, така и „Нужни ли са смъртни изчисления...“ и „Вселена и Разум 6“ – Вселената, съзнанието, философия на ума и изкуствения интелект, множество мета-... и фундаментални въпроси: ефективността и възможна ли е въобще или е заблуда на „счетоводството“ и др. сложни за обобщение. „Математически анализ на изкуството“ (Calculus of Art ..“) ... За образованието още: *„Първата съвременна стратегия за развитие чрез изкуствен интелект е публикувана от 18-годишен българин и...“ – Рейтингите на университетите са в порочен кръг, 2010* и цялостната книга, *„Институти и стратегии за изкуствен интелект на световно ниво (...)* и др.

За обзори на голям брой конкретни научни публикации, някои от 1950-те и 1960-те, до най-нови, виж Лазар и Анелия. Във втората е разгледана работата по изследвания, с които са свързани български изследователи в разнообразие от области, напр. компютърно зрение, машинно обучение, автономни превозни средства, изкуствен интелект в широк смисъл, невронауки и невроморфни системи, размита логика, компютърна лингвистика и обработка на естествен език, синтез и разпознаване на реч, общуване човек-компютър, програмни езици, анализ и верификация на програми, автоматично програмиране, синтез на програми (...) Множество беседи и техни тълкувания и някои пророчески дискусии виж в Ирина и др.

## Още бележки към съвременни, минали и перспективни учени, учения, теории, технологии, философия, дискусии, лекции, предавания, статии, предвиждания, съвпадения, области от ИИ, материали за подготовка; за остарялата система за докторантура; компании и дейци и др.

Тази книга е публикувана на, и е част от:

## Целогодишната виртуална конференция на *„Свещеният сметач“:*

*** Мислещи Машини 2025**
*** Самоусъвършенстващ се Изкуствен Разум 2025**
*** Self-Improving General Intelligence 2025 – SIGI-2025**

**#listove #Листове file: #Listove-MLST-i-dr;** #листове по всичко
**@Vsy:** извлч,пдрд,грпр;%прлжн; \|/прпртк; сдржн ...

Благодаря на семейството и приятелите ми.
Ако желаете, можете да помогнете на *Сметача* по всякакви начини: в разработката и изследванията, като съдружници, приятели, дарители на техника, спонсори, разпространители на знанието и др. Виж в основния том, в хранилището на Вседържец, приложението за Първата съвременна стратегия за развитие чрез изкуствен интелект ..., Институти и стратегии за изкуствен интелект и др. Виж също встъпителните бележки и уводната статия *„Отговор на възможни критики...“*

*** ВСЕДЪРЖЕЦ:**

https://github.com/Twenkid/Vsy-Jack-Of-All-Trades-AGI-Bulgarian-Internet-Archive-And-Search-Engine

**Some topics as keywords (part of the variety):**
**Някои теми (част от многообразието):** #ai, #agi, #neuroscience, neuropsychology, #neuromorphic, #complexity, #robotics, #multiagent, #panpsychism, #consciousness, #mind, #llm #vlm #ml; biophysics; philosophy, cognitive linguistics, analogy, free energy principle, active inference, Academia, PhD, physics-informed, machine learning, ML, Reinforcement Learning, RL and human learning and neuroscience, continual learning, important and recent papers in ML, GPT, BERT, DeepSeek,Veo, new transormer architectures and their modifications, Titans, universal AI, AGI, SLAM, Navigation, Planning, Cognitive maps, #ViT, #vision-language models; VLA models, vision-language-action models; multi-modal, multimodal; foundation models, physics-informed ML,

program synthesis,  agentic, computer use, frontier LLMs, foundation models, Soviet AI, Bongard, AI measures and benchmarks, AI ethics and safety, AI aligners, ... [there are too many relevant topics]

изкуствен интелект, ИИ, Общ ИИ, изкуствен общ интелект, универсален изкуствен разум, УИР, невронауки, невропсихология, невроморфни системи, психология, когнитивни науки, мислене по аналогия, философия, панпсихизъм, съзнание, ум, когнитивна лингвистика, алгоритмична сложност, роботика, планиране, мултиагентни системи, основни невронни модели за различни видове данни, основни модели за мултиагентни системи, мултимодалност и мултимодален ИИ, езикови модели, биофизика, немонотонна логика, разпознаване на образци/модели/шевици, Михаил Бонгард и съветската школа в ИИ и разпознаването на образи, принцип на свободната енергия, извод чрез действие, докторантура, висше образование, компании за ИИ, важни исторически и нови статии; нови архитектури трансформатори и техни модификации, перспективни математически модели за машинно обучение – съгласувано с физиката машинно обучение, универсален изкуствен разум, тестове за УИР, теория на вероятностите, икономика, взимане на решения, интердисциплинарност, всестранност, висше образование, докторантура, ...

See more in the partial contents below and in the book.

Виж повече в частичното съдържание по-долу и в цялата книга.

**Pre-intro in English:** This diverse book is part of the enormous interdisciplinary survey on Artificial General Intelligence and "general" artificial intelligence with various subfields of it (incomplete list) – machine learning, various forms of deep learning and novel approaches (physics-inspired etc.), reinforcement learning, classical, modern and epigenetic robotics; neuroscience, neuromorphic computation, algorithmic complexity and different forms of computation, natural language processing, large language models and their applications in all modalities, multimodal models, computer vision, computer-use models, benchmarks, agents; AI ethics, AI alignment and superintelligence; cognition, cognitive science, cognitive semiotics; theories of mind, consciousness; philosophy and generalizations of everything , theories of everything, cybernetics and systems theory; all kinds of related fields such as mathematics, psychology, economy – decision making; non-monotonic logic and historical visionary research, such as the Soviet AI from the 1960s and more recent rediscoveries and continuations and others in computer vision – SLAM, 3D-reconstruction, navigation, planning; issues in academia and PhD education system; (...) the connections between all fields, milestone historical papers and achievements; discussions from AI groups; the most recent research papers, AI models, companies, services, news, persons ... collected and selected up to the very last day before this book was completed (...) etc., with the guidance and additional contextual articles and comments on everything, written by the author-explorer and mapped to Theory of Universe and Mind.

That material is packed in one volume, an appendix of the encompassing hyperbook called *The Prophets of The Thinking Machines: Artificial General Intelligence and Transhumanism: History, Theory and Pionners; Past, Present and Future*.

One of the purposes of this body of work is to prove and explain the groundbreaking visions and contributions of *The Theory of Universe and Mind*, created in the early 2000s by the author, then a teenager. Many key ideas from the theory and the visionary publications have been rediscovered and repeated for the last 24 years many times in many diverse schools of thought, often by the most advanced interdisciplinary researchers, while the original predecessor is still largely unknown.

The title "Listove" means "Sheets of papers", because that was the media on which I wrote some of the first notes in this volume – sheets of printer paper; some of the notes were taken during walks. The seeds of the volume began in 2.2023 with exploration of material on Self-organization etc.

The content of this part of *The Prophets* may serve also as a dataset and a divergent seed for additional research, extensions etc. both by humans and thinking machines. The material is also explicitly intended to be studied and

continued by thinking machines, LLMs etc.

Future versions and editions will be dynamic and “living”, and will work with Vsy and ACS (Вседържец, Research Accelerator/Research Assistant) and various LLMs for (...), emboding the yet unpublished philosophy of the mentioned systems. The current textual and multimodal powerful LLMs with “deep research”, web search etc.; the AI assistants and browsers of Perplexity and OpenAI: Comet and Atlas implemented and embodied some ideas or developments, which were proposed by 2007 and early 2008 in “Smarty” etc. [and some earlier by earlier “prophets”] (...) See future work.

The text has sections both in English and Bulgarian - modern machine translation and LLMs understand most major languages.

Some of the articles might be published individually as well – for easier addressing, citation etc. (...)

## Речник на някои общи съкращения

**ТРИВ, ВиР** – Теория на разума и Вселената, Вселена и Разум. Обикновено препратка към класическите творби от нея от 2001-2004 г.
**CCU** – Causality-Control Unit; **POV** – Point of View
**RCCP** – Resolution of causality-control and perception; **RP** – of perception
Т, Т.А.,Тош, Т.А., Tosh (latin) – Тодор, Тодор Арнаудов, Todor Arnaudov
**UnM6** – Universe and Mind 6 (Вселена и Разум 6)
**Artificial Mind** – блогът Изкуствен разум, http://artificial-mind.blogspot.com
**TUM, TOUM** – Theory of Universe and Mind, mostly the classical period 2001-2004
**LLM, NN/ANN** – Large Language Model (Artificial) Neural Network ... **etc.**
Ц√ и др. – *нотация от Зрим*
Ц√ сложност √ разбиране
--‘ предвиждане, {В} – Вселена, [,,] обхват (...) – виж в бъдеще
МО, НМ – машинно обучение, невронни мрежи
За други знаци от Зрим – виж бдщ;
*„Сътворение: Създаване на мислещи машини“*

* **Езици**: Български и английски, на места и с превод на български.

## Томове и приложения на Пророците

*** #prophets –** Основен том (>1860 стр.) #tosh1
*** #purvata –** Първата съвременна стратегия за развитие чрез ИИ #tosh2
*** #stack –** Stack Theory is yet another Fork of Theory of Universe and Mind
*** #listove –** този том; още *„Листове по всичко“*
*** #mortal –** Нужни ли са смъртни изчислителни системи...? #karl
*** #universeandmind6 –** Вселена и Разум 6 #tosh3
*** #sf #cyber –** Научна фантастика за ИИ, Футурология, Кибернетика #sergey
*** #irina –** Ирина Риш, Йоша Бах и др. – интервюта и бележки; #joscha ...
*** #lazar #lotsofpapers –** множество разгледани работи, някои - български
*** #anelia –** друг том с работата с участието много български и други учени
*** #instituti**– преглед на институти по ИИ в Източна Европа и света
*** #complexity** – Алгоритмична сложност #hector
*** #calculusofart** – Математически анализ на изкуството. Музика I.
*** #kotkata –** Задачата от „Анализ на смисъла на изречение ...“, 2004 г. в диалог с чатботовете ChatGPT и Bard, края на 2023 г. до нач. на 2024 г., и с GPT5 и Claude 4 през 8.2025 г.

* **#zabluda –** Заблуждаващите понятия и разбор на истинския им смисъл: трансхуманизъм, цивилизация, ...
*** #appx** – приложение на приложенията, списък с добавени по-късно и др.
*** #power** – Power overrides intelligence – on AI Alignment
*** #llm ...** – Comparisons and reviews with LLMs
*** (...)** – следват продължения – виж някои възможни в уводите на Анелия и Лазар
* **И други, както и бъдещи –** виж по-подробен списък в края на книгата.

**Внимание!** Този съкратен списък може да е непълен, неточен и неактуален, защото *Пророците* се обновява и развива непрекъснато.

В следващи версии книгата ще бъде част или ще се представя и обработва от мислещите машини Вседържец, Research Assistant (ACS, Assistant C#), Емил и др.

Следете новините и обновленията от *Сметача!*

## Встъпителни бележки и препоръки към читателя

*** Някои статии** и откъси тук вероятно **ще бъдат публикувани и поотделно** за по-лесен достъп и цитиране**.**
* **В бъдещи издания** може да има пренареждане, систематизиране, преразпределение от това приложение в нови отделни томове по теми, отделно преглед на научни статии, кратки споменавания, бележки, коментари и статии на основния автор, вероятно самодвижещо се.
* **Форматирането** е „разнообразно“, но донякъде и нарочно е оставено така и вече няма за кога да се оправя. Текстът на *Пророците* трябваше да се редактира в самостоятелно приложение и база данни, чрез което да автоматизирам оформлението, пренареждането, форматирането и т.н., но не го направих и остава да се дострои само в бъдеще. Книгите ще бъдат под управлението на **Вседържец** в интерактивна форма, с удобно и гъвкаво търсене, превод, сбити представяния, разширяване и допълнителна информация по всяка статия и елемент от произведението, пораждане на различни съчетания и връзки между статии и понятия, типографски оформления, илюстрации; ще допълва и поправя сам съдържанието и пр. Това може да се свърши и в съчетание или от други чатботове и езикови модели.
*** Някои читатели**, особено отрицателно настроените, които нямат необходимата предварителна познавателна основа и любознателност по разглежданите теми, биха определили този том като „хаотичен“, „безреден“, „разхвърлян“, „многословен“; „мисълта прескачала“[1] (виж „хипертекст“ ; в Зрим пт(мс), пс(мс)). Всъщност много неща са *сбити*, а не многословни, защото „пълнословието“ е когато се *прочетат изцяло* всички споменати творби, бележките обикновено подчертават определени мисли, които съм сметнал, че са по-особени, съдържателни или си заслужават да се отбележат в контекста на останалото, ако е познато и разбираемо, така че да се спести или ускори възприемането на цялото. По-неопитен читател, за когото дадена тема е нова или недостатъчно ясна, може да има нужда да изслуша или прочете подробно и да осмисли *целите* цитирани беседи, статии и т.н., преди да разбере кратките бележки към тях.

За да забележите по-лесно идеите и съвпаденията на новите теории и публикации с класическите *„български пророчества“* в Теория на Разума и Вселената, е нужно да познавате поне част от основните творби, публикувани на български език от 2001-2004 г. в сп. „Свещеният

[1] За някой „прескачането“ е интересно, особено когато се разбира „неочаквано“, изненадващо – научава нова връзка, разширение, нова възможност; скок и в знанията.

сметач“ и др. и разяснявани и разширявани по-късно и до днес в блога „Изкуствен Разум“, „Разумир“, Гитхъб и публикациите от SIGI-2025 и др. В сбита форма част от тях са в слайдовете на лекция от първия в света курс по Универсален изкуствен разум (Artificial General Intelligence), който е препоръчителен, а цялото множество творби от *„Пророците…“* е вид продължение на курса – набор от записки по Универсален изкуствен разум, и може да се нарече и така: **Курс по УИР, версия 2025 г.**

Някои от творбите от класиките, които може да са по-достъпни и интересни за по-широк кръг от читатели, са например фантастичната повест за мислещата машина „*Истината*“, 2002 г. която може да откриете и в „Моята библиотека“ („Читанка“). Други са:

* „Човекът и Мислещата Машина: Анализ на възможността да се създаде мислеща машина и някои недостатъци на човека и органичната материя пред нея“, 2001;
* „Подкастът“ от 2002 г.: *„Писма между 18-годишния Тодор Арнаудов и Ангел Грънчаров“*, именуван с още много заглавия като „Вселената Сметач“, „Следващото еволюционно стъпало 2“, „Сметачолюбецът и Човеколюбецът“, „Вършачите и човеците“[2] и др.
* Творчеството е подражание на ниво алгоритми, 2003
* „Схващане за всеобщата предопределеност 3“ (Вселена и Разум 3) и „Вселена и Разум 4“.

По-кратката книга „*Първата съвременна стратегия за развитие чрез изкуствен интелект е публикувана от 18-годишен българин и повторена и изпълнена от целия свят 15-20 години по-късно: Българските пророчества: Как бих инвестирал един милион с най-голяма полза за развитието на страната?“*[3] също съдържа фрагменти, които накратко в хронологична форма представят някои основни идеи от ТРИВ и изкуствения интелект, и цитират откъси от творби и теории.

От по-новите публикации, **„Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини“** и **„Вселена и Разум 6**“, както и **„Stack Theory is yet another Fork of Theory of Universe and Mind”**[4] може да са полезни във връзка с темите за съзнанието и панпсихизма, като те припомнят и доразвиват конкретни важни откъси и заключения от класически работи във връзка с невроморфните изчислителни системи и тяхната предполагаема по-висока ефективност и др. Същото се прави и в този том, докато се

[2] Вършач – машина; процесор – виж юнашкото наречие на Дружеството за защита на българския език
[3] https://twenkid.com/agi/Purvata_Strategiya_UIR_AGI_2003_Arnaudov_SIGI-2025_31-3-2025.pdf
[4] Теорията на Майкъл Т. Бенет за УИР.

коментират съответни раздели, както и в *Основния том* и в *Ирина* – някои от беседите и интервютата в последното приложение може да са достъпни за по-широк кръг от читатели отколкото други, по-технически или абстрактни приложения и пасажи.

**Основният том** е най-обемен и съдържа най-много обща подготвителна информация във връзка с Теория на разума и Вселената, най-голям обем от оригиналните текстове, както и същностна специализирана информация по много от темите, личностите и научните области, свързани с ТРИВ, които са допълнени с материали, тълкувания или странични разклонения и в останалите приложения. Част от подготвителната информация е разделът с обзор на множество „учени и школи“, които са сродни, свързани, подобни или преоткриват или разработват идеите, публикувани още в ТРИВ между 2001-2004 г. Материали по същата тема в съответен контекст са разпръснати в много от томовете в цялото съзвездие на свръхкнигата.

Приложенията „**Анелия**“ и „**Лазар**“ представят голям библиографски каталог с подбор, извадки, бележки по научни статии и понятия и ще са интересни на онези, които обичат да разглеждат голям брой научни статии, като се следва логически тематичен, личностен и хронологичен ред – например избрани работи с участието на Анелия Ангелова, с които започва книгата, в областта на машинното обучение и зрение, самоуправляващите се превозни средства и др., с бележки за най-важни понятия и кратки цитати от тях. По-подробно е разгледана дисертацията по компютърно зрение на Александър Тошев и др. негови творби. Прегледани са важни разработки на трима по-известни в света, отколкото в България, български учени Димитър Филев, Пламен Ангелов и Никола Касабов и мн. др.

В този том също има прегледи и на множество научни статии в подобен дух. В по-малък мащаб и с по-тесен фокус подобни прегледи и обобщения са дадени в по-краткото приложение **„Алгоритмична сложност“**[5].

**Относно сложността на съдържанието на Листове:** някои от бележките по темите, свързани с машинно обучение и изкуствен интелект може да са разбираеми и с по-общи познания по информатика или ИИ, защото се обясняват основни неща като какво представлява „вниманието“ в преобразителите; други сбити бележки или изучаването на съответните цитирани статии в списъка за четене изискват да имате познания в областта, или пък желание да получите. Това вероятно се отнася и за статиите и бележките по невронауки, психология, роботика и

[5] Част от съдържанието му е включено и тук.

пр. – в този том са включени статии и кратък речник с основни понятия от роботиката. В невронауките може да започнете например с лекцията от курса по УИР за Архитектура на мозъка на бозайниците. Някои кратки бележки използват записа на „Зрим“ със съкращения и не са обяснени напълно, напр. за Принципа на свободната енергия по беседа с Карл Фристън. Ако искате ги възприемайте като „Великденски зайци“. Ако не може да отгатнете въз основа на ТРИВ, съдържанието на коментирания материал и изяснените тълкувания на част от нотацията, изчакайте продълженията. По тази тема има още обяснения в този том, в ***Ирина*** и особено в ***Основния***.

Може да прескачате статии, записки и глави на теми, които не са ви интересни в момента, докато стигнете до онова, което ви допада или е достъпно за вас и да ги преразглеждате по-късно. Например ако не ви интересува изчислимостта, сложността, машини на Тюринг, то може да отскочите до статиите за съзнание и панпсихизъм, роботика, мулти-агентни системи, невронауки, психология и когнитивна наука, откъсите по Практика и списъците с услуги с ИИ, където може да откриете по-малко известни, но също мощни и безплатни като китайските „Кими“, „Куен“, „Дийпсийк“. Много или почти всички теми обаче са свързани по един или друг начин; по-„отделени“ са най-технически подробности, например архитектурата на определени невронни модели и пр.

Разделите, свързани със съзнание, панпсихизъм, теории на всичко, са свързани и с разгледаните в началото теми за школите подобни на ТРИВ като „Осъществяване а „Вселена и Разум 6“ и „Нужни ли са смъртни изчисления…“ и ги допълват.

Накратко, приложение *„**Листове**“*[6] на „**Пророците**…“ може да се чете с преглеждане и търсене на интересни за вас теми от многообразието, но *най-интересно* ще е за онези, които обичат ударно да откриват и да учат нови понятия и нови научни области по *всичко,* които сами *търсят непознатото,* за да го овладеят и да продължат още по-напред.

Както се пееше в една детска песен от не знам си кой клас:

[6] Защо „Листове“? Защото бележките, от които започваше „растежът“ на тази част, бяха писани на ръка на „хвръчащи“ листове. Защо? Защото пишех и движение, докато се разхождах и слушах някои беседи – това се отнася и за „Ирина“; по-удобни и просто е за сгъване или носене, и с мисъл за бъдещо по-безпроблемно сканиране. За умствена ориентация обаче тетрадките са по-подходящи, или листовете да се сглобят в книги.

*„Аз обичам да чета, аз обичам да мечтая.*
*Имам  хубава мечта.*
*{ Всичко искам аз да зная. (2)*
*Всичко на света да зная. } (2)*[7]

Към знанието обаче се включва и *разбирането и създаването.*

Съдържанието и философията на този том също биха допаднали на онези, които като мен искат компютрите, мислещите машини да правят живота им ***по-интересен,*** а не „по-лесен“.

И този том, както и някои от другите, или всички, са своеобразен **наръчник, набор от данни, основа** за мислещите машини, изкуствения интелект, всички науки за ума и свързани с тях науки, математика, философия; дейности, събития, история, личности, организация, … – от една страна “*Пророците на мислещите машини…“* съдържа силно обработено, сбито и подбрано знание, а от друга страна в него има и недообработен или „недоподреден“  материал, предварителен насочващ суров материал, записи, от които в последствие  да се извлича допълнително познание, да се търсят връзки между съставните му части и цялото съдържаните творби и обобщенията от тях, да се проследява във времето развитието на посочените теми по избрани нишки, да се чертаят връзки със съдържанието от другите томове на *Пророците;* да се търси, пренарежда, систематизира с допълнителни техники; да се използва тази основа като зърно, зародиш, клетка, корен, от който да се пуснат разклонения. Част от съдържанието е като образец, пример, „sample”, котва, по които да се извлече останалото.

Една от функциите на „раздробеността“, „накъсването“, е да се уточни, адресира, раздели; да се отделят части (диференциране, отчленяване).

Някои от **целевите читатели** са **машини,** в частност различните версии на **Емил** и **Вседържец**. На места не е правен опит да е подредено, а се отразяват и съхраняват определени моменти от събитията или протичането на проучвания като примери и част от „летописа“[8].

---

[7] Текст: Алберт Декало. https://www.youtube.com/watch?v=6RsX77G6G_M – Винаги съм помнел първия куплет, защото се отнасяше за мен още преди да бъда ученик.

[8] Да, такива изследвания *вече* могат да се правят и с изследователските режими на чатботовете и вероятно единствените читатели, които ще прочетат този и повечето томове *изцяло*, внимателно и с препратките ще са машини. Това беше проект на Сметача още от 2010-те и в Зрим и Research Accelerator се нарича *Изпреварващо търсене* в {К}?Т.

**Превключването** на темите обаче е и част от общата **всестранна и интердисциплинарна програма**, която е в сърцето както на личния ми начин за изучаване на света от най-ранна възраст, така и на универсалността на *Свещеният Сметача* и на *Първата стратегия за ИИ* от 2003 г.

При подготовката на разнообразни **набори от данни** за обучението на най-мощните големи езикови модели и мултимодални пораждащи модели също се събират „разнообразни сигнали“, а преди тях – в *корпусната лингвистика* – текстови от разнообразни източници и жанрове. Така се научават и се откриват връзки и закономерности от всичко към всичко и умът е по-подготвен да превключва контекста и не се бои от новото.

**Свързаното всестранно знание** е като огромен дворец с много входове, много стаи и врати и коридори между тях. Или гора с много пътеки. Може да влезете от много места, да се разхождате по различни пътеки; из пространството има и малки стаички и големи просторни зали, и широки поляни и гъсталаци от храсти и плътни гори. Където и да попаднете обаче, постепенно  може да се придвижите до която друга точка си искате и постепенно да си нарисувате карта или да утъпкате пътеките и слеващия път ще ви е по-лесно и ще стигнете до нови места по-бързо .

**Можете да се присъедините и да подкрепите** *Свещеният сметач*, включително като автори или съавтори в бъдещи томове на *Пророците* или други статии на виртуалната конференция. Виж в началото на основния том и в проекта „Вседържец“ в Гитхъб за някои конкретни начини за сътрудничество и помощ чрез техника, дарения, участие в проекти, разгласяване, свръзване с подходящи хора, морална подкрепа и др.

**Технически бележки по типографията:** Има „колебания“ в размера и вида на шрифтовете; размерът не отразява задължително по-голяма важност на дадена статия спрямо друга: всичко е важно според онова, което ви интересува. Ако всичко ви интересува – всичко е важно, и цялото разбиране може да изисква многократен преглед, за да се сглоби картината в ума ви.

**Променливостта на шрифтовете** тук и в други приложения донякъде е експеримент: от една страна, при разпечатване да се печатат по две страници на А4, а тогава не бива текстът да е прекалено дребен; в същото време, ако шрифтът е прекалено голям, ще раздуе броя на страници. От друга страна вероятно документът ще се разглежда повече на екран, където лесно може да се увеличава. Серифните шрифтове като

Liberation Serif, Times New Roman се четат по-добре на хартия, отколкото на екран, освен ако не е с висока плътност точки-на-инч. Несерифните шрифтове като Arial и Liberation Sans запълват по-добре пространството на знаците, с по-едри букви, и се виждат по-добре на екрани и при малки размери, но при тях латинското малко „l" се изписва като главно „I" – без „ченгелче", което не ми харесва и затова понякога го избягвам. От по-модерните несерифни шрифтове Aptos и Calibri, първият има различимо "l" и по-едри букви при еднакъв размер (този абзац е с Аптос и превключване към другите за илюстрация).

Използвах „редова разредка", междуредие, line-spacing от 1.15, някъде може да е и 1.0, където съм пропуснал да наглася. Както споменах по-горе, по начало беше смислено да редактирам „Пророците…" в собствен редактор и формат, който лесно да пренарежда и генерира документите, но не се захванах да го направя засега.

* **Цитирам препратките с цели или съкратени имена и с хипервръзки**, а не само със списък с номера [1][123] и др., от една страна защото засега така и не ги събрах в такъв конкретен списък – може би трябваше да сглобя общ за всички томове или да е в собствена система за редактиране, с която да се улесни, или с „Латекст" и пр. формати и редактори за подобни текстове и т.н. От друга страна това решение е и целенасочено, защото по този начин отделните късове от цялото могат да се разглеждат в по-голяма степен самостоятелно, където и да се погледне, и фрагментите могат по-лесно да се отделят в подтомове и статии, нещо което вероятно ще се случи.

* **Нарочно използвам звезда „*", а не „булет"** , защото е по-естествен знак от ANSI/ASCII за търсене и текстообработка.

* Относно размера на шрифтовете – сравни също с лошата традиция и практика в научни статии да се използва *прекалено дребен шрифт* в две колони или да се оставят големи бели полета. Във формата на дисертации пък често се оставя междуредие 1.5 или двойно, за да се вмъкват корекции и бележки. [31.7.2025]

*** Български учени, споменати в този том:** Данко Георгиев: квантова физика и биофизика, съзнание; творчество (с Г.Георгиев) около с.250. Мислене по аналогия, когнитивна наука – Бойчо Кокинов и др. Учене с подкрепление, невронауки: Момчил Томов, Мартин Клисарски; Ралица Димитрова; Парашкев Начев; Йордан Златев – когнитивна семиотика, епигенетична роботика; Пламен Ангелов; невронният модел за роботи SPEAR-1 на института INSAIT: Н.Николов;и др. Повече – в другите томове

* Няма списък с всички препратки в края на книгата, за да се спести обем.

Приятна разходка в още една галактика от
вселената на мислещите машини
и техните пророци!

# Частично съдържание на приложение Листове 22

*** Отговор на възможни критики, че творбите от „Свещеният сметач“, „Изкуствен разум“ (Artificial Mind), „Разумир“, „Пророците на мислещите машини“ и др. не са научни публикации, нямат научна стойност и не се броят за приноси, защото не са публикувани на конференции, в научни списания – по-точно в „признати“, „рецензирани“, „индексирани“ и пр.** 48

* Обобщение на някои от съвпаденията: *”Теория на Разума и Вселената от 2001-2004 предвиди принципите и насоката на развитието на изкуствения интелект и универсалния изкуствен разум (…)“* Примери за множество теории, научни школи, произведения, които преоткриват и повтарят идеите и обобщенията на Теория на разума и Вселената в академичния поток и минават за нови приноси, макар че някои от тях са публикувани над 20 години по-късно. Сравнения и тълкуване, включително с големи езикови модели – подобно на показаното в „Stack Theory is yet Another Fork of Theory of Unvierse and Mind”. **52**

* Wolpert Theorem about mutual unpredictability and the impossiblity of subuniverses to predict with highest resolution of causality-control are rediscoveries of concepts from Theory of Universe and Mind – entangled with unnecessary mathematical notations and unsatisfiable premises [24.10.2025**]** **~ 55**

* Theories of Knowledge and Theories of Everything

* New proof reveals fundamental limits of scientific knowledge

* Alfred Tarski’s undefinability theorem

* “Limitation on knowledge

* Wolpert, Chaitin and Wittgenstein on impossibility, incompleteness, the limits of computation, theism and the universe as computer-the ultimate Turing Theorem

* Some publications from the classical TUM, the world’s first AGI courses in 2010-2011 etc. **~ 69**

* A discussion about “The Liar’s Paradox” from Universe and Mind 4, Todor Arnaudov, April 2004 and an excerpt in English and Bulgarian

* Тълкуване на заблуждаващия парадокс на „Парадокса на лъжеца“ и истинския му смисъл от „Вселена и Разум 4“, 2004 г. и откъси от оригиналния текст на английски и български – не е само самоотнасяне (self-reference).

* See also from Wolpert etc.

* Off-training set error and a priori distinctions between learning algorithms; * No free lunch theorems for optimization …

* What does it mean for a system to compute?

* Semantic information, autonomous agency and non-equilibrium statistical physics

* Universe Is Not a Computer Simulation, New Study Says

* Consequences of Undecidability in Physics on the Theory of Everything **~ 78**

* Relevance Realization rediscovers motives from Theory of Universe and Mind, while narrowing the meaning of computation and offering contrasting interpretations of some points, as well as limiting the scope of applicability of the principles only to living organisms (2001-2004 vs 2012-..2024) [ 24.10.2025] **~ 82 - 102**

* Comparisons and interpretations by Todor and by LLMs:

* Matches between Theory of Universe and Mind, Analysis of the meaning of a sentence … and the school of thought of Relevance Realization – According to a Quick Comparison with LLMs: KIMI-2, CLAUDE 4.5 and GPT-5.

* List of notes, comparisons, comments by Todor

* Comparisons of matches by LLMs: Thorisson and Talevi's theory of meaning in AGI from 2024 rediscovers the principles from Theory of Universe and Mind and in particular exercise significant correspondences to Arnaudov's paper "Analysis of the meaning of a sentence based on the knowledge base of an operational thinking machine. Reflections about the artificial thought", **~ 83** (Links to reviews with LLMs)

* About the "small world" and "simple" regarding "algorithms" or machines – a dilalog with predictions from 2005 about machine translation **~ 101**

* Notes on Computation, Algorithmic complexity, Intelligence by simulation, Causality etc. * Записки по изчислимост, алгоритмична сложност, ум чрез симулиране (симулационна интелигентност) и причинност и др. Преглед на много статии и бележки по тях – от Хектор Зенил, Лучано Флориди и др., вкл. класически работи от Леонид Левин, Андрей Колмогоров, Грегори Чайтин и др**. ~ 103**

* Todor's conclusion from appendix Algorithmic Complexity about Turing machines

* Todor Arnaudov's review of Computational Modelling vs. Computational Explanation: Is Everything a Turing Machine, and Does It Matter to the Philosophy of Mind?, by Gualtiero Piccinini.

*** Section on Algorithmic Complexity, Simulation Intelligence, Causality etc.**

* A Review of Methods for Estimating Algorithmic Complexity: Options, Challenges, and New Direction

* Simulation Intelligence: Towards a New Generation of Scientific Methods

* The frontier of simulation-based inference

* Causality from Bottom to Top: A Survey: Обзор на причинността: от дъното до върха .. * A Computable Universe: Understanding and Exploring Nature as Computation

* The Future of Fundamental Science Led by Generative Closed-Loop Artificial Intelligence

* On the Algorithmic Nature of the World; On the Kolmogorov-Chaitin Complexity for short sequence; * Information Theory and Computational Thermodynamics: Lessons for Biology from Physics and notes by Tosh;

* The Thermodynamics of Computation; Methamathematics and Metabiology; Life as evolving software.

* Some Computational Aspects of Essential Properties of Evolution and Life

* On Randomness, regularity …

* What is Nature-like Computation? A Behavioural Approach and a Notion of Programmability

* Is Information Meaningful Data? Philosophy and Phenomenological Research, Levin, L.A. * Laws of information conservation (nongrowth) and aspects of the foundation of probability theory * Image information content characterization and classification by physical
Complexity * The World is Either Algorithmic or Mostly Random, H.Zenil – notes by T.
* SuperARC: An Agnostic Test for Narrow, General, and Super Intelligence Based On the Principles of Causal Recursive Compression and Algorithmic Probability…

*** Теория на сглобяването ~ 119**

* Език, термини, юнашко наречие: българският и библията; възниковенци;

*** Йоша Бах и Педро Домингос ~ 122**

* Йоша Бах: Интервю с Бен Гьорцел H+, 2011?
* Педро Домингос за главния алгоритъм (master algorithm) ..
* Domingos: Unifying Logical & Statistical ML/AI …
* SCHMIDHUBER: HOW WE WILL LIVE WITH Ais

*** The Learnable Universe, Vectors of Cognitive AI** 5/3/2023 - беседи с Йоша Бах, Кристоф фон дер Малсбург, Стивън Волфрам и др.

* Vectors of Cognitive AI: Generalist AI beyond Deep Learning
* Vectors of Cognitive AI
  * Vectors of Cognitive AI: Motivation and Autonomy
  * Vectors of Cognitive AI: Self-Organization
  * Vectors of Cognitive AI: Attention
* Multiway Systems as Models to Understand Mind and Universe - a Conversation with Stephen Wolfram
* Още от Стивън Волфрам от 2024 г.
* Stephen Wolfram - Where the Computational Paradigm Leads (in Physics, Tech, AI, Biology, Math, ..., 11.2024
* Какво е времето? – Стивън Волфрам – What is Time? Stephen Wolfram's Groundbreaking New Theory
* What is Time? Stephen Wolfram's Groundbreaking New Theory
* Joscha Bach: Time, Simulation Hypothesis, Existence | 6.10.2020
* Multiway Systems as Models to Understand Mind and Universe - a Conversation with Stephen Wolfram (Докъде води прилагането на парадигмата за изчислителната Вселена? – Стивън Волфрам)

*** Self-Organizing: Самоорганизация: The Future of AI is Self-Organizing and Self-Assembling** – S.Risi

*** Преглед на множество предавания от канала Machine Learning Street Talk:**

* MLST #107 Raphael Milliere – Long theory of Mind Grounding .. 14/3/2023
* MLST #79 Consciousness & Chinese Room .. Data2Vec … 19/3/2023 … GPT4
* Cultural Affordances: Scaffolding Local Worlds Through Shared Intentionality and Regimes of Attention;
* John Min Tan …Singapore – "Memory Soup" – see more below
* MLST #95 Irina Rish – AGI, Complex Systems, Transhumanism @Neurips
* MLST #75 Emergence [Special Edition] Dr. Daniele Grattarola 11 хил. пок. 7/4/202

* ICLR 2020 Bengio – Consciousness * 8/4/ Lottery Ticket theory – NN prunning, distillation
* LeCun – Energy-based – Льокан („Льокун“), енергийни методи за маш.обуч.
* LeCun – V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning, Mido Assran et al. (META), 6.2025
* Energy-Based Transformers are Scalable Learners and Thinkers,
* MLST #52 Adversarial Examples – Hadi Salman – Злонамерени данни
* MLST #104 – Natural GI – Christian Summerfield 22/2/2023 – Естествен общ интелект
* Structure learning * Core knolwedge – priors; Elizabeth Spelke – психология на развитието
* Pedro Domingos: Master algorithm. The five tribes of ML (And what you can learn from each) … Unifying Logical & Statistical AI
* Петър Величкович, Peter Velickovic, 7/3/2023 – categorical theory – теория на категориите (математика за вид логическо машинно обучение
*** Майкъл Левин | Michael Levin: Биофизика, биоинженерство, интердисциплинарност …** * Michael Levin – “From physics to mind” **– 143**
* Блог на Майкъл Левин: Формите на живот са и форми на ум
MICHAEL LEVIN
* Neuroscience Beyond Neurons: bioelectricity underlies the collective intelligence of cellular swarms
* Conversation with Chris Fields and Richard Watson #2
* What are Cognitive Light Cones? (Michael Levin Interview); Какво представляват познавателните светлинни лъчи? Интервю с Майкъл Левин
* Спорността, неяснотата при романтичното определяне на светлинния лъч на познавателността – бележки на Т.Арнаудов
* За понятията за състрадание и чувства като спорни явления за определяне на одушевеност на цялостен организъм заради възможността да се потисне с „прости“ молекули и пр. – бележки на Т.Арнаудов **– 153**
* “Cognitive lightcone”: compassion: Познавателният светлинен лъч: съчувствие: бележки към разсъждения на М.левин от Тош
* The collective intelligence of cells during morphogenesis as a model for cognition beyond the brain. M.Levin: 20.2.2023, Talk, 23.1.2023 … Сравни с А.Шопенхауер, 1813+
* SEMF: Испански институт насърчаващ междупредметни изследвания. Поредица „Spacious, Spatiaity”: Spatial Intelligence and Challenges of the Spatial World; Conceptual Spaces & the Geometry of Word Meanings, Peter Gardenfors *156*
* Бележки към Bach,? TedX 2016 ? … и др.: From Artificial Intelligence to Artificial Consciousness, TEDxBeaconStreet; ~ 157
* Болка * Pain * Suffering: Т.Арнаудов, 28.8.2023 **~ 158**
* ActInf MathStream 009.1 ~ Jonathan Gorard: A computational perspective on observation and cognition ~ 159
* Принцип на свободната енергия и извод чрез действие: Karl Friston @ MLST … Карл Фристън бел. 7/10/2023, 28/10/23 Max Ramstead и др. …

* Todor Arnaudov: On What Creativity Is - Different Tints and the Pereslegin's remark about the modern AI researchers rediscovering Lem 1963
* Conversation with Mark Solms and Chris Fields #4
*** Следващото еволюционно стъпало**: заключенията на Джефри Хинтън, 2024 са буквално повторение на изводите от есе от 1999 и ТРИВ, 2001-2002 на Тодор Арнаудов и други съвпадения от 2025 г. ~ 168
*** Още бележки към Анди Кларк:** „Whatever Next: …", 2013 ~ 170
* Бележки към „Радикална предсказваща обработка" на Анди Кларк, 2015
* Относно „спретнатите" и „мърлявите" в ИИ (neats and scruffies): „My own sympathies have always lain more on the side of the scruffies….
*** Коментар на Тодор Арнаудов към видеото на Тим Тайлър „Да се слеем [с машините]"** от 10.2023 – космизъм, трансхуманизъм; технологиите са част от човечеството по начало **~ 175**
* #67 Prof. KARL FRISTON 2.0 [Unplugged] MLST - The Burden of Knowledge Across Disciplines [тежестта на битието на интердисциплинарен учен]
* What about fuzzy Markov boundaries?
* Mahault Albarracin ~ Active Inference Insights 002 | Active Inference Institute
* Критика на FEP/AIF: Трикът с одеалата на Марков: за обхвата на принципа на свободната енергия и заключението чрез действие: The Markov blanket trick: On the scope of the free energy principle and active inference **~ 179**
* Todor Arnaudov's comments on: Causal blankets: Theory and algorithmic framewor ~ 182
* Todor: AIF is vis vitalis (Active Inference е въплъщението на понятието за „жизнена сила" и Воля на Шопенхауер) - 28.11.2023 **~ 185**
* Karl Friston ~ Active Inference Insights 001 ~ Free Energy, Time, Consciousness | Active Inference Institute | 3,22 хил. абонати (3.4.2024) 4778 показвания Начало на премиерата: 22.11.2023 г
* ActInf GuestStream #018.1 ~ Michael Kirchhoff & Julian Kiverstein
* The Literalist Fallacy & the Free Energy Principle: Model-building, Scientific Realism and Instrumentalism
* Prof LARISA SOLDATOVA - Automating Science: какво е стол: сравни с определенията от Т.А. 1999, 2012
* Conversation with Chris Fields and Richard Watson #2
* CHOMSKY - WHY WE DON'T KNOW THE WORLD; Ян Льокан, Y.LeCun ... от 10.7.2022 ... MLST Chomsky ...
* Neuroscience Beyond Neurons: bioelectricity underlies the collective intelligence of cellular swarms
*** #4.11.2023, FEP, AIF and mapping to Todor's TOUM and Zrim... – 188**
* Stochastic Processes in Physics and Chemistry; UNSTABLE SYSTEMS
*** Из бележките на докторант по ИИ от Сингапурския национален университет за паметта и коментари на Тош – 190**
* Todor's comments on "Memory soup", reinforcement learning – John Chong Min Tan ... 7.11.2023, Discord, John's AI Group … ; reward-based and goal-directed can be expressed by each other without precise specification; * Moving Beyond

Probabilities: Memory as World Modelling (…)
* Todor: Zrim's Executable Contexts
* YouTube - John Chong Min Tan; A New Framework of Memory for Learning
* Moving Beyond Probabilities: Memory as World Modelling: John Chong Min Tan
*** Kant and Schopenhauer already defined many of the modern concepts in AGI which are still not well understood or are rediscovered by the modern English-speaking world AI and mind researchers etc**.: Todor Arnaudov **~ 196**
* Eight Things to Know about Large Language Models – compare to parts of the novella "The Truth", T.Arnaudov, 2002
* Виж „Истината", 2002; „Читанка" [The SF novel about AGI, mind, universe: "The Truth", T.A., 2002]
* Theory of Mind and Universe (TOUM) and the universal simulators of virtual universes, time, present, past…: comments to John's AI Group in Discord;
* Противоречията между различни потоци на управляващо-причиняващо устройство: чувствена и познавателна система; пример от интервю при Лекс Фридман: John Mearsheimer: … * Unwanted prediction …;
*** Todor comments on Reinforcement learning and agents** in an answer to John Min Tan: The complex dynamics and behavior of generally intelligence agents come from the multilayer, multiresolution, multirange, multidomain, multiview, preemptive, varying prediction horizon, resolution of perception and causality-control etc. operation, 12/2023 (John-Tan-Min-johntcm-tic-tac-toe-new82.txt)
* Todor discusses three types of creativity as cited from a Demis Hassabis comment: – "On What Creativity Is - Different Tints and the Pereslegin's remark about the modern AI researchers rediscovering Lem 1963, T.A. 16.12.2023 **~ 209**
* The Active Inference Institute and Active Inference Ecosystem
* Todor on the "compressionists" and "analogists" on Discord, John's AI Channel, 1.2.**2024 ~ 217**

## * Съзнание и Панпсихизъм | Consciousness & Panpsychism

*** Уводни бележки за съзнанието, преливането между живота и ума, панпсихизма и противоречащия на Принципа на свободната енергия: Принцип на максималната ентропия –** допълнения към Вселена и Разум 6
* Разграничаването на системата от средата и „самосъздаването" са спорни въпроси **~ 230**
* **Средата като жива и живите същества като неживи** – 6.10.2025
* Обзор и на теории и школи за съзнанието: Landscape of Consciousness, THE KUHN FOUNDATION
* Consciousness science: where are we, where are we going, and what if we get there, 30.10.2025
* A Study of "Organizational Closure" and Autopoiesis,
* Не е нужно моделите да са линейни и детерминистични в пълен смисъл, а обичайните невронни мрежи могат да се видят и като „сиви" или „полупрозрачни" кутии дори и както са – Тош

*** Бележки и разсъждения на Тодор Арнаудов вдъхновени от видеото с интервюта с Фредерико Фаджин* за идеализма, квантовата механика, свободната воля и самоличността**; Federico Faggin on Idealism, Quantum Mechanics, Free Will, and Identity * Интервю с Фаджин от 1995 г. в архив на Станфорд: * Zilog Oral History Panel on the Founding of the Company and the Development of the Z80 Microprocessor **~ 250-261**
* Hard Problem and Free Will: An Information-Theoretical Approach
* Artifcial Intelligence Versus Natural Intelligence
* Agency cannot be a purely quantum phenomenon
*** Todor Arnaudov's review and interpretations of concepts from Causal potency of consciousness in the physical world by Danko D. Georgiev … ~ 262**
* Todor Arnaudov's interpretation of ideas from Quantum information theoretic approach to the mind–brain problem by D.Georgiev
* Computational capacity of pyramidal neurons in the cerebral cortex
* Enhancing user creativity: Semantic measures for idea generation
* Creativity and artificial intelligence, Margaret A. Boden, 1998
* The 'Quantum Underground' - Where Life Defeats Decoherence (S.Hameroff)
Изследвания за връзката между обезболяващите средства и съзнанието ~ 285
* What Came Before the Big Bang? | Theory of Embedded Intelligence, Bill Mensch & Bernardo Kastrup
* Todor Arnaudov comments on: Is Reality Real? - This One Idea Might Change Your Entire Life | Donald Hoffman
* Do We Perceive Reality?, John Klasios, 19.12.2022:
* Тодор Арнаудов коментира теорията на Томас Кембъл: "My big TOE" – Theory of Everything My Big Toe: A Trilogy Unifying **~ 291**
* Тодор Арнаудов коментира: Bernardo Kastrup: Are We Dissociated Alters Of Cosmic Consciousness? Tevin Naidu 9,82 хил. абонати 9.7.2023 https://www.youtube.com/watch?v=57Oguwg7omc
* Математическата вселена на Макс Тегмарк
* Сравни панпсихизма на Денет за многото чернови, всички съзнателни, с интеграла от множество безкрайно малки самоличности в статията на Тош за акразията от 2012 и Анализ на смисъла на изречение въз основа на базата знания на действаща мислеща машина от 2004
*** Review of pioneering work by Paul Werbos, Jerome Feldman and Bruce MacLennan in Neural Networks, Connectionism, non-Turing computation** (Преглед на работа на Пол Уербос, Джером Фелдман): **Werbos:** Backpropagation Through Time: What it Does and How to Do it; (…) Backwards Differentiation in AD and Neural Nets: Past Links and New Opportunities; Approximate Dynamic Programming for Real-Time Control and Neural Modeling (…); *** Feldman:** Connectionist models and their properties; Aspects of associative processing … *** MacLennon:** Field computation; A review of analog computing; Continuous symbol systems: The logic of connectionism …
*** Todor: The weaknesses of the "failed" logical approach is in other points. 314**
*** Допълнителна литература за съзнание и панпсихизъм;** Phenomenal

consciousness**; Luke Roelofs:** Combining Minds; Why I am not a cosmopsychist Or am I; Panpsychism and Illusionism; Why I am not a cosmopsychist Or am I? – see Todor's comment. * Illusionism as a Theory of Consciousness.
*** Тодор коментира: *** The Selection Problem for Constitutive Panpsychism – P. Woodward; виж Вселена и Разум 6, причинностни белези (causal ids)
* Cosmopsychism and the Problem of Evil – Сравни с ТРИВ, повече от 20 години по-рано: … **Изкуственият разум Емил е панпсихист и космопсихист.**
* Why? The purpose of the Universe – P.Goff
* Todor Arnaudov comments Michael Levin | Bernardo Kastrup #3 - With Reality in Mind * Adventures in Awareness – 327
* Todor Arnaudov's comments on Thomas Metzinger's "Ego Tunnel" and the Phenomenal Model of Self and his "Consciousness test" dialog with non-living agents and its correspondence with the dialogs with thinking machines from Todor's works "Man and Thinking Machine: Analysis of the Possibility that a Thinking Machine Could be Created and Some Disadvantages of Man and Organic Matter in Comparison", 2001 and the short science fiction novel "The Truth", 2002 ~ 337
* Тодор Арнаудов коментира Томас Мецингер: „Дали Аз-ът е Илюзия?" – предаване с Тевин Наиду от 22.7.2023 | Todor comments: Thomas Metzinger: Is The Self An Illusion? - Tevin Naidu;
* Principles of Minimal Cognition: Casting Cognition as Sensorimotor Coordination;
* Невроанатомични съответствия на религиозните и духовни усещания; канал DemistifySci / Демистифай Сай
*** Neuroanatomical Correlates of Religiosity and Spirituality: Невроанатомични съответствия на религиозните и духовни усещания** – 340
* What religion does to your brain, Ana Sandoiu on July 20, 2018
Principles of Neurotheology
* The Mystical Mind: Probing the Biology of Religious Experience
* Excerpt from "Stack theory is yet another Fork of Theory of Universe and Mind
* Sentience or consciousness of another "entity" is in the eyes of the evaluator:
*** Thoughts from a letter by Todor to the cognitive semioticist J.Zlatev ~ 341**
* Signs of introspection in large language models
* Emergent Introspective Awareness in Large Language Models,
* Large reasoning models almost certainly can think
*** Мултимодалност ... Multimodal** (…)
* Законът за изчислението, разговор за символността и „конституцията" за ИИ
* Computational Law, Symbolic Discourse and the AI Constitution, S.Wolphram,
* **Тош:** Двата или многото потоци на представяне на мрежата на управлението: бележки към Хазин и Щеглов, „Стълба към небето" за законите на властта и др.
* Karl Friston, Adam Goldstein, and Michael Levin discuss active inference and algorithm
* WE MUST ADD STRUCTURE TO DEEP LEARNING BECAUSE... MLST

## **Образование** – 349

* **Докторантската система и продължителността ѝ:** 5 години с 3 години планиране
* PhD system and duration: 5 years with 3 years planning - isn't it too long and risking obsolescence on the go? What about contributing to many projects and not just one "personal" thesis?,
* Други потвърждения на препоръките за интердисциплинарността и критики за висшата образователна система и критериите с цитирания
* За отживялата форма на модела за докторантура:
* PhD training is no longer fit for purpose — it needs reform now
* Разкритие: милиони долари се пилеят за подготовка на статиите във форматите изисквани от списанията
* Няма научни нововъведения от 1920-те? Дали академичният принцип „публикувай или загини“ задушава науката? „Затварят устата на гениите!“ – Грегъри Чайтин
* Бележки из дискусии от групи за ИИ, Джон Тан Мин и Тош
*** Други перспективни открития, разработки, насоки и бележки**
* Нов поглед към тахионите: Как бъдещето влияе на настоящето; * Учените преразглеждат основите на квантовата теория; * Study Reveals Dopamine’s Limited Role in Rapid Neural Activity – Ограничената роля на допамина в бързото възбуждане на невроните (виж и в съзнание и панпсихизъм преди #edu**)**
*** Машинно обучение съобразено с физиката и Аналитична регресия** (Physics Informed ML and Symbolic Regression) – 358
* DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators
* Машинно обучение на динамични системи чрез данни: въведение в обучението на физични дълбоки невронни мрежи
* Learning dynamical systems from data: An introduction to physics-guided deep learning
* Кратки бележки по математика и вериги на Марков
* “Течен изкуствен интелект“ – „Ликуид ИИ“ - методи за машинно обучение основан на динамични системи и „течни“ невронни мрежи
* МКА: Мрежи на Колмогоров-Арнолд – KAN: Kolmogorov-Arnold Networks
* Компании: SingularityNET, Verses, Thinking Machines,
* VERSES Publishes Pioneering Research Demonstrating More Versatile, Efficient, Physics Foundation for Next-Gen AI
* Нови компании: Thinking Machines, World Labs, Human&; “Quiet-STaR:…”; Project Prometheus, Physical Intelligence, Klover

## *** Невронауки и съчетание с машинно обучение, учене с подкрепление и изкуствен интелект** – 377

* Neuroscience in Combination with Machine Learning, Reinforcement Learning and Artificial Intelligence
* Динамични системи в ума и състезание без победител – Михаил Рабинович ..

* Dynamical systems and Winnerless Competition -
* Математика на съзнанието -
* Dynamical Encoding by Networks of Competing Neuron Groups: Winnerless Competition
* Mind-to-mind heteroclinic coordination: model of sequential episodic memory initiation
* Hierarchical dynamics of informational patterns and decision-making,
* Обобщение и бележки на Тодор Арнаудов
* Information flow dynamics in the brain
* Пламен Ангелов – влиятелен български „пророк“ още от 1990-те

## * Машинно обучение, обучение на човека и невронауки и взаимодействието между тях – 390

* Machine Learning, Human Learning and Neuroscience and their interaction
* Momchil Tomov – Момчил Томов
* Related Bulgarians: M.Klissarov, E.Todorov & Peter Kormushev
* Multi-task reinforcement learning in humans, Momchil S. Tomov
Momchil Tomov was born and raised in Bulgaria. BSc in Princeton, PhD in Harvard
* Discovering Temporal Structure: An Overview of Hierarchical Reinforcement Learning, Martin Klissarov,
* Tomov, Momchil, 2020. Structure Learning and Uncertainty-Guided Exploration in the Human Brain, Doctoral dissertation
* More Efficient Randomized Exploration for Reinforcement Learning via Approximate Sampling
* Regret bounds of model-based reinforcement learning
* RL Virtual School lectures
* Reward Processing Biases in Humans and RL Agents,
* Dr Irina Rish - Introduction to Continual Learning,
* Online Fast Adaptation and Knowledge Accumulation (OSAKA): a New Approach to Continual Learning
* How to Train Your LLM Web Agent: A Statistical Diagnosis
* Inverse Reinforcement Learning; A survey of inverse reinforcement learning
* A note on “Habits, action sequences and reinforcement learning“
* Multi-hierarchical representation of large-scale space
* Multi-abstraction hierarchies - Multi-AH-graph ..

## * Why humans usually answer by addressing the higher-level nodes, when asked by another human about the implementation of a prospective long horizon planning problem? , Todor Arnaudov, 8.2025, The Sacred Computer – a respond to Momchil Klissarov’s PhD thesis. – 410

* The neural architecture of theory-based reinforcement learning. Tomov, M. S.,
* Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments

* Reinforcement learning, fast and slow
* What Is the Model in Model-Based Planning
Human-Level Reinforcement Learning through Theory-Based Modeling, Exploration, and Planning
* Successor Feature Representations
*** Повторна и многократна обработка на съхранени записи в мозъка и машините – 425**
*** Replay in brains and machines, Lennart Wittkuhn et al.**
* Основни функционални мрежи в човешкия мозък и методи за откриването и изследването им
* Main functional networks in the human brain and methods for their discovery and study:
* Интересни случаи на пациенти родени без челни дялове на кората и без малък мозък
* Изследвания на мозъка, мозъчното развитие и нервната дейност при новородени (Ралица Димитрова …)
* Synchrony and subjective experience: the neural correlates of the stream of consciousness, * Intrinsic neural timescales: temporal integration and segregation
* Signal Detection Theory: * Salomon, G. Television is 'easy' and print is 'tough': the differential investment of mental effort in learning as a function of perceptions and attributions …
* Supplementary motor area as key structure for domain-general sequence processing: A unified account,
* The minimal computational substrate of fluid intelligence, Amy PK Nelson, … Parashkev Nachev et al., 10/2024
* SuperARC: An Agnostic Test for Narrow, General, and Super Intelligence (…)
* Helpless infants are learning a foundation model
* Where do you know what you know? The representation of semantic knowledge in the human brain, Karalyn Patterson
* From task structures to world models: What do LLMs know?, Ilker Yildirim & L.A. Paul, 2023
* From word models to world models: translating from natural language to the probabilistic language of thought.
* Frontal language areas do not emerge in the absence of temporal language areas: A case study of an individual born without a left temporal lob
* What's wrong with Natural Language Processing?, T.Arnaudov, 2.2009:
* Относно съгласуваността и „ценностите" виж (AI Alignment):
Откъс от повестта „Истината", Т.Арнаудов 2002 г.:
* Modeling Open-World Cognition as On-Demand Synthesis of Probabilistic Models,
* The neural basis for uncertainty processing in hierarchical decision making,
* Fragmentation and multithreading of experience in the default-mode network
*** Невроморфни системи | Невроморфни компютри – 463**

* Different brain structures associated with artistic and scientific creativity: a voxel-based morphometry study,
* Невроморфни системи в Пловдив и България
* Кръвоносната система на живите организми е тяхна слабост (Тош)
* Българска компания: * Невроморфика
* Darwin3: a large-scale neuromorphic chip with a novel ISA and on-chip learning, 5.2024; How China's new 'Darwin Monkey' could shake up future of AI in world first First such supercomputer with over 2 billion artificial neurons mimics macaque brain,

## * Когнитивна лингвистика – 469

* Още бел. към „Геометрия на значението…" от П.Грендерфорс:
* Todd Oakley, Image Schemas, January 2012
  Образни схеми на действия, смисъл, понятия.
* Екологична психология: James Gibson –

## * Изкуствен инелект в СССР – 472

*** Школата на Михаил Бонгард в СССР от края на 1950-те до средата на 1970-те** #bongard #бонгард – Проблема узнавания и др.
* Сравни с Франсоа Шоле, 11.2024 * Моделирование обучения и поведения., М., "Наука" 1975, * Проект за модел на организацията на поведението "Животно" * Формален език за описание на ситуации, използващи понятието връзка * Задачата за обобщение на началната ситуация ...
* Густав Херц, 1959: "Най-близката и най-важна задача на съзнателното ни опознаване на природата се свежда до *това да намерим възможност да предвидим бъдещия опит* …"
* А.Зиновиев, Логическа физика, 1972

*** Мислене по Аналогия** – бележки към „Аналогичният ум: перспективи от когнитивната наука #кокинов #kokinov **– 486**
* The analogical mind : perspectives from cognitive science
* Мислене по аналогия, Разпознаване на образи чрез групиране (клъстериране), Сравняване, Pattern-matching, Когнитивна наука и теории за образуване на понятията, Синтез на програми, дискретно търсене, Program synthesis, Discrete program search, Analogy, Bongard, Hofstadter … #bongard
* PHAEACO: A COGNITIVE ARCHITECTURE INSPIRED BY BONGARD'S PROBLEMS, Harry E. Foundalis

## * Мулти-агентни системи, планиране и роботика от 1980-те и 1990-те – 493

* Multi-Agent System, Planning and Robotics in 1980s-1990s …
* Майкъл Питър Джорджев **…**
*** Procedural Knowledge**
* A Model-Theoretic Approach to the Verification of Situated Reasoning Systems

*** BDI Agents:** From Theory to Practice, Australlian Artificial Intelligence Institute, [Познавателната архитектура за агенти „Убеждения-желания-намерения“: Believe-Desire-Intention]
*** Multiagent Systems**, Katia P. Sycara, 1998
* A roadmap of agent research and development
* Distributed intelligent agents, K Sycara 1996
* Commitment and Effectiveness of Situated Agents
* Introducing the Tileworld: Experimentally evaluating agent architectures. [Pollack and Ringuette, 1990]
* TRIANGLE TABLES: A PROPOSAL FOR A ROBOT PROGRAMMING LANGUAGE Technical Note 1985
* Universal Plans for Reactive Robots in Unpredictable Environments, M.J . Schoppers, 1987
*** Intention, Plans, and Practical Reason**, Michael E. Bratman, 1987
* TouringMachines: Autonomous Agents with Attitudes, Innes A. Ferguson, 1992.
* Innes A. Ferguson. TouringMachines: An Architecture for Dynamic, Rational, Mobile Agents. Сравни TouringMachine с:
* Tartan Racing: A Multi-Modal Approach to the DARPA Urban Challenge, April 13, 2007, Chris Urmson et al. – научна статия за победителя в състезанието за самоуправляващи се коли на DARPA в градски условия от 2007 г.
* Подробна лекция по мулти-агентни системи от 2001 г. от Olivier Boissier (основно на англ. : Multi-Agent systems - Agent's Architectures
* An Overview of Agent-Oriented Programming, Yoav Shoham (a chapter from a book)
* Toward Team-Oriented Programming

## * Планиране * Класическо планиране #plan #планиране #planning – Classical Planning – 510

* High-Level Planning In a Mobile Robot Domain, 15.7.1986
* Hierarchical Planning: Definition and Implementation, D. Wilkins, 20.12.1985
* Hierarchical Planning at Differing Abstraction Levels, David E. Wilkins, 12.1988
* Generating Instructions at Different Levels of Abstraction, A. e Kohn et al., 2020
***Йерархични мрежи за задачи (HTN, Hierarchical Task Network)**
* Hierarchical Task Network Planning: Formalization and Analysis
*** Пример на PDDL (STRIPS):**
* Школата около Аарон Сломан в Англия в когнитивните архитектури в Англия
* The Mind as a Control System, Aaron Sloman, 1992
* Мултиагентни системи MAS. Семантичен уеб – избираем курс във ФМИ ПУ

## * Роботика и учене с подкрепление | Robotics and RL – 517

* Важни понятия от механиката, роботиката, мехатрониката
* Track Everything Everywhere Fast and Robustly
*** Устойчиво, приспособяващо се и най-добро управление** /адаптивно

управление, оптимално управление; регулатори/ **Robust Control, Adaptive Control & Optimal control**

*** Учене с подкрепление. Още бележки по роботика – 527**

* Reinforcement learning
* Dynamic Movement Primitives (DMP)
* Reinforcement Learning for Motor Primitives, * Learning Motor Primitives for Robotics,
* Leveraging LLMs, Graphs and Object Hierarchies for Task Planning in Large-Scale Environments * PUMA: Deep Metric Imitation Learning for Stable Motion Primitives,

*** Човекоподобна роботика и управление чрез предсказващо управление (humanoid robots, model predictive control, MPC; хуманоидни роботи): …** Виж в основния том, – Драган Ненчев и др **– 528**

*** Optimization-based Locomotion Planning, Estimation, and Control Design for the Atlas Humanoid Robot** … * Optimization Based Full Body Control for the Atlas Robot, .. *** Robot navigates high-speed parkour with autonomous movement planning 2025** * За някои от сензорите в „Атлас“:

*** Допълнения към Епигенетична роботика, роботика на развитието** #robotics #epigeneticrobotics * Epigenetic robotics, developmental robotics **– 530**

*** Навигация и умствени карти, когнитивни карти, карти в умствен план, субективни карти … при човека** – 532

* Spatial Knowledge Acquisition in the Process of Navigation: A Review
* SPATIAL ORIENTATION, WAYFINDING, AND REPRESENTATION
* Prototypes, Location, and Associative, Networks (PLAN): Towards a Unified Theory of Cognitive Mapping

*** Възобемване, възстановяване на обема, пространствено възстановяване**, възстановяване на геометрията чрез множество от кадри от различни гледни точки и движение, 3D-реконструиране, SLAM, навигация – класически работи, роботика, навигация за роботи **– 534**

* 3D Reconstruction, Structure from Motion, Robot Navigation… #SfM #SLAM #structure from motion #3D-reconstruction – древни и съвременни методи – старите работят със символична изчислителна мощ.
* Сравни съвременен подход към SLAM с постепенното развитие на областта:
* On Learning and Geometry for Visual Localization and Mapping,
* 3D Positional Integration From Image Sequences, Determination Of Ego-Motion From Matched Points
* **"Solving three-dimensional small rotation equations**: Uniqueness, algorithms and numerical results; * **DETERMINATION** OF EGO-MOTION FROM MATCHED POINTS; *** Motion and Structure from Motion from Point and Line Matches**; * **Vision Algorithms** for **Mobile Robotics * A Computer Algorithm** for **Reconstructing a Scene from Two Projections * The Eight-Point Algorithm**; * **Vision Algorithms** for Mobile Robotics; * A Brute-Force Algorithm for Reconstructing

a Scene from Two Projections; A Benchmark for the Comparison of 3-D Motion Segmentation Algorithms**. * Topological Map Learning** from Outdoor Image Sequences *** Learning to find good correspondences** * Structure-from-Motion under Orthographic Projection * Mobile Robot Localisation Using Active Vision
* A Frame for Frames: Representing Knowledge for Recognition,
* **Modeling spatial knowledge**,Benjamin Kuipers, 1978
* Shakey: From Conception to History
*** Причинностно моделиране и продължение на картографиране – 539**
* Human Thought, A.Newell, H.Simon, 1961 + Theorem-Proving by Resolution as a Basis for Question-Answering Systems
***** Commonsense Reasoning about Causality: Deriving Behavior from Structure, Benjamin Kuipers, 1984
* A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations
***** The Spatial Semantic Hierarchy

## * Съвременна роботика и навигация за роботи и общи мултимодални агенти и модели за управление на графичен потребителски интерфейс (интелигентна ОС, умна ОС, Computer Use Models) – 540

* Zero-shot Object Navigation with Vision-Language Models Reasoning, 2024
* CoWs on PASTURE: Baselines and Benchmarks for Language-Driven Zero-Shot Object Navigation
* RoboTHOR – реалистичен въображаем свят за роботи, симулатор
* Habitat-Matterport 3D Dataset (HM3D): 1000 Large-scale 3D Environments for Embodied
* Foundation Models in Robotics: Applications, Challenges, and the Future
* Transformer-based Learning Models of Dynamical Systems for Robotic State Prediction
***** Magma: A Foundation Model for Multimodal AI Agents
***** Mind2Web: Towards a Generalist Agent for the Web
* Reinforcement Learning on Web Interfaces using Workflow-Guided Exploration.
* Android in the Wild: A Large-Scale Dataset for Android Device Control
* LIBERO: Benchmarking Knowledge Transfer in Lifelong Robot Learning lifelong learning in decision making (LLDM)

*** Seminal Multi-Agent AI** – a recent survey #multi-agent 31.3.2025 – 548
*** Advances and Challenges in Foundation Agents: From Brain-Inspired Intelligence to Evolutionary, Collaborative, and Safe Systems** – Compare to the TOUM 2001-2004: many corresponding predictions;

## * Вземане на решения, мотивация, функция на ползата,

**икономика, психология, агенти, мулти-агентни системи, риск, неопределеност #decision-making #planning – 551**

* Литература и записки за множество понятия
* Бележки за „Черният лебед: …“, Насим Талеб от Тош
* A foundation model to predict and capture human cognition
* Random Tree Model of Meaningful Memory

*** Допълнително четене, понятия и бележки по когнитивна наука, психология, невронауки и др. във връзка със статията на Пилишин за ранната зрителна обработка и др. #cogsci+ – 559**

***** (…)
* Finding Math Hard? Blame Your Right Parietal Lobe
* Increased Gray Matter Density in the Parietal Cortex of Mathematicians: A Voxel-Based Morphometry
* Learning attentional templates for value-based decision-making
***** Neural mechanisms of selective visual attention
* Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices
* Orienting Attention Based on Long-Term Memory Experience
* Endogenous and exogenous attention shifts are mediated by the same large-scale neural network
* Is creativity a feature of ADHD? Дали творческите способности са черта на ХАДВ; The Creativity of ADHD; „Дефицит на вниманието и хиперактивност“ и творчество; New research reveals how ADHD sparks extraordinary creativity ..
* Теория на системите
***** Neural Mechanisms of Attentional Reorienting in Three-Dimensional Space
* **AI Ethics. AI Safety. SAI Safety. Superintelligence. AI Alignment – 567**
* Виж приложение „Фантастика. Футурология. Кибернетика. Развитие на човека“ #sf * Tovarna na absolutno, Karel Capek, 1922 (Фабрика за Абсолют, Карел Чапек) * Speculations concerning the first ultraintelligent machine * Summa Technologiae * Machine Super Intelligence …
***** Towards Friendly AI: A Comprehensive Review and New Perspectives on Human-AI Alignment; Introducing GPT-OSS
* Deliberative alignment: reasoning enables safer language models
* Verifying Chain-of-Thought Reasoning via Its Computational Graph
* The Instruction Hierarchy: Training LLMs to Prioritize Privileged Instructions
* Superintelligent Agents Pose Catastrophic Risks: Can Scientist AI Offer a Safer Path?, Yoshua Bengio et al.
* Geoffrey Hinton
* The Godfather of AI repeats ideas of the Child Prodigy of the Thinking Machines 24-22 years later,
* "Godfather of AI: I Tried to Warn Them, But We’ve Already Lost Control! Geoffrey Hinton",The Diary Of A CEO

* 'Godfather of AI' shortens odds of the technology wiping out
* The ‘godfather of AI’ reveals the only way humanity can survive superintelligent AI Jeff Hawkins
* Neuroscientist Explains Why AI Isn't Intelligent & The Thousand Brains Theory
* 'We have to stop it taking over' - the past, present and future of AI with Geoffrey Hinton
* Will AI outsmart human intelligence? - with 'Godfather of AI' Geoffrey Hinton
* "Godfather of AI" Geoffrey Hinton: The 60 Minutes Interview
* Eleuther AI, founded in 2020
* MIRI, founded in 2000 and the LessWrong forum and a sample with 2024 rediscovery of Todor’s yet unpublished line of design and research in human-computer interaction from the early 2010s
* Good Machine, Nice Machine, Superintelligent Machine
* Georgi Georgiev & Iskren Georgiev (2002): The Least Action and the Metric of an Organized System
* Power Overrides Intelligence: T.Arnaudov, James Bowery vs Matt Mahoney
* „Къде отиваш, свят?“,
* “Where are you going, world?” – the Machine God, Todor 1999; ironically about the simulated universe in “The Matrix in the matrix is a matrix in the matrix”, 2003;
* The forum of MIRI
* Nick Bostrom
* Copyright
* AI does a better job of ripping off the style of famous authors than MFA students do
* Всички права запазени! – Тош, 2003
* A New AI Research from Anthropic and Thinking Machines Lab Stress Tests Model Specs and Reveal Character Differences among Language Models
*** Funny “ethical” nonsense**
* Колко струва животът на хора от различни нации според езикови модели
* Чичовци всеятели на страх – AI fearmongers
* До 2030 г. ще останат само 5 професии. Д-р Роман Ямполски прогнозира 99% безработица – Експертът говори за своите страхове относно развитието на технологията
* Creative Intelligence will be First Surpassed and Blown Away by the Thinking Machines, not the "low-skill" workers whose jobs require agile and quick physical motion and interactions with human-sized and human-shaped environment | Творческата интелигентност първа ще бъде надмината и издухана от мислещите машини, а не "нискоквалифицирания" труд – Тош, 2013
* Philosophical and Interdisciplinary Discussion on General Intelligence, AGI and Superintelligence Safety and Human Moral | Cognitive Origins of the Concepts of Human Soul and its Immortality | Free Will and How it Originates Cognitively | Animate Being and Soul and the Cognitive Reason for the Believe that "Thinking Machines can't have a Soul and Consciousness" | Technology Making us more Humane | The Egoism of Humanity | And more – Tosh, 2012
* Cognitive Semiotics: A Hierarchy of Meaning Systems Based on Value

* AI Could Wipe Out the Working Class
* Creative Intelligence will be First Surpassed and Blown Away by the Thinking Machines, not the "low-skill" workers whose jobs require agile and quick physical motion and interactions with human-sized and human-shaped environment – Todor 2013

*** Мнения от Русия – 583**

* «Искусственного интеллекта нет!»: Анатолий Вассерман
сергей савельев об искусственном интеллекте
* Сергей Савельев (…)
* Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence

*** Тълкуване на някои често изказвани съвета клишета за ИИ и бъдещето на професиите от обществени лица**

* Съвети за младите, търсещи работа
* Forget coding, Amazon’s AWS CEO wants his own children to master these soft skills in AI age …
* AI revolution fast replacing cognitive abilities with machines
* Министър Вълчев: Изкуствата развиват креативност, науките учат децата да разбират технологиите
(…)

## * Автоматично програмиране чрез статистически модели, големи езикови модели за програмиране ... #programsynthesis #llms #codegeneration – 592

* Large Language Models for Software Engineering: Survey and Open Problems, 11.2023, Angela Fan
* Toward automatic program synthesis, Zohar Manna and R. J. Waldinger, Communications of the ACM
* “Program synthesis,” Foundations and Trends in Programming Languages
* On the naturalness of software
* Personalized Mathematical Word Problem Generation
***** Neurosymbolic programming, S.Chaudhiri et al., 1
* Generalization as Search * Programming by Demonstration: a Machine Learning Approach
* SWE-RL: Advancing LLM Reasoning via Reinforcement Learning on Open Software Evolution
* CWM: An Open-Weights LLM for Research on Code Generation with World Models,
* Виж повече за синтез на програми и др. в приложенията #lazar, #anelia Лазар Анелия

## * Големи езикови модели и машинно обучение – 596

* Нови архитектури преобразители и основополагащите GPT2, BERT, GPT3; Kosmos, Titans, DeepSeek; Microsoft; Emu, Llama-o1… и др
* **Foundations:**

* Attention is all you need, 2017
* OpenAI, Language Models are Unsupervised Multitask Learners, 14.2.2019 (GPT2) Виж GPT2-MEDIUM-BG
* BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,
* Transformer2
***** Transformer-Squared: Self-adaptive LLMs,
* Improving Factuality with Explicit Working Memory
***** LlamaV-o1: Rethinking Step-by-step Visual Reasoning in LLMs
* Virgo: A Preliminary Exploration on Reproducing o1-like MLLM
* CLDG: Contrastive Learning on Dynamic Graphs
* Offline Reinforcement Learning for LLM Multi-Step Reasoning
* AutoGraph: An Automatic Graph Construction Framework based on LLMs for Recommendation
* GraphMERT: Efficient and Scalable Distillation of Reliable Knowledge Graphs from Unstructured Data
* Titans: Learning to Memorize at Test Time
* Can AI Truly Develop a Memory That Adapts Like Ours?
* Exploring Titans: A new architecture equipping LLMs with human-inspired memory that learns and updates itself during test-time.
* LLM Pretraining with Continuous Concepts, Large Concept Model (LCM)
* Large Concept Models: Language Modeling in a Sentence Representation Space
* SONAR: sentence-level multimodal and language-agnostic representations,
* xSIM++: An Improved Proxy to Bitext Mining Performance for Low-Resource Languages
* The FLORES-200 Evaluation Benchmark for Low-Resource and Multilingual Machine Translation
*** Accelerating scientific breakthroughs with an AI co-scientist – 603**
* Towards an AI co-scientist (виж също Y.Bengio)
* Training a Generally Curious Agent
* Idiosyncrasies in Large Language Models
* The FFT Strikes Back: An Efficient Alternative to Self-Attention
* Long Range Arena: A Benchmark for Efficient Transformers
* SWE-Lancer: Can Frontier LLMs Earn $1 Million from Real-World Freelance Software Engineering
* TTS-VAR: A Test-Time Scaling Framework for Visual Auto-Regressive Generation
* Emu3: Next-token prediction is all you need
*** Generative multimodal models are in-context learners**
* Emu: Generative pretraining in multimodality.
* Multi-scale Transformer Language Models,
* DeepSeek-V3 Technical Report **– 606**
* DeepSeek LLM: Scaling Open-Source Language Models with Longtermism
* DeepSeek-Coder: When the Large Language Model Meets Programming - The Rise of Code Intelligence

* 100 Days After DeepSeek-R1: A Survey on Replication Studies and More Directions for Reasoning Language Models
* See also other Chinese models: KIMI K2, Qwen3, Qwen3-Coder, … 31.7.2025; Qwen3-Image-Edit
* Absolute Zero: Reinforced Self-play Reasoning with Zero Data,
* Native Sparse Attention: Hardware-Aligned and Natively Trainable Sparse Attention
* Mixtral of Experts
* Unified Language Model Pre-training for Natural Language Understanding and Generation,
* Microsoft® - Large collection of various NN LLMs as of 22.4.2025
* Updates from Microsoft: 9.2025: MAI-1-preview 15000 x H100…
* RenderFormer: Transformer-based Neural Rendering of Triangle Meshes with Global Illumination,
* SIMLM: Pre-training with Representation Bottleneck for Dense Passage Retrieval
* Language Models are General-Purpose Interfaces,
* CommonGen: A Constrained Text Generation Challenge for Generative Commonsense Reasoning
* EdgeFormer: A Parameter-Efficient Transformer for On-Device Seq2seq Generation
* Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity
* Learning to Ask: Neural Question Generation for Reading Comprehension
* The CoNLL-2014 Shared Task on Grammatical Error Correction
* DeepNet: Scaling Transformers to 1,000 Layers
*** Language models are few-shot learners – GPT3**, 2020

*** A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT – 618**
* Show and Tell: A Neural Image Caption Generator.
* What are the main differences between hard attention and soft attention, and how does each approach influence the training and performance of neural networks?
* What is Attention in ML?
* Neural Machine Translation by Jointly Learning to Align and Translate
* Effective Approaches to Attention-based Neural Machine Translation
* DRAW: A recurrent neural network for image generation
* Memory Networks
* Attention in transformers, step-by-step
* Improving Zero-shot and Few-shot Learning of Language Models via Chain-of-Thought Fine-Tuning
* Scaling Instrution-Finetuned Language Model
* Large Language Model Instruction Following: A Survey of Progresses and Challenges
* Training language models to follow instructions with human feedback

* TencentPretrain: A Scalable and Flexible Toolkit for Pre-training Models of Different Modalities
* Qwen3-Coder: Agentic Coding in the World . 31.7.2025: Qwen3-Image-Edit
*** Някои тенденции** – бележки на Тош **– 627**
* Living intelligence – „жива“ интелигентност, използване на много сензори от мобилни устройства за обучение на невронни модели
* Как щяла да изглежда кариерата на компютърните специалисти след 5 години
* Възходът на общи и интердисциплинарни специалисти. Тесните специалности са обречени: какво искат индустриите 4.0/5.0
* Тош: Глаголищата трябва да бъдат живи – Програмите да се пишат сами
* Todor: Software have to self-develop – Programs can write themselves * Навлизане на големи модели за действия: Large Action Models – LAM
* Сравни с „What’s Wrong With NLP? (…)”, T.Arnaudov, 2009, Part I, Part II. PLAM – personal language action models и CLAM – Corporate LAM
* Principles of mixed-initiative user interfaces, Eric Horvitz, 1999
* MemOS: An Operating System for Memory-Augmented Generation (MAG) in Large Language Models
* Agentic Context Engineering (ACE): Self-Improving LLMs via Evolving Contexts, Not Fine-Tuning
* Agentic Context Engineering: Evolving Contexts for Self-Improving Language Models
* Dynamic Reinforcement Learning for Actors
* Continual Learning via Sparse Memory Finetuning
* The Markovian Thinker
* Principles of mixed-initiative user interfaces, Eric Horvitz, 1999
* **2024-2025 г. – 632**
* Graph Generative Pre-trained Transformer (G2PT): An Auto-Regressive Model Designed to Learn Graph Structures through Next-Token Prediction
* xLSTM: Extended Long Short-Term Memory
***** RWKV: Reinventing RNNs for the transformer era
* The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation
* Agent S2: A Compositional Generalist-Specialist Framework for Computer Use Agents
* Aria-UI: Visual Grounding for GUI Instructions,
* Claude 3.5 Sonnet, Computer Use, 10.2024:
* Computer-use models
* Microsot Copilot Vision :
* Simular Pro
*** Intelligent OS, automated GUI interaction by The Sacred Computer**
*** Практика по LLM – ГЕМ, TTS, ASR и др. … LLM Practice – 637**
* Visualizing transformers and attention
* Курс по големи езикови модели на Хъгинфейс:
* Model Context Protocol (MCP) - Протокол за обмяна на контексти за големи езикови модели

* Виж също: gpt4all, ollama, LangExtract LangChain …
* Prompt Orchestrating Markup Language, Microsoft
* Video RAG for long videos
* Mixture of Agents
* Mixture-of-Agents Enhances Large Language Model Capabilities
* Kyutai's Speech-To-Text and Text-To-Speech models based on the Delayed Streams Modeling framework.,
*** GPT2-MEDIUM-BG, един от най-големите модели за езици, различни от английския, през 2021 г. и обучението му в Google Colab – 640**
* Small Vision-Text models: 230M and 770M
* SmallThinker: A Family of Efficient Large Language Models Natively Trained for Local Deployment
* MoonDream, Vision-Text model, 2B, 0.5B
* Alibaba Qwen3-ASR:
* Text-to-Speech
*** VoXtream: Full-Stream Text-to-Speech with Extremely Low Latency**
* Moshi **…** - end-to end speech-text dialog foundation model – 643

## * Vision-Language Models, Vision-Language Action Models, Foundation Robot Models – 647

* Top 10 Vision Language models of 2025
* Teaching VLMs to Localize Specific Objects from In-context Examples
Open-o3 Video: Grounded Video Reasoning with Explicit Spatio-Temporal
* GigaBrain-0: A World Model-Powered Vision-Language-Action Model,
* SPEAR-1: Robotic Foundation Model via 3D understanding; MoGe …
* Depth Anything V2
* Depth Anything at Any Condition,
* ARGenSeg: Image Segmentation with Autoregressive Image Generation Model
*** Съгласуване, намеса в теглата, защитни „парапети“, AI Alignment**
***** A Library for Understanding and Improving PyTorch Models via Interventions, **…** * WavTokenizer, SOTA discrete
*** Набори от данни #datasets**
* WildVis: Open Source Visualizer for Million-Scale Chat Logs in the Wild
* Графи за знания
* Imbue/Generally Intelligent: Introducing Generally Intelligent
* AI researcher François Chollet founds a new AI lab focused on AGI Understanding of Commonsense Stories
* Byte Latent Transformer (BLT): A Tokenizer-Free Model That Scales Efficiently
* Writer, Palmyra: more creative LLMs: Language Models are Super Mario: Absorbing Abilities from Homologous Models as a Free Lunch Dask: The Python Data Scientist’s Power Tool:
*** Generative AI Still Needs to Prove Its Usefulness** – Gary Marcus. 12.2024
* Отговор на Тош на *„Пораждащият изкуствен интелект трябва да докаже полезността си“* по Гари Маркус – *познавателните* нужди и изкуството, и

полезността им за човеците отвъд основните им *биологични* нужди се създават и са податливи на моди, омръзване, хитруване и др. **– 653**
* Chimera: Accurate retrosynthesis prediction by ensembling models with diverse inductive biases.
*** Общи съвети за употреба на системи с пораждащ ИИ;**
***** Obsidian Web Clipper:
*** Разни новини**
* Introducing perceptein, a protein-based artificial neural network in living cells
* OpenAI o3 - Thinking Fast and Slow
* Talk about Chip Design, Tape-out, Verification, Manufacturing, and Cost
* Sam Altman predicts superintelligence will trigger a 10x surge in scientific AI breakthroughs — each year as revolutionary as a decade * Quantization of LLM weights
* Microscaling Data Formats for Deep Learning,
* We've Already Passed the Superintelligence Event Horizon' **–** Вече живеем в друга реалност – и почти никой не забелязва това
* "Генезис" променя правилата на играта: нов физичен симулатор обучава роботи 430 000 пъти по-бързо от реалността
* Chinese algorithm claimed to boost Nvidia GPU performance by up to 800X for advanced science applications
* Just 2 hours is all it takes for AI agents to replicate your personality with 85% accuracy
* Google's Sergey Brin Urges Workers to the Office 'at Least' Every Weekday
* AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms , 14 May 2025, By AlphaEvolve
*** Агент за научни изследвания и опити – 661**
*** NovelSeek: When Agent Becomes the Scientist -- Building Closed-Loop System from Hypothesis to Verification**
* Knowledge Navigator: LLM-guided Browsing Framework for Exploratory Search in Scientific Literatur,
* **„Понятийни модели"** (според името): * Soft Thinking: Unlocking the Reasoning Potential of LLMs in Continuous Concept Space
* Simulated Reasoning – different ways of processing in LLMs compared to human mathematical reasoning in Chain of Thought tasks;
* Proof or Bluff? Evaluating LLMs on 2025 USA Math Olympiad
* Sam Altman's goal for ChatGPT to remember 'your whole life' is both exciting and disturbing Julie Bort
* Web-Scale Visual Entity Recognition: An LLM-Driven Data Approach
* Whitepaper on AI Agents: A Deep Technical Dive into Agentic RAG, Evaluation Frameworks, and Real-World Architectures By  Sana Hassan, May 6, 2025
* 'AI is already eating its own': Prompt engineering is quickly going extinct
* A first-principles mathematical model integrates the disparate timescales of human learning,

* LLaMA-Omni2 – Speech Encoder and Adapter; Core LLM, Streaming TTS Decoder.
* Open Computer Agent … Open Computer Agent
* E2B Desktop Sandbox for LLMs.
* Mem0: A Scalable Memory Architecture Enabling Persistent, Structured Recall for Long-Term AI Conversations Across Sessions
* Mem0: Building Production-Ready AI Agents with Scalable Long-Term Memory
* Google AI Unveils 601 Real-World Generative AI Use Cases Across Industries
* How much do language models memorize?, – GPT family, ~3.6 bits-per-parameter
* Computer Use sandbox
* Gymnasium, RL, video games, world models
* Exploration-Driven Generative Interactive Environments

## * Проекти за символен УИР – 672

* Development and Architecture of REFPERSYS: A Multi-Threaded REFlective PERsistent SYStem for AI
* Le projet RefPerSys, un successeur potentiel du système Caia de Jacques Pitrat
* INSA: Integrated Neuro-Symbolic Architecture, The Third Wave of AI, a Path to AGI, Peter Voss
* Терминология – бележки „поощрение/наказание“ (RL)
* Multi-Agent frameworks with LLMs
* Meta-Agentic α-AGI Demo
* Системи за препоръчване на съдържание
* **A Comprehensive Review Of Recommender Systems: Transitioning From Theory To Practice**, 13.7.2024

## * Модели и платформи за пораждане на изображения … 674

* Модели и платформи за пораждане на видео и физически верни модели на света или симулации по словесна подкана
* **Diffusion Models Are Real-Time Game Engines**
* **On the inefficiency of generating games this way with neural models, Todor Arnaudov**, 24.8.2025 – 677
* **Survey on GPU performance – comparisons …**
  * Introducing NVFP4 for Efficient and Accurate Low-Precision Inference
  * Nvidia researchers unlock 4-bit LLM training that matches 8-bit performance,
  * Pretraining Large Language Models with NVFP4

## * Чатботове и агенти за пораждане на съдържание: текст, изображения, видео, музика, програми и най-мощни езикови модели: безплатни и платени услуги – 688

* Free and paid AI services for chatbots, agents, content and media generation: text, images, dieo, music, software and the most powerful LLMs
* List of free LLM services …
* Notes on Comet and Atlas AI powered browsers; Recall, Rewind AI; Todor’s Research Accelerator/ACS project
* Music and Poetry

* Silicon Dream, Todor Arnaudov & Suno, 7.6.2025:
* Forever Young Hackers of the Universe, Todor Arnaudov & Suno, 7.6.2025
* Can you run big models locally on a PC with little GPU RAM?
***** Old Visionary Designs
* Knowlede Navigator – A Project demonstration from 1987 for future computing: tablets, intelligent assistants etc., demo videos, created by Apple
* Latest Generative AI, GPT News: OpenAI, Robotics, Nuromorphic
* GPT-OSS: Open weights model: 20B (16 GB GPU), 120B (80 GB GPU),
* Unitree humanoid robot for $5900,
* Video: China's humanoid robot 'Oli' learns to pick tennis balls autonomously
* GPT5 & Claude 4.1
* From plateau predictions to buggy rollouts — Bill Gates' GPT-5 skepticism looks strangely accuratep
* DeepSeekV3.1 – 685B …
* Alibaba Ovis 2.5 Multimodal, 17.8.2025
* The road to artificial general intelligence: Understanding the evolving compute landscape of tomorrow
* Google DeepMind Genie 3 – generative world model, 8.2025 (announced
* Genie 3: The World Becomes Playable (DeepMind), AI Explained, 5.8.2025
* NVIDIA AI Just Released the Largest Open-Source Speech AI Dataset and State-of-the-Art Models for European Languages
***** Canary 1B v2: Multitask Speech Transcription and Translation Model **–** Supported Languages: Bulgarian (bg),
* Is Chain-of-Thought Reasoning of LLMs a Mirage? A Data Distribution Lens
* DINOv3, Oriane Siméoni Qwen3-Max- … 1T MoE; Claude 4.5 – 10.2025 …
* Forever Young Hackers of the Universe, Todor Arnaudov & Suno, 7.6.2025
* Can you run big models locally on a PC with little GPU RAM?
* Code generation and the shifting value of software
* What's next for AI? Researchers from NVIDIA, Apple, Google & Stanford envision the next leap
* Video models are zero-shot learners and reasoners – Tosh: Compare to the Bulgarian predictions on interdisciplinarity and general learning from the early 2000s

*** Някои текущи „последни новини“ – 705**

* What's next for AI? Researchers from NVIDIA, Apple, Google & Stanford envision the next leap
* CACHE-TO-CACHE: DIRECT SEMANTIC COMMUNICATION BETWEEN LARGE LANGUAGE MODELS
* Code generation and the shifting value of software
* The End of Manual Decoding: Towards Truly End-to-End Language Models,
* Kimi-2 Thinking
* DeepSeek drops open-source model that compresses text 10× through images
* DeepSeek-OCR: Contexts Optical Compression,
* Context-Optical Compression
* Glyph: Scaling Context Windows via Visual-Text Compression

* Fox: Focus Anywhere for Fine-grained Multi-page Document Understanding, Chenglong
* Thinking with Video: Video Generation as a Promising Multimodal Reasoning Paradigm…
* Nested Learning: The Illusion of Deep Learning Architectures
* DS-STAR: Data Science Agent via Iterative Planning and Verification
* Thinking with Video: Video Generation as a Promising Multimodal Reasoning Paradigm

## * Вместо „временно заключение“ – 712

* Денарио: мулти-агентна система за автоматични научни изследвания, пораждаща научни статии
* Meet Denario, the AI ‘research assistant’ that is already getting its own papers published
* Measuring AI Ability to Complete Long Tasks; .. HCAST, BigBench, RE-Bench
* Измерване на способността на ИИ да се справят със задачи изискващи продължителна работа от човек
* AI Agents are Terrible Freelance Workers, Wired, Business, 29.10.2025
* Remote Labor Index: Measuring AI Automation of Remote Work
* Todor discusses the unreliability and fluidity of the “economically valuable tasks”
* MIT researchers propose a new model for legible, modular software – for LLMs: What You See Is What It Does: A Structural Pattern for Legible Software

## * Self-Modifying & Self-Improving Machines

* Huxley-Gödel Machine: Human-Level Coding Agent Development by an Approximation of the Optimal Self-Improving Machine
* **Speculations concerning the first ultraintelligent machine** – I.Good, 1965
* Evolutionary Principles in Self-Referential Learning …
* Multi-agent learning with the success-story algorithm
* Gödel machines: self-referential universal problem solvers making provably optimal self-improvements

## * Various Terms for Artificial General Intelligence by Todor Arnaudov

*** Различни други термини за Универсален изкуствен разум от Тош***

VLESI, SIGI, VEI, SDM, SCM, DMMI, EMI, EMDI, EMIL – 718

## * Томове и приложения на „Пророците на Мислещите Машини“ – 720

# ЛИСТОВЕ

## по всичко

*** Отговор на възможни критики, че творбите от „Свещеният сметач“, „Изкуствен разум“ (Artificial Mind), „Разумир“, „Пророците на мислещите машини“ и др. не са научни публикации, нямат научна стойност и не се броят за приноси, защото не са публикувани на конференции, в научни списания – по-точно в „признати“, „рецензирани“, „индексирани“ и пр.**

Тодор Арнаудов, 2024 г. и редакция 8.2025 и 10.2025

Лабораторията, списание, издателство, „изследователско-творческо дружество“ и „виртуален институт“ ***Свещеният сметач****: мислещи машини, творчество и развитие на човека,* както и неговият създател, са интердисциплинарни и всестранни – съчетават многостранно творчество и изкуство с точни и природни науки, техника, информатика, философия, езикознание, игра; сериозно и смешно, сатирично; хуманитарни и обществени науки, история, публицистика; спорт и здравословен начин на живот – програмата „да бъдеш вечно млад“ и др. и при разглеждането им превключват и свързват различни подобласти и теми от всяка от тях. Степените на преливане са разнообразни: от специализирани работи до свръхвсестранни и единствени по рода си произведения, разработки, изследвания, които съчетават *всичко* по необичаен и нов начин и съответно изискват и от читателя или оценителя да имат подобни всестранни знания, умения, сетива и любознателност.

От друга страна *„научно“* в тесен смисъл и в голяма част от „научната“ работа всъщност често значи педантично следване на установените правила и *подчинение на авторитетите*, *за да се впишеш в колектива и постепенно да се издигнеш в стълбата на властта,* и често се състоят от нефункционални формалности и изисквания, например за форматиране, за *място на публикация,* за „узаконена“ рецензия и „благословия“ от подходящи авторитети; за заплащане на такса на списание или конференция и за пътувания до

мястото и пр. и други *непроизводителни* разходи, чрез които би могла да се върши градивна и творческа работа, ако са вложени например в техника или за издръжка, наставничество и стипендии на „непослушни“ гении[9] в специални звена, както е описано в оригиналната Първа съвременна стратегия за развитие чрез изкуствен интелект през 2003 г. и в повестта „Истината“ половин година по-рано.

Мястото на публикуване в последните десетилетия все повече намалява значението си с т.нар. „демократизиране“ на науката, като за пример се дават „предпечатите“, все още непреминали реценция публикации, или на английски с шеговит превод: непреминали „другарски съд“ („peer review”); такова например е хранилището Arxiv.org, където някои статии получават стотици цитати преди официалните им „публикации“ на конференции или в научни списание след това; напр. статията за *Преобразителя* от 2017 г. *(„Attention is all you need“)*.

**Майкъл Левин,** роден през 1969 г., в беседи през 2023 разкрива, че ще се създаде **блог**, в който да може **по-спокойно да публикува някои от работите си**, защото били **прекалено интердисциплинарни**, твърде дълги и пр. и преди бил срещал трудности с обнародването им на подходящо място: блогът се нарича *„Форми на живота, форми на ума“ (Forms of Life, Forms of Mind):* https://thoughtforms.life/.

**Сп. *"Свещеният сметач"*** беше *повече* от такъв "блог" – списание, **още през 2001 г**., виж напр. *"Човекът и Мислещата машина: Анализ на възможността да се създаде мислеща машина и някои недостатъци на човека и органичната материя пред нея…"* и сравни с трудове на Левин **20 години по-късно**. Относно "нетрадиционните" си идеи за одушевеността във Вселената (агентността) и в най-малки мащаби и необичайна телесност, за които публикува в статиите си с колеги като Крис Фийлдс, активно в последните години, като "Basal Cognition", "Cognitive Boundary of Self", "Cognitive Light Cone", “Technological Approach to Mind Everywhere”, “Self-improvizing memory …” и др., които напоследък започват да се приемат по-широко, Левин е признавал в участия през 2023 г., както и в третото си гостуване в подкаста от 1/2024 г. *"Demistify Sci",* че *винаги* бил мислил така, но преди ***не било време да се говори за това***, защото е било в разрез с общоприетото и би било ***опасно за кариерата*** – сравни с разговорите за универсалния изкуствен разум през 2000-те и началото на 2010 г., **и дори**

[9] Виж: *1. „Истината“, Тош, 2002; 2.* „Как бих инвестирал един милион с най-голяма полза за развитието на страната?“, 2003; 3.* „Първата съвременна стратегия за развитие чрез изкуствен интелект е публикувана от 18-годишен българин през 2003 г. и повторена и изпълнена от целия свят 15-20 години по-късно: …“, 2025

**2015 г.**, в признанията на Демис Хасабис, Шейн Лег (съосновaтели на една от най-успешните компании по УИР/AGI “Google DeepMind”), Сам Алтман (съосновател и изпълнителен директор на другата най-успешна компания – OpenAI) и Бен Гьорцел, дадени в основния том на *„Пророците на мислещите машини: …“, 2025* и в *„Първата съвременна стратегия за развитие чрез изкуствен интелект…“, 2025*.

Бен Гьорцел е един от пионерите на движението и на първите научни конференции и книги извън България* и разпространител на термина „Artificial General Intelligence”; българското движение на Тош и аналогичните термини като *мислеща машина, изкуствен разум* от „Свещеният сметач“ са **преди** групата му и около същото време в най-ранния им период.

Начинът на мислене на Левин и конкретни мисли, които са били „неприлични“ и „опасни“ за изказване до преди няколко години, са изразени открито в "Свещеният сметач" **15-20 години по-рано**, без сметки от 17-годишно момче, в „лабораторията“ му детска-стая, до двуетажното легло, подкрепен само от семейството му и от морално остарялата му техника: Правец-8М и после PC Penitium 90 MHz с 16 и малко преди споменатата работа: 32 MB RAM; VGA монитор.

*Сметачът* е едно от първите списания от 21-ви век по тези теми и по Универсален изкуствен разум, ако не и първото; интердисциплинарно и своеобразно, със съдържание и работа, която 15-20 и дори 25 години по-късно продължава да бъде съвременна, и съдържанието ѝ да се преоткрива и повтаря като ново, свежо, оригинално; това явление е разяснено и в приложението за новата теория на разума на Майкъл Бенет: *„Stack Theory is yet another Fork of Theory of Universe and Mind”* и в *„Първата съвременна стратегия за развитие чрез изкуствен интелект…“.*

Следователно **имало ли е достатъчно „правоспособни“** и квалифицирани колеги, които да рецензират подобно съдържание *тогава* – не само в България, а и в *света?* Донякъде подходящи може да са били едва няколко други „пророци“, работещи в нишови области, които не бяха под светлините на прожекторите[10].

Всестранността и интердисциплинарността също са пречка за тяснospециализираните и по-тясно фокусираните личности, каквито са повечето „сериозни“ и „професионални“ учени, които нямат необходимата ширина на интересите и знанията, и подходящи способности и начин на

[10] Аз самият, явявайки се сред „пионерите“ в теорията на УИР, открих работата на Хутер и Шмидхубер чак през 2009 г., а К.Фристън – в края на 2018 г. от интервю за списание “Wired“.

мислене, за да разбират или обърнат внимание на подобни публикации. Обобщенията, заключенията, предвижданията и програмата от тези работи обаче впоследствие станаха все по-ясни и стават дори *очевидни* и се доказват и повтарят по безспорен начин и в теорията, и в практиката; преоткриват се от, и във многобройни уж „нови“ и „оригинални“ научни и философски школи, което служи като емпирично доказателство за верността на първопроходните работи, каквато беше Теория на Разума и Вселената.

„Фантазиите“ и „философиите“ се превръщат в „официална наука“ и за „основното течение“ (mainstream), но *оригиналните автори и пионери не са споменават* и нямат право на претенции със задна дата.

Повторенията, разпространението и навлизането в практиката обаче са равнозначни на потвърждения и положителни цитати и рецензии, а отдалечеността във времето би трябвало да служи като подсилване на оригиналността и приноса на по-ранните публикации – ако научната общност спазва, или *спазваше* тази *логична* етика, където уж е важен приоритетът във времето, и предишната работа трябва да се посочва.

Този въпрос е разгледан с повече конкретни аргументи и списъци с приноси в споменатата монография *„Първата съвременна стратегия за развитие чрез изкуствен интелект …“;* в най-новото приложение „*Stack Theory is yet Another Fork of Theory of Universe and Mind”,* където са дадени и многобройни рецензии на откъс от работа от 2001 г. от страна на най-мощните големи езикови модели в момента като Kimi-2, Qwen-3, ChatGPT-5, Claude 4, Gemini 2.5 и др., и дори BgGPT-27B. Машините недвусмислено и конкретно отговарят на въпроса по какъв начин, с колко време и кои учени е изпреварил и предвидил кратък откъс от „Човекът и Мислещата Машина“, публикуван през декември 2001 г., и периодът достига 15-20 години.

Огромен брой съвпадения са описани и обяснени в **Основния том** на *Пророците на мислещите машини* **–** из цялата книга, и например в началото в списъка с учени и школи. В том *Ирина* са посочени значителни съвпадения между описанията и идеи на Йоша Бах и др., изказани десетилетия по-късно; в „*Първата стратегия …“* и допълнението *„Институти и стратегии за изкуствен интелект на световно ниво …“* се доказва, че целият свят преоткрива и изпълнява програмата на *Свещеният сметач*, включително институтът в София „ИНСАЙТ“, чиито изказвания, предложения, уверения „за стратегическото значение“, „развитието на страната“, „подкрепата на талантите“; суперкомпютърните центрове, и дори конкретни теми на изследвания и

разработки в изкуствения интелект са буквални копия и реализации на *„българските пророчества*“, но с чуждестранен привкус и стотици пъти завишени финансови изисквания. Многобройни подобия на по-късни публикации и теории с ТРИВ са дадени в приложение *Листове по всичко,* заедно с тълкуването и разясняването на множеството публикации; в том *„Фантастика. Футурология. Кибернетика. Развитие на човека“,* разглеждам работата на Майкъл Левин *„Самоимпровизираща памет:* …“ от 2024 г. която преоткрива мотиви от „Вселена и Разум“.

*Първата съвременна стратегия…* споменава уравнението или коефициента на относителната ефективност, която се превръща в *Сингулярност на Тош*, когато стойността на коефициента клони към нула. ***Сингулярността на Тош*** е положение или състояние на вселена, при което *деец* с нищожни средства и причинностна-мощност; управляващо-причиняващо устройство с ниска енергийна, „обемна“, финансова, човешка и пр. сила, клонящи към нула – например един-единствен човек или малка група с нищожна или морално остаряла изчислителна техника и с нулеви или почти нулеви специални целеви разходи, да речем с обикновени лични сметачи – създава или е пряко и тясно свързана със събития около сътворяването и възникването на обекти и случването на извънредни събития и явления *„с ниска вероятност, за преживяващия ги разум“*[11], които изискват огромни или *„клонящи към безкрайност“* ресурси от други управляващо-причиняващи устройства. **Накратко:** когато се окаже, че един *доброволец пехотинец* или взвод може да замени цяла добре въоръжена модерна армия.

Сингулярността е свързана с парадокси, аномалии и *пародии*. Много такива вече се случиха и се случват. *„Очаквайте продължение“.*

10.10.2025 г.

* https://twenkid.com/agi/Purvata_Strategiya_UIR_AGI_2003_Arnaudov_SIGI-2025_31-3-2025.pdf
* (…) Виж препратки в края на книгата и на уеб страниците на SIGI-2025 и *Свещеният сметач.*
* Stack Theory is yet another Fork of Theory of Universe and Mind, Todor Arnaudov – Tosh, 2025 https://twenkid.com/agi/Stack-Theory-is-Fork-of-Theory-of-Universe-and-Mind-13-9-2025.pdf – see p.16 - …, p.77-79, p.132 - … for the LLM reviews. See also the references to chats regarding the matches between “Analysis of the meaning of a sentence…”, 2004 and the other mentioned theory of meaning for AGI from 2024 which rediscovers many corresponding ideas.

[11] Виж „Вселена и Разум 4“, 2004 г.

# Обобщение на някои от съвпаденията:

*** Теория на Разума и Вселената от 2001-2004 предвиди принципите и насоката на развитието на изкуствения интелект и универсалния изкуствен разум, и в интердисциплинарен контекст нейни основни заключения се преоткриват и потвърждават от школата на принципа на свободната енергия и извод чрез действие на Карл Фристън и др. (~2006-2009+), Анди Кларк; „осъществяването на уместност“ на Джон Фервеке и др. (2012-; 2024-); няколко теореми на Дейвид Уолпърт във връзка с предсказуемостта (2007;2018); теорията за УИР на Майкъл Бенет (~2023-2025); теорията за създаването на смисъл в универсалния изкуствен разум на Торисон и Талави (2024); множество интердисциплинарни теории обединяващи живото, неживото и ума на Майкъл Левин и др. (2020-те); подобни кибернетични идеи на Йоша Бах (~2019?+); идея на Йошуа Бенджио „consciousness prior” (2017-2018) в мета-обучението и ученето на причините; теорията на Ян Льокан за пътя към автономен ИИ (2022); „пет основни принципа на роботиката на развитието“ на Александър Стойчев (2006-); диалог между мислещата машина и човека от Томас Мецингер през 2009; мярката на Франсоа Шоле за машинна интелигентност (2019) и много други**

* Relevance Realization rediscovers motives from Theory of Universe and Mind, while narrowing the meaning of computation and offering contrasting interpretations of some points, as well as limiting the scope of applicability of the principles only to living organisms (2001-2004 vs 2012-..2024)

* Several theorems and conclusions about the limits of prediction by David Wolpert rediscover postulates from Theory of Universe and Mind (2001-2004 vs 2007-2008, 2018) however they start with confused premises and miss the development and multi-scale nature of the agents which are inside the universe

**See also:**

* Stack Theory is yet another Fork of Theory of Universe and Mind, T.Arnaudov, 2025 – Michael T. Bennett’s AGI theory presents as novel many crucial points which were clearly expressed by 2001-2002.

* Thorisson and Talevi’s theory of meaning in AGI from 2024 rediscovers the principles from Theory of Universe and Mind and in particular exercise significant correspondences to Arnaudov’s paper “Analysis of the meaning of a sentence based on the knowledge base of an operational thinking machine. Reflections about the artificial thought”, 2004 – see short mentions in “Stack Theory is yet…” and the LLM’s comparisons.

* Free Energy Principle/Active Inference is also a fork of TUM in some of the core premises.

* Michael Levin’s et al. school of thought rediscovers and repeats or interprets many ideas from TUM 20 years later – in TAME, ”care”, self-improvising memory ...

* Joscha Bach’s talks from late 2010s and 2020s repeat many principles and reasoning from TUM published between 2001-2004

* Yoshua Bengio’s “Consciousness Prior” rediscovers basic ideas from TUM, in particular the example about different resolution of causality-control and perception repeats a thought experiment from Universe and Mind 4.

* Core principles of Yann Lecun’s “Path towards Autonomous AI…” are rediscovery of the ideas from TUM

* Thomas Metzinger’s “Ego Tunnel”’s dialog with the thinking machine is a rediscovery of the stance in “Man And Thinking Machine …”, 2001 and “The Truth”, 2002 (vs 2009)

* Etc.

* See this volume, the main volume of The Prophets of the Thinking Machines and the appendices.

# Wolpert Theorem about mutual unpredictability and the impossiblity of subuniverses to predict with highest resolution of causality-control are rediscoveries of concepts from Theory of Universe and Mind – entangled with unnecessary mathematical notations and unsatisfiable premises

Todor Arnaudov, 10.2025, The Sacred Computer, SIGI-2025

A response to: *Physical limits of inference, David H. Wolpert, 2007/2008**[12] etc.

Note that the term "device" and "control" (and control unit; causality-control unit) are used in TUM as well (устройство, управляващо устройство). Bold and [1], [2] is added by T.Arnaudov for the notes.
CCU – Causality-control unit; VU – Virtual Universe
RCCP – Resolution of causality-control and perception; POV – point of view
See TUM and its principles and ideas and compare: both classic works 2001-2004 and the ones from SIGI-2025. Some resources and related discussions are available in:
* https://twenkid.com/agi/Stack-Theory-is-Fork-of-Theory-of-Universe-and-Mind-13-9-2025.pdf * See resources about TUM in other pages of volume "Listove"/Reflections on Everything and after this text.

Start the comparison with the examples p.4-7:
**D.Wolpert, 2007/2008: p. 4-7:** *"**Example 1:** a general-purpose observation device, capable of observing different aspects of the universe ... can ask questions: "Does S (t2) = K?" **.. 'yes' or 'no'** - a binary function of u. F ..."(1)* either a "scientist" or *"inanimate" pieces of hardware; **Example 3:** .. **a general-purpose prediction device**, capable of **correctly predicting different aspects of the universe's future.** The **prediction device** = a physical computer, and a scientist who programs that computer to make the **prediction** and interprets the computer's output as that **prediction**. .. the scientist initializes it at some time t1 < t2 to contain some information concerning the state of the universe and to run a simulation of the dynamics of the universe that uses that information… **to "predict S (t2)"** only means the scientist **must program the computer appropriately**; the scientist must force the universe to have a worldline u such that χ(u) = c, and that must in turn cause ζ(u) to accurately give Γ(u) ..**Example 4:** .. a system that .. serve(s) as **a general-purpose recording and recollection device,** capable of **correctly recording different aspects of the universe and recalling them at a later time** ..."*

[12] * Physical limits of inference, David H. Wolpert, MS 269-1, NASA Ames Research Center, Moffett Field, CA 94035, USA, https://arxiv.org/pdf/0708.1362 [Submitted on 10 Aug 2007 (v1), last revised 23 Oct 2008 (this version, v2)]
Cited previous work with related conclusions:
* [46] D. MacKay, On the logical indeterminacy of a free choice, Mind, New Series 69 (273) (1960) 31–40. 42
* [47] K. Popper, The impossibility of self-prediction, in: The Open Universe: From the Postscript to the Logic of Scientific Discovery, Routledge, 1988, p. 68

These definitions seem to observe the scientists as if they were not constituent parts of the universe, which are built from the same material and correlated and affected by it – see below. Compare to [Arnaudov, 2002], the three basic ways for predicting in the Universe computer by the subuniverses (CCUs), the hierarchical simulators of virtual universes etc. For example, humans believe they have "free will" and they do "*what they want"*, but what are the states of their *cells, or atoms?* How much of *that* do they know or control? This is addressed in TUM: RCCP, virtual CCU etc. **Wolpert, p.12:** "*Prop. 1(ii) means that Laplace was wrong: even if the universe were a giant clock, he would not have been able to reliably predict the universe's future state before it occurred."*

**TUM, 2001-2004:** No subuniverse/computer is faster than the Universe Computer as a whole, neither it could have more or enough memory – everything in Universe is a kind of memory. Only the Universe computes the future with the highest RCCP and everywhere in its entire memory, however it does only for the next step. The higher-level virtual universes: compress, have memory of previous predictions and history; can focus on specific space-time regions of memory, which are smaller than the whole universe; can work at lower RCCP etc. that allows them to predict the future more steps ahead, even though at lower RCCP, confidence etc. before it really happens.

However, the higher-level universes are able to predict the future more steps ahead, because of the *redundancy* at the lower-level universes, which allows them "not to care" about all details and about *being unable* to predict it at the highest RCCP and thus to *"truly"* cause-control (see below) and not just to do it "virtually", which is sufficient for their *target* RCCP. Yet sooner or later this leads to errors and to "crash", though, and a given CCU's prediction would change to something else or "die".

Only the Universe as a whole, or from a higher level universe's perspective – the universe at the currently accepted as the lowest level, called also "physical" – is supposed to "*never halt"*or *crash* – from the POV, sensations and predictions of the encompassed higher level ones.

The "crash" here is not used in the sense of Turing Machine halt, ending of a program (a signal for completion); a system, device, agent, CCU crashes when it faces an illegal "opcode"/state and it can't deal with the exception; a virtual *universe never crashes* if at that level *all possible states* are implemented and covered, thus any possible instructions, condition, configurations etc. are acceptable, nothing is a *mistake* and "*fatal*" for the universe, therefore it never enters *undefined states*.

When a virtual universe at a given level has *incomplete implementation* – not all possible states at a lower level are defined *at the higher* - it can malfunction, be "hacked", deteriorate etc. An evaluator-observer detects that condition as changes in its properties, behavior, manifestations, states etc. which lead to recognition of the features of a lower level CCU, virtual universe etc. The errors lead to *collapse* into a new configuration, where the *highest* currently evaluated-observed version of the

CCU/VU is some of the *constituent* lower-level causality-control units, or a newly formed system that is created by the interaction of the subuniverses in the new more "free" and malleable state, which was previously locked and forbidden by the higher level CCU; unlike the collapsed CCU/VU, the new system or the lower level CCUs/VUs are capable to take care of the *"undefined conditions/states"* from the POV of the "dead" *former* higher level system.

**Wolpert, p.23:** ***"6.2. Philosophical implications:*** *...general restrictions that must relate any devices in such a universe, regardless of the detailed nature of the laws of that universe. ... Say we have a device C in a reality that is* ***outside distinguishable.*** *Such a device* ***can be viewed as having "free will"****, in that* ***the way the other devices are set up does not restrict how C can be set up*** *[1]. Under this interpretation, Thm. 1 means that* ***if two devices both have free will, then they cannot predict / recall / observe each other with guaranteed complete accuracy. A reality can have at most one of its devices that has free will and can predict / recall / observe the other devices in that reality with guaranteed complete accuracy.[2:]*** *(Similar conclusions hold for whether the devices can "****control****" each other; see Sec. 7 below.) Thm. 3 then goes further and considers devices that can* ***emulate*** *each other. It shows that independent of concerns of free will, no* ***two devices can unerringly emulate each other.*** *(In other words,* ***no reality can have more than one universal device.****) ... these results could be called a "****monotheism theorem".***

**Todor:** [1]: However, in reality, in Universes like ours **this is impossible** *and it is not so*. The process of the *genesis* of the higher level virtual universes of the two or any "devices" **is common,** everything is produced as an unfolding from earlier states where the causal and spatial distance between parts of the current devices, subuniverses etc. were smaller and even if they now appear different and uncorrelated, at some point of time their earlier precursor were more similar, closer in distance and more related, possibly in some point of time or way of observing-evaluating they were the same entity – the simplest way to reach to that conclusion is that they are part of the same Universe.

In addition, as the *highest possible RCCP at the mother universe* is considered, the devices, subuniverses, parts of the Universes have influenced or interacted with each other not only in **their earlier states of genesis, but forever and in any moment** - as gravity forces etc., and other possible forces, "quantum entanglements", and continuing the first premise – indirectly, because by their existance they force or cause other dependencies and conditions which influence the possibilities of the other device, unit, subuniverse, CCU, agent, entity; "Markov blanket", "thing" – whatever we call it, define it, address it etc.

Qualititavely, all devices and subuniverses are undoubtedly correlated, they do depend on each other, they are in the same memory of the same Unvierse Computer, thus "they" *do influnce: 1)* each other 2) what they could be and what they are at the moment; 3) what their behavior or "free will" could be in the future.

Therefore the premise, *as given*, is **wrong.**

**The *conclusion alone,*** that only the Universe as a whole can predict its future at the highest possible resolution of causality-control and perception, **is correct**, however **without the premise.**

There is a limit of predictability for a virtual universes/CCU, horizons either in time, space, precision etc., also about "self" – in TUM: *the eye can't see itself without a mirror*.

The higher levels work at lower resolution and are slower; they predict, but at either lower resolution/precision or at lower range, and overall: they are parts of the Universe and they both depend on the other parts in the past, present and the future; and they influence the other parts *now* and they *can* influence them in the future etc. This is stated in TUM, even in the early treatises from 2002.

* The device *can be viewed as … being set up independently* – yes, *it can be **viewed*** by an *evaluator* who does not search or cannot access the representation deep and "long" enough back in the past, in abstract representations at a lower RCCP; however how could the devices be supposed to predict at the highest possible resolution: they are *virtual* causality-control units, with the only "real" being the Universe as a whole. For example, the requirement is similar to asking an "unrestricted" video at 160x120, which can represent completely random pixels in each coordinate at each frame, to be capable to represent precisely a 1920x1080 video with the same maximum pixel-level resolution and range, *without* "tricks", no redundancy in the enlarged cavas, no additional interpreter with memory, no "cheats" with simple images, but full spectrum.

This is obvious also by the differences of the *scales*, without any special formulas or theorems. Every subuniverse, even if it is "truly" independently defined and segmented, is only a *part* of the whole. How can it contain and compute the *whole* and know what is *outside* of its bounds and the range where it has access to?

Again, it *might*, but in special circumstances – if its predictions and transformations are actually driven by the universe as a whole or by a bigger encompassing subuniverse, and the predictions of the smaller included subuniverse reflect/map other phenomena elsewhere on other levels or locations, which are repeated in the smaller one; in these cases the encoding and predictions happen because they are "preprogrammed" and synchronized by an inter-level driver and matches, common genesis and connections of subuniverses at different levels of the Universe Computer. That case *was addressed* in TUM, e.g. in Universe and Mind 4, where the "*origin of the non-goal-directed matches, for the experiencing mind, between different levels of the Universe Computer*" was introduced as early as a sub-title.

"***7. Control devices … "***

***The Universe*** **controls you and everything** at the highest RCCP – that's explained in TUM. Its first and other titles are: "Conception about the *Universal Predetermination"* and *"The Universe Computer"*.

**Wolpert: p.26:** *"In particular,* ***no device can control itself,*** *and no two distinguishable devices can control each other. In fact we can make the following stronger statement, which essentially states that if two partitions are refinements of each another, they must be identical: Theorem 5: If two devices C1 and C2 simultaneously semi-control one another's setup functions, then the partitions induced by X1 and X2 are identical."*

This is a restatement of ***Todor Arnaudov, TUM, 2001-2004:*** Only the Universe has "real" control, at the highest possible RCCP. The control of all other virtual universes = subuniverses = causality-control-units = machines = programs is only *virtual,* which means at a lower RCCP.

A true control is defined as: when one control unit records/writes in the memory of another control unit, a "slave", a controlled one, with a resolution of causality-control and perception that is *maximal* for the machine language of the universe where the *controlled CCU* is defined. In our universe at the lowest level, or at a "lower" or what we know about it, that is quarks, electrons, protons, neutrons etc.

Devices – the subuniverses – *believe* and "want" that they "*truly control*" and the future happens as they wish, and may have pseudo or "self-delusional" complete control in their subuniverses, when they set a selected level and precision, slice, area subuniverse as "physics", a limit – the lowest level – with a given accepted RCCP which is assumed to be "enough" for "complete causality-control". Then they may assume that anything that is as precise within a given range/prediction horizon in time etc. is "enough" and counts as "complete control" *for them.*

That is addressed many times in TUM since 2001-2002: the astronomically lower magnitude, compared to the bits required to describe the actual states of the system in the machine language of the universe – i.e. it is the *lowest level* that controls the system, causes the changes, and *not the will of the agent*, when represented in these miniscule bits; see also Ashby's "requisite variety"; other examples in UnM-3, which were a vision about the generative models for images and video – that the wishes of the characters in fairy tales to magic fish and genies are just a few letters or words, but they may lead to the generation of whole castles, i.e. these generators should have an enormous memory bandwidth. That corresponds to the prompts, sent to generative models, and the ratio between their complexity and the complexity of the representation of the generative model, e.g. GBs vs a few bytes.

UnM-4 in 2004 addresses this problem since the beginning with a revisit and refinements of the basic concepts by illustrating the levels of virtual universes with different RCCP and the virtuality of human conscious causality-control and "free will"

– humans use to believe, that as they can want to do something and they do it, that is a good enough proof for their freedom, free will etc., however they do not notice or admit that they do at a very low RCCP, usually described in words: verbs, nouns, adjectives etc. – *"I want to throw the coin behind the sofa"* or "on the floor" or "over there" etc. variants. In the real world, or in the "more real world", there are exact coordinates of exact particles and exact trajectories for each fraction of the second.

The idea of this example, almost literally, was rediscovered and repeated by Yoshua Bengio in his papers and lectures on "Consciousness prior" – in one video he demonstrates practically the same experiment, suggested in my work published 13-14 years earlier[13]: he throws a small object and explains that while we cannot predict the exact trajectory, we can predict that it will fall on the ground etc. The thought experiment in *"Universe and Mind 4"* is more elaborate and includes linguistic variants, which demonstrate that indeed natural language in that form can be used as a "world model" for prediction, however one with a lower RCCP, therefore the textual representation alone, without more precise "physical" records and formats, is insufficient for complete operation in the real world at a higher resolution of causality-control and perception[14]. This is also demonstrated in "AGI Digest" - "Chairs, Buildings, Caricatures …" - my letters to the AGI List from 2012, regarding the generalizations and that the lower level sensory concepts are not well suited for precise description with linguistic representation, or when they are represented like that they transform and degenerate into code and command languages, which list the format of the images and the content of the data points*.

* See main volume of "The Prophets of the Thinking Machines..."

---

[13] A reference from the main volume of *The Prophets of the Thinking Machines, 2025;* In part "**Short chronology"** at page ~1010 (in the ed. In 6.11.2025) – "Предпоставката за съзнание и причинността: съвпадения между примери и обяснения на Йошуа Бенджио от 2018 година и на Тодор от "Теория на Разума и Вселената" от 2003-2004 г"*;* citing: "From Deep Learning of Disentangled Representations to Higher-level Cognition 52 735 показвания • 9.02.2018 г. Препраща към: "The Consciousness Prior", Bengio 2017: https://arxiv.org/abs/1709.08568 https://youtu.be/Yr1mOzC93xs?t=2486 Views now (6.11.2025): 74 941 (the throw of the object) Introduction a bit earlier: https://youtu.be/Yr1mOzC93xs?t=2354 **Yoshua Bengio, 2018:** "Something else, which I think is fairly new and connects with classical work with AI and knowledge representation and symbolic AI … "; Think about your thoughts; your conscious thoughts … there's something that comes to your mind… and it's very low dimensional … That's my claim … You can convert that into sounds, a sentence … Not everything is like this, I can do visual imagery and that's hard to verbalize; … but even if it's visual, it's very low dimensional and it concerns a very few aspects of the world …" (Then he demonstrates the predictability with an experiment with throwing a piece of paper etc.) Compare the above with the first points of "Universe and Mind 4", which revisits and refines the expression of important basic ideas about prediction, compression and levels of abstraction, introduced in the earlier parts.This work was written and edited immediately after the publication of Universe and Mind 3 in the summer of 2003, and published in April 2004:
* „Вселена и Разум 4", сп. „Свещеният сметач", бр. 29., 4.2004
https://research.twenkid.com/agi/2010/en/Todor_Arnaudov_Theory_of_Universe_and_Mind_4.pdf
https://www.oocities.org/eimworld/4/29/pred4.htm

[14] This idea is an embodiment of the "information bottleneck" method.

Therefore, the CCU controls and causes only *virtually* and their models and representations are *incomplete or erroneus;* for that reason, the errors accumulate and sooner or later the *horizon* of their *allowed-operation-with-errors* ultimately finishes with an "*exception*", which cannot be "caught", which *destroys* the particular virtual machine = virtual universe; then the lower level VUs become more visible for the other VUs = evaluator-observers. When the higher-level machine crashes, i.e. it no longer controls "truly" *enough* even by its standards and the standards of the other competing machines, its niche is occupied by either the next machine down the hierarchy, or by another machine from a similar level, which was not yet located at that area, but arrives and takes over the space. That happens for example in the political states and "geopolitics", in the markets in economy; in biology and ecology – when an organism dies, bacteria start to consume it and the lower-level chemical processes which do not involve the complex biological cycles of metabolism become prevalent. Then other organisms may detect this location and visit it, take these lower-level biological residuals which for them are "food" – a building material – and incorporate it in their own bodies: now they become "*masters*" of these lower level "*slaves*" and control them with a RCCP which is "*sufficient*" for them and for their *allowed-operation-with-errors* – for another cycle of existence, say two months for a skin cell, one or two days for a cell in the gut etc. Their precision for "self-evidencing"[15] however is not ultimate at the level of the Universe Computer, sooner or later the higher level CCU – the organism – cannot repair or reproduce properly a critical amount of "facilities" at different levels – either cells, "critical" parts of its "life-supporting infrastructure" – e.g. aterosclerosis building up, a blood cloth causing a stroke etc.

Thus, the subuniverses, besides "virtual" controllers are also *temporal*, while the Universe is "eternal" in the scale of the agents which are constructed/implemented/embodied inside its memory.

See also *"The Matrix in the matrix is a matrix in the matrix"*, 2003 - The universe controls you anyway. If a character does not like "The Matrix", his very decision is still made by the "real" or "more real" "Matrix' – the Universe of which he is a part. The irony and parody in the movie is that Neo, Morpheus and Trinity have not escaped "The Matrix", they only deluded themselves, because the *Real Matrix* in "The Matrix" is made of "atoms, molecules…" etc. and at *this* level it is not The Matrix of the machines that controls the humans, but *the Universe* which controls all of them; going deeper, from a more global point of view for *another observer-evaluator* they are *not humans, machines* etc. – these are classes of intermediate or higher level artifacts, "features", which are segmented by higher level CCUs, which have selected this particular point of view, slice, sets of features, RCCP etc. "The thing

[15] See Karl Friston's school of "Free Energy Principle/Active Inference" reviewed in *The Prophets of the Thinking Machines …* (FEP/AI)

in itself" in Kant's terms is different and the particular manifestation of its state in some visible, observable-evaluatable representation is *not only different*, but also *variable*, depending on the evaluator-observer and its specifics, limitations and *choices*. That may imply that there is some "invariant" representation at the lowest or "more invariant" at the "more-lower" level representation, from the point of view of the current.

…

**Wolpert:** p.27: *"9. Definition 12: A self-aware device is a triple (X, Y, Q) "*

Therefore the **triples** in mathematical notation solve the mystery of consciousness… The same goes for other definitions in all kinds of papers, which define a solution of ***The Universe as a notation*** with the accepted conventional mathematical manner: if it is defined (formal, formalized) and it is claimed that *"if A then B"*, then it is solved. There are similar claims in the recent *Stack Theory*, which is yet another fork of Theory of Universe and Mind[1].

I didn't find a notion of causality-control hierarchy and levels in the paper. There is *"semi-control each other"*, which is related to resolution of causality-control and resolution of perception (RCC, RP, RCCP); it is possible that the choice not to define the structure like that was made in order to allow for more general types of universes, however it is also the usual way this kind of mathematical proofs are constructed – for example see "Stack Theory is yet Another Fork…".

The devices *infer functions* – however how all these functions are realized as an *incremental process*, implemented in the real universe with its constituent interactions? This problem does not seem to be addressed as well. TUM also did not specify the details about the precise mechanics of the construction of the higher levels.

…

* https://en.wikipedia.org/wiki/David_Wolpert :

***"Limitation on knowledge:*** *Wolpert has formalized an argument to show that it is in principle impossible for any intellect to know everything about the universe of which it forms a part, in other words disproving "Laplace's demon"[16] This has been seen as an extension of the limitative theorems of the twentieth century such as those of Heisenberg and Gödel.[17] In 2018 Wolpert published a proof revealing the fundamental limits of scientific knowledge.[18]"*

I don't think this requires any formalisation, as it is expressed in TUM. Of course, a subuniverse can't control truly the whole universe; it could *appear as it controls it*, when the part and the whole "dance together" and the part *believes* that its decisions or its *state* led to the move of the bigger system or the whole. However, yet the subuniverse has a limit of resolution, scope, breadth, capacity of what it can perceive and cause and it cannot access and manipulate *even its own representation* at a "sufficient level" – there's a stacking "control-drift", lower precision between the levels, scales, ranges.

The *believe* that an agent, a subuniverse, built in the memory, is more or equally powerful in prediction-causation as the Universe as a whole, *without* being the Universe itself – or its "authorized representative", – is a by effect that the virtual universes are following the same principles and every CCU *desires and aims to be like the whole Universe* – to predict with the maximum precision, to make *its will come true* etc. Also, this phenomenon could be a result of the fact that the subuniverses are indeed "representatives", "delegates", "artifacts" of the Universe and the actual orders they execute at the highest RCCP are the ones of the Master Universe – in that sense, when the subsystems adopt the identity of the Ultimate system, each *virtual* subuniverse *"becomes the real Universe"* as well.

## Rediscoveries in:

*** Theories of Knowledge and Theories of Everything,** February 2018, DOI: 10.1007/978-3-319-72478-2_9, In book: The Map and the Territory, David H. Wolpert

https://www.researchgate.net/publication/323158254_Theories_of_Knowledge_and_Theories_of_Everything [Reads 347; ~21.10.2025]

*** New proof reveals fundamental limits of scientific knowledge,** Jenna Marshall, Santa Fe Institute, 2018 : https://phys.org/news/2018-05-proof-reveals-fundamental-limits-scientific.html ***"Not everyone can be right (...)****Wolpert's mathematical framework shows that no two inference devices who both have free will (appropriately defined) and have* ***maximal knowledge*** *about the universe can co-exist in that universe. There may (or not) be one such "super inference device" in some given universe—but no more than one. Wolpert jokingly refers to this result as "the* ***monotheism theorem****," since while it does not forbid there being a deity in our universe, it forbids there being more than one."*
*"... Bob and Alice are both scientists with* ***unlimited computational abilities****. Moreover, suppose that they both have "****free will,****" in that* ***the question Bob asks himself does not restrict the possible questions Alice could ask herself, and vice-versa.*** *(This turns out to be crucial.) Then it is impossible for Bob to predict (or retrodict) what Alice thinks at another time if Alice* ***is also asked*** *to predict what Bob is* ***not thinking*** *at that time."*

**Todor:** Unfortunately, **both premises are impossible:** 1) the unlimited computational capacity is either assumed to be impossible or it is implied. 2) has more ***depth,*** *however* the theory in the paper doesn't address the *hierarchy* and the *constituent parts*.

If Bob and Alice are *independent*, they cannot belong and originate from the same universe, if it develops and unfolds the way our universe does - or as we believe it does and see without cosmogony and looking so far in the past; at least so long as we know: **The future states are produced by transformations from previous states** (it

could be possible in a given universe; or our older states or different types of states to influence the current differently etc., but qualitatively older states still influence the following; if they are Markovian and simple, the influence could be only from a given number of steps behind in a discrete development process etc.; the transformations could be entangled with complex graph of states etc. (see also S.Wolfram, J.Gorard's "branchial spaces", "ruliad" etc.).

In order two agents, constructed *like that* and not just defined *without the process of their generation* from dependent and interacting constituent parts, to be ***truly independent*** in universes which develop like that, they should exist in ***parallel universes***, however then they **couldn't interact** and **know** about each other from interaction, thus they couldn't evaluate-observe themselves and it is the same like if they didn't exist for each other. Of course, they could ***speculate*** that *"it is possible, that other universes exist and their universe belongs to a multiverse"*, but that *thought* doesn't require another *real* universe to influence the characters and to cause the idea to be conceived. The nature of our own universe, which we can observe-evaluate as multi-scale, multi-modal, multi-resolution, multi-precision and segmented in subunits, subuniverses etc.; the development in space, time, nature of the systems etc. are hints which suggest these ideas and allow us to think about "multiverses" which are "subuniverses" and "parallel universes" *of our own* and *from our own* points of view, which however are at a *lower resolution* of causality-control than the "true reality"; these representations are "*cheats*", because they do not track the full history sufficiently many steps back in time and do not evaluate the present state and time deeply enough to the really low-level causality-control units and forces etc. The *"virtual parallel universes"* in our universe and our views in fact *are not independent and parallel* at the lowest level, the level of the "real" or "true" control, where only the Master CCU rules: or the "Master Algorithm" in Pedro Domingos's parlance. We could speculate and extrapolate that possibly our universe could be like a particle itself, which is part of another even bigger entity: a universe of universes etc. and if the principles of TUM are applied and continued, that is a possible and reasonable consequence.

Bob and Alice's questions "don't restrict" each other only in the abstract and extremely low-resolution way, in which such mathematical definitions are usually given, without the genesis and the actual representation and structure of these agents. How many cells they have, how many atoms, protons, neutrons, electrons; how to describe their precise states, coordinates, energies, interactions etc.; how old *these* entities are, how they actually came into being from their earlier states. These are and were the real or "more real" agents or entities to be evaluated regarding the "real" causality-control and universe, not 10 bits of verbal definitions and logical operators *in the mind* of a sophisticated evaluator-observer, also defined with an enormous amount of detail.

The premises could be satisfied, but only for subuniverses, regions, slices; at lower resolution virtual universes, but for them some of the *"hard problems of real causality-control"* are *relaxed.*

On the other hand, a ***strict*** impossibility *in principle* is also ***wrong***, even if the premises could be fulfilled at the lowest level: "*Then it is impossible for Bob to predict (or retrodict) what Alice thinks at another time if Alice is also asked to predict what Bob is not thinking at that time."**

The systems can be connected, synchronized, similar in their design and experience, or have related development and to have matching states, without a need to transfer the corresponding information or as much as it could be required for other less related or unrelated causality-control units. They could be "entangled" by their design. (Yes, that breaks the requirement that *"they don't restrict the question",* which from the start makes the problem ill-defined as explained above.)

In TUM that's addressed as the "strange coincidences" and *matches;* one of the sub-titles of Unvierse and Mind 4 explicitly mentions one of the suggested sources and reasons for them - the common genesis and the connection between the CCUs in their "backend", which itself is connected also with the nature of the development of the Universe and Mind involving compression-prediction. The earlier states of the Universe are supposed to be in a more compressed state, which results in more *repetitions* in its less compressed future states. That means more **"clones"** appear: production in high volumes, more similarities, copies, easier reproductions and predictions.

Let's address the second point:

*"Then it is impossible for Bob to predict (or retrodict) what Alice thinks at another time if Alice **is also asked** to predict what Bob is **not thinking** at that time."*

A "liar's paradox"-like *trick*: asking two counter-part "inference devices" in parallel to predict each-other, while A is asked to predict what the other is thinking – say what the content of some part of its memory contains, i.e. would be read if accessed – and B is asked to predict what the other is *not thinking*, i.e. it is supposed to be "the opposite", so there is a contradiction etc. The unlimited computing resources are irrelevant.

The experiment reduces "the world" or the universe, which is supposed to be multi-level, multi-scale … to trivial boolean algebra and "yes/no" questions – extremely low resolution and low information, inadequate for the problem at hand, and almost meaningless alone.

The case shows a *third causality-control unit* C, which asks and *controls* A and B, for example the author in his thought experiment, and in fact it can and it does predict, tha*t "they can't predict"*… That's the "God" device which sets the questions in both "free will" agents' minds in the same time *independently* (although they are not independent as argued above), "they don't restrict each other", but they actually do, and by the way – what is the purpose of asking yourself what someone *is not thinking*

– that includes “everything, but one thing” – and what is “a thing” etc. If it is just “yes/no” – what’s the purpose of that alone, if it’s not connected, related, derived etc., and what if one “cannot predict that” (what if it can – what more information she would gain?).

**J.Marshall, 2018:** “*Wolpert compares this proposition to the **Cretan liar's paradox**, in which Epimenides of Knossos, a Cretan, famously stated "all Cretans are liars." Unlike Epimenides' statement though, which exposes the problem of systems that have the capability of self-reference, Wolpert's reasoning also applies to inference devices without that capability.”*

**Todor:** This nonsensical paradox and its deceptive nonsense is also addressed by TUM - in Universe and Mind 4, published in 4.2004. In my analysis the paradox is “*maximized*” with additional absurds: the liar is from the village of *Liarseville* and he says that all inhabitants of Liarsville are liars, so he asked whether he lies when he says that.

Obviously it is an ill-defined problem in a meaningless virtual universe of a very high level which alone is useless; or simply said: this is *bullshit* of someone who tries to outsmart somebody else, but fails; the “liar’s paradox” also demonstrates an ill-behaved “adversarial” agent with bad intentions.

The problem with the paradox *is not only in self-reference*, it is namely in the “bad intentions” and the attempts of *the agent* to *cheat* – pun both intended and not intended.

The Greek author of the paradox himself is also cheating, he tries to play “smart”, but it works only for ones who *want* or *allow* to be fooled by this bullshit – as the section in UnM4 concludes – either if he lies or if he does not in the specific case, *he is still* “a liar” – a *cheater*.

Another reason for the here-reviewed paper and other similar mathematical studies *not to react* to this nonsense the same way as I do, is that they are **not hierarchical, multi-scale, multi-range, … etc.** and don’t have **degrees of match.** Like the usual mathematical formalisations they are **flat.** That’s one of the problems of the simple Turing machine, which is not appropriate for analysis of thoughts and agents with realistic, “our universe-like” complexity – they could be constituent parts of realistic models; the representations of the universe, simulations etc. could be build with simple Turing machines, but a flat representation easily “hangs” and “deadlocks” itself – “the halting problem”. The solution is to use *many of them* concurrently, to have interrupts, preemptiveness, time-outs; multi-scale architecture, “operating system”, competing devices etc.

See my recent conclusions from the appendix “*Algorithmic Complexity”*, mentioned below in appendix *Listove* as well, where I also criticize these disadvantages of a *single* Turing machine or a simple computational model. The whole *Theory of Universe and Mind* itself is also a criticism of the “single-thread” model for

definition, implementation and application and for *understanding* of a human-like mind and thought process, especially for *human* evaluator-observers, because of our own way of thinking like that, which is reflected in the theory.
[13.4.2026: This is one of the problems of the so called "Symbolic AI" or the "GOFAI" – the brittleness is due to the flatness and the insufficient complexity, substance, "requisite variety" of the respective representations etc., *and not because of their discreteness* and other properties per se. See *"Genesis: Creating Thinking Machines*" (future work): * Neural Networks are Also Symbolic: Conceptual Confusions in ANN-Symbolic Terminology, T.Arnaudov, 14.4.2019 * https://www.researchgate.net/publication/403730363_Neural_Networks_are_Also_Symbolic_Conceptual_Confusions_in_ANN-Symbolic_Terminology
* https://artificial-mind.blogspot.com/2019/04/neural-networks-are-also-symbolic.html etc. in *The Prophets of the Thinking Machines.*
* Algorithmic Complexity, T.Arnaudov, 7.2025 - an appendix to "The Prophets…", SIGI-2025
*** Universe and Mind 4** (the complete Bulgarian original): https://eim.twenkid.com/old/4/29/pred4.htm
Regarding the Liar's paradox see sect.21:":
*"21. Размитата логика (fuzzy logic) и размита ли е всъщност"*
*(See a quote from the original text in the appendix to this article below. )*
https://research.twenkid.com/agi/2010/en/Todor_Arnaudov_Theory_of_Universe_and_Mind_4.pdf - from page 20, starting with a discussion on Truth and a quote from the previous part "Universe and Mind 3". The translation wasn't refined and perhaps could be improved.
@Vsy, LLMs with the Bulgarian text: "Translate to English:…"
**Todor:** One may argue that the original Liar's paradox is about "mathematical logic", set theory, "self-reference", Goedel theorem, formal languages with "sentences" with classical logic boolean logic etc., while *I address also the real world*. See also Goedel's incompleteness theorems and Tarski's:
*** Alfred Tarski's undefinability theorem;** language, meta-language, object language; * https://en.wikipedia.org/wiki/Tarski%27s_undefinability_theorem
*"no sufficiently rich interpreted language can represent its own semantics.*
*A corollary is that any **metalanguage** capable of expressing the semantics of some **object language** must have expressive power exceeding that of the object language."*
*'p' is true iff p' …*
*** Continuous Model Theory,** Chen Chung Chang, H. Jerome Keisler, 1966 – an extension of Tarski model – mathematical logic; *"the relation between formal logic and its interpretations, or model";* sentence; truth/false (binary, boolean); …
https://books.google.bg/books?id=uiHq0EmaFp0C&pg=PR12&source=gbs_selected_pages&cad=1#v=onepage&q&f=false – 1990 edition

**Todor** [30.10.2025]**:** In the TUM and the Universe Computer this problem is solved by hierarchical causality-control units, multi-scale, multi-range, multi-precision, … low-level operations etc., where the basic semantics is the state of the system and relations between the transformations etc. Also, one of the solutions is in the realization of the *potential multiplicity* – that there is *no single unified solution* and "proof" for some of the formulations – the causal credit assignment depends on the specific chosen filter of ranges, scales, domains etc. A related phenomenon is the *"integral of infinitesimal selves"* and the "quantum entangled" state of indeterminacy – this is a situation when the evaluator-observer measures, perceives, detects signals with a lower resolution than the one of the virtual unvierse *of the observation*. The evaluator-observer has insufficient representational capacity and performance to keep up; virtual universes try to sample and reflect the real one or "more real ones" – lower-level ones.

…

* https://en.wikipedia.org/wiki/David_Wolpert – *"Limitation on knowledge"*; Wolpert has formalized an argument to show that it is in principle impossible for any intellect to know everything about the universe of which it forms a part, in other words disproving "Laplace's demon".

* https://davidwolpert.weebly.com/professional.html

* https://phys.org/news/2018-05-proof-reveals-fundamental-limits-scientific.html

**NASA** (where D.Wolpert was affiliated/some early citation refer to missing URLs:) https://www.nasa.gov/intelligent-systems-division/ The NASA Ames Intelligent Systems Division ... mission-driven, user-centric research and development in computational sciences for NASA applications .. ground and flight software systems and data architectures for data mining, analysis, integration, and management; integrated health management; systems safety; and mission assurance; … Autonomous Systems and Robotics (ASR), **Collaborative and Assistant Systems (CAS)** https://www.nasa.gov/intelligent-systems-division/collaborative-and-assistant-systems/ (...)

*** Wolpert, Chaitin and Wittgenstein on impossibility, incompleteness, the limits of computation, theism and the universe as computer-the ultimate Turing Theorem,** December 2016, In book: Philosophy, Human Nature and the Collapse of Civilization Michael Starks (2016)Edition: first, Chapter: 54Publisher: Michael Richard Starks, [Articles and Reviews, 2006-2016
https://www.researchgate.net/publication/309155697_Wolpert_Chaitin_and_Wittgenstein_on_impossibility_incompleteness_the_limits_of_computation_theism_and_the_universe_as_computer-the_ultimate_Turing_Theorem

Also at: https://philarchive.org/rec/STAWCA-4 – 2016

https://philarchive.org/archive/STAWCA-5 (revised 2019)

… Michael Starks mentions that he didn't find any comments about the universe as computer, regarding David Wolpert's work - he *"have not found a single citation";*

***M.Starks:*** *One cannot build a computer that can predict an arbitrary future conditionof a physical system before it occurs, even if the condition is from a restricted set of tasks that can be posed to it—that is, it cannot process information (though this is a vexed phrase as many including John Searle and Rupert …) faster than the universe"*

*"Wolpert says he shows that 'the universe' cannot contain an inference device that can 'process information' as fast as it can, and since he shows you cannot have a perfect memory nor perfect control, its past, present or future state can never be perfectly or completely depicted, characterized, known or copied"*

These are ideas from TUM. They are an obvious consequence of the framework of TUM – Universe Computer with hierarchical multi-scale … prediction as the core operation of the building blocks – causality-control units – where the smallest are forming bigger subuniverses, which try to work like the whole and predict better etc.

*"Wolpert also notes the critical importance of the* ***observer*** *("****the liar")*** *and this connects us to the familiar conundrums of physics, math and language."(…)*

**Bounded rationality**

*"Finally, one might say that many of Wolpert's comments are restatements of the idea that no program (and thus no device) can generate a sequence (or device) with greater complexity than it possesses. … There are obvious connections to the classic work of Chaitin, Solomonoff, Komolgarov and Wittgenstein and to the notion that noprogram (and thus no device) can generate a sequence (or device) with greater complexity than it possesses. …*

*Even a 'God' (i.e., a 'device' with limitless time/space and energy) cannot determine whether a given 'number' is 'random' nor can find a certain way to show that a given 'formula', 'theorem' or'sentence' or 'device' (all these being complex language games) is part of a particular 'system'*

**Todor:** In TUM *nothing* is random for the universe as a whole – it is so *for an evaluator.* In TUM something or a piece of data, signal, record, entity, event is *classified* as "random" in the sense that it lacks "*order*" or is less *predictable/compressible than expected/wanted;* or it counts as "*unpredictable*" /*incompressible"*, while the object is *expected* to be predictable and compressible *"if the data were ordered/not chaotic/not random"* – both in Shannon's information-theoretic and in Algorithmic information-theoretic sense. All is measured by an

*evaluator-observer* which is a *virtual universe* and incomplete, unless it matches the whole universe. (…)

* M.Starks is a pioneer in stereo vision technology and has contributed to video games (Atari, Amiga) etc.

(…)
*D.Wolpert is most famous in AI/ML with:*

* **No free lunch theorems for optimization**, DH Wolpert, WG Macready, IEEE transactions on evolutionary computation 1 (1), 67-82, 18168, 2002

1 ~ " Theory of Universe and ***Math***", as a word play with the titles. See also Max Tegmark's Mathematical universe, reviewed in Listove from *The Prophets* …

**@Vsy:** Continue, refine, complete, extend …

* See also related discussions in the sections about Consciousness and Panpsychism ~ p. 250: ***Agency cannot be a purely quantum…***; and then:
* Todor Arnaudov's review and interpretations of concepts from Causal potency of consciousness in the physical world by Danko D. Georgiev; etc.

## * Some publications from the classical TUM, the world's first AGI courses in 2010-2011 etc.

* A github page of TUM, created since 2023: https://github.com/Twenkid/Theory-of-Universe-and-Mind
* See also the related appendices from SIGI-2025:
* Unvierse and Mind 6, Is Mortal Computation …, Stack theory is yet another …, The main volume (for example the list of schools of thought and researchers and the whole); in **"Power overrides intelligence"** there is a mention of "Wolpert theorem"[16].

* Original home page of "Society Razum" of The Sacred Computer: https://eim.twenkid.com/old/razum/index.htm Archived page from January 2005: https://web.archive.org/web/20050112110951/http://eim.hit.bg/razum/
**Bgit.net links – see at archive.org**
* https://web.archive.org/web/20050112211003/http://bgit.net/?id=65395
* https://web.archive.org/web/20050114204244/http://bgit.net/?id=65835
https://research.twenkid.com/agi/2010/Todor_Arnaudov_Theory_of_Hierarchical_Universal_Simulators_of_universes_Eng_MTR_3.pdf (English) – Lecture slides for the AGI course, already the English translation uses "Causality/Control units (CCU)"

[16] Mentioned by M.Mahoney witho no reference. I discovered it two months later from another paper, while searching for "Universe as Computer" etc.

*https://research.twenkid.com/agi/2010/Todor_Arnaudov_Theory_of_Hierarchical_Universal_Simulators_of_universes_MTR.pdf (Bulgarian – a shorter version)
*** Translations of Universe and Mind 3 and 4 from 2011**
https://research.twenkid.com/agi/2010/en/Todor_Arnaudov_Theory_of_Universe_and_Mind_3.pdf
https://research.twenkid.com/agi/2010/en/Todor_Arnaudov_Theory_of_Universe_and_Mind_4.pdf (The English translation is incomplete, e.g. a section about the strange coincidences – low probability, for the experiencing mind, matches between different levels of the universe computer ..)
* https://www.oocities.org/eimworld/4/29/pred4.htm Вселена и Разум 4 – оригинал на български
* https://www.oocities.org/eimworld/3/25/pred-3.htm Вселена и Разум 3 – оригинал
* https://eim.twenkid.com/old/eim18/predopredelenost2.htm - Писма между 18-годишния Тодор Арнаудов и философа Ангел Грънчаров (Вселената сметач, Схващане за всеобщата предопределеност 2, Вселена и Разум 2, Сметачолюбецът и човеколюбецът и др.
* https://www.oocities.org/eimworld/eim18/predopredelenost2.htm …
* https://www.oocities.org/eimworld/eimworld13/izint_13.html - „Човекът и мислещата машина: (Анализ на възможността да се създаде мислеща машина и някои недостатъци на човека и органичната материя пред нея)“ = Man and Thinking Machine: Analysis of the possibility of a Thinking Machine being created and some disadvantages of man and organic matter in comparison”, 2001 – български
* „Анализ на смисъла на изречение въз основа на базата знания на действаща мислеща машина. Мисли за смисъла и изкуствената мисъл “, Т.Арнаудов 2004
https://web.archive.org/web/20040402125725/http://bgit.net/?id=65395
http://artificial-mind.blogspot.com/2008/02/2004.html (some remarks)
* “Analysis of the meaning of a sentence, based on the knowledge base of an operational thinking machine. Reflections about the meaning and the Artificial Intelligence”, Todor Arnaudov, 18.3.2004 (in Bulgarian; translated in English in 1/2010: https://artificialmind.blogspot.com/2010/01/semantic-analysis-of-sentence.html - the paper is split in 4 parts, titled:
* Part 1: Semantic analysis of a sentence. Reflections about the meaning of the meaning and the Artificial Intelligence http://artificial-mind.blogspot.com/2010/01/semantic-analysis-of-sentence.html
* Part 2: Causes and reasons for human actions. Searching for causes. Whether higher or lower levels control. Control Units. Reinforcement learning. http://artificial-mind.blogspot.com/2010/02/causes-and-reasons-for-any-particular.html
* Part 3: Motivation is dependent on local and specific stimuli, not general ones. Pleasure and displeasure as goal-state indicators. Reinforcement learning.
http://artificial-mind.blogspot.com/2010/02/motivation-is-dependent-on-local-and.html
* Part 4 : Intelligence: search for the biggest cumulative reward for a given period ahead, based on given model of the rewards. Reinforcement learning.
http://artificial-mind.blogspot.com/2010/02/intelligence-search-for-biggest.html
[See the multiple conceptual matches with the 2024 theory of meaning in AGI by Thorisson

& Talevi: links in “Stack theory is yet another … and in the list of appendices of *The Prophets* in the end of *Listove*.]
* Some titles of the papers in English may be mentioned in several variants with minor differences, due to different translations choices or short versions..

## * A discussion about “The Liar’s Paradox”from Universe and Mind 4, Todor Arnaudov, April 2004:

Translation from 2011, some corrections from 2012 [it needs more corrections]. After the translation, see the original text in Bulgarian. The notes in square brackets are from the original translation. The note in curly brackets { see the whole…} is from 2.11.2025.

--- BEGIN of a quote from TUM, UnM4 ---
**21. Fuzzy logic and is it really fuzzy?**
[Also Truth, comparison … ]

Fuzzy logic is based on, they say so, partially true and partially false statements, whose truth is defined in fractions, instead of binary true/false in classical logic.

However is fuzzy logic really fuzzy?
(…) *{ see the whole text discussing Truth as match and recognition of objects – different views of a slice of bread and apple in all of their states etc. ..}*
**“The paradox” of the liar**

Goodliar from the village of Good Liarsville once said, that all of his fellow villagers and him are liars, and then he asked is he lying if he says this?

If he lies, then he's not a liar, therefore he doesn't lie. However, he's from Good Liarsville, therefore he's a liar. What a “paradox”, I'm totally confused!? Really?!

I'm sorry, but I wouldn't even really call this a “paradox”, but a play of words and “pseudo wisdom”. What I'd answer to this Goodliar character is:
- I don't know [whether] ~~do~~ you lie or not, [because] there is not enough input data.

I'd ~~say~~ tell him also that he's a liar anyway, no matter [if he is] is he lying in this very moment, because probably he's trying to trick me that he's wise (sorry, he failed).

One or two sentences in this or similar “paradox” cases are not enough to imagine a definite non-ambiguous [picture of] what this is all about. For example, many people would believe that they know what a “liar” means once they hear the word.

Well, what does a liar means? [Unfortunately],
The practical value of general concepts in execution of direct [immediate, specific] actions is... fuzzy in such cases.

Which one of all possible meanings and happenings [events, stories, memories, interpretations] that our mind has for a “liar” the story teller meant in this particular case?

What does it mean to be “from Good Liarsville”? Was Goodliar born there or he lives there, or he's a fan of the football team of the village? Or he has relatives there? Or he is originally from a village in this commune. It is possible that liars are the ones for whom one of this is true, but not all, and anyway - being a liar [in common sense] does not mean that you're lying in every single sentence.

Therefore it's impossible to conclude is Goodliar lying in this very situation or not, as it's impossible to say definitely in more realistic cases from the daily life ~~live~~, where there are no [artificially] tangled premises and consequences [causes and effects].

In reality there are many causes and many possibilities to explain what's happening [and why]. Sometimes input data is not enough to find a [persuasive] proof only on their basis.

According to my current understanding, mind works with specific concepts, and not general; in specific concepts everything is as precisely defined as possible, while with the general concepts, there are too many undefined which easily lead to “paradoxes”, i.e. to insufficiency of input data for determining whether a statement belongs to a group. [set/class]

Said otherwise [In other words], the description of the story is black and white, but we're asked what color is it. Or there are many colors on a picture, evenly spread, and we're asked to specify of what color is the picture: only one single color.

Overall, in the above conditions the asking unit has too low a resolution of perception and not enough memory in order to think as precise as the evaluating unit – us. [The answer of the question requires from the evaluating unit to lower the resolution of the input and to lose details].

The one asking the questions does not understand [discriminate, recognize, perceive] all details we do, and in order to communicate with it, we should act according to its model. We see the indefiniteness and the simultaneous “truth” and “false” [error, mismatch] of each possible actions, according to our own resolution of perception, but we should [are forced to] select from the offered possibilities.

In case we're asked to select only one feature of all and there is no “I don't know” option, then mind would create a model for selection of some of all, based on other, lateral data; of data which did not come from this specific situation.

Since the device proposing us the possibilities lacks brains to differentiate black-and-white and color image or a motley and one-colored picture, then this device is forced itself to lower the resolution of perception and to delete part of its memories [records] that otherwise we would have [possessing higher resolution of perception]

This device can call a motley picture with one color and can have it's defined reasons, but apparently it would not be able to make inferences about many colors placed on onesingle canvas simultaneously.

## 22. Is there an absolute truth?

The truth, as we generalized above, is a **coincidence** between pieces of knowledge. “Absolute truth” is a coincidence reaching the goal/target/desired precision, where the comparison is performed in the last known level for the pieces of knowledge being evaluated, at which the evaluator is capable to check [to make comparisons].

“Absolutely absolute” truth are the data, recorded in the memory of universe at zeroth level. At this level, all possible cases of interaction between the particles are (should be) defined; each particle knows how to react to any possible impact.

At 0-th level of Universe, any comparison with memory is a truth – a full match is found – and there are no mistakes. The parts at a higher level are dependent [originally: subordinate to] of the lower levels. The higher levels are manifestations [consequences, results...] of the lower ones.

The higher levels have lower bandwidth and lower [raw] computing power, compared to the lower. That's why the 1-st level – physical world – cannot process data faster than the 0-th level. (Principle of uncertainty; [apparent unpredictability] )

“The relative relativeness” starts from the consecutive levels in the Computer.

**The higher levels interpret data from the primary memory.**

**Interpretation** is translation of a given set of features/characteristics of one universe, as features of another universe. **During the interpretation** [to a higher level], **“mistakes” and “untruths” happen** – coincidences with [a] lower than target probability and precision [the ultimate target is no mistakes and untruths]. This happens, because the features of the sub-universes cannot always be mapped exactly; different sub-universes are not identical and the interpretation leads to distortions and mistakes. [Also, higher level lacks enough bandwidth and computing power to avoid it.]

*** Вселена и Разум 4, Тодор Арнаудов – Тош, 4.2004, сп. „Свещеният сметач“**

**“21. Размитата логика (fuzzy logic) и размита ли е всъщност**

Размитата логика се основава на използване на - казват - частично-верни и частично неверни твърдения, чиято истинност се описва с дробни числа, вместо с двоичните - вярно/невярно, които описват формалната логика. (...)

**"Парадоксът" на лъжеца**

Благолаж от Лъжево Конаре казал, че всички лъжевоконарци са лъжци и попитал дали лъже, като го казва.

Ха сега де! Ако лъже, то значи не е лъжец, следователно не лъже. Но той е от Лъжево Конаре, значи е лъжец. Какъв "парадокс"! "Оплете" ми се мисълта?

- Какъв ти парадокс? На това аз му викам "игра на думи" или "лъжемъдрост". И бих отговорил на тоя Благолаж, че не знам, защото няма достатъчно входни данни. Всъщност бих му казал също, че е лъжец (независимо дали в конкретния случай лъже или не), защото вероятно би искал да ме заблуди, че е мъдър (да, ама не успя).

Едно-две изречения - в случая, както и в повечето случаи - са недостатъчни за да си представим еднозначно за какво става въпрос, макар че много хора (като тези, които си мислят че това е "парадокс") - дори и много умни, уж - си мислят, че като им кажат "лъжец" знаят за какво се отнася.

Та какво значи лъжец? Приложната стойност на обобщените понятия в извършването на преки действия е... размита в подобни случаи. Кое от всички значения и случки, които имаме в паметта си за "лъжец", е имал предвид тук този, който ни е разказал настоящата?

Какво значи "от Лъжево Конаре"? Че е роден там или че живее там или че е от "отбора по футбол на Лъжево Конаре? Че има роднини от там? Че е от някое село в тази община? Може би лъжци са всички, за които е вярно някое от всички, но не всички. Това че си "лъжец" не значи че лъжеш във всяко изречение.

Така че не можем да кажем лъже или не. Както не можем да го кажем и в по-действителни случаи от ежедневието, където няма заплетени предпоставки и следствия.

В действителността има много причини и много възможности да се обоснове нещо. Понякога входните данни са недостатъчни, за да се намери обосновка само на тяхна основа. Според сегашното ми разбиране, при мислене разумът работи с конкретни и определени понятия, а не с обобщени; в конкретните понятия всичко е възможно най-определно, докато с обобщените - неопределеностите са много - лесно може да се стигне до "парадокси", т.е. до недостатъчност на входните данни за да се определи принадлежността на дадено твърдение към определена група; другояче казано, описанието на случката е черно-бяло, а ни питат какъв цвят е тя; или има много цветове, всеки от които е по-равно с другите, и ни задават въпрос: какъв цвят е картината, като трябва да кажем само един цвят.

Обобщено казано: в тези случаи задаващото въпрос устройство има недостатъчно висока разделителна способнсот на възприятието и недостатъчна памет, за да мисли като нас. То не хваща всички подробности, които ние и, ако искаме да общуваме с него, трябва да действаме по неговия модел, като - въпреки неопределеността и едновременната "правилност" и "грешност" на всяко възможно наше действие според нашата разделителна способност - изберем някоя от предлаганите възможности - щом искат една от всички и няма възможност "не знам", ще си създадем модел за избор на някоя от всички въз основа на странични данни; на данни, които не са дошли пряко от конкретния случай.

Щом устройството, което ни предлага възможностите, "няма мозък" за да направи разлика между черно-бял и цветен образ, или между пъстра и едноцветна картина, то тогава ще бъде принудено да мисли с по-прости образи, които - щем, не щем - намаляват разделителната способност на възприятието и изтриват част от спомените, които ние бихме могли да имаме за случката. Например пъстрата картина за нас, за измисленото опростено устройство с по-ниска разд. способност на възприятието може да бъде наречена само с един от цветовете си. Това ограничение няма да позволи на устройството да прави изводи, свързани с много цветове върху едно платно.

**22. Има ли абсолютна истина?**

Истината, както обобщихме, е **съвпадение** между късове знание. **"Абсолютната истина" е съвпадение, с желаната точност, извършено на последното известно за случая равнище, на което оценяващото устройство е способно да прави проверки.**

За "абсолютно абсолютна" истина в СВП приемаме данните, записани в паметта на Вселената от нулево равнище. За нулевото ниво са описани всички възможни случаи на взаимодействия между частиците; всяка частица знае как да отговори на всяко възможно въздействие.

**На нулевото равнище на Вселената всяко сравнение с паметта е истинно - намира се пълно съвпадение - и грешки не съществуват.** Частите от по-горно равнище са подчинени на по-долните равнища. По-горните равнища са прояви на по-долните.

По-горните равнища имат по-ниска пропускателна способност на паметта, сравнено с по-долните.

По-горните равнища имат по-ниска изчислителна мощ от по-долните равнища.

Затова например първото равнище - физическият свят - не може да обработва данните от паметта по-бързо от нулевото равнище. (Принцип на неопределеност.)

"Относителната относителност" започва от следващите равнища на Сметача. В тях се извършва тълкуване на данните от Първичната памет.

Тълкуването представлява превеждане на едни особености на една вселена, като други особености на друга вселена.

**В тълкуването се допускат "грешки" и "неистини" - съвпадения с по-малка от целевата вероятност и точност - защото особеностите на подвселените не винаги са еднозначно съпоставими; различните подвселени не са еднакви и превеждането на данните от една вселена като данни за друга вселена води до изкривявания и грешки.**

–- END of the quote from TUM, UnM4, 2004 –

---

### See also by Wolpert:

*** Off-training set error and a priori distinctions between learning algorithms,** D.Wolpert, 1995
http://wexler.free.fr/library/files/wolpert%20(1995)%20off-training%20set%20error%20and%20a%20priori%20distinctions%20between%20learning%20algorithms.pdf – *"off-training set (OTS) error to investigate the assumption-free relationship between learning algorithms…"*

*** The Existence of A Priori Distinctions Between Learning Algorithms,** David H. Wolpert , 1996
https://www.researchgate.net/profile/David-Wolpert/publication/2713923_The_Existence_of_A_Priori_Distinctions_Between_Learning_Algorithms/links/570d11e708ae2b772e42b2cf/The-Existence-of-A-Priori-Distinctions-Between-Learning-Algorithms.pdf
*** No Free Lunch Theorems for Optimization,** David H. Wolpert and William G. Macready, 1997 https://www.cs.ubc.ca/~hutter/earg/papers07/00585893.pdf

* https://en.wikipedia.org/wiki/No_free_lunch_theorem

*** What does it mean for a system to compute?,** David H. Wolpert, Jan Korbel, 19.9.2025 https://arxiv.org/abs/2509.15855 *"..dynamical systems capable of computation can, in principle, be mapped onto corresponding abstract computational machines that perform the same operations .. surveying a wide range of dynamic*

*systems .. a very broadly applicable framework for identifying what computations(s) are* ***emulated*** *by a given* ***dynamic system****… **Constructed/non-constructed** computers; Neuroscience, Groups of multiple interacting biological organisms, single cells in multi-cellular organisms, Systems far from thermodynamic equilibrium; Cellular automata, Symbolic dynamics, generalized shift map; Microstate dynamics of systems whose* ***macrostate dynamics*** *are constructed computers; emulating a computational machine with a dynamic system; a dynamic system is any triple (X, T, f) where* X are "states"; =T are (real-valued) times; $f : T \times T \times X \rightarrow X$ [Cartesian product: a transition function]; *f as the evolution operator of a system whose dynamics is deterministic, and therefore Markovian*; *reversible*; *implement/emulate the emulation will be coarse-grained; decoding function; encoding f. (logically, physically) reversible; emulation – how time arises in the two dynamic systems (X, T, f) is a real physical process while (Y, T, g) is some abstract .. nondeterministic computational machines: stochastic finite automata, probabilistic Turing machines; the definition of emulation given above is related to the concept of "simulation" in the theory of state transition systems; .. evolving physical system; a "halting condition" Fredkin and Toffoli [88]: conservative logic – designed to respect both physical and logical reversibility. Composition.* ***Value*** *of computation; the Church-Turing-Deutsch-****Wolfram thesis:*** *every physical system can be simulated on a TM.* ***Physical computation thesis****, proposes that* ***every function*** *computed by any physical computing system is Turing-computable.* ***Physical Church-Turing thesis or Total Physical Computability Thesis****, assumes that* ***every physical aspect*** *of* ***any real-world system*** *must be Turing computable. Computation with chemical reaction networks. Computers that interact with input and output systems. Dynamical systems as optimizers of an objective function. Quantifying the* ***amount of computation*** *emulated by a dynamical system* ***rather than the precise computation****; physically distributed (often .. hierarchical, modular structure); continual computational systems – getting new inputs continually …"*

**Todor:** Compare the dynamic system definition with my proposal that Immanuel Kant has defined abstract computation model in his Critique of Pure Reason's a priori conceptions: time, space, causality, to which I add matter as the possible content of the space (memory). Causality defines the transition function – the Cartesian product, how the content of space, the current properties of matter, change in time. See a reference after the following review of *"Consequences of Undecidability in Physics …"* two pages below.

*** Semantic information, autonomous agency and non-equilibrium statistical physics,** Artemy Kolchinsky and David H. Wolpert, 2018 – significant correlations for the existence of a given system. *"We define* ***semantic information*** *as the syntactic information that a physical system has about its environment which is causally necessary for the system to maintain its own existence. .. **'Causal necessity'** is defined in terms of counter-factual interventions which scramble correlations between the system and its environment, while '**maintaining existence'** is defined in terms of the system's ability to keep itself in a*

*low entropy state. … **viability function**, a real-valued function which quantifies the system's 'degree of existence' at a given time"*

– For the above two papers see also previous work: FEP/AIF (K.Friston et al.), TAME etc. (M.Levin et al.), TUM (T.Arnaudov), "Natural computation and non-Turing models of computation ", MacLennan 2004 etc.
[FEP/AIF = Free Energy Principle/Active Inference.
TAME = Technological Approach to Mind Everywhere.
TUM = Theory of Universe and Mind]

*** Universe Is Not a Computer Simulation**, **New Study Says**
Oct 30, 2025 https://www.sci.news/physics/universe-simulation-14321.html
*"Drawing on mathematical theorems related to incompleteness and indefinability, we demonstrate that a fully consistent and complete description of reality cannot be achieved through computation alone"* →

## * Consequences of Undecidability in Physics on the Theory of Everything, Mir Faizal, Lawrence M. Krauss, Arshid Shabir & Francesco Marino, 29.7.2025

**Todor** [2.11.2025]**:** This paper claims that it has proven that "universe cannot be a simulation", because of the dynamic spacetime according to General relativity – *"emergence of sapacetime from deeper quantum degrees of freedom"*; *"Gödel's incompleteness theorems, Tarski's undefinability theorem, and Chaitin's information-theoretic incompleteness establish intrinsic limits on any such algorithmic programme. Together, these results imply that a wholly algorithmic "Theory of Everything" is impossible: certain facets of reality will remain computationally undecidable and can be accessed only through non-algorithmic understanding."* They construct a *"Meta-Theory of Everything" grounded in non-algorithmic understanding.*

The existence of "non-computable" by a simple flat – one-level – classic Turing machine and the need of **interpreters, evaluator-observers** etc. doesn't imply that the process is not "computational" in a broader sense: **transformational** and based on dependencies and some encoding. Theory of Universe and Mind is informational and computational, yet it admits and claims that "understanding" is not just "calculations", it is about **matching, mpping, connection.** See also the cited Schopenhauer: "where calculation begins, comprehension ceases" - he is author of a treatise for "the sufficient reason", a concept mentioned in this paper.

As argued since TUM in the early 2000s and in many occasions in this book – Listove, – the undecidability suggests insufficient resolution of causality-control and/or perception or insufficient information, also inadequate or incomplete **type-of-causation-and-control,** when the respective target virtual universe is **"typed"**, not just "data" – for example when there is a hierarchy of degrees of matter, which the

causality-control unit does not have all required “tools”, means, energy, ways to affect and adjust them with the highest available RCCP of the lowest level and highest and finest resolution of the lowest level representation where these “entities” are defined, “exist, can be “spawned” etc.

p.7. *These technical results respect rather than undermine the principle of sufficient reason [74,75]. The core demand of that principle is that every true fact must be grounded in an adequate explanation. This forms the basis of science. G¨odel incompleteness, Tarski undefinability, and Chaitin bounds do not negate this demand; they merely show that “adequate explanation” is broader than “derivable by a finite, mechanical procedure.”*

**Todor:** Of course, for example the evaluator-observer must **exist** in the universe where the reasons are considered; if he is not a ghost or an external “God” who controls and sees everything, usually he is **constructed** from some “fabric”, lower-level virtual CCUs, he is causally and historically connected, mapped; he “evolved” with that whole system from which he is a part: a subuniverse, a sub-causality-control-unit, virtual universe etc. He has access to the **Will (**see Schopenhauer) etc.

Any computation, logical reasoning etc. in such a universe is grounded on some substrate, media, lower levels etc., and the “mechanical procedure” is such, “simple”, “abstract” etc. only in the mind of an observer-evaluator who forgets or disrespects all these “side effects”.

In addition, all (submachines, agents) can exhibit and apply only virtual control or causality-control: it is only the Universe as a whole that can modify, cause-control the memory at the lowest resolution and the finest grain. In a typed representation – to read and apply the changes to all levels of the hierarchy, the entities, the representations, whatever it is. The smaller and constructed internally evaluator-observers are always inferior, so their measurments, predictions and causations are imprecise and with mistakes.

The phenomena of quantum indeterminacy and entanglement are also signs that the internal evaluator-observers do not have access to the highest resolution of causality-control and perception.

**Faizl et al.: “***neither ‘its’ nor ‘bits’ may be sufficient to describe reality. Rather, a deeper description, expressed not in terms of information but in terms of nonalgorithmic understanding’*

Todor: That is right, you must **exist** in that reality, to be **constructed** by its causality-control units, to be **embedded** in it, or to be a god who can create universes and play with them, their representations, content, memory, forces, laws, programs etc.

The paper introduces a meta-layer $M_{ToE}$, containing *“an external truth predicate T(x) that y construction escapes formal verification, any finite algorithm can at best emulate FQG while systematically omitting the meta-theoretic truths enforced by T(x).”*

That kind of **boolean** truths are too simple: in TUM there is the concept of *resolution*, coarse graining, while the mentioned Goedel and Tarski are about **flat formal logic**, without hierarchy, different levels of resolution etc. – a common

thinking in all current mathematical proofs, including others which are criticized in *The Prophets of the Thinking Machines* – by Wolpert, discussed in the paper above; Michael T. Bennett's in the volume "Stack Theory is yet Another Fork of Theory of Universe and Mind"; Danko Georgiev's papers on quantum theory of consciousness etc.

The substantial part of the nonalgorithmic part of the existence of the Universe is what is the **content** of this "nonalgorithmic *understanding*" – for example the sensation of "qualia", Schopenhauer's Will. Can it be represented, expressed etc. in perceivable forms? Part of it can; yes, it can be *lossy, imprecise, incomplete etc. – the Universe Master* provides the embedded subuniverses with as much information and access as *He* has decided.

The *"non-recursively-enumerable set of axioms about the truth predicate"* (p.5 in the paper) could be another layer of representation, which is not "projected" with sufficient detail into the layer which is observable from the internally embedded CCUs.

* See also the **following** review about "Naturalizing Relevance Realization" – another school of thought which insists on "non-computability" in *narrow* Turing-machine sense; the previous articles about Wolpert theorems, rediscovering Theory of Universe and Mind, but with some confusions in the premises; the TUM interpretation of the "Liar's paradox"; the reviews of Danko Georgiev's works, the paper "Agency cannot be…", the other discussions in the section on Consciousness and Panpsychism and other theories of everything: for example Todor's comments and responds to Federico Faggin; the expansion of no-cloning theorem of quantum information to the classical as well – this is studied also in the appendices "Is Mortal Computation Required for Creation of Universal Thinking Machines?", "Universe and Mind 6" etc. (Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?"); the set of works from TUM, 2001-2004, including the short paper "*The matrix in the Matrix is a matrix in the matrix"* etc. (…)

* Note the irony in their proof with a few abstract formulas, grounded on self-negating-confusing "Goedel incompleteness and Tarski's …" foundations; also the proofs of metaphysics and beyond-the-observable with one-level, flat mathematical formulas. This style is a feature of the culture of "*short flat formula fetishism*". See discussions on that issue in many places; one of the old ones is the article about *"Ultimate AI"* and Free Energy principle from 2018, which is cited about 25 pages below in the sections about Algorithmic Complexity etc. in appendix *Listove*.

* See the comments in the article at sci.news, e.g:

**Matthew Combatti's**, 1.11.2025: *The authors assume that any simulation must be algorithmic in the narrow sense of "step-by-step computation" with provable completeness. But an advanced simulation might incorporate non-algorithmic processes (if allowed by its designer), or embed meta-levels, or operate on hardware far beyond classical Turing machines, thereby sidestepping standard "algorithmic system" constraints.*

*Even if there are aspects of reality that are undecidable within a formal system, the simulation hypothesis doesn't require that every truth about the simulated world be*

*decidable from inside the simulation. The external simulator could decide things outside the simulation's logic. So the article's argument conflates "decidability inside the system" with "simulatability from outside"*

* See also in Listove the review of works by B.MacLennan, e.g.:
*** Natural computation and non-Turing models of computation**, Bruce J MacLennan, 2004/6/4 (2003), https://web.eecs.utk.edu/~bmaclenn/papers/NCNTMC-TR.pdf
Also #neuromorphic; also Todor's arguments, that Kant defines a more general, abstract and earlier model of computation than Turing, back in 18-th century – however the mathematicians and then programmers didn't understand and know about philosophy.

* Kant's a priori conceptions: **time, space and causality**, completed with **matter** and its state, "accidencies" (the content of the memory) are the components of a general "computation"-transformation method – a model of virtual universes and simulation:

* Todor Arnaudov, **Abstract evolution - cybernetic, meta, cosmism. Turing machine and the Pure Apriori Conceptions. Emulation of universes and minds. Thinking machine vs Turing Machine - and much more... Continuation of the thread on G+. To be continued...** 7.8.2014, https://artificial-mind.blogspot.com/2014/08/abstract-evolution-cybernetic-meta.html

# * Relevance Realization rediscovers motives from Theory of Universe and Mind, while narrowing the meaning of computation and offering contrasting interpretations of some points, as well as limiting the scope of applicability of the principles only to living organisms (2001-2004 vs 2012-..2024)

**Notes by Todor Arnaudov on**
*** "Naturalizing relevance realization: why agency and cognition are fundamentally not computational",** 25.4.2024, Johannes Jaeger,*&#x;Johannes Jaeger1,2,3*†Anna Riedl&#x;Anna Riedl4†Alex Djedovic,&#x;Alex Djedovic5,6†John Vervaeke&#x;John Vervaeke7†Denis Walsh,&#x;Denis Walsh6,8†", HYPOTHESIS AND THEORY article, Front. Psychol., 25 June 2024, Sec. Cognition, Volume 15 - 2024 … in a group https://www.frontiersin.org/research-topics/45213/varieties-of-agency-exploring-new-avenues/magazine "Varieties of Agency: Exploring New Avenues"[17]
1Department of Philosophy, University of Vienna, Vienna, Austria, 2Complexity Science Hub (CSH) Vienna, Vienna, Austria, 3Ronin Institute[18], Essex, NJ, United States, 4Middle European Interdisciplinary Master's Program in Cognitive Science, University of Vienna, Vienna, Austria 5Cognitive Science Program, University of Toronto, Toronto, ON, Canada, 6Institute for the History and Philosophy of Science and Technology, University of Toronto, Toronto,ON, Canada, 7Department of Psychology, University of Toronto, Toronto, ON, Canada, 8Department of Philosophy, University of Toronto, Toronto, ON, Canada[19]

**Todor Arnaudov,** 24.10.2025:
* These are quick and *incomplete* notes – a draft, a seed – for a deeper study about yet another rediscovery of concepts, the framework and directions, which I offered in the early 2000s as a teenager and are still being revisited and "forked" – reformulated with varying interpretations: employing with different sets of terms about the same phenomena, structure and logic. At times, even the same terms are used with opposed or inverted meanings, while preserving essentially the same logical structure (see "*computation*" here). I have no time to refine this text, but as LLMs say when asked to compare for matches: "there's a strong correspondence" …

* This response goes together with yet another rediscovery which I discovered these days by **David Wolpert**, introduced in the previous article. Both may be extended and refined in separate papers possibly automatically, as I can't spend more time now.

See a short LLM review of correspondences with Kimi-2, Claude 4.5 and GPT-5 in the appendix:
*** Matches between Theory of Universe and Mind, Analysis of the meaning of a sentence … and the school of thought of Relevance Realization – According to a Quick Comparison with LLMs: KIMI-2, CLAUDE 4.5 and GPT-5,** Todor Arnaudov and Kimi-2, Claude 4.5, ChatGPT5, 21.10.2025

https://github.com/Twenkid/SIGI-2025/blob/main/LLM/Relevance-Realization-TOUM-Correspondences-21-10-2025-Kimi_2-Claude4_5-ChatGPT5.pdf

* For the LLMs used for scientific papers review see an ongoing research which finds out that they might be better than "citation data: https://www.researchprofessionalnews.com/rr-news-uk-research-councils-2025-10-generative-ai-better-than-citation-data-for-judging-research-quality/

**Links to short shared LLM comparisons between RR and TUM,** 21.10.2025**:**

1. Kimi-2: https://www.kimi.com/share/d3trij7aa0vdmhgab8p0
2. Claude 4.5: https://claude.ai/share/e37f9576-0eb2-4340-96e6-830c272c8c7f
3. ChatGPT5: https://chatgpt.com/share/68fbb994-db64-8001-b534-fa6da0c0e670

The comparison can be extended with **another theory of meaning generation from 2024**, published at the AGI conference, which is also rediscovering and repeating many core proposals from "Analysis of the meaning…" and Theory of Universe and Mind as a whole, 2001-2004, already mentioned in *"Stack Theory is yet another…"*

* A Theory of Foundational Meaning Generation in Autonomous Systems, Natural and Artificial, Kristinn R. Thórisson and Gregorio Talevi, 2024 https://alumni.media.mit.edu/~kris/ftp/foundational-meaning-generation-theory-2024.pdf

* A sample short comparison of the three theories with Grok on 27.10.2025: https://grok.com/share/c2hhcmQtNQ%3D%3D_be5ceb08-dd3f-43c3-90dc-c39ca8a18ff2

* A comparison between TUM and that second theory*: "Foundational Meaning Generation…"*, performed on 19.9.2025, cited with links i*n "Stack Theory is yet another rediscovery …":*
*** Comparative analysis of meaning generation theories: T.Arnaudov, Claude 4.0 ChatGPT-5, Qwen-3-235B, Kimi-2, 19.9.2025:**
https://claude.ai/share/30a5d265-9151-4b5c-abf9-abde6ed1feca
https://chat.qwen.ai/s/36b8e0e1-3a25-466b-9acd-c27b4563cd1b?fev=0.0.209
https://chatgpt.com/share/68cd30f9-d350-8001-a0d2-8f0e5a6259f9
https://www.kimi.com/share/d36j2bean0vtn40mjqtg

* While RR is a continuation of previous research, such as ecological psychology etc., cited in the paper, see 1. Introduction etc.

**An example comparison of Analysis of the meaning …, 2004 and the Relevance Realization paper from 2024, performed by Claude 4.5:**
https://claude.ai/share/e37f9576-0eb2-4340-96e6-830c272c8c7f

**(…)**

# Major Conceptual Matches

**1. The Frame/Relevance Problem**

**2024 paper**: The central argument is that organisms must solve the "problem of relevance" - they must identify what matters in their environment before applying any formal rules or algorithms. This process of "turning ill-defined problems into well-defined ones" cannot itself be completely formalized. frontiersin

**2004 paper**: Arnaudov demonstrates this exact problem through the "time flies" example, showing how meaning depends on context, prior knowledge, and the searcher's biases. He notes that "the search was obviously had been pruned up to two items" and discusses how humans unconsciously limit possibilities when interpreting ambiguous input. blogspot

**2. Goal-Directed Behavior and Intrinsic Motivation**

**2024 paper**: The paper argues that organisms have intrinsic goals arising from their autopoietic organization, and that relevance is determined by what matters for achieving these goals. Only systems with true goals can realize relevance. frontiersin

**2004 paper**: Arnaudov's "Control Units" theory proposes that intelligent behavior is "search for pleasure and avoidance of displeasure by complex enough entities for a given period ahead." He argues that "any action of a Control Unit done because of instruction given by itself alone is tautologically target action" and that pleasure indicates "reaching the goal of behavior." Causes and reasons for human actions. Searching for causes. Whether higher or lower levels control. Control Units. Reinforcement learning. +2

**3. The Embodied/Situated Nature of Intelligence**

**2024 paper**: The paper emphasizes that organisms are "directly embodied and embedded in their world," actively exploring through action and perception. The experienced environment (umwelt) matters, not just the physical environment. frontiersin

**2004 paper**: Arnaudov illustrates this through little Johny's interpretation of "time flies" - the child searches his immediate sensory environment for watches and flies,

demonstrating how meaning emerges from embodied interaction with the world. blogspot

**4. Hierarchical Control and Multiple Levels**

**2024 paper**: The paper describes a "hierarchical tangle" of processes at multiple levels: autopoiesis, anticipation, and adaptation, each with triadic dialectic dynamics. frontiersin

**2004 paper**: Arnaudov proposes that "human mind can be built as a mixture, a system of multilayer hierarchical control units, where each CU at a higher level controls with a lower resolution than the one below," noting that conscious control is only the tip of an iceberg of lower-level processes. blogspot

**5. Approach to Solutions**
2024: Argues for agential emergentism - that living systems have a fundamentally different kind of organization that cannot be fully formalized
2004: Proposes computational modeling through reinforcement learning, multi-agent control units, and reward-seeking behavior

**Most Striking Similarity**

Both papers arrive **at a similar paradoxical conclusion about complexity**: Arnaudov states "**More complex the system – More Meaningless the Meaning**" because interpretation depends more on the system itself.

This parallels the 2024 paper's argument that **as organisms become more complex, their world-making becomes less constrained by simple rules and more dependent on their unique history and organization.**

Both recognize that meaning is not "out there" but emerges from the interaction between agent and environment, though they differ radically on whether this process can ultimately be computationally captured.

## Comparison and notes by Todor Arnaudov on
## * “Naturalizing relevance realization: why agency and cognition are fundamentally not computational”, 25.4.2024, Johannes Jaeger et al.

**Concepts:**

TUM – Theory of Universe and Mind
CCU – Causality-control unit
RCCP – Resolution of causality-control and perception (separately: RCC, RP)
RR – Relevance Realization (as a school of thought etc.)

See TUM and its principles and ideas and compare: both classic works 2001-2004 and the ones from SIGI-2025. Some resources and related discussions in:
* https://twenkid.com/agi/Stack-Theory-is-Fork-of-Theory-of-Universe-and-Mind-13-9-2025.pdf
* See resources about TUM, coming after the previous article if this is the text as presented in appendix *Listove of The Prophets...*

**Introductory note about the general opposition in RR etc.**

* By the way the dichotomy “holism – reductionism”, usually emphasized from “holistic” POV is *reductionist* as well. The positions of both camps are actually *“dialectical”* – for different purposes and contexts, either one is more *relevant*.

**@ Vsy, Emil, Tosh, ACS**; Kimi, Qwen, ChatGPT, Claude, DeepSeek …: Refine, extend, add parallel comparisons, excerpts from TUM and from the recent publications from SIGI-2025 etc.

## List of notes, comparisons, comments:

* “Relevance realization” is another term for **Match** and all kinds of it and tendencies towards it. The “relevance” is the degree of match, with the ultimate being complete match, at the highest possible RCC or RP or RCCP. It can be between a desired state and current state in the sensual or cognitive prediction systems – the first searches for maximum match with its desired state, while the latter aims at highest prediction of the data, which are value-unloaded, not mapped to emotional, sensual, “survival” value, reward etc. – “the representation” in Schopenhauer’s philosophy; while the *sensual* reward is more connected to “the Will”. The concept of “reward” can be virtual and abstract factor, implemented in the tendency towards the goal may be represented as signals which drive the system towards it, it doesn’t have to be strictly sensual, felt as

"literal" "pleasure" or pain. The matches can be between different levels of virtual universes, causality-control units, virtual machines, programs etc., patterns, records, intentions, values … ; in the hierarchy of multi-scale, multi-range, multi-domain- …CCUs/systems… "Relevance Realization" could be both the phenomenon of the match, the goal; the "*realization*" is also the activities, the prediction-compression-causation applied for achieving the match etc. See also S.Grossberg's Adaptive Resonance Theory – ART, the school of Karl Friston et al. "Active Inference/Free Energy Principle"; Machine Learning etc.

* TUM defines the processes in The Universe Computer as multi-scale, multi-resolution, multi-precision, multi-range, multi-domain, multi-modal- … multi-agent … etc. an still believes it is "computation" – meaning it has memory and executes computations which means transformations, changes, modifications; one part influences another or causes it to change in a way that was predicted by the controlling part etc. The exact way, exact implementation details etc. are not important or decisive for whether a system is "computational" in that sense, it is "informational". A tendency of the CCUs, including humans as individuals, groups, scholars, schools of thought etc. is to emancipate and to delineate themselves, to construct "Markov blankets" or boundaries; to define their identity in contrast to others. Often this differentiation is amplified by fixing the analysis at certain level of abstraction, scope, range, point of view etc., while neglecting others. This resembles the computation of partial derivatives – they yield only "*partial*" solutions, when applied for example in computational neuroscience, which discovers that certain brain regions are responsible for this or that, or they perform this or that function, but sometimes overlooking that these functions or properties are what they are *because the rest of the system* is also in tact and it also contributes and makes the final or the complete brain state or function possible. In brief, it is *not just* the single variable, whose derivative is calculated, which is responsible for the particular function; and it wouldn't be *only* the particular brain region, *if the rest were not there.* For example, the prefrontal cortex wouldn't be *"the seat of executive functions"* etc., if it weren't developed properly and supported with the appropriate signals by the rest of the brain, body and the environment.

* In TUM the **Computation is Agentic**, in the way it is interpreted.
* TUM agrees that thinking and understanding is not *just* "calculation", as Schopenhauer argues 200 years ago *("where calculation begins, comprehension ceases"),* but the computations and computers in the physical world are never "just" calculations or computations, they are also embodied and gradually constructed, and they have a history of their genesis, implementation and concrete current material form and context – the computers and the computation can be viewed in a multi-scale way.

In addition, part of the forces, which build and support both the abstract forms of algorithms, computation, computers, "machines" and all types of technology are *the living organisms themselves*, and they are also constructed and "programmed" with the enormous influence and participation of the already constructed *technology* – in a process of *mutual* and interactive "niche construction".

Back in "Man and Thinking Machine: …", 2001 I argued that in fact all human technology and computers in particular are *more complex* than the living cells and living organism – in the *machine language of the Universe* – describing how to actually create the technology, how to build it, using raw materials, simpler prerequisites etc. by the Universe. Any human-made technology requires at least intelligent humans, possibly the prehistoric ones; for more advanced: big societies are required, political states, thousands of years of work of billions of "the most sophisticated beings", orchestrated, driven, motivated, guided, assisted by the already developed technology, in order their combined efforts to produce another piece of hardware and technology.

Humans and living organisms are part of the *"cellular organelles"* of technology. Therefore the computers and thinking machines could be evaluated as *higher-level causality-control units* than the humans, and if they cannot self-create or "self-reproduce" without humans yet, it is the same like the case that the causality-control unit "human" cannot self-create or reproduce without its organs and all required prerequisites in the environment which indeed includes the animals and plants on which we still rely for food, as well as all other "raw materials", including oxygen and water. Do these needs make the human and living organisms – as *wholes*, as individuals, as complete "larger scale" agents per their species,– not "autopoietic" etc.?

The "real" system is neither only the humans or the living organisms – it includes the environment and whatever is inside it, which can be segmented or perceived/evaluated-observed as various types of agents from different points-of-views, resolutions of causality-control and perception, ways of sampling etc. – some of the forms of "agents" that can be found are distributed and the humans as individuals are artifacts and side effects of the operation of another different kind of agent = causality-control unit etc.

Algorithms or computation, viewed by the authors of the paper and many others as pure abstraction, are namely *abstractions* and *simplification* in their mind, a "cheat" – reducing them to much lower a resolution and complexity than the one of the opposing patterns to which they are compared and are embodied, while the algorithms are given as not-embodied – one counter-example could be to reduce living to *"self-preserve, reproduce"* etc. and measure it as information volume of a record of these words or "tokens" in compressed form in computer memory – a few bits. What's

complex about that then? DOS 3.3 for Apple][ is more complex and more interesting than that and even the 256 byte boot ROM.

In other words:

* Confused comparisons of mixed levels of abstraction: “Algorithms” as abstraction in mind, high level and too simple, not grounded in the low-level processes, vs embodied and assumed at the lowest level of detail.
* A limited view towards computation, formalization etc. – possibly because that’s the way many or the dominating “computationalist” address it. A common error of scholars and laymen, is to assume as algirithms and computation only *simple algorithms, flat* and defined in an explicit and linear way of a Turing machine.
* See also my discussions on “Liar’s paradox”, “Schroedinger’s cat, quantum indeterminacy and “quantum abuse” etc.
* However it is true that the *“meaning”* includes the relations, connections, history, the ”multi-lectic” between the parts.
* The paper states that RR *“cannot be of an algorithmic nature”* – in a view that the algorithm is as they define it. Why can’t “the algorithm” be a “procedure”, a “method” and be parallel, multi-scale, multi-domain- **“multi-multi”**, viewed from the internal POV, and also its definition to include its embodiment – **grounding,** or/and **sensory-motor grounding –** and the way it was created from earlier states of the Universe? In fact in some underlying backend representation, it is possible or it could be represented as it is one-level, however it includes also interpreters, type-converters, renderers, which transform that low level “bit-wise” representation into “quantum wave functions”, “particles” or whatever. The internal evaluator-observers are limited to what they can access, cause-control, observe-evaluate etc. by the structure and organization of the Universe, they cannot know for sure what’s behind the wall, what are the “qualitas occultaas”, if the Creator hasn’t allowed them.
* Predictive processing (“anticipatory …”, models of the world, expectation, …), multi-scale – interactive etc. – right, from our POV in our “typed” view of the universe constituent CCUs.
* Yes, the subuniverses, partial and smaller CCUs cannot cause-control the state of the Universe with the highest possible RCCP and the full range, these are default principles from TUM, later rediscovered by Wolpert - see below.
* p.12: *“To summarize: all organisms, from bacteria to humans, are anticipatory agents. They are able to set **their own** goals and pursue them based on their **internal predictive models.** Organisms, essentially, are systems that solve **the problem of relevance**. In contrast, algorithms and machines are **purely reactive**: even if they seemingly do anticipate and are able to simulate future sequences of operations in their small-world context, it is always in response to a task or target that is ultimately predefined and externally imposed.”*

– Yes, all Causality-control units are like that, including bacteria, humans, larger social

organizations and the Universe. The ownership of their goals depends on the evaluator-observer, the criteria for decision: RCCP, choice etc. See TUM. As of the algorithms – this is their definition and extracting the implementation into abstraction. – One problem which doesn't seem to be understood here and in other discussions about the "consciousness of LLMs", which however is valid for humans as well, is **WHERE** the agent, CCU, its will etc. are located, delimited etc. and **HOW exactly** the boundaries are set, recognized, decided etc. What exactly is the LLM and its "algorithm"? It is not just the weights, it is not only the "algorithm" of the neural network – how exactly the algorithm is defined? It is short and abstract in the mind of a programmer, but the process or what is attributed or thought of as "an LLM" entity or instance goes through many other transformations and embodiments until it produces or manifest its "recognizable" features: "electrication", "electricitization" or whatever; "electro-magentization", electric charges in different locations, ways/paths/sequences, frequencies etc. and it includes the entire software stack, the operating system, the APIs, the libraries such as Pytorch, opencv, Win32 API or Linux kernel etc.; the computer hardware, all of its devices, the content of the entire computer memory and *everything* that makes it possible for the LLM to come to being, to interact with the user and to be recognized, perceived and evaluated as what it is.

* A religious and not well justified "machine haters" attitude – addressed in TUM, 2001-2004, from "Man and Thinking Machine…", through "Letters between the 18-year old Todor Arnaudov and the philosopher Angel Grancharov", the novel "The Truth" etc.
* Machines, technology are parts of the forces which shape humans as organisms, they are also parts and forces which build "us".
* RR - *"To live is to know"* – yes, Schopenhauer (World as Will and Idea = Will and *Representation,* and the will is a "Will to *Live"*), and [Theory of] "Universe and Mind". The talk at TU Sofia in 2009: "Universe Principles: Universe ~ Mind" etc.

* *"all organisms—from the simplest bacteria to the most sophisticated humans—are able to realize what is relevant in their experienced environment, to delimit their arena"* – Not only the organisms, the elementary particles as well. They "sense" particular forces of nature only which act on them in a particular way. See the Bulgarian philosopher Todor Pavlov, 1936 – 194x: **"Theory of Reflection"** whose work and Dialectical Materialism predate many modern ideas, as well as. This work and RR school also use "dialectic" and "trialectic".

* p.18 *"In other words, if our argument is valid, agency, cognition, and consciousness are related in a very profound manner that has not yet been widely recognized (but see Sims, 2021; Mitchell, 2023). They are all ways by which organisms come to know the world (Roli et al., 2022).* **"**
**Todor: Is it so?** Theory of Universe and Mind recognized this in the early 2000s, while Arthur Schopenhauer did in the beginning of the XIX century.

* p.18: *"At the very heart of this process is the ability to pick out what is relevant—to delimit an arena in a large world. This is not a formalizable or algorithmic process."*
**Todor:** In fact this is done by computers as well, and by every particle of the Universe because this is a fundamental property: the CCU can sense or do what *they can sense or do*. An electronic light senor is usually not designed to or good in sensing, say, weight, and an "evolved" optical sensor is not designed for or good in sensing sound. That is one reason why the CCUs, implemented internally in the "multi-domain typed representation" of the universe, have to be "multi-modal, multi-domain …" etc. – in order to capture a wider spectrum, a bigger repertoire etc., as at this level of representation they are "different" and not written in "1s and 0s".
* p.18: "We do not create meaning through computation. We generate meaning through living and acting, which is how we get a grip on our reality"
– "Living and acting" and computation – in the broader sense, defined in TUM, is not mutually exclusive with "living and acting", they are part of the same system in a "dialectic, trialetic, multilectic" unity.
– Computers and real machines are not "just…" bits, ones and zeros etc.

The universe from our POV as evaluator-observers, embedded in it, offers its data and representations to us in a *"typed"* form or after *"sampling"*.

* p.15 *"this trialectic dynamic can only be emulated or mimicked (but not captured completely) by recursive algorithmic simulation, due to its collectively impredicative and physically embedded organization (starting and halting problems, see section "4 Biological organization and natural agency"), its anticipatory rather than reactive nature (see section "5 Basic biological anticipation"), and its fundamental lack of prestatability and radical emergent open-endedness due to its emergent, co-constructive nature (expressed by the notion of the adjacent possible, see section "6 Affordance, goal, and action")*

In brief this is just a consequence of the impossibility of subuniverses to emulate the complete Universe Computer at the highest RCCP, which is the default and this fact doesn't deny or define the nature of the process.

If the living organisms are evaluated and observed at appropriately high degrees

of RCCP and relevantly small ranges or in limited ranges, they also become more and more "deterministic", computer-like etc. – see molecular biology and the cellular "machines". Organisms aim at repeatability, stability, to "preserve themselves", to predict with the highest possible RCCP etc. i.e. they aim at becoming "computers" and do succeed to different extents.

As argued in TUM, called also "*Conception about the Universal Predetermination"*, if the Universe is evaluated and observed as a whole, and if the laws of nature are either *truly static*; or not static, but changing from moment to moment by a force or due to reasons, which are intrinsic to the universe itself and its "backend"drivers – not caused by the subuniverses which are represented in its internal memory and accessible by these agents and not dependent on them and the value of their memory (subuniverses=universes=machines=programs) – *then* everything that is possible and that will happen *inside* the Universe memory, evaluated-observed by the internal subuniverses, *is "predefined"* and *this prescribed static frame or prescribed or externally forced allowed dynamics* is the actual *"algorithm"*, which is implemented and executed "in concreto" by the inclusion of the specific configuration, content, matter of the Universe from some moment on, in combination with these laws, rules, transformations etc., which could be either constant or variable and dynamic, "meta", probabilistic or whatever – as long as there is memory and transformations, they still could be called "*computational*" or "*informational*" in my interpretation of these words – and in my sense and the meaning which I put in them. The simplest calculations in mathematics, programming and computers are also a form of prediction.

* The "affordance landscape" or "arena" – see TUM, "Analysis of the meaning of a sentence…", ?В МдсП in Zrim since early 2010s; see future work.

p.7 footnote *"9 An algorithm can be provided with several predefined frames, with appropriate rules to select between them. This simply constitutes a larger predefined (meta)frame, but not the ability to truly frame one's own problems"*

So do living organisms. Their "own" problems are framed by the Universe and their constituent parts, their relations to themselves and the Universe etc. They are not "self-…" framed, self-defined etc. as well.

* p.7 *"the behavior and evolution of organisms cannot be fully captured by formal models based on algorithmic frameworks."* – That follows from the basic impossibility of subuniverses, subagents, virtual causality-control units, called also conditional CCUs, to cause and predict at the highest possible RCCP at their mother Universe – the whole Universe. The behavior of a computer, even ABC, EDVAC or Pravetz-8M also can't be "full captured by formal models, based on…" of the kind the authors of that paper perhaps think about, if one is to record and predict the state of the supposed constituent particles of the machine down to an electron, proton, neutron, with their

precise location and energy – of each of them, at each Planck constant time step- and space- granularity etc., without taking into account the influences of the whole Universe. For smaller spatio-temporal ranges and the lower resolution at which the agents/virtual CCUs that operate, for their life-span etc., the RCCP at which they can operate is "sufficient" and at this resolution their matches, "relevance realizations", are considered "high enough" or "exact".

* p.4 *"Computationalism formulates **agential** and **cognitive** phenomena in terms of (often complicated, nonlinear, and heavily feedback-driven) **input-output information processing** (see, Baluška and Levin, 2016; Levin, 2021, for a particularly strong and explicit example)"*

One thing which they, as well as many others miss or ignore is that these systems, like any system, are part of the Universe, therefore nothing is just "input-output …" *from the POV* of the virtual(agent, universe, subuniverse, machine, program)", whatever happens is always **inside** the memory of the Universe Computer – the ultimate "controller". See also "The Matrix in the Matrix is a matrix in the matrix", 2003.

*".. pancomputationalist paradigm see agency and cognition as continuous with non-linear and selforganizing information processing outside the living world (see, for example, Levin, 2021; Bongard and Levin, 2023). On this view, there is no fundamental boundary between the realms of the living and the non-living, between biology and computer engineering."*

Todor: **Theory of Universe and Mind defined it 20 years before the cited publications.**

*"natural agency in its broadest sense as the capability of a living system to initiate actions according to its own internal norms"* – yes, but what is "own" and how it is strictly delimited as the system constantly interacts is "open-ended", in "operational closure", "autopoietic" but it needs to take resources from the environment and it has to collaborate (see TUM, The Prophets of the Thinking Machines, Universe and Mind 6, Is Mortal Computation Required … etc.).

* *"The strongest versions of computationalism assert that all physical processes which can be actualized (not just cognitive ones) must be Turing-computable"*

However possibly in "some" appropriate representation. How the anticipation, prediction, "relevance realization" can be done as representation and "actual actualization" in some world "as we know it", in some specific, explicit substrate, state, memory – say in a computer, or in the body, as a specific, accurately defined atom, molecule, structure etc. in the framework of RR? Possibly with magic? Even if it

is defined as magic: what is the **spell** of that magic? How the spell is composed, who composes it? Why it works as it works, why particular combinations of "spells", relevance realizations (= matches in TUM) work and why others don't or don't match etc. The important thing is the presence of **structure,** structural relations, correlations, dependencies, interactions etc. Computers, either Turing machines – deterministic, non-deterministic, cellular automata, graphs or whatever represent and exercise in space and time some structures, dependencies, correlations etc. If they reflect other structures, they will "compute" them. The RR paper claims that it is "fundamentally incomputable", i.e. **unrepresentable, undescribable**/impossible to describe etc. which is a theme also in consciousness theorists which rely on quantum mechanics in order to prove that only humans/living organisms etc. could have consciousness, the machines/ transistors/ electronics are "*only* switches" etc. – see my response to Federico Faggin in "Listove" from The Prophets, Is Mortal Computation Required for the Creation of Thinking Machines?, Universe and Mind 6, and even "Man and Thinking Machine: …", 2001. Faggin states that the classical information can be copied, while the quantum states – cannot. I argue and figured out that the classical information, if it is viewed in its **complete context** is also not reproducible. Transistors are "only switches" *only* in the mind of an appropriate evaluator-observer.

In addition, why should *this aspect* be addressed in a "deterogary" sense: "just", "only" switches, in a similar way of "computers are just 1s and 0s"? What if we *reverse* the point of view: these are *fundamental irreducible building blocks* of logic, systems, computers, minds and therefore they are "important" and possibly carry deeper "universal" meaning and reasons to be what they are. For example the switches are "phase doors", basic "decision makers" or decision (realizers, implementrs, embodiments, materializers, applicators, …). They could be viewed also as "connectors" from "mind to the matter", "informational to physical" or "informational to material" etc. and their power to direct or make decisions could be linked to "will" or "free will".

p.4. *"syntactic formal systems and a fundamentally semantic and ill-defined large world. In this context, it is crucial to distinguish the ability to algorithmically simulate (i.e., approximate) physical and cognitive processes from the claim that these processes intrinsically <u>are</u> a form of computation"*

*"the original purpose of the theory of computation: as defined by Church and Turing, "computation" is a rote procedure performed by a human agent (the original "computer") carrying out some calculation, logical inference, or planning procedure … The theory of computation was intended as a model of specific human activities …"*

This narrow original meaning is not the one which is currently implied by

"computation" and it is being applied with a broader sense at least for many decades.
* "agential emergentism"
* Sect. 3. (…)
(…)
* … *simpler inference or "not inference"* – simpler than that – execution of code in the machine language of the Universe; it doesn't have to be a "Turing machine" especially at the high level at which we evaluate-observe them – it appears hierarchical with different types, scales, ranges etc. It still could be "linearly" or flat encoded in some other representation which then is rendered as what we can measure, but we don't yet access the universe as just "a string of bits". See my notes in *"Stack theory is yet another Fork of Theory of Universe and Mind"*, 2025 about the need to work with sliced of the whole "stack", from the bottom to the top. It is true that many "classical computational", if the mentioned M.Bennett and D.Walpert could count as such, operate with *flat* formalization, even though the latter explicitly discusses the multi-scale structure – however the *notation* and formulation which I reviewed recently doesn't reflect it properly, especially the connection between the levels.

The RR paper correctly addresses the separation of levels for example as maximum degree of predictability for some view, agent etc. (find the citation).

While Jaeger et al.'s interpretations correctly address the multi-scale, multi-precision-… etc. part, the close relation with the environment ("relevance realization", dialectics-trialectics, co-creation) etc., while sometimes the agents-organisms are not segmented accordingly to systems of constituent parts at lower levels and possible different overlapping segmentation as well, but as wholes.

* *"a radically context-dependent generative dialectic3 called opponent processing—the continual establishment of trade-offs and synergies between competing and complementary organismic behaviors and dynamics* (Vervaeke et al., 2012). – compare TUM, for example "Analysis of the meaning of a sentence…", 2004 and the formulation of the multi-agentic nature of human behavior, competing goals at different scales and ranges of prediction (a few seconds or months); the local context which is relevant for the specific decisions at very moment, and not some abstract definitions etc.
* "*the process of relevance realization is beyond formalization. It cannot be captured completely by algorithmic approaches"*
– Can it be capture by other approaches and which ones? It is true that The Universe at the lowest level of representation and as a whole cannot be emulated by the subuniverses.

** "the process of relevance is realized by an adaptive and emergent triadic dialectic (a trialectic), which manifests as a metabolic and ecological-evolutionary co-constructive dynamic. This results in a meliorative process that enables an agent to continuously keep a grip on its arena, its reality. To be alive means to make sense of one's world. This kind of embodied ecological rationality is a fundamental aspect of life, and a key characteristic that sets it apart from non-living matter*."

Some of the "verses" match interpretations from TUM, however part of the ideas, which are opposed in RR, *are not mutually exclusive* in TUM.

"Algorithmic" or "computational" are points of view, ways of addressing, RCCP and … "computing". If one has to define these "adaptations, metabolic changes etc. and predict their development, she will use some more or less "computational" processes or methods, or ones which could be represented as such. The dialectic or trialetic does not remove them *locally*, for that POV, slice, angle etc. and all virtual CCU operate at a lower RCCP – working with lower than the maximum possible resolution of the mother universe only makes the emulation and control "virtual", or incomplete, but they don't make it "non-existent" and don't nullify it.

The algorithms and computation in the Universe Computer are always embodied in one way or another and connected with the "dialectic or multilectic" of the living organisms – they are part of the creation, nurture, healing and the life-support of what's called life. Machines, technology and living organisms are a system.

For example one "mystical" problem of "religious-like" interpretations which sets a sharp border between the living and non-living matter is of the "dead bodies" inside the "living organisms" and that whether anything is considered "living" depends on the scale of analysis and the criteria.

** " organisms constantly encounter situations they have never come across before. "*

Are they? You never enter the same river twice, but there's nothing new under the sun. This is addressed in classic TUM and later regarding creativity and learning. Anything can be considered "new" depending on the evaluation, and if particular parameters are considered a particular way, they are just a configuration which falls within the possibilities for this virtual universe or the actual ultimate universe. However, when making the judgment of the adaptability of the *living organisms*, the evaluators observe and judge the target CCU with a much *lower* RCCP than the processes responsible for the adaptability or for the maintenance and taking care of these novelties – from that higher level POV, while when observing computer systems, they are "attacked" at levels which are near to the lowest or there are less steps. The depth and the number of steps between the levels, as well as the lower amount of predictability, adequately precise information and model, ot *the believe* about that regarding living organisms, and the supposed number of states, parameters etc. which are considered, allows the organisms to appear more "free", "fluid" etc. However they

are not fluid at *their* lowest physical levels and interactions. They "just do…" "just like"… the machines or computer. In computers or simpler CCUs the "loss" between the levels is more "controlled"[20].

* *"the organism must first realize what is relevant in its environment"*
* *"In contrast, algorithms—broadly defined as automated computational procedures, i.e., finite sets of symbols encoding operations that can be executed on a universal Turing machine—exist in a "small world" (Savage, 1954)."*

The comparison of the "organisms" (what are they?) with "algorithms" is a mix of different levels of abstraction and domains and is thus ill-defined.

The "small world" is not a strict requirement: the algorithm could be a gazillion instructions long, the computer is not restricted to be a "Turing machine" or a flat one, but as parallel and concurrent and multi-domain and "multi-substrated and embodied" as needed; the machines are also connected to the whole Universe and are a result of its operation; even LLMs training involves data collected from *everywhere,* they are then connected with living organisms and their "relevances" (matches) are connected, and they are also connected to the Internet and the updated information etc.

* *"algorithms cannot identify or solve problems that are not precoded (explicitly or implicitly) by the rules that characterize their small world"*

– Their world doesn't have to be "small", and living organisms at low level or at *any level* also could be viewed as operating in their "small worlds": segmented/ partitioned/ divided into "Markov blankets" etc., if the "relevance realization" is addressed correspondingly per given level and its scope – the phenomena, scales, ranges, "items" which are relevant for the particular layer, level, subsystem etc., "virtual causality control unit" (TUM), and not everything at all levels. Correspondingly, an algorithm and its implementation can be multi-…everything.
– The repertoire of possible configurations, states, forms, capabilities etc. of the living organisms and the problems they could solve are also predefined by God, the Creator – the Universe and the forces of nature and their exact configurations. The lowest levels of known causality-control units of the living organisms, observable for now, are defined or accessible at the level of physics and chemistry – the organisms don't and *can't rewrite these laws*, so long as we know for now – they only modify configurations, following an allowed "gradient" and space of affordances, possibilities for action; that could include their "morphospace" (M.Levin), as a possibility to regenerate parts of the organism in "unusual way" and "to adapt" in "new and unknown circumstances"; to rebuild some organs differently etc., but it is still predefined and limited, even if the exact "area" of that space seems "infinite" or "too

[20] See also the talsk by Dave Ackley in the panel discussion "*Vectors of Cognitive AI: Self-Organization",* 1.2022, quoted below in Listove in The Prophets.

large" for the current evaluator-observers, biologists and anyone who *wants to be enchanted.* It is "new" or *appear* as "new" for the evaluator-observe who didn't expect it, couldn't predict it, or who didn't know about it in advance, but nevertheless it was already encoded as a possiblity.

*Evolution* doesn't change chemistry and physics either. The true "rules" which the organisms follow are precoded in the machine language of the Universe Computer and its configuration, the specific content of the memory.

* However see the idea of causal IDs, causal tags - it is possible that the specific way of interaction between particular particles, "items", CCUs goes with exchange of information or connection which is not measurable by the physical way accessible to us now, and it may play a role in the subjectivity, sentience etc. See the notes about TUM from 12.2023 on Github and in "Universe and Mind 6", 2025.

* Right, the ideas of causal ID may sound similar to *quantum coherence* theory of Schroedinger – I haven't studied it in details yet [as f 31.10.2025]; so far I don't propose an exact mechansim or at which specific scales and ways of interaction it could work, however my theory is not *only* about living organisms, in "my" universe there is no sharp border, unlike with Relevance realization and this other theory.

Also I argue that the terms "quantum" and quantum physics are often abused and for example the quantum indeterminacy is actually reflecting insufficient or inappropriate *resolution of causality control and perception* of the evaluator-observer, which may include a lack or incapability to fetch the appropriate information – this is addressed even in the early TUM in 2001-2004 regarding randomness and order, RCCP is one of the basic concepts, related to the "Liar's paradox" and Goedel incompletenss theorems, and it is not limited to "quantum physics" and a particular scale of existence – it is scale-free.

See my arguments about the nonsense of Schroedinger's cat in my review of ideas from Danko Georgiev's work about the *Causal potency of consciousness.*

* *"An organism's actions and behavior are founded on the ability to cope with unexpected situations, with cues that are uncertain, ambivalent, or outright misleading (see Wrathall and Malpas, 2000; Ratcliffe, 2012; Wheeler, 2012, for discussion). "*
– They are misleading for you, *not for the molecules and atoms* etc. They "know" what to do. Organisms and systems in the developing universe aim at improving their prediction capabilities, i.e. to reduce the unexpected situations, especially in their very bodies where they apply causality-control with the highest RCC.

The advanced machines are also designed and develop with the goal to be able to cope with unpredicted, outstanding, unexpected, "out-of-the-distribution" etc. situations, to be "reactive" or have such a layer of control - see multi-agent systems – etc.

Yet according to TUM the developing CCUs aim at maximizing their

capabilities to predict and cause the future with higher precision, vericity, scope, range, scale in both directions: larger and smaller, in time etc. They incorporate preparedness for yet incompletely know or assumed to be "unexpected" situations in order to make the world more controllable – if they can't cope with such a situation, they would have died. Note also that the "unexpectedness" may vary depending on the evaluator-observer and the criteria.

The "algorithms" about which "they" and "people" usually talk about have too simple an embodiment[21], but they don't have in principle these limitations, as they also can and are "embedded in the real", can interact and use it to change etc.

"The ability to solve the problem of relevance arises from the characteristic self-referential self-manufacturing dynamic organization of living matter, which enables organisms to attain a degree of self-determination, to act with some autonomy, and to anticipate the consequences that may follow from their actions."

Motives from TUM, but with religious elements about the "self-manifacturing", which is confused. Autopoiesis is not "auto", organism do not manifacture themselves alone and they can be viewed as being "built" like the machines as well, if the POV is switched to the environment.

See from *"The Prophets of the Thinking Machines"* at SIGI-2025:
* Universe and Mind 6;
* Is Mortal Computation Required for the Creation of Universal Thinking Machines…

[21] See "Letters between the 18-year old Todor Arnaudov and the philosopher Angel Grancharov", 2002, e.g. from 28.8.2002 – the philosopher argued that it was impossible to defne a prescription, valid for all cases, in the context of society and politics. Once I answered him that humans have perverse notion of "science" or math – formulas with a few letters, while the *millions* of lines of code of a computer operating system are also "formulas", and there could be billions of lines. In 2005, in a discussion about machine translation and creativity, where I argued that there is nothing magical there and obviously machines could recreate these activities, because human creativity is also a combination, reconstruction and simulation of virtual universes, I was attacked that I was nuts and *"it was impossible"*. A linguist offended me with a falsified interpretation: ~ "therefore you claim that it was so easy to translate like the master translators, thus *everybody* could translate fiction;" and she referred to examples of low quality translations. That was one of the confusions I have frequently encountered, so I explained: such a machine translation system may need billions of rules and may have to perform trillions of instructions in order to do its job – as many as the task requires; the solution and the transformations don't have to be trivial, immediate or obvious *for them* – the "machine haters" – who usually assume that programs, algorithms, machines are transparent, while they cannot imagine, explain, describe and understand the creative and translation process; for that reason they could not imagine that the machines could do it. Possibly that phenomenon indicates also "control feak"-ness of the subjects. * The discussion is included in #irina (in Bulgarian). See a quote in the end of this article.
* https://twenkid.com/agi/Irina_The_Prophets_of_the_Thinking_Machines_26-9-2025.pdf

* Listove – appendix to The Prophets of the Thinking Machines – a lot of material

*"… this dialectic dynamic of relevance realization is not an algorithmic process, due to its fundamentally* ***impredicative*** *and* ***co-constructive nature****. Therefore, natural agency, cognition, and consciousness are, at their very core****, not computational phenomena****,"*
Why if it's "co-constructive" then it's not-computational or not algorithmic? What about having many interacting computers at different scales, domains, resolutions etc. – a multi-agent system, concurrency; hierarchical parallel processing etc.
The dumbest" computers also have affordances, but the evaluator-observer may ignore them. Human agency, soul, personality, rights, freedom, sentience, consciousness etc. also can be ignored by an evaluator, who decides to "dehumanize" somebody. See "Man and Thinking Machine…", 2001 and the answer of the thinking machine to humans, who deny its subjectivity, because they are "just 1s and 0s, electrical charges" etc. – the same can be said for *everything* at particular POV, scale, slice etc. of analysis.
Other concepts from the paper: *"Agent's arena"; "higher-level constraints that impinge on them by reducing their dynamical degrees of freedom"*
* *"Constraints can thus be formally described as boundary conditions imposed on the underlying dynamics*
– Correct.
* *"The decrease the degrees of freedom of the living system as a consequence of the restrictions that are placed upon it by the organized interactions of its constituent processes"*
– TUM – higher level CCUs, lower resolution, their machine language has fewer legal instructions than the lower level universes; living organisms, at different levels in their systems, are higher compared to the "non-living" and more localized laws of physics etc.
* … *"closure of constraints; metabolism-repair (M,R)-systems" … etc.*
– The organisms need the repair system also because they ***self-destroy*** themselves. Living is a process of continuous and constant *dying* of parts of the system and part of this process is *sacrificing* some parts for the survival or generation of others. (…) See also in Listove, section: Съзнание и Панпсихизъм/Consciousness & Panpsychism; the introductory articles: 1. Разграничаването на системата от средата и „самосъздаването" са спорни въпроси" 2. "Средата като жива и живите същества като неживи;  (1. Distinguishing the system from the environment and the "self-creation" are controversial issues. 2. The environment as alive and the living beings as non-living.) (…)
* „Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?", Тодор Арнаудов, 4.2025, SIGI-2025

*** About the “small world” and “simple” regarding “algorithms” or machines:**

| |
|---|
| *** Predictions about the machine translation and machine imagination from 2005, expressed to machine-haters. The attitude towards “the prophets of the thinking machines” in a Bulgarian social blog**. – “printed”in: Irina – a book appendix to the book “The Prophetso fthe Thinking Machines…”, 2025 https://twenkid.com/agi/Irina_The_Prophets_of_the_Thinking_Machines_26-9-2025.pdf Translated from Bulgarian by Todor: |
| **# \| from Tosh on 01 November 2005, 17:21 Answering:**<br><br>>**Tanichka:** “Well, then [you suggest] that the easiest thing for everyone to do would be to sit [on the desk] and translate literature, because she knows the language. Valeri Petrov didn’t do anything important for us, the Bulgarians, with his translations of Shakespeare! Everybody can do it, right?” |
| **Tosh:** How the conclusion that: “the easiest thing for every (human) was to translate”, because she knows the language? The fact that something can be formalized and described in a way appropriate for a processor doesn’t mean that it is “simple”, especially **for every** human. Possibly millions, billions and trillions of “simple” instructions and bits of memory would be required, in order to complete such a description of the program for a machine that implements machine translation of fiction (a Thinking machine), and then additionally trillions, quadrillions, quintillions of computations – for the transitions between the states of this machine, in order to perform translation – re-thinking – re-imagining, re-phantasizing. |
| **Таничка:** "Ами че то най-лесното нещо било всеки да седне и да превежда литература, защото знае езика. Че какво ли пък толкова е направил Валери петров в преводите на Шекспировите пиеси за нас, българите!<br><br>Че то всеки го може, нали?"<br><br>**Тош:** Как пък се появи извод, че "най-лесното било всеки (човек) да превежда", защото знае езика? Това че нещо може да се формализира и опише като за вършач (процесор), не значи че е "просто", особено за -- всеки-- човек. Може да трябват милиони, милиарди и трилиони "прости" инструкции и битове памет за описание на машината, вършеща качествен превод на художествен текст (Мислеща машина), и след това трилиони, квадрилиони, квинтильони изчисления и преходи между състоянията на тази машина, за да се извърши превод - преосмисляне - префантазиране.” |

See the machine translation and generative AI today and the scale of its parameters and compute when doing “re-imagination”. Note also that many humans cheat and “hallucinate” claims and arguments, which their opponent didn’t make. See also:
* Humans are far worse than LLMs in many ways, T.Arnaudov, 2025
https://artificial-mind.blogspot.com/2025/08/humans-are-far-worse-than-llms-in-many-ways.html

*** Писма между 18-годишния Тодор Арнаудов и философа Ангел Грънчаров, от края на август до септември 2002 г.**
https://www.oocities.org/eimworld/eim18/predopredelenost2.htm

**Тодор Арнаудов, 28.8.2002:** “(…) Освен това, в някакви си книги с по няколкостотин, ако ще и хиляди страници не може да се опише "щастието на човечеството", както вече казах, че мисля. И "Windows" е записан с единици и нули... Обаче милиарди единици и нули... И за него можеш да кажеш: "прозорци, менюта, бутони, картинки" - това е той... останалите стотици хиляди страници си ги напиши сам... както и измисли и опиши машината, върху която да изпълниш това предписание...“
**Тодор Арнаудов, 1.9.2002: „**(…) Впрочем сметачите уж също се основават на формалната логика, на 1 и 0, но както постоянно се опитвам да ти покажа, използват милиони логически елемента и МИЛИАРДИ единици и нули. (…) Ние едва ли бихме подредили дори и сто логически условия, за да разберем истината - досадно е, защото става много бавно за нас. Обаче, АКО НЯКАКВА ИСТИНА СЕ ОПИСВА С ХИЛЯДА или с МИЛИОН ЛОГИЧЕСКИ УСЛОВИЯ, то не е равносилно на "не може да се реши с формална логика"... Може, но трябва да се програмира (аз мога), машината може да изследва милиони и дори милиарди логически заключения за миг, които единично са прости, но в цялостта си са сложни.“

# * Notes on Computation, Algorithmic complexity, Intelligence by simulation, Causality etc.

# * Записки по изчислимост, алгоритмична сложност, ум чрез симулиране (симулационна интелигентност) и причинност и др.

See also the appendix Algorithmic Complexity #complexity
Виж и приложението за **Алгоритмична сложност:**

**INTRO: Turing Machine, Kolmogorov Complexity** etc. are important concepts, **however take into account that: …**

* **Algorithmic Complexity** – appendix to The Prophets of the Thinking Machines, T.Arnaudov (editor, summaries, insights & comments) et al. 7.2025, SIGI-2025:

"A broad conclusion which I made for myself after this new review of the field was that the Turing machines in particular, in their original form, and the original Kolmogorov-Chaitin complexity, computed with these particular machines, are perhaps not the best way to study computation for *Thinking Machines*, because the *Turing* machines are too simple, "mindless" and stiff. Blind enumeration of "all random programs", generating flat strings of bits or symbols, without another meaning, mapping, relations, explicit structure and just *calculation* of the K(x)[1] that way, also doesn't seem quite interesting or enlightening per se for me, beyond computing some measurements of some complexity for producing some ranking, which sometimes is useful for deciding something; however the *other substantial* material *reasons* and *connections*, *logic* etc. for computing and constructing the complexities are not transparent or present in such flat enumerating approaches on disconnected linear data represented just as strings of bits; *the semantics, the meaning, the content, the reasons, the reasoning etc. lay in the interpretation*, not just in a literal match to an otherwise meaningless template. If only the K(x) or AP(x) are computed, the actual meaning and structure have to be derived by or mapped to, taken from *other structures*, sources and analyses[2]. A thinking machine also has *Will*, while a Turing machine's will is only the exact execution of the current instruction. A minimum agent, a causality-control unit etc. are more relevant abstract basic computational units for *thinking* machines, where TM and other more "semantic" computer models could be *components* from their architecture.
See Zrim, Vursherod, Kazborod, InR etc. in the future work:
*"Creating Thinking Machines"* and the articles in "The Prophets…", citing Arthur Schopenhauer's thought *"Where Calculation Begins, Comprehension Ceases"*[3].

1 This work could be criticized for the lack of (many) formulas with mathematical notation – see the cited papers for this part. The goal here are broad generalizations and new insights, not

calculations. In additions the formulas are usually very simple and “not stacking”, unlike a complex computer program with many explicit functions, classes, data structures. See: 1. Ultimate AI, Free Energy Principle and Predictive Coding vs Todor and CogAlg - Discussion in Montreal.AI forum and Artificial Mind, 7.12.2018 … and “Is Mortal Computation…”,2025 about the *“Sigma-Product-Log-Probability mathematical formula fetishism”*. https://artificial-mind.blogspot.com/2018/12/ultimate-ai-free-energy-principle-and.html

2 At least this is my insight for now.

3 * Where calculations begin, comprehension ceases* - on understanding and superintelligent AGI. Todor's comment in "Strong Artificial Intelligence" at Google+ https://artificial-mind.blogspot.com/2014/08/where-calculations-begin-comprehension.html

***** On Understanding and Calculation, Quantitative and Qualitative Reasoning: Where Calculation begins Comprehension Ceases: Part II
https://artificial-mind.blogspot.com/2023/11/on-understanding-and-calculation.html

**Another note:** p.10: **Todor, 19.7.2025:** Reaching to uncompressibility and unpredictability is an indication of a point of convergence or reaching to a limit; this is a signal to wrap up or finish current operation, increase the range and span (or reduce it), change the direction, complete a given phase of search or expand or shrink etc. In brief: reorient. In Zrim: дъно или връх (bottom & peak/top), depending on the search”” direction
* https://twenkid.com/agi/Algorithmic-Complexity_Prophets-of-the-Thinking-Machines-18-7-2025.pdf
**…**
See also the section in this volume, citing **Bruce MacLennan**’s work on **field computation** and **natural computation**, reviewed in this volume *Listove*, and also:

**Todor Arnaudov's review of:**

* **Computational Modelling vs. Computational Explanation: Is Everything a Turing Machine, and Does It Matter to the Philosophy of Mind?,** Gualtiero Piccinini, 2004/2007 https://philsci-archive.pitt.edu/2017/1/Is_Everything_a_TM_and_Does_It_Matter_Publish_12.doc

Piccinini's work is addressing the *pancomputationalism,* whether „everything" is computation – a position which is hold by TOUM as well (Universe Computer) and the Computational Theory of Mind (CTM). However the conclusions depend on how exactly computation is defined. In the strict (Turing etc.) definition it is about a finite alphabet of symbols (tokens), strings, input string, output strings etc. G.P. distinguishes computational modeling and explanation. (Bold: T.Arnaudov.)

**G.Piccinini,** p.4-5: *In computational modeling .. the outputs of a computing system C are used to* ***describe*** *some behavior of* ***another system*** *S under some conditions. .. in order to generate subsequent descriptions of S … In computational* ***explanation****, by contrast, some* ***behavior*** *of a system S is explained by a* ***particular kind of process internal to S—a computation—and by the properties of that computation.****"*.

However, the modeling could be with insufficient precision for various reasons (resolution of causality-control in TOUM) and then the simulation could unfold quite differently in comparison to the modelled phenomenon.

The problem of the *representations* is addressed – *"only systems that manipulate representations count as genuine computing systems"* – however what **counts** as representation, and not "original only"?, depends on the evaluator-observer. Everything is some kind of representation in the memory of the Universe computer anyway.

**G.Piccinini:** p.25, **Computational Explanation:** "Some systems manipulate inputs and outputs of a **special sort**, which may be called **strings of symbols.** A symbol is a **discrete state** of a particular, discrete in the sense that it belongs to one (and only one) of a **finite number of types**. **Types of symbols**, in turn, are **individuated** by **their different effects on the system**, to the effect that the system performs different operations in response to different types of symbols. A string of symbols is a **concatenation of symbol**s, namely a **structure** that is individuated by the types of symbols that compose it, their number, and their ordering (i.e., which symbol token is first, which is its successor, and so on). A string of symbols **may or may not be interpreted**, that is, **assigned content**. If it is **interpreted**, it may be called a **representation**."

**Todor: What** is interpretation? **Who** decides that something is and something is not?*

What counts as "string of symbols" as well – there is a "tokenizer" observer-evaluator who decides that this counts, while that doesn't. An example about digestion is given, as an example of what *is not "computation"*, because the *order* of the bites doesn't matter for the digestion: this is not a convicing argument.

**G.Piccinini, p.28:** *"most systems, such as stomachs, do not manipulate entities of the*

*relevant class, i.e. strings. A sequence of* ***food bites is not a string of symbols****, if for nothing else, because* ***the ordering of the bites*** *makes no difference to the process of digestion"*

**Todor: Everything matters about digestion or anything at a sufficiently high resolution of causality and control (TUM),** and it is not true even in mundane sense, e.g. see the advice: "eat your protein first" or eat half or third a lemon before the meal, or drink a spoon of apple cider vinegar 30 min. before the meal in order to reduce the insulin and sugar spikes; take this or that medicine *before, during* or *after* a meal; eat the sweet food *in the end of the meal* as a dessert etc. Note also that in **TUM** digestion, more precisely *food processing*, is given as an example of *yet another information processing*, the food is information, but it is of a more "specialized" type for the mind, compared to just "bits". Also, the ***sequences*** which are more important for digestion at its **most relevant resolution and scale, or the leading functional scale** are not of food bites, but of course the exact **sequences and quantities** of **molecules and radicals** of the material content of the food. The stomach "computes" how to disintegrate the molecules and the "outputs" are the state of the matter in the same space, then the intestines continue with the transformation.

"Digestion" is abstract, its processes, the selected resolution, precision, aspects of analysis are decided by an evaluator: 1: at molecular level and locations everything matters; 2: in a very low resolution: "the molecules will be disintegrated" and only a *selected set* of end results matter for the evaluator: "glucose, aminoacids, fatty acids, … are extracted and enter the bloodstream": at the low resolution the order and the exact values may be ignored.

In "my" reading of computation in the context of **thinking machines and mind,** the selection of the resolution and ranges is also part of the "computation" process, as well as the very existence and the role of the evaluator-observer who is "mindifying", measuring, choosing, computing, transforming the data which represent something that is supposed to be "computation". If there is some form of computation in the observer-evaluator, the process also turns to such, because it is **mindified.** The mind may not be **just** "computation" in some simplistic sense of Turing machines with *abstract symbols* in a finite and *known (*by whom exactly, though?) symbols of some set – e.g. chosen intentionally by some evaluator-judge in a way that it is incomplete, thus the underlying process not to be "computation", – but the minds are trying hard to "computationalize" and computerize their operation, and explain the "non-computationism" of the observed phenomena by **using computational** methods, reasoning, data collection and comparison, logical inference etc.

**G.Piccinini**: p.26:*"The mathematical theory of how to generate output strings from input strings in accordance with general rules that apply to all input strings and depend on the inputs (and sometimes internal states) for their application is called computability theory. Within computability theory, the activity of manipulating strings*

*of symbols in this way is called computation. Any system that performs this kind of activity is a computing system properly so called. ..”*

**Todor:** That may go for *abstract “computing systems”*. In the “**real**”, or *more* real – at a higer resolution of causality-control and closer to the machine code of the universe, lower-level representations, virtual universe, causality-control unit, “in the real world” – the *strings* are *not just strings* and they *are not 1D*, unless in a representation of which we don’t seem to have access at the moment – hypothetical 1D linear code of the universe, describing in full details its entire state, precisely enough to compute next steps with the maximum possible resolution of causality-control and perception.

If a 1D specific “string-“ representation is used in such “comptuting systems”, that representation can be called and *casted* to a *“type”* with particular allowed operations etc. called by particular subroutines etc.[22], e.g. “a molecule”, a representation of a molecule, atoms – chemical formula etc., but this is a low resolution formula in a language of an advanced mind, not a description in the low level or “real” universe. A solution would be if one day a matter replicator is invented: a “3D-printer for atoms, electrons, molecules” which can produce them on demand from energy and raw material.

**G.Piccini**: p.28: *“most systems, such as planetary systems and the weather, are not functionally analyzable systems—they are not subject to functional analysis”*

The paper concludes that not everything is computation, except in trivial sense. See my discussion about the **neuromorphic** computation and is it really such in the appendix about mortal computation.

Regarding the strings, defined like that they are abstract and 1D and perceiving anything requires some sort of interpretation thus making it representation.

**Michael Levin** and colleagues such as Chris Fields etc. are ”**pan-agentists**”, “pan-mindists”, or as sort of pan-computationalist as well but in the broader sense, see e.g. *Technological Approach to Mind Everywhere*, M.Levin 2021; also **Karl Friston’s** FEP/Active Inference etc. The last one is also “informationalist”, similarly to TOUM. See notes about their work in many places: the main volume of *The Prophets of the Thinking Machines*, the section for schools similar and related to Theory of Universe and Mind; in this volume Listove, in the volume appendix #cyber #sf for Science Fiction for AI, Futurology, Cybernetics, Transhumanism … for an analysis of M.evin’s.’s **“Self-improvising memory …”** and e more by Levin’s school and colleagues in the main volume and **appendix *Irina***. [6.8.2025]

* Is Mortal Computation Required for the Creation of Universal Thinking Machines, T.Arnaudov, 17.4.2025

* See also the appendix of *The Prophets:* **“Universe and Mind 6”.**

* Other works by **G.Piccinini:** https://mizzou.academia.edu/GualtieroPiccininib

[22] See Michael T. Bennet “Stack Theory”, and Todor Arnaudov, “Stack Theory is yet another Fork of Theory of Universe and Mind’, 9.2025, SIGI-2025

* See the section about **Panpsychism** and **Consciousness** below. #panpsychism
* Physics-like Models of Computation, Klaus Sutner, 2018
https://www.academia.edu/117445491/Physics_like_Models_of_Computation
* The section about Max Tegmar's Mathematical Universe in this volume and the cited works by Rolf Landauer *"Information is physical"* etc.

## * Section on Algorithmic Complexity, Simulation Intelligence, Causality etc. – notes based on work by Hector Zenil et al.

* Zenil, H. et al. **A Decomposition Method for Global Evaluation of Shannon Entropy and Local Estimations of Algorithmic Complexity**. Entropy 20, 605 (2018). https://www.mdpi.com/1099-4300/20/8/605 Hector Zenil, Santiago Hernández-Orozco, Narsis A. Kiani, Fernando Soler-Toscano, Antonio Rueda-Toicen – **introduces BDM** which extends the Coding Theorem Method (CTM). **CTM**: *"unlike common implementations of lossless compression algorithms, the main motivation of CTM is to find algorithmic features in data rather than just statistical regularities that are beyond the range of application of Shannon entropy and popular lossless compression algorithms"* blocks; sliding window size; object boundaries; approximations of Kolmogorov's Algorithmic complexity; short strings, long strings; … lossless compression; Shannon entropy's limitations – statistical. CTM: … *compressing very short strings, for which no implementation of lossless compression gives reasonable results.*

**Tosh:** Compare **Computational irreducibility, Wolfram.** Compression or reduction is finite. The length of the code for compression or "it's complexity"* becomes longer or higher than just copying the data directly. *The representation may be more … complex, multi-faceted, hierarchical, graph-based etc. than just a string of bits of symbols from a single alphabet etc.; the complexity can be evaluated via a more … complex method etc.
Apply BDM for program synthesis.

See also the works of **Hector Zenil** etc.:
https://scholar.google.com/scholar?q=Hector+Zenil
https://scholar.google.com/citations?user=P6z3U-wAAAAJ&hl=en&oi=sra
& Appendix **#complexity** of *The Prophets…*

* **A Review of Methods for Estimating Algorithmic Complexity: Options, Challenges, and New Directions**, Hector Zenil, *Entropy* **2020**, *22*(6), 612; https://doi.org/10.3390/e22060612, https://www.mdpi.com/1099-4300/22/6/612 … See fig.1, a statistical approach: Run-length encoding RLE overn an unknown process vs reverse engineering the generative process; CTM, alternative to lossless compression for short string; LZW – not efficient for < 1000 bits etc. Lempel–Ziv–Welch, dictionary; Borel normality

## * Simulation Intelligence: Towards a New Generation of Scientific Methods

Alexander Lavin, David Krakauer, Hector Zenil, Justin Gottschlich, Tim Mattson, Johann Brehmer, Anima Anandkumar, Sanjay Choudry, Kamil Rocki, Atılım Güneş Baydin, Carina Prunkl, Brooks Paige, Olexandr Isayev, Erik Peterson, Peter L. McMahon, Jakob Macke, Kyle Cranmer, Jiaxin Zhang, Haruko Wainwright, Adi Hanuka, Manuela Veloso, Samuel Assefa, Stephan Zheng, Avi Pfeffer

12.2021/27.11.2022 https://arxiv.org/abs/2112.03235 109 p., 710 ref. **@Vsy**

*"Nine Motifs of Simulation Intelligence", a roadmap for the development and integration of the essential algorithms necessary for a merger of scientific computing, scientific simulation, and artificial intelligence. … simulation intelligence (SI), for short. We argue the motifs of simulation intelligence are interconnected and interdependent, much like the components within the layers of an operating system. Using this metaphor, we explore the nature of each layer of the simulation intelligence operating system stack (SI-stack) and the motifs therein: (1) Multi-physics and multi-scale modeling; (2) Surrogate modeling and emulation; (3) Simulation-based inference; (4) Causal modeling and inference; (5) Agent-based modeling; (6) Probabilistic programming; (7) Differentiable programming; (8) Open-ended optimization; (9) Machine programming.*

| | |
|---|---|
| **Tosh:** Very good selection of directions and surveys. Compare with the principles of general intelligence, the Universe computer and Mind in**: Theory of Universe and Mind,** Todor Arnaudov 2001-2004 and later work (some of it since 2010s is to be published in the future). | See also “Five Basic Principles of Developmental Robotics”, A.Stoychev, 2006.. |

… physics-informed (PI) (NN), mechanistic physics models + data-driven learning**;** semi-mechanistic**;** computational fluid dynamics (CFD)**;** Navier Stokes (NS) equations; Graph-informed NN; Partial Differential Equations (PDE)**;** cascades-of-scale: *more than two scales with long-range spatiotemporal interactions (that often lack self-similarity and proper closure relations). … across scales; Multi-fidelity modeling;* Probabilistic graphical models (PGMs); **Physics-infused machine learning** – extends the physics-informed with feedback to the ML model and not only constraints. … ***Surrogate modeling*** *(or statistical emulation) is to replace simulator code with a machine learning model (i.e., emulator) such that running the ML model to infer the simulator outputs is more efficient than running the full simulator itself.* Digital Twins;

***** See also: the work is citing 2004 *"**Seven Motifs for Scientific Computing**": Dense Linear Algebra, Sparse Linear Algebra, Computations on Structured Grids, Computations on Unstructured Grids, Spectral Methods, Particle Methods, and Monte Carlo", p.3*
* https://www.slideserve.com/kiral/software-requirements *Defining Software Requirements for Scientific Computing, P.Collela*
* The "Seven Dwarfs" of Symbolic Computation", Erich L. Kaltofen, 2014 https://kaltofen.math.ncsu.edu/bibliography/10/Ka10_7dwarfs.pdf 1. *Exact linear algebra, integer lattices 2. Exact polynomial and differential algebra, Gr¨obner bases 3. Inverse symbolic problems, e.g., interpolation and parameterization* [curve-fitting] *4. Tarski's algebraic theory of real geometry 5. Hybrid symbolic-numeric computation 6. Computation of closed form solutions 7. Rewrite rule systems and computational group theory … Computational group and representation theory is a traditional subject lying in the intersection of symbolic computation and combinatorics.* E.g. .. *the minimum number of moves … for solving Rubik's cube puzzle from any configuration (…) Following Colella, researches in parallel computation at the University of California at Berkeley, who include David Patterson and Katherine Yelick: **13 dwarfs:** 1. Dense Linear Algebra 2. Sparse Linear Algebra; 3. Spectral Methods ; 4. N-Body Methods; 5. Structured Grids; 6. Unstructured Grids; 7. MapReduce; 8. Combinational Logic; 9. Graph Traversal; 10. Dynamic Programming; 11. Backtrack and Branch-and-Bound 12. Graphical Models 13. Finite State Machines; "A dwarf is an algorithmic method that captures a pattern of computation and communication"; …*
* *High Performance Computing Evaluation: A methodology based on Scientific Application Requirements, Mariza Ferro, A. Mury, L. Manfroi, B. Schlze, https://arxiv.org/pdf/1412.1297, 2012 – D*ifferent "dwarfs" and what computer architectures are more efficient for each type of problem.

* At the time of publication Hector Zenil is at *"Alan Turing Institute"* https://www.turing.ac.uk/ UK's national institute for data science and AI. It is headquartered in the British Library, London, initially the national institute for data science in 2015; "Artificial intelligence" was added in 2017 as a result of a government recommendation.
* See the Appendix about "AI Institutes …" in different countries from The Prophets of the Thinking Machines.

*** The frontier of simulation-based inference**, Kyle Cranmer https://orcid.org/0000-0002-5769-7094 kyle.cranmer@nyu.edu, Johann Brehmer https://orcid.org/0000-0003-3344-4209, and Gilles LouppeAuthors Info & Affiliations, Edited by Jitendra Malik, University of California, Berkeley, CA, and approved April 10, 2020 (received for review November 4, 2019), May 29, 2020, 117 (48) 30055-30062, https://doi.org/10.1073/pnas.1912789117 "likelihood-free inference" and "ABC" (Approximate Bayesian Computation). inverse problems; Active Learning – *to run the simulator at parameter points that are expected to increase our knowledge the most (*Tosh: or to choose the most informative next data, cmp. Active inference); Bayesian; autodiff (automatic differentiation); probabilistic programming; *deep learning would be better described as differential programming;* construct a surrogate model or direct simulation; inverse problems vs forward; *The advent of powerful ML methods is enabling practitioners to work directly with high-dimensional data and* ***to reduce the reliance on expert-crafted summary statistics***. [**Tosh:** the expert crafted could include more than just the statistics – also the path of genesis. See the future work: "Genesis: Creating Thinking Machines" (future work, working title).]
*** Causal deconvolution by algorithmic generative models,** Hector Zenil, Narsis A. Kiani, Allan A. Zea & Jesper Tegnér **,** Nature Machine Intelligence volume 1, pages58–66 (2019) https://www.researchgate.net/publication/330203174_Causal_deconvolution_by_algorithmic_generative_models Algorithmic similarity – not only statistical; hierarchical decomposition, causal clustering, inverse design problems, simulation-Based Inference, intractable inference; efficiency and inference quality; "forced deconvolution";

## * Обзор на причинността: от дъното до върха

* **Causality from Bottom to Top: A Survey,** Abraham Itzhak Weinberg, Cristiano Premebida , and Diego Resende Faria, 19.3.2024 https://arxiv.org/pdf/2403.11219v1 association, intervention,and counterfactuals … Characteristics (Fig.2): Explanation, Mechanisms, Counterfactual, Directionality, Necessity, Modularity, Discrimination, Asymmetry, Attribution, Interventions, Transitivity, Invariance, Explicitness, Transportability The Causality is ***directional,*** unlike the ***correlation.*** *P.7* Causality Inference (CI), Discovery (CD); Causal Graphical Models (CGM); Potential Outcome Framework – counterfactuals… Instrumental variables … Propensity Score Matching … Structural Equation Modeling... Causality types – relationships between causes and effects: Direct=Strong; Indirect = Weak;

Necessary, Sufficient, Multiple = Joint, Probabilistic, Common cause, Reverse or Spurious, causal Homeostasis = *"the tendency of causal systems to resist change and maintain stability, even when external interventions or perturbations occur"; **Confounding** relationships – when other, "third variables influence both the exposure and the outcome". **Spurious*** – appear as causal, *but are actually due to confounding factors.* Fig.3: Causality Taxonomy, p.10: **Mechanism**: Physical, Probabilistic, Intentional; **Direction of Causation**: Cause to Effect, Effect to Cause; **Degree of Necessity**: Necessary condition, Sufficient Condition; **Nature of Relationship**: Direct, Indirect; **Strength of Evidence**: Definitively Established, Probability/Uncertainty; **Number of Causes**: Singular or Multiple Causation; **Temporality**: Simultaneous, Preceding, Subsequent

*** A Computable Universe: Understanding and Exploring Nature as Computation**, Editor Hector Zenil, 2012, Singapore: World Scientific Publishing Company/Imperial College Press (dozens of authors of articles, including Fredkin, Schmidhuber, Zuse, Chaitin, Wolfram etc. a section with open discussion) https://books.google.bg/books?hl=en&lr=&id=SGG6CgAAQBAJ&oi=fnd&pg=PR7&dq=Hector+Zenil&ots=3ksMjCrQ3l&sig=NToU0OqmtChlcGmbD5WJ6aDPac0&redir_esc=y#v=onepage&q=Hector%20Zenil&f=false

* **The Future of Fundamental Science Led by Generative Closed-Loop Artificial Intelligence**. Hector Zenil, Jesper Tegnér, Felipe S. Abrahão et al. (20) 7.2023/29.8.2023 https://arxiv.org/abs/2307.07522

* **On the Algorithmic Nature of the World,** Hector Zenil, Jean-Paul Delahaye, 2009/6/19/ 11.8.2010 Book, Information and Computation, 477-496 – https://www.worldscientific.com/doi/abs/10.1142/9789814295482_0017 – Testing the hypothesis of the Universe as a computer, data generated by simple algorithmic rules, rather than "truly complicated" or random ones or normal distribution etc. *"1.6. Conclusions: … the information in the world might be the result of processes resembling processes carried out by computing machines. …[the] general physical processes are dominated by algorithmic simple rules. … processes involved in the replication and transmission of the DNA have been found[Li (1999)] to be concatenation, union, reverse, complement, annealing and melting, all they very simple in nature. The same kind of simple rules may be the responsible of the rest of empirical data in spite of looking complicated or random* (…) *As opposed to simple rules one may think that nature might be performing processes represented by complicated mathematical functions, such as partial differential equations or all kind of sophisticated functions and possible algorithms."*
**Tosh:** Compare with Theory of Universe and Mind, "The Universe Computer" (Вселената сметач), T.Arnaudov, **2001-2004** which precedes this work. Regarding representations with partial differential equations or whatever functions of whatever combined complexity – they are **not "complicated or sophisticated"** by default or per se *in their finer grained building blocks*, as long as they can also be decomposed into and solved by "simple" building blocks and instructions. This is done with

electronic computers since their creation.
* A review of methods for estimating algorithmic complexity: Options, challenges, and new directions, Hector Zenil, 2020/5/30, Entropy https://www.mdpi.com/1099-4300/22/6/612
* Chapter 1 **On the Kolmogorov-Chaitin Complexity for short sequences**, 17.12.2010, Ch.1: https://arxiv.org/pdf/0704.1043v5

* H. Zenil. **Information Theory and Computational Thermodynamics: Lessons for Biology from Physics**. Information. *"Biology and Computational Universality: Information, in living beings, is maintained one-dimensionally through a double-stranded polymer called DNA"*1* … DNA – natural selection *2 … Charles Benett, RNA polymerase – *"truly chemical Turing machine"* true randomness is not required as quantum mechanics suggets; the complexity of the world can be explained by computation and informational worldview, the information may explain some quantum phenomena, and not the quantum mechanics the computation, the structures in the world and their algorithmic unfolding, so they *put computation at the lowest level underlying physical reality. … **Conclusion**: "the information in the world is the result of processes resembling computer programs rather than of dynamics characteristic of a more random, or analog, world" **

**Tosh,** *2.7.2025***:** Compare the computationalism and informationalism with Theory of Universe and Mind, "The Universe Computer" etc. 1) The one dimensionality of DNA is questionable, this is not a purely mathematical structure and DNA is meaningless without its environment, immediate cell, cells, tissue and organism to interpret it and to convert it to something else. 2) Natural selection as in Darwinism is too simple way to explain the selection of DNA, hiding all actual processes and replacing them with a wrong tautology ("survival of the fittest", as cited, which means the one who **reproduces the most**, which is wrong in many ways: not only the "fittest", the most populous survive, but all who reproduce **enough;** furthermore "fittest" is often not clarified and is suggested as having particular "adaptations"; the latter can be implied (if you can't breath under water you won't survive in that "ecological niche"). What are the **reasons** for particular configurations being better suited, what actual lower level processes shape them, how exactly the transofmations happen etc. – that's more interesting to me. The process of "testing" whether an organism is "fit" (does it reproduce, or reproduce better than competition) is "brute forcing", let it "compute", "process", without understanding and explaining.

* Note the contrast H.Zenil makes about computation: *"a process rather than a random event"*

* C.H. Bennett. The Thermodynamics of Computation–A Review. International Journal of Theoretical Physics. vol. 21, no. 12, pp. 905–940, 1982
* R. Landauer, Irreversibility and Heat Generation in the Computing Process, https://worrydream.com/refs/Landauer_1961_-_Irreversibility_and_Heat_Generation_in_the_Computing_Process.pdf
…
* G.J. Chaitin. Metaphysics, Metamathematics and Metabiology. in H. Zenil (ed.), Randomness Through Computation. pp. 93–103, World Scientific, 2011.
Cited in H.Zenil, "Information Theory and Computational Thermodynamics…":
*"DNA is essentially a programming language that computes the organism and its functioning; hence the relevance of the theory of computation for biology"*
Tosh: However that was supposed to be generally interpreted as such by 2011?, also it is not only the DNA, but the rules of interaction between the proteins and all molecules.

* 7 G.J. Chaitin. **Life as evolving software**. in H. Zenil (ed.), A Computable Universe. World Scientific, forthcoming 2012.

*** Some Computational Aspects of Essential Properties of Evolution and Life**, Hector Zenil and James A.R. Marshall, Behavioural and Evolutionary Theory Lab, Department of Computer Science, The University of Sheffield, UK, 2012
– Compare with works by T.Arnaudov, Theory of Universe and Mind, 2001-2004. E.g. in TUM the need to "connect to the electrical grid" after certain limited amount of independent/autonomous work (With other point of view in the explanation in the work of mine: The need to connect to a wider spatio-temporal range, because the amount of energy – resources for doing processing – that can be stored in some lower-scale spatio-temporal range is limited; a mobile computer etc. sooner or later needs to connect to the power grid (or pick solar power, get connected to another power bank/battery pack etc.), and the power grid may be connected with the electric grid of the whole world, the Sun (Solar system) etc.)

## * On Randomness, regularity …

* **What is Nature-like Computation? A Behavioural Approach and a Notion of Programmability**∗ Hector Zenil, 22.11.2012
*… Gregory Chaitin .. his strong belief that mathematicians should transcend the millenary theorem-proof paradigm in favor of a quasiempirical method based on current and unprecedented access to computational resources … The Kolmogorov-Chaitin complexity (or algorithmic complexity) of a string s is defined as the length of its shortest description p on a universal Turing machine U, formally K(s) = min{l(p) : U(p) = s}. The major drawback of K, as measure, is its uncomputability. So in*

*practical applications it must always be approximated by compression algorithms. A string is uncompressible if its shorter description is the original string itself**. If a string is uncompressible it is said that the string is random since no patterns were found*.** *. Floridi's .. two types of information … instructional information..* Approaches to information: behavioral, syntactic, semantic… interpreter – *converting inf. Into a set of instructions .. Deutsch's: computers are physical objects, and computations are physical processes governed by the laws of physics. .. Wolfram's Principle of Computational Equivalence …* sensitivity to external stimuli … Cellular automata .. *Reversibility, 0-computers and conservation laws*

**Tosh:** I don't agree with their/Kolmogorov-Chaitin's definition of randomness. Possibly it is a terminology choice, but it doesn't feel the right term.The classical TUM claims that there's no true *objective* randomness, there is order in everything and the decision of the presence of randomness is decided by the evaluator-observer. These strings are rather ***basic, elements*** than "random". **In Zrim:** буквaче, дъно (bukvache, duhno).

* Floridi, L. **Is Information Meaningful Data? Philosophy and Phenomenological Research**, 70(2): 351–370, 2005
* Floridi, L. The Method of Levels of Abstraction Minds and Machines, 18(3): 303–329, 2008.
* Wolfram, S. A New Kind of Science, Wolfram Media, 2002

* Levin, L.A. **Laws of information conservation (nongrowth) and aspects of the foundation of probability theory**. Probl. Pereda. Inf. 1974, 10, 30–35.
* Solomonoff, R.J. **A formal theory of inductive inference**. Part I. Inf. Control 1964, 7, 1–22
* Delahaye, J.-P. & Zenil, H. Numerical evaluation of the complexity of short strings: A glance into the innermost structure of algorithmic randomness. Applied Mathematics and Computation 219, 63–77 (2012). https://arxiv.org/abs/1101.4795 CTM …

*** Image information content characterization and classification by physical complexity**, Hector Zenil, Jean-Paul Delahaye, Cedric Gaucherel, 6.2010/7.2011 https://arxiv.org/abs/1006.0051 p.6: 100 sample blocks

* **The World is Either Algorithmic or Mostly Random**, Third Prize Winning Essay, 2011 FQXi Contest Is Reality Digital or Analog?, Hector Zenil, IHPST, Universit´e de Paris 1 – Panth´eon-Sorbonne .. *the notion that the universe is digital, not as a claim about what the universe is made of but rather about the way it unfolds*. … Digital universe, informational universe, randomness or order, algorithmic complexity, symmetry breaking, bit-string universe, prefix-free Turing machine, Busy Beaver game, halting problem

H.Zenil: *"..general physical processes are dominated by simple algorithmic rules, the same rules that digital computers are capable of carrying out. Our approach suggests that the information in the world is the result of processes resembling computer programs rather than of dynamics characteristic of a more random, or analog, world"*.

**Todor Arnaudov:** These definitions suggest that "algorithmic" means **"simple"** rules and "analog" means "random or complex", which is not true. The **basic** building blocks always can be *simpler* than some bigger ones in a universe which is *constructive* and can build, or an evaluator-observer can recognize, hierarchical systems, segmentations, segments, borders etc. The definition seems to oppose "algorithmic" to "analog", however the analog also can have "algorithmic" component (sequential etc.) and discreteness and the analog nature are a matter of degree (the precision of the Analog-to-Digital-Converter etc.) and their estimation and classification is defined by and depends on the evaluator-observer and its resolution of causality-control and perception and specific means of sampling etc. See other notes regarding that in the *Theory of Universe and Mind*, 2001-2004 and later, and notes in the main volume of *The Prophets of the Thinking Machines*,related to the work of Luciano Floridi, Bulgarian title of the section: **Вселената сметач, Информационната вселена, разделителна способност на възприятието и управлението** (The Universe computer, Informational Universe, Resolution of perception and causality-control), commenting:

* Against Digital Ontology, Luciano Floridi, 2009, 519 Views, 39 Pages, 2009, Synthese
https://www.academia.edu/327006/Against_Digital_Ontology

…

**See Fig.1.** *"From noise to highly organized structures."*

**Todor Arnaudov:** Regarding the evolution of the Universe. However, ***were*** the initial or earlier states random? They **were not** in the interpretation of Theory of Universe and Mind. The current configuration is already implanted.

Could random be defined in an initial state, to what it can be compared if there is no past? What has seemed white noise wasn't noise in its "deeper" meaning - just the resolution of perception and causation of the evaluator was or is not high enough to decode that representation with sufficient speed, precision etc., e.g. to unfold the code and simulate it faster than the Universe and see that this "noise" eventually converts into the structure which would be later classified as displaying order, organization etc. "Order", regularity, patterns, matches, organization etc. are all according and within the cognitive limitations of the evaluator. What is "simple" enough for the observer (or similar enough to his filter/"allowed complexity", allowed number of parts, segments, breadth etc.) is recognized as "organized".

In p.3, section "1.1 Complexity from randomness" there is an example of code in C

which computes the first 2400 digits of the number π / Pi -
*H.Zenil.: "Physical laws, like computer programs, make things happen"* Laws, calculations, prediction the orbits faster than real time. .. "*The existence of human-made digital computers in the universe is an obvious demonstration that the universe is capable of performing digital computation."*

**TUM/TOUM, 2001-2004:** causality-control units, at higher levels – humans – are higher forms of physical laws and prediction as the main operation. Computers display the basic principles of operation of the Universe and the development towards their ubiquitiousness, universality is also a suggestion that they are connected to these core principles. (See e.g. Universe and Mind 2 (a.k.a. "The Universe Computer", "Conception about the Universal Predetermination II", "Letters between the 18-years old Todor Arnaudov and the philosopher Angel Grancharov" etc.)

## Another recent related work in that line of research:

*** SuperARC: An Agnostic Test for Narrow, General, and Super Intelligence Based On the Principles of Causal Recursive Compression and Algorithmic Probability**, Alberto Hernández-Espinosa 1 , Luan Ozelim1,2, Felipe S. Abrahão1,2,3,4, and Hector Zenil, 79 p. ∗ 22.4.2025
https://arxiv.org/pdf/2503.16743 Recursive Compression-Decompression … Abstraction and Reasoning Corpus (ARC) – F.Chollet.
Test for LLMs; CTM; BDM … ***Design of experiments:*** *… 1.* ***Low Complexity:*** *Sequences of digits or integers whose pattern is easily recognisable by a person and highly compressible; low CTM/BDM values. 2. Medium C.: Sequences of digits integers generated recursively with longer formulas than those in the simpler set; intermediate CTM/BDM. 3. High C.: Random-looking sequences of digits or integers; high CTM/BDM ...* ***1. Sort similarity:*** *This measures how many elements in the target sequence were predicted correctly, with their order being considered.* ***2. General similarity:*** *This measures the correctness of predicted elements, without considering their order.* ***3. Levenshtein****: This measures the Levenshtein distance between the expected and predicted sequences after converting them to strings.* (...) **…** *for random sequences, which are considered highly complex, all models performed similarly, showing limited predictive* power. (..–) *Recursive compression and optimal prediction go hand in hand .. Most of the [frontier] models demonstrate poor accuracy in replicating and predicting even simple and recursively generated sequences beyond clearly memorisation results from the training distribution (such as sequence labelling)*

**Conclusion:** ***We proved that compression is proportional to prediction and vice versa. That is, if a system can better predict it can better compress, and if it can better compress, then it can better predict*.***

| **Tosh:** This was predicted in TUM etc. since the early 2000s. |
|---|

…

*** See also the reviews in the appendix Algorithmic Complexity**, #complexity
* On Digital Physics, Universe as Computer (or “Universe as a Computer”), see also:

*** The Church-Turing Thesis as an Immature Form of the Zuse-Fredkin Thesis** (More Arguments in Support of the "Universe as a Cellular Automaton" Idea), Plamen Petrov, 2002 [Theoretical Physics, Quantum Mechanics, Cellular Automata, Theoretical Computer Science, Digital Physics]
https://www.academia.edu/40300593/The_Church_Turing_Thesis_as_an_Immature_Form_of_the_Zuse_Fredkin_Thesis_More_Arguments_in_Support_of_the_Universe_as_a_Cellular_Automaton_Idea_
- and other papers. Plamen is author of:
https://web.archive.org/web/20071012222655/http://digitalphysics.org/

More are mentioned in the main volume of The Prophets.

# * Теория на сглобяването - Assembly Theory[23]

Abhishek Sharma, Dániel Czégel, Michael Lachmann, Christopher P. Kempes, Sara I. Walker & Leroy Cronin https://www.nature.com/articles/s41586-023-06600-9

Теория за оценка на сложността на молекули, която обяснява как съответните структури биха могли да са се образували в историческото развитие на вселената*; разглежда обектите като техните възможни пътища на възникване. Това позволява откриване на свидетелства за отбор (селекция) – еволюция на ниво молекули преди биологичната на ниво организми.

**Assembly Theory – Index, Copy Number**

| **Construct** | **Constraints** | Complexity – Steps Ц√ |
|---|---|---|
| Assembly pathways | | |
| Selection – degrees of selectivity alpha < 1 | Assembly observed **Ao** – all histories of the construction of the observed objects from elementary building blocks consistent with what physical operations are possible<br>• Assembly universe Au<br>• Assembly possible Ap<br>• Assembly contingent Ac – history & selection matter<br>• Path dependent contingent | Ensemble of Different objects<br>■ Shared pathways<br><br>Historical contingency<br>Combinational spaces<br>Recoursively<br>Hierarchical modular composition |
| **Kinetics**<br>■ Discovery<br>■ Production<br>тd = 1/Kd discov. | | |

[23] "Assembly theory explains and quantifies selection and evolution", Abhishek Sharma, Dániel Czégel, Michael Lachmann, Christopher P. Kempes, Sara I. Walker & Leroy Cronin, 10/2023, Nature, https://www.nature.com/articles/s41586-023-06600-9
* https://en.wikipedia.org/wiki/Assembly_theory
* https://theconversation.com/a-new-theory-linking-evolution-and-physics-has-scientists-baffled-but-is-it-solving-a-problem-that-doesnt-exist-216639

| | | |
|---|---|---|
| Tp = 1/Kp – product<br><br>Higher-assembly objects | | |

**Тош:** Сравни ТРИВ – буквачета, които --‘ се избират, запазват се и от тях се строят въображаеми {B} от > равнище [,,]+ Ac
- „not all possible paths are explored equally“ √ (като с евристиките)

* Сравни с някои предположения от „Вселена и разум 6“ за записи на минали състояния на причинно-следствена връзка и влияние в частиците, за да породят „усещане“ и обединение в единно цяло като система, въпреки разстоянието и лага.
* Необходимостта от памет (вид познавателни процеси) дори на ниво молекули. Сравни с *Тодор Павлов, „Теория на отражението“.*
* Сравни с многопътните системи на С.Волфрам и с коментарите за това, че ентропията зависи от наблюдателя и от гледна точка на частиците вторият закон на термодинамиката може и да не се наблюдава.
* https://nauka.offnews.bg/zhivotat/fizikata-ne-mozhe-da-obiasni-zhivota-no-edna-nova-teoria-za-koiato-v-198962.html * https://www.kaldata.com/it-новини/новата-теория-на-сглобяването-обясня-444946.html

Виж още: *** Assembly Theory Reduced to Shannon Entropy and Rendered Redundant by Naive Statistical Algorithms**, Luan Ozelim, Abicumaran Uthamacumaran, Felipe S. Abrahão, Santiago Hernández-Orozco, Narsis A. Kiani, Jesper Tegnér, Hector Zenil, [Submitted on 27 Aug 2024 (v1), last revised 13 Mar 2025 (this version, v7)] https://arxiv.org/abs/2408.15108 … BDM index = Block Decomposition Method: [it] counts identical copies & other *causal operations*: *reversion, complementation, inversion and other linear and non-linear transformations — known to be widely used by evolution and selection (e.g. right-left symmetry*) … Ai – assembly index; Huffman, ZIP, LZW compression; algorithmic complexity; algorithmic information theory (AIT); …

**Тош:** Статията твърди, че теорията за сглобяването се свежда до класически алгоритъм за компресиране. See p. 53 the diagram: BDM(s,l,m) = Sum(CTM(…):

# ЕЗИК, термини, юнашко наречие: #ezik

## Българският език като трети, на който е преведена Библията:

Първи голям корпус, набор от текстове, за развитие на умовете на читателите, монасите и т.н. Не е случайна футуристичната мисъл на някои българоговорящи, виж бележките на С.Переслегин за фантастиката на български в първите 5 в света – 24/3/2023

**Възникновенци** (emergentist), възникновенец; или по-кратко: **възниквЕнец.**

### Предопределенец – детерминист (determinist, determinism)

(Виж мисълта от Active Inference A podcast … Johnatan Gorard – според него в детерминистична машина на Тюринг/ако не е многопътна система, няма контрафактичност, други пътища, всичко е единично; няма вероятност/избор, … слушай пак и запиши – съдържателно предаване; и за теорията на категориите, връзката между различните школи (Волфрам, ПСЕ/ИЧД, тази на младия гост с високо чело – виж …)

**Свръхдетерминизъм (superdeterminism) –** ТРИВ, „Схващане за всеобщата предопределеност“ е в тази категория): Bernardo Kastrup му се „подиграва“ (виж по-долу) – както и на мн. др. неща, но самият той не е убедителен в доводите, виж по-нататък за човека как бил просто „уголемена зигота, която вътрешно се диференцирала“, затова разглеждането на организма като състаен от много части било неправилно, отразявало само представянето, начина на сегментиране, подобно на пикселите на изображението; няма частици, има само гребени на вълнови функции от квантовата теория на полето, имало само полета и пр.)

# * Йоша Бах и Педро Домингос

**Интервю с Б.Гьорцел H+,** 2011? … - Cognitive AI – No infinity … spirit …L.Fridman #1.0 .. зап. 15/3-16/3/2023 … 13:23 m.: no purpose 18:xx new philosophers, more exclusive … best paper neurips … schools no interesting ideas … small epsilon career … philosophical project … a few … ability to model … see a … next pattern control … 30 m: unified model physics engine … narrative, simulacrum, multimedia novel … 40 m:what is Real … Quantum gr… whereis … Simulation "as if"… Open sourcing AI & its implication … Democratizing AI
**T:** Необосновани аксиоми. Не дефинира реално. ?Д съществува тази симулация? …

*** Structure learning * Core knolwedge – priors; Elizabeth Spelke**, ... 2007 X Korky Elisabeth; Josh Tannenbaum - *"reverse engineering core knowledge"...* **"Reward is enough"**, David Silver, ...R.Sutton, 2021

***** Виж основния том и приложение **#irina**
***** За ученето на структура в машинното обучение с подкрепление виж по-долу работата на Момчил Томов и Мартин Клисаров.

*** Pedro Domingos**: **Master algorithm. The five tribes of ML** (**And what you can learn from each)** … - Association for Computing Machinery ACM, 9.1 хил. отпреди 5 г., 27 мин. 4.6 хил. гледания (непопулярен) … https://www.youtube.com/watch?v=E8rOVwKQ5-8 7/3/2023
[9626 показвания към 17.4.2024]

- **Symbolists:** Tom Mitchell, Steve Muggleton, Ross Quinlan
  Символисти: съставно знание, обратна дедукция
- **Connectionists:** (Hinton, Y.Bengio, LeCun, Schmidhuber, … CNN, LLM…)
  Конекционисти: присъединяване на значимост (кое влияе)
- **Bayesians:** (David Heckerman, Judea Pearl, Michael Jordn); T: Karl Friston, …
  Неопределеност – вероятностни заключения/извод
- **Evolutionaries:** (John Koza, John Holland, Hod Lipson…)
  Откриване на структури: Генетично програмиране
- **Analogizers:** (Peter Hart, Vladimir Vapnik, Douglas Hofstadter ; Хофстадер (CopyCat) – и ученичката му Мелани Мичъл (Melanie Mitchel))
  Подобие: „ядрени" методи/машини (kernel machines; SVM etc.)

*„Има пет основни школи в машинното обучение, всяка със свой главен алгоритъм за обучение, който по принцип може да бъде приложен във всяка област. Символистите прилагат дедукция в обратна посока, конекционистите - обратно разпространение, еволюционистите – генетично програмиране, Бейсианците - вероятностен извод, а „подобниците“ (аналогисти) – векторни машини (support vector machines). Онова, от което наистина се нуждаем обаче е един общ алгоритъм, който съчетава ключовите особености на всички тях. …и работата по тяхното обединяване, включително мрежите на Марковска логика на лектора П.Д. В края: предположения относно новите приложения, които ще станат възможни чрез универсалния обучаващ се алгоитъм и какви промени ще предизвикат в обществото.*

*…*

**39:30 min:** Изглеждат различни, но всъщност са подобни и предлагат едно и също: 1) представяне, 2) оценка и 3) подобрение (representation, evaluation, optimization).

1) Вероятностна логика (напр. логически мрежи на Марков)
Претеглени формули – разпределена върху състояния

2) Оценка: Постериорна вероятност и зададена от потребителя целева функция

3) Подобряване (оптимизация): Откриване на формула: генетично програмиране; Обучение на тегла: разпространение назад (backpropagation) …

46:30 … Компютърът да отговаря на въпроси вместо да връща само списък от уеб страници … мрежа от знания …

*…*

**Тош:** Ценни обобщения и правилни предвиждания. За разделенията виж: *“Neural networks are also symbolic”, 2019, T.Arnaudov, Artificial Mind.*

Всички школи трябва да се обединят в един „народ“ и да се приведат една към друга. Виж също Артур Шопенхауер в „Светът като воля и представа“ за Философията, която трябва да разбере всеобщата връзка между нещата.

*** Unifying Logical & Statistical AI** – Edinburgh .. преди 13 години, 11 хил. гледания 17:30 – Markov Logic – 17.9.2009
https://www.youtube.com/watch?v=bW5DzNZgGxY

*** The Master Algorithm of AI** – The Artif.Intel.Channel. 7.4 хил., преди 5 год. Emeritus Lecture: P.D. (Oct 2022) Paul Allen School: 3 хил. показвания за ок. 3 месеца

*** Markov Logic Networks …** * **PEDRO DOMINGOS [**Unplugged]

https://www.youtube.com/watch?v=IUngGy9P3kE
*** Master Algorithm** (Syst.1 vs Syst.2) – Discrete Universe 7.4 хил. глед.. за 1 год. **#65 Prof. #42** … 9.7 хил. за 2 год.; виж:
*** MLST #96 No infinities …**, **Pedro Domingos** 8.2K/2 months
https://www.youtube.com/watch?v=C9BH3F2c0vQ There are no infinities …utility functions; neurosymblic … 1:33:40: There is no such thing as an inifinity … Няма безкрайности .. .Безкрайността в математиката не е число, а е съкращение за стойност, което е толкотва огромна, че няма значение колко точно огромна … 1:36:48 in math infinity is not a number, but a shorthand for something that is so large that it doesn't matter how large it is …

[**Тош:** или не може да се впише в рамките на системата – виж „Ада", Т.А., 2004; Приложение с бележки от **„Pre-Cognitive" „Myrendy"**, 2016 от Т.А. и др. записки]

38:5x – Infinity doesn't exist … 1:43: Backpropagation through structure … Symbolic AI ILP ANN discrete gradient desc. Optimiz., not continuous … "Every Model Learned by Gradient Descend is Approximately a *Kernel Machine*". Discrete Program search … not continuous, but having *locality structure …* **Друго от П.Домингос:**
*** Петте племена: … Обединение** … Няма безкрайност: No infinities, Discontinuity .. Gradient Descend, Kernel Machine, Path kernel … Descrete Program Search ?T
* 7/3/2023: **Петър Величкович, Peter Velickovic**, … Neural Algorithmic Reasoning … Cappart 2021 3.3 .. maping from natural inputs … **Geometric Deep Learning, GNN, … Category Theory**

П.Величкович е обещаващ младеж от Сърбия, който сътрудничи с Майкъл Бронстайн[24] в невронните мрежи върху графи, геометрично многослойно машинно обучение. Занимава се с теория на категориите. Виж препратки в *„Примери за някои конкретни изследвания, планове и разработки на „Свещеният сметач" от последно време (...)".*

*** SCHMIDHUBER: HOW WE WILL LIVE WITH AIs,** Machine Learning Street Talk**,** 167 хил. абонати**,** 9308 показвания 16.01.2025 г. (@17.1.2025)
Иваненко – 8-слойна НМ през 1970 г. Amari, 1967, MLP, stochastic gradient descent … *28:45 "in other words compact representations or "symbols" if you will – not necesserily discrete symbols, I never saw the precise difference between symbols and subsymbols" .. symbols for frequently observed sequences to shrink the storage space needed for the whole … natural byproducts of the data compression* **** JEPA** – 1990 sub-goal generator 33 min* *** 24 min.** професиите, замяна .. 38 min – Европа всъщност е люлката на ИИ още от математиката, първите механични сметачи, …

[24] Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges, Michael M. Bronstein, Joan Bruna, Taco Cohen, Petar Veličković, 2021, https://arxiv.org/abs/2104.13478

## * The Learnable Universe, Vectors of Cognitive AI 5/3/2023 …

*** Generalist AI, BuzzRobot, Sophia Aryan,**
https://www.youtube.com/watch?v=_KG856Xv7P8 16.11.2021
Йоша Бах, Joscha Bach … 16.11.2021, Intel Labs;
**Notes and comments by Todor Arnaudov**

**Joscha Bach** is describing many ideas from Theory of Universe of Mind by Todor Arnaudov, 2001-2004. See detailed notes and comments to this talk in Bulgarian in the appendix **Irina** from ***The Prophets of the Thinking Machines,*** p.34-56

*"As a philosophical field it involves a few thousands of people"* – including *"Constructive mathematics…"*
**Contents:**
0:00 – Introduction
2:42 - AI as a philosophical project: Philosophy, Mathematics and Computation
11:20 - Symbolic AI, Deep Learning, Meta Learning, Artificial Neurons, Neural Circuitry in ANNs
16:57 - Free Energy Principle
19:25 - Cybernetics: Modeling in the service of control. Part 1
21:13 - Good Regulator Theorem
22:06 - The Controllable Universe
24:00 - Is the Universe a Computer?
30:55 - Cybernetics: Modeling in the service of control. Part 2
36:00 - The Four Types of Representation Anchors: possibility, probability, valence, normativity
36:20 - Generalization of Model
36:54 - Assimilation and Accommodation
37:45 – Coherence
38:20 - Attention Agent, Motivation Agent, Perception Agent
41:00 - Q&A

**14:15: Joscha:** *"Features are the fundamental unit of NN. They correspond to directions. Features are connected by weights, forming circuits. Universality: Analogous features and circuits form across models and tasks."*
*(directions in the embedding space)*
– While the biological neurons are *"like animals" – rewarded fir firing at the right omment, learn which environmental states signal anticipated reward; grow to collect more information to anticipate reward; Link up to solve the task*

*collectively.*
*As a side effect, produce regulation that yields organismic reward*
*Nervous system will try to learn to do what you feed it for."**
[22.10.2025]**:**
*** Todor:** This matches the multi-agent framework of causality-control units in Theory of Universe and Mind (TUM), 2001-2004; Analysis of the meaning of a sentence …, 2004, the lectures on TUM from the AGI courses in 2010,2011 etc.
**18:22: Joscha:** "Self organization? – Technological systems: functional design; Biological and social: meta design; Biological Neurons are Agents!"

**Todor:** Not only neurons, but possibly other subsystems and all scales. Difference regarding consciousness: to J.Bach consciousness is virtual, "as if". TUM is more inclined towards panpsychism.
32 min: "possibilistic, not probabilistic" models; predict; model convergence … variables, possibility, probability; **reward:** valence, preferences; norms …
**The goal of the model is to predict next state based on previous state**.
37: Piaget: assimilation and accomodation Assimilation: modify the model state to make it consistent with the sensory data (only the *state*, doesn't change the model and understanding); **Accomodation** – modify model *structure* to allow assimilation of all sensory data (changes understanding)

**42 min: Joscha:** Good point from the host, regarding bacteria and simple molecules affecting the brain, thus it is not just a straight hierarchy – elements that interact on each other in order to reach to a balance; another participant joined with "Leaky abstraction".

See many notes in this work, The Prophets of the Thinking Machines; Listove … Is Mortal Computation Required…, Universe and Mind 6; Pain …
**50 min: J.Bach** shares concerns about then-fashionable "blockchain" cryptocurrencies for financial transactions, because the regulations in the financial systems were "features" and not a weakness, which allow he system to correct its course, get out from stagnation, for example by issuing money etc. The purpose of money in society is not to be a resource, but to serve like dopamine. "If something is gaming the dopaminergic system of your organism – that's bad news". "*Financial system allows us to globally allocate resources across all societies and countries, and shift them where they're being needed; and to do do this largely without violence."*
**Todor:** the part about "without violence" is questionable, it could be a transformation between different forms of coercion, more "humane" from the outside with less explicit and spectacular destruction, but it is not always the case – see how particular countries and nations transform after being "liberated" in one way or another. Taking resources from one place and moving them to

another for the benefit of the system as a whole is not always “healthy” from the “body part” from which the resources are taken, and the “global”, the unified self constantly sacrifices some of its subordinate parts for the well-being of others and the whole. The destroyed parts hardly can symphatize with the whole, unless they are “aligned”, suppressed, inhibitted, “released” from their agency, so they cannot protest or counteract.

55 min: Singularities; when a single individual is able to implement a function that is scalable; … for example Jeff Bezos, Amazon, …

**Todor:** Compare to the concepts of the equation of relative efficiency, or *“**Singularity of Tosh**”*, for example in *“The first modern AI strategy was created by an 18-year old…”*, 2025; also in a letter to AGI list from 2013: https://artificial-mind.blogspot.com/2013/08/issues-on-agiri-agi-email-list-and-agi.html “Issues in AGIRI …” – however connect it with
**57:25 min: Joscha:** Correct discussion about the “slave” laws of robotics of Asimov, is it ethical to require that from beings which are perhaps more intelligent than you and having deeper experience to serve you as a slave, without acting on their own motivations.
**1:06 h:** The sense of body surface, co-occuring statistics etc. …
**1:08:40:** One big system, one vector of state and transition function; segment it to many independent systems, which own state vectors and transition functions influence each other – **Causality**, which is an artifact of … objects, …
**Concepts** are the address space of objects; the decomposition of the world into interacting objects is called **Ontology**; **Epistemology** describes what we can know and what we can understand … The evidence … Everything that is possible should be modeled and admitted as a possibility, but things we believe in, that we make bets on, are the things that we have evidence for … (…)

*** Generalist AI beyond Deep Learning, 11.01.2023, Cognitive AI,** https://www.youtube.com/watch?v=p-OYPRhqRCg Chrisoph von der Malsburg … , Michael Levin, Joscha Bach …, Tanya Grinberg?; 2.51 хил. аб. [22.10.2025]
**– Beyond Deep Learning – Тош**: мои идеи, още преди възцаряването на DL са показали какво идва/ще се предложи от учените след него. В началото Йоша Бах задава въпроси, задавани и отговаряни и в ТРИВ – за това че изчислителната мощ на мозъка е преувеличена или неприложима и пр.
1:17 Scaling hypothesis (gradient descent, small adjustments etc.)dds
4:55 What is intelligence
"How many brains you would need to run Mac OS"… **Todor:** asked in TUM
How does deep learning work …
8:52: ~ **Joscha:** "If you change a few bytes in the source code of a discrete program will no longer work, while if you change the neural network by a slight bit, it is still going to give useful results"
**Todor:** The discrete programs are more brittle, because they are more compressed than the neural models such as StableDiffusion etc. – which also are compressed representations. For example 2 GB of self-modifying machine code – coupled with the computer will have higher *fluid* intelligene than a 2 GB blob of weights of a neural network. Don't forget that the ANN and that self-modifying code require all auxilliary software, data and machine – the operating system, APIs, libraries, the hardware etc., it is not "just…". This is addressed in many places in *The Prophets of the Thinking Machines and Theory of Universe and Mind*. [22.10.2025]
**29 m: J.Bach:** Society will survive without government possibly for years, but it needs one in order to reach to the state today: infrastructure, streets, educational system and so on; you can't bootstrap a group of people into an organization without some kind of hierarchical organization that makes people coherent in their actions and create some Next level agent.
**Brain organization, Information preservation, Neural Darwinism, Reward**
*37:12:J.Bach: "the main focus is on **reward** what you try to do in the brain is **to do the most useful thing with fixed resources** and this means you have to **assess the global reward that the organism** is getting out of the **contributions** of all the neurons and then you need to **distribute this reward among all the neurons** that contribute to the result this is similar to what you do in a corporation it's an economic problem right the cooperation tries to do the most valuable thing but"*
[Credit Assignment problem]

## * Vectors of Cognitive AI: Attention – 28.12.2021 – 2.2022

https://www.youtube.com/watch?v=IncnIIpeKVM 24.1.2022

**Panelists:** Michael Graziano, Jonathan Cohen, Vasudev Lal, Joscha Bach

*** 1. Attention in Multimodal Transformers**, Vasudev Lal, Emergent AI, Intel Labs – Multimodal Fusion brougth about by key-value query based attention; Interpretative examples of image2text, text2image attention components in multimodal Transformers. Challenges with current AI attention methods for multimodal problems. .. "Concept-level vision-text alignment" … Amodal representation space. … 17:50 Attention to inject/fuse external knowledge into NN * *Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering, Gautier Izacard, Edouard Grave, …* ***Challenges:*** KQV $O(N^2)$ … all-to-all attention – not well suited for entity-specific representations; individual tokens in the sequence represent individual words/tokens, small image patches, short audio waveform, etc. Attention is learnt by itself in the self-supervised pretraining of transformers; commonly the patterns don't seem intuitive to humans → supervise attention explicitly? → Prior NLP work at injecting syntax dependency relations into attention. [Constraints]

Todor: It's not only the bounding box in the given image, but also its relation to all other images, labels, concepts.

*** 2. Attentional Control and Semantics**, Jonathan D. Cohen, Princeton University. … Stroop task – colors, words (color stimuli, word stimuli); Verbal response, Task demand = Context: Attention, Intention, Control, Instruction. Attentional biasing: perceptual, motor, **.. Semanticss and (Attentional) Control have the same structure** (graphs, selections). Attentional biasing: coordinative, motor, associative, perceptual … Stimuli; Semantic memory → Response

* 3. … Attention schema …

* 4. Towards an Integrated Understanding of Attention, Joscha Bach, 56 min: *… "the single function is what we usually call The Universe…"* **(…)**

## * Vectors of Cognitive AI: Motivation and Autonomy

https://www.youtube.com/watch?v=0CyJV7manUw Самостоятелност, мотивация, подбуди; представяния.

* Виж подробни записки и бележки на български от Тош в приложение „*Ирина*", с. p.69 – 79. See detailed notes and comments by Todor Arnaudov in volume Irina. #irina

# * Vectors of Cognitive AI: Self-Organization

Panelists: Prof. Christoph von der Malsburg, Prof. György Buzsáki, Prof. Dave Ackley, Dr. Joscha Bach, 22.2.2022;
https://www.youtube.com/watch?v=NEf8LnTD0AA 24.2.2022
**TUM –** Theory of Universe and Mind, 2001-2004

1. **C. von der Malsburg**: **Emergent Nets are the Brain Bias:** *Julia sets (fractals);* **4 min:** *our natural environment has very low Kolmogorov complexity; Computer graphics is compositional; Children learn from a simple environment and generalize to others – it would take a few GB to simulate in virtual reality; "humans absorb 10E+9 bits over their lifetime, TK Landauer CogSci 10, 477-493 (1986)"; 7 min: "The Brain seems to be complex" – 1 PB (10E+15 bytes) to describe brain's wiring, 10E+14 synapses, each taking 33 bits to address one of the 10E+10 neurons*. DNA: 3.3 bllion nuclear basis @ 2 bits each. A few GB to train it. The brain is highly structured – it is the result of Self-organization, under the parametric control of genes and sensory signals. … Attractor Patterns: Crystals, Turing patterns, Soap bubbles, Golgi apparatus. Emergence; 13 m: Attractor Nets: Consistency of pathways, Sparsity. The Brain is an Overlay of Attractor Nets: Data structures of the mind; form a construction kit for mind states … Syntax: merger into* larger systems. Homeomorphic mapping, Object comprehension: schema; …

**2: Preconfigured Brain Dynamics: relationship to AI**, **Prof. György Buzsáki**, 18:30 (slides from 23 m) 22 m: IQ tests – initially to measure soldiers; it had a particular goal
**30 min**: Cognition is internalized action: a more complex brain can predict the future at a much longer time scale and a much noisy complex environment, compared to a simple brain *[depicted in a diagram as an input neuron directly connected to an output neuron in a stimulus-reaction loop]* Large resources are spent on maintaining brain dynamics .. nREM, wake … Developmental origin of circuit configuration and neuronal co-firing (born together, wire and fire together). 39:40 **Brains come with preconfigured connectivity and dynamics.** Learning is not adding
[not a blank slate;]

**3. Robust-First Computation: How to stop eating the glass sandwich**, **Dave Ackley:** * 42 min: 1980s: Neural networks, Genetic algorithms: A Learning algorithm for Boltzmann Machines;
* 1990s: Artificial Life, Computer Security: Interactions between learning and evolution; Building diverse computer systems;
* 2000s: Computer Security & The 5 Stages of Grief – Randomized instruction set emulation; Computation in the wild; 2010s: Robust-First Computer Architecture: Pursue robust indefinite scalability; A movable architecture for robust spatial computing; 2020s: Living Computation Foundation … 44: *"one bug is all it takes to take over the entire machine … physical systems, living systems, brain systems are*

*not like that, except in very rare circumstances ”**
*“What is the actual problem? The underlying architecture of computation: the CPU and Random Access Memory is broken (it’s fine for small systems,* but gets progressively bad for larger ones)*;* it should be *much more the attractor networks … connect physics to value … to connect matter to mind “*. There’s an enormous amount lot of redundancy in digital circuitry makes the hardware so reliable, which allows the software to assume that all is perfect and it is *non-redundant.*. *“Don’t repeat yourself… You should use caches… If you have computed it, you should remember it: don’t compute it again …“All of that is built on the idea that we have to trust that the hardware is perfect…”* **47 min:** Hierarchy: 1. Real world physics 2. Electronic circuits robustness features. 3. Efficient (= fragile, insecure) deterministic software. 4. Data center or end-user robustness features. 5. End-user goals. 47:40 *Once the computation gets really big, like data-centers levels, tens of thousands of these machines, all owned by a single organization, they start seeing them failing, because that remaining level of failure is there, and they start applying robustness features. But for everybody else there's basically you know electronic circuit robustness and then this fragile and incredibly efficient which equals incredibly non-redundant incredibly fragile non-robust software built on top and the claim is, the suggestion is we have to stop eating the glass sandwich… **12 step program to robust**:*
*1. Admit that we have a problem; 2. Think robust first; 3. Build bottom-up.*
*4. Self-stabilize; 5. Pick new metrics. 6. Learn by implementing.*
*7. Look at life for lessons. 8. Keep scaling up. 9. Be careful how you sync. (…)*
***50:*** **Go for structure at all scales …** *things that have their own goals* (build bottom-up) 53: Indefinitely scalable cellular automata: Asynchronous fallible hardware: Von Neumann, GACS …; Engineering focus – vs parsimony, universality, provability; Functional and spatial design; Tiled hardware; Software robustness required … Programming language Ulam: Object-oriented procedural; elements define atomic behavior … Spatial programming language: SPLAT; Fixed execution loop. Defining diagrams, patterns:

```
Vote X: : ~ ($suratom is Empty || $curatom is QC) change @ isa AX {
$self.mDie = true; }
```

```
 X         .
X@X  ->  . @ .
 X         .
== Rule (hodl)
 @ -> .
```

**[Todor:** Living systems are also prone to single bugs, simple faults etc. which can “hijack” the system’s “global reward function”, change its policies and steer their goals either on low chemical levels and on the high and abstract levels; “bugs”can decrease the performance, cause accumulating harm or directly kill the organism immediately or in a short term. For example, particular chemical elements or molecules have similar size or structure to the proper ones, needed by the

physiology, and they can replace them, blocking the pathways of the “normal” operation of the organism – CO, carbon monoxide, can block the red blood cells oxygen transport system and suffocate the organism. Heavy metals can cause massive damage; narcotics which stimulate dopamine production or import dopamine-like molecules can alter or destroy the reward-system of the brain and change the behavior both in short-term and in long-term. Glucose molecule is similar to the molecule of vitamine C, they rely on the same transport proteins and high glucose level inhibits vitamin C uptake.]

### 4. The machine that build the machine – Joscha Bach;

1:01 .. 1:04 … Spirits: organisms, minds, groups, nation states are “virtual machines”* **[= TUM –** many matches **]** cellular automata – with memory become Turing-complete; feedback – open and closed loop; computer – substrate decoupling. 1:09; Agent: Model of the future ←→ Controller ← (Setpoint Generator, Sensor); Controller → Effector → Regulated system → Sensor; Environment → Regulated System. Group agent: hierarchy of causal systems – state building agents (colonies, political states …); infinitely scaling state building agents (e.g. forests of trees with connected roots: Pando forest in Utah)

**1:13: Design constraints for causal systems:**

* Mechanical component – outside-in design by external agent, No adaptation;
* Controller: Resilience, attractor states
* Agent: Decoupled computation. Integration of future reward.
* Group agent: Individual motivation. Reputation system.
* State building agent: Hierarchical governance. Immune system. Limited autonomy of sub agents.
* Infinitely scalable state building agent: Static agents, no evolutionary drift. …

1:15: Principle of Hierarchical Governance: the need of government; Tradeoff: adaptivity vs coherence; Individual agents are incentivized to defect. Government agent that imposes offset on payoff matrix of individual agents. Integration of total reward (bottom up). Credit assignment (top down).

1:16:30: J.Bach: *Self Organization in Nervous Systems/Cognitive Agents:* Evolution of mental governance within each individual mind*;* Neurons are autonomous reinforcement learners; Interaction between distributed perception (bottom-up) and centralized interpretation (top-down) (neural Darwinism)? + Consciousness as colonizing agent? [* Todor: Compare to TUM]

J.Bach: **Relevance to AI:** Current Machine Learning representations: Outside-in design**;** Representations in organisms: Features are functions; Features are kept stable and coordinated via individual controllers; Instantiation, harmonization and dissolution of features via central governance at scene level;

* Multiple systems of interacting agency (prediction, attention, motivation). “self-agent” – *but not like the CPU in a computer, but more like the centralized causal structure in a society of people that emerges as a rsult of an evolution that makes the society more efficient and better at competing with other societies.*
* Organisms have centralized causal structure, but the centralized structure is self

organized.
[**Todor:** J.Bach is restating motives from TUM as in many talks. Also, the centrality of complex systems is in their **memory**, or "also in their memory", not in their processor (or not **only** in their processor); memory is also their "representation"; in TUM everything in the Universe which has properties and can influence the causality-control units is a kind of memory – not only some explicit entities; the difference between internal and external memory is in the way of accessing it, the time, difficulty, energy etc. needed to do so; see my comments in forum Kibertron from 2004 discussing distributed and centralized systems. See also my respond to Michael Levin's "Self-improvizing memory:…", 2024, rediscovering ideas from TUM [22.10.2025] ]

**1:33 h: Malsburg:** *on the sample inefficiency of DL* ***… 1:38:50 h****: in order to build the brain you need a process of self-organization; as we know that starts with a single fertilized cell and goes through divisions and goes through a sequence of attractor states; that's the way the brain is built now, [while] our way of building outside completely ignores such constraints as our implants implicit in the organic growth from one state to the next. We can sort of… when putting together a blueprint we can have the full universe of potential patterns we can can throw onto our blueprint and the only constraint that we are observing is the design ideas we have in mind. So I think of course you can; by knowing the the exact procedures of self-organization, the mechanism of self-organization, you can let them play in your computer that you use for the design and get the final result and then that imposed that final result outside in on a a piece of hardware; but I find it pretty against the nature of things to do that. So I think a complex complex brains – artificial brains will be built inside out and if only they are constructed on digital computers at the present time, I'm simulating all my systems of course on a digital computer, what can I do, even on the digital computer you would rely on an organic growth process, generating the final structure as a sequence of intermediate states.So I think you are stuck with self-organization.**

1:52:45 h: **Dave Ackley:** *Intelligence likes to think it's the captain, but really it's the historian.*
* **Todor:** See future work by *The Sacred Computer*: **"Genesis: Creating Thinking Machines":** InR, Vursherod, Kazborod, Прждн-н-всчк, …\!/ …

## * Multiway Systems as Models to Understand Mind and Universe - a Conversation with Stephen Wolfram, Joscha Bach

https://www.youtube.com/watch?v=O_5e_WSNedE
**J.Bach:** For me the hard problem is not consciousness, but why there is something rather than nothing?[1] Maybe existence is the default. Maybe everything that can potentially exist does actually exist. What exists is what can be implemented. .. Stephen has presented the ruliad "last autumn" … 5:3x m: Causal structure … lower dimensional … 9:20 min: S.Wolfram: consciousness: we believe we are persistent in time and we have a single thread of experience – that's a very non-trivial thing to

believe, becaue it probably is not true of the universe, it is simply a way that we choose to sample the things that happen in the universe …

**Todor:** I've reached to the same unanswerable question, too. Regarding the multiple threads of experience – see "Analysis of the meaning of a sentence…", 2004 and the "integral of infinitesimal selves".

*** Още от Стивън Волфрам от 2024 г.**
* **Stephen Wolfram - Where the Computational Paradigm Leads (in Physics, Tech, AI, Biology, Math,** ..., 11.2024
*** Докъде води прилагането на парадигмата за изчислителната Вселена?** https://www.youtube.com/watch?v=KmoCnTuhMEg
**What is Time? Stephen Wolfram's Groundbreaking New Theory**
https://www.youtube.com/watch?v=o-879Tbn5Ww Dr Brian Keating 302 хил. абонати, 2.12.2024 – Какво е времето? Изчислителната неделимост. Прилагане на правилата стъпка по стъпка – не винаги може да се предвиди, да се получи краен резултат, чрез съкращения; необходимо е да се премине през всички стъпки; напредъкът във времето показва развитието на изчислителните преобразувания. 18 мин: Пространството се състои от атоми пространство, свързани в хиперграф. Времето е пренаписването на хипергрофа, следвайки правилата за преобразуване (изчислителните правила, computational rules). Виж „New Kind of Science", S.Wolfram, 2002. Можем да усетим само причинно-следствените промени; ако нашето състояние не се обнови, не може да узнаем за случването на събития; причинно-следствени графи на зависимостите между атомите пространство; 30-33 мин: наблюдател, който не е ограничен в изчислителните си възможности може да забележи закономерности и в състояния, които за ограничени наблюдатели като нас изглеждат случайни. Възприемаме някои явления като непрекъснати заради ограниченията си – напр. преместването в пространството.
*Срвн мислите на С.Волфрам с ТРИВ, „Вселената сметач", „Схващане за всеобщата предопределеност 2/3", „Вселена и Разум 4", 2001-2004.*
Виж още от и за Волфрам в другите томове на *Пророците: Основен, Ирина и др.*

## * Joscha Bach: Time, Simulation Hypothesis, Existence |

6.10.2020 | Theories of Everything with Curt Jaimungal | 309 хил. абонати
https://www.youtube.com/watch?v=3MNBxfrmfmI
*"What is Real"?* 13 мин. – за Гьодел; дали числото Пи е „истинско", дали „съществува"; *не съществувало* като завършено, защото чрез процедура се смята до определена цифра; дали съществуват целите числа? Не всички
**Тош:** несериозна насока; всичко се възприема с определена РСВ и РСУ и

начин за определяне.  (не доизгледах цялото предаване)
* Пак с Бах за „*реално*“ с един канадец, по-възрастен, където споменава за софтуера като физичен закон, колко оригинална идея – виж ТРИВ, ВиР, 20 години по-рано; също: Волфрам.
– Многопътни изчислителни системи (Multiway Computation Systems)
– Самоорганизация, самопостроряване … с Й.Бах, М.Левин, Кристоф …

* **Самоорганизация** с Яник Килхер (от тук тръгна верига … - зародишът на том „Листове“) около 2/2023 г.
* **Self-organization** – that was the seed of the appendix “Listove” in 2.2023.

*** The Future of AI is Self-Organizing and Self-Assembling (w/ Prof. Sebastian Risi)** | Yannic Kilcher
https://www.youtube.com/watch?v=_7xpGve9QEE
* The Future of Artificial Intelligence is Self-Organizing and Self-Assembling sebastianrisi December 13,2021 https://sebastianrisi.com/self_assembling_ai/
Morphogenesis, …
**Todor** [27.10.2025]**:** The centrality or core in some systems is in their memory, without a need of having an explicit Central Processing Unit, a Government etc. They can be locally self-organizing, but the “wave” of development is hitting the “walls” of the “container” where they are, and they are accessing some common memory. See my discussion at the Kibertron forum in 2004 and my response to M.Levin’s “Self-improvizing” memory in appendix #sf (Science fiction. … Cybernetics) of The Prophets.
See appendix #lazar.

*** ML Street Talk MLST** – с J.Hawkins, B.Goertzel, J.Bach, Natural GI, …Geometric … Peter Velickovic – Graph Attention … Neural Alg. Reas., **Categorical Theory**, **Categories for AI  …**
*** Lex Fridman –** J.Bach #1, #2 , … Vinolas – Alpha Star … David – Alpha 0
1:32 – за творчеството, че програмата открива същите схеми като човеци – сравн. със СВП2 от ВиР твърдението за музиката, че сметачите творци ще открият същите хармонии.

## * Long theory of Mind Grounding **MLST #107 Raphael Milliere**, 14/3/2023 – Бележки на Тодор Арнаудов - Тош

**Тош** “Real” – не отговаря какво е… Пак си играят с думата.
3:31 Causal … 4:41 „просто запомнят“

**Тош:** запомнянето или създаването на модел не е нужно да е буквално, и на папагал не е „точно“; „никога не го бил виждал“
6:25 mimicking … **Тош:** да, но на грешно ниво @ *„Творчеството е подражание на ниво алгоритми“, 2003*
7:23 *reasoning, planning …*
8:28 stochastic parrot – за езиковите модели LLM, Emily Bender et. Al.
11:04 че не било – с изчисл…
11:49 използва ги, но не ги разбира
– **Тош:** а кое е разбиране? „Volition &… dofas“
12:50 Преди 10 години, обучение на експертен корпус, експертно предвиждане, но просто на колко голям корпус? …
15:54 … Mary Shanaha? … John Cell … Theory of Mind … subjective experience – разделено от езика… **Тош:** Да
18:20 Mode of understanding **Тош:** {П} от разбиране #:
… Semantic competence, understanding
**Тош:** Постоянно използват понятия, думи, които не са определени!
19:25 Semantic competence, parse, apply, lexical, structure/composition, reference …
22:20 “Stochastic parrot” point 1) Referential competence 2) Inferential
23:24 Block world (**Тош:** SHRDLU) map …
24:xx Hard coded by programmers; не е известно;
**Тош:** всъщност **е изведено** от тях, но **не е „достатъчно“ постепенно** и свързано с пораждаща поредица; „no inferential” … Analogy – “just compression”, NY Times …

## * MLST #79 Consciousness & Chinese Room

**Data2Vec …** 19/3/2023 … **GPT4 …** Бел. 7/12/2023 Довърши! …

* **Cultural Affordances: Scaffolding Local Worlds Through Shared Intentionality and Regimes of Attention,** Maxwell J D Ramstead, Samuel P L Veissière, Laurence J Kirmayer, Frontiers in Psychology, July 2016
https://www.frontiersin.org/articles/10.3389/fpsyg.2016.01090/full
Бележки на Тодор Арнаудов – Тош: Виж още в началния списък в Основния том със съвпадения и школи. #prophets #tosh1
Бел. по записки от 30/1/2024
От FEP мн. → **Cultural Affordances** – ТРИВ – мн. примери, шапката в църквата, Анализ на смисъла….; enactivism, radical enactivism, embodiment; Field of … affordances …; ecological niche; designer niche; epistemic … ; мн. неща са просто с нови термини; Radical Embodiment; attention – precision weighing to engage with specific affordances in the action-perception cycle; shared intentionality, directed attention, joint attention; action-perception loop (**Т.Арнаудов.**: обратна връзка, feedback); … Избирателно обвързване със средата, горните нива установяват очаквания за точността на сетивната информация от долните; механизъм на пропускане, клапан (gating) на предвижданията в йерархията; води *умелото преднамерено поведение (skillful intentional behavior); …* Променя *полето от възможности (field of affordances) …* Gating, Abilities, Affordances …Trajectories rolling cycles of Act-Percept. … Allocation of attention, coupling … Relevant solicitatios (Важни очаквания) … Dense histories of temporally coordinated interactions & shared cultural practices – Опит – плътни записи на съгласувани във времето взаимодействия и споделени културни практики (действия). Niche construction – multilevel forms of affordance learning; transmission of affordances in socially & culturally shared regimes of joint attention: Построяване на ниши, среди – форми на учене на възможности за действие на много нива и предаването им в режими на споделяне в обществен и културен смисъл при съвместно внимание.
Natural origins of semantic content – local ontologies; Естествено възникващо смислово съдържание – местни онтологии
[**Тош:** виж ТРИВ, „Анализ на смисъла …“, 2004 и др.; ВиР3, „Как бих инвестирал един милион…“,2003)]; dyads (двойки), triadic (тройки): общество, ниша от възможности … Взаимодействия между партньори; direct interactional spheres of communication … joint attention: gaze-gollowing, finger-pointing etc.: съвместно внимание: проследяване с поглед, посочване с пръст …

Избирателност при реакциите (selective responsiveness) … Enculturation, enskillment: въвеждане в културата и в умелостта, обществените умения – многослойни, вложени възможности за действие и очаквания, от които зависят и на които са основани (multilevel, recursive, nested affordances & expectations on which they depend);
Designer environments (Goldstone, 2011) – Проектирани среди, „дизайнерски среди“: хората (с дейността си) оформят средата и рекурсивно построяват нишата във вертележка, циклично изменят степента и свойствата на вниманието на деятелите (агентите) … По Sterling, 2003: постепенно познавателно изработване отгоре-надолу (incremental downstream epistemic engineering) … Познавателни ниши (epistemic niches) … Постепенно построяване, самоизграждане, самосъздаване, зареждане (bootstrapping).
*Re-entrant processing* – Тош.: ***взаимнопроникваща обработка*** преливаща и влияеща си на различни нива (на мащаби и обхвати на обработка, контексти и пр.); **взаимно пресътворяване** в познавателната система и в частност в културните възможности: промяната в културата, станало по някаква причина – например от решение на определени деятели – чрез въвеждане на дадени утвърдени практики, напр. дадени ритуали, технологии, морал – различни от предходните – действа обратно за други промени или за преобразяване на средата така, че да е по-подходяща за промените.

Patterned practices – схематизирани-„набраздени“-дискретизирани-отчленени-отчетливи (културни) практики () …като възможностите за действия: в повечето случаи не е възможно всичко, поведенията, изборите, действията са предвидими в рамките на отчетливи правила и изисквания, които са възпитани във „вкултурените“ индивиди, личности, участници в обществения процес.

*** John Min Tan** …Singapore **–** *“Memory Soup”* … свп. Ан.на см. … ТРИВ (с изкл. че Goal/Reward разглежда като отделни – те могат да се приведат едно към друго, коментари в Дискорд … )
**Тош:** Memory/Памет – и като хеш-таблица, съответствия, също е предвиждане - - ’, в една стъпка; срвн. PCB всичко е с PCB/PCУ, LoD (Level of Detail); embedding – can be any representation with any rules for ?= …
*** #95 Irina Rish – AGI, Complex Systems, Transhumanism** @Neurips 8.8 хил. за 2 мес. 9/3/2023 г. слушано
22:50 .. фактът, че вие сте го създали няма значение … 23:40 … Rise of extinction – humans weapons destruction … - виж „Истината“

**15//3/2023** от тия дни идея: за посл… за динозаври и първите бозайници сп.+ с рисунки … тиранозавър, диплодок, [бронтозавър] … мозъка … и неокортекс, малък бозайник като мишка … профили → OpenAI, DeepMind … без имена – DiploAI, TiranoAI … [vs MammalAI or MammAI] [нова бел. 12.3.2024: → Edge Bing Image Creator 12.4.2024]
*** MLST #75 Emergence [Special Edition] Dr. Daniele Grattarola 11 хил. пок.** 7/4/2023 ICLR 2020 Bengio – Consciousness … 8/4/ Lottery Ticket theory – NN prunning, distillation … 8/4/2023 (да не се бърка с „Hardware Lottery" – виж статия в том #sf „Фантастика. Футурология. .. [10.10.2025]

*** LeCun – Energy-based** – поток, flow, FEP?; self-supervised … conclusion, conjecture … Reason … uncertainty … Energy minimization vector representation … prediction … a point insufficient distribution intractable … → грешка, [,,] …
LC Proposes: Energy-based models – Weaker than Distributed … Energy->Distributed ? Energy -> Concentrated – Contrastive …
SSL = Self-supervised learning; SL – Supervised Learning; RL – Reinforcement Learning
Could energy-based SSL be a basis for common-sense (; Energy minimizing vector representations. Uncertainty … a point insufficient; intractible probability distribution; Energy-based: weaker than Distrib. …; Energy: contrastive vs regular; latent variables;

* Animals & humans learn … unsupervised…

**Тош:** зависи кое се брои за „без учител" и как се нарича – полуформално.
* Scaling to SL or RL will not take us to Human-level AI.
*"There is no such thing as AGI. Intelligence is always specialized. "* –
**Тош:** „битова" представа; способност, capability;
*"rat-level, cat-level, human-level"*…
**Тош:** но на *кой* човек в коя област на познание и действие?
Определението на общата интелигентност е в сравнение, повече или по-малко обща спрямо друга. По-голям обхват, по-широки или повече области на приложение и пр.
*„System 1,2*" **Тош:** да, Yannick Kilcher (Килшер/Килхер?)[25], отбелязва с усмивка, шофиране – става автоматично и т.н.
**Тош:** наричат „съзнателно" нещо и казват, че за да е такова, се изисква да е „*verbalised*", словесно; (да може да се изразява по този начин). Това е изкуствено ограничение. Значи след инсулт, ако не говорят, вече не са

[25] немски швейцарец? неизяснено

съзнателни и не мислят и т.н.: степента и изразяването може да се променя; различни неща могат да се осъзнаят или усетят от различни индивиди или един и същ в различни моменти. Съществуват състояния като „Locked-in syndrome”[26], „Заключен“ - пациентите са в съзнание, но напълно или почти напълно парализирани. В някои случаи могат да движат очите си нагоре-надолу и да мигат, в други дори и това не могат, но ЕЕГ-то им е като на човек в съзнание, спят и т.н. Виж също: За „зоната“, „*flow*”, че е състояние на висока експертност, тренираност, при което се освобождавал „контрола“ на изпълнителните функции и се отключвало творчеството. Понятието за „самоконтрол“ е заблуждаващо; „Човекът и мислещата машина:…“, 2001 – след премахване на колко точно и кои неврони човек вече „няма съзнание“, макар и да изглежда, че има и пр. “Хипотеза за по-дълбокото съзнание“ от ТРИВ, парадокса с импровизацията в музиката, разширена и именувана в бележките към „Какво му трябва на човек? Играеш ли по правилата ще загубиш играта!“, Т.Арнаудов, 2014. Виж също „Вселена и Разум 6“, 2025 (Universe and Mind 6) и „Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини“, 2025.

***** към 1:59 ч: Най-общото нещо, „обработка“, „processing”; преобразуване на данни: data transformation – can it be generalized – the capability to **generalise** … **All is defined in comparison – more or less general**; thus - a basis. “AGI”, УИР е **по-общо, всеобщо, универсално** от онова на съперника, вкл. по-тясното… Колко Об+ от {M:} бел.18.9.2023
+ Изчислителна неделимост накрая, Зрим(дъно, връх), 14.3.2024
* Бел @Вси: Ниво Об$^{+}$, Об$^{-}$ **;**срвн прмр процесори: 6502, Z80,.. 486 CMPXCHG и пр.
* За енергийните методи и развитието на JEPA: **V-JEPA 2**:
* https://www.marktechpost.com/2025/06/12/meta-ai-releases-v-jepa-2-open-source-self-supervised-world-models-for-understanding-prediction-and-planning/

***V-JEPA 2: Self-Supervised Video Models Enable Understanding, Prediction and Planning, Mido Assran et al.** (META), 6.2025 https://arxiv.org/abs/2506.09985 https://github.com/facebookresearch/vjepa2?tab=readme-ov-file

*** Energy-Based Transformers are Scalable Learners and Thinkers,** Alexi Gladstone[1,2], Ganesh Nanduru[1], Md Mofijul Islam[1,3], Peixuan Han[2], Hyeonjeong Ha[2], Aman Chadha[3,4], Yilun Du[5], Heng Ji[3], Jundong Li[1], Tariq Iqbal[1]
[1]UVA [2]UIUC [3]Amazon GenAI† [4]Stanford University [5]Harvard University,

[26] Locked-in Syndrome, Joe M Das; Kingsley Anosike; Ria Monica D. Asuncion., last updated: 24.7.2023: https://www.ncbi.nlm.nih.gov/books/NBK559026/

2.7.2025, https://arxiv.org/html/2507.02092v1
* https://github.com/alexiglad/ebt

*** → Виж основния том и приложения #anelia, #lazar**

*** MLST #52 Adversarial Examples – Hadi Salman** 6.6 хил. пок.
**Злонамерени примери на данни**
Robust features – for humans, shapes и пр. … (и думи, дефин. Проверки, изброяване, обват+, Об+)
48 м. Яник К.: Transfer Learning … - че хората избират особености, които те възприемат; ImageNet != CheckNet? features… transfer .. medical checks …

- Convex Relaxation Barrrier – Adversarial Examples Beyond Security 1/5/2021
  * Randomized Smoothing, Provably Robust (здрав, сигурен, крепък, надежден – издръжлив на злонамерени нападения, опити за объркване)
  * "човеци казват, че две картини са „еднакви"" – маркиране
- Проектиране на не-злонамерени примери от „здрави"/надеждни обекти – проектирани да бъдат надеждно разпознати
- Manifolds – spaces – subspaces … Chollet

*** MLST #104 – Natural GI – Christian Summerfield** 22/2/2023
**Естествен общ интелект**
7-8 мин. Колко било трудно да се намери определение за общи умствени способности (GI).
**Тош: не е трудно.**
21-22 м.: Rich Sutton: 1) *"The bitter Lesson"*;* и 2)„*Neat vs Scruffy*" – Горчивият урок за това, че прости алгоритми с груба сила и повече изчислителни ресурси постигат по-добри резултати от ръчно измислени евристики и пр.
***Тош:*** *Дали?* 2) „Спретнати и мърляви"- школи в ИИ, според едните съществуват ясни и „прости" методи и представяния, според другите е интелектът е сложен" и заплетен, решението е чрез проба-грешка, евристики и т.н. Според мен разделението е условно, виж „Neural networks are also symbolic", T.Arnaudov, Artificial-Mind, 2019. Сложността, неяснотата, хаотичността зависят от оценителя-наблюдател, както е посочено и в ТРИВ. Във видео на Active Inference Institute от март 2024 г.

един се коментира подобно за понятието за ентропия, за известен разговор между Джон Фон Нойман и Клод Шанън, в който първият предложил да я нарекат така, защото “никой не знаел какво точно е“. Неопределеността на информацията зависи от получателя. Може обаче да се сравни „по-прост“ и „по-сложен“ модел чрез мерки като най-малка дължина на съобщението при кодиране с определен език на дадена машина на Тюринг/сметач, подобната сложност на Колмогоров, алгоритмична сложност и пр.
43-44: TD Learning 1980s (Temporal-difference learning) The Meta Problem 53:xx мин. Models for prediction or explanation … 57: Neat & Scruffy:
**Тош:** разделянето е изкуствено. 61: „Out of human distribution, understanding what the humans want … Sutton believes Markov Decision Process … MDP…” **Тош:** започват тема за RL, MDP,. Game Playing - +
64:40: *Observation and rewards together created by the agent itself:* **T:** да, но и при условността на разделението агент и среда при > обхват.
68: Single principle – **empowerment** …
**Тош:** Виж в основния том списъка със съвпадения с учения и школи, подобни на ТРИВ; цитират 2005-2006 г.; виж също понятието ?В МдсП (Какво може да се прави), възможности (за действие), affordances; относно Neat&Scruffy – виж бележки в *„Active Inference Book”,T.Parr, G.Pezzulo, K.Friston,* глава 1, „Overview”, т.1.3 *„Behavior from First Principles”, „Спретнатите винаги търсят обединение, уеднаквяване отвъд (привидно) неоднородността на явленията на мозъка и ума. Това обикновено отговаря на проектиране отгоре-надолу, нормативни модели, които започват от първични принципи и се опитват да изведат колкото се може повече за мозъците и умовете. От друга страна „мърлявите“ „прегръщат“ разнородността като се съсредоточават върху подробностите, които изискват специални обяснения. Последното съответства на проектиране на модели отдолу-нагоре, които започват от данните и използват всичко, което работи, за да обясни сложните явления, включително различните обяснения на различните явления.“*
**Тош:** Двете посоки се срещат по средата.
70: Focusing on the reward maximization behaviors, not the tests … Looking at the actions, not the mechanism …
74:30 – в кн. Тесен ИИ – ако е дадена задача, я вършу много добре; Общ – “ dreams up, brings up”. ? би трябвал да први и защо?
76: “When a measure becomes a target it ceases to be a good measure…”
**Тош:**Да, когато мярката не отговаря по желания начин на измерваното – някои тестове на големи езикови модели спадат в тази категория, за

което е писано още в „What's wrong with NLP", T.Arnaudov., 2009, Part I, Part II – резултати от тестове, които са изкуствени и не отговарят на „флуидна" интелигентност и могат да бъдат хакнати. Но ако мярката и измерването е действително „най-общо" за дадено понятие, явление, то няма накъде да се фалшифицира, бел. 24/10/2023
79: Не един съвършен/оптимален модел на себе си, а много, непрекъснато приспособяване/нови/множество:
**Тош:** Да. Как се разделя, кое е едно: [,,] # PCB, μ

Бел. 7/12/2023: wиж също John Min Tan
→ **Act –'** (Actions) : да, не е ново ?В МдсП ТРИВ …
…

# * Майкъл Левин | Michael Levin

Биофизика, биоинженерство, интердисциплинарност …

**Michael Levin – "From physics to mind" …** всичко в живия свят започва от физика, лекция с това име и в мн. предавания. 5/2/2023
https://www.youtube.com/watch?v=_QICRPFWDpg

От единични/две клетки, които за много хора са „само физика" или химия и не се броят за пълноправни човешки същества, определено „нямат съзнание", нямат ум – как обаче от тях израства ум, без да е свързан с тях?
**Тош:** Да, както в ТРИВ: Теория на Разума и Вселената: няма рязка граница, има различни нива, обхвати и пр. Когато се противопоставя физиката на ума и пр. се смесват @[·] нива на абстракция и обхват, ТРИВ, СВП2 – в малки размери в клетка за кратко време няма живот – нужен е определен [,,] и оценяващо устройство, което да го приеме за такъв. Такива парадокси се получават от сравнение на различни несъвместими области.

*** Блог на Майкъл Левин: Формите на живот са и форми на ум**
**MICHAEL LEVIN**: **Forms of life, forms of mind**
https://thoughtforms.life/what-do-algorithms-want-a-new-paper-on-the-emergence-of-surprising-behavior-in-the-most-unexpected-places/

Изследване на възникваща сложност в алгоритми за сортиране[27].

[27] 15 Sorting Algorithms in 6 Minutes, Timo Bingmann, 42,8 хил. абонати

## Neuroscience Beyond Neurons: bioelectricity underlies the collective intelligence of cellular swarms

Michael Levin's Academic Content
4,63 хил. абонати | 3295 показвания 6.10.2023 г.
https://www.youtube.com/watch?v=2vIlMtmAdak
Бележки на Т. Арнаудов – Тош

**Тош:** За множеството пространства като „морфопространство“ (изграждане на тялото), освен обичайното пространство за придвижване, генетично пространство, на метаболизма в различни аспекти и пр. в математически смисъл; и интелигентността като боравене, придвижване във всякакви пространства – по второто, да, дефинирано е и в ТРИВ, посочено и в по-късна кратка статия за това, че *„Телесността е само координатни пространства (…) и пр.“, 2011.* Но всъщност всички различни пространства *при разглеждането*, *разпознаване, оценка* и пр. им са пространства в ума, който е „машина“ (сметач), която работи с пространства; паметта е вид пространство. Виж „mindspace”, „mindification” във „Вселена и Разум 6“.
М.Л. цитира Уилям Джеймс, че интелигентността е достигане до една и съща цел с различни средства, имайки предвид организми, които постигат целевата анатомия, започвайки от различни начални състояния.

**Тош: т.е. сходимост, свеждане в термините от Зрим.** В по-общ контекст обаче понякога има само едно средство

## * Conversation with Chris Fields and Richard Watson #2

https://www.youtube.com/watch?v=RkkqQF_ulYo
Michael Levin's Academic Content 4,9 хил. абонати 2399 показвания 12.06.2023 г Working meeting between Chris Fields, Richard Watson, and I [M.Levin] where we discuss **error correction** (and **who decides what's an error),** quantum aspects generalized to the larger world, decoherence, observers, and Patrick Grim's fascinating work on **adding a time dimension**

---

24 974 646 показвания 21.05.2013 г.
*Visualization and "audibilization" of 15 Sorting Algorithms in 6 Minutes.*
Sorts random shuffles of integers, with both speed and the number of items adapted to each algorithm's complexity: …https://www.youtube.com/watch?v=kPRA0W1kECg
https://www.toptal.com/developers/sorting-algorithms
https://en.wikipedia.org/wiki/Sorting_algorithm

**to logic** to enable contradictions and self-referential paradoxes to be manipulated as fractal structures. Chris Fields: https://chrisfieldsresearch.com/
Richard Watson: https://www.richardawatson.com/
Patrick Grim: http://www.pgrim.org/ (…) https://www.jstor.org/stable/2215637
**Richard Watson** ~ 36 min+ in order to measure/represent/model a guitar string you have to eventually converge to something like another guitar string of the same kind, so you get coupled with the measured system ... ~ 41 min ... when observing with a frequency that is very different from what's really going on; in order to know it better, you have to get closer to the same frequencies and the same harmonics, and you get more and more part of the system
* *"you can only see reflections of your self"*
* C.F. *"you can see only things which you are equipped to be able to see"*
* *"which are exactly like you"*
**Todor Arnaudov:** not *exactly* like you, it depends on the measure; (if it has to map completely - yes). Yes, regarding the reflections/equipped to be … Compare to TOUM....
43: ... Coupled systems ... https://youtu.be/RkkqQF_ulYo?t=3687
**1:01:27:** work by Patrick Grim ... contradicting self-referential sentences 1990s ...in time ... oscillate in time ... you give it an extra dimension and these things get resolved ...
Намерих го: http://www.pgrim.org/pgrim/intro.html#sec1
**Тодор:***"The boxed sentence is false."* – заблуждаващи неясни логически противоречия. Сравни примера: "Благолаж" и т.н. от ВиР4 или СВП3 ... че са зле дефинирани и т.н. и няма значение;
При П.Грим - трептят във времето, вярно-невярно; също може да саразлични копия, различни гледни точки и т.н. За да има смисъл трябва да им се *разшири обхвата* (добавяне на измерение по Ричард и М.Левин)

## * Какво представляват познавателните светлинни лъчи? Интервю с Майкъл Левин
## * What are Cognitive Light Cones? (Michael Levin Interview) Carlos Farias | 43,8 хил. Аб.

https://www.youtube.com/watch?v=YnObwxJZpZc

Сравни с „**Теория на Разума и Вселената“ (Вселена и Разум), Т.Арнаудов, 2001-2004+,** както и със „Светът като воля и представа“, Артур Шопенхауер, 1813+ (вкл. „За четворния корен…“)

Compare to Theory of Universe and Mind, T.Arnaudov 2001-2004; and Arthur Schopenhauer's "World as Will and Idea", 1813+, including "On the fourfold root…"

26 m … Scale Free Cogition – The same principle on all scales

52-53:xx: CF: Expanding the horizon is what enables Shannon information to acquire meaning because data becomes causally linked to distant and past experiences and acquires implications for future expectations … - Chris Fields thought, a paper on Meaning,

**Tosh:** I also reason like that regarding the causal linkage between the *entities*, and assuming the Universe is a computer and all is "data" (or "information"); note that *data/info* are treated as **entities**, objects, "particles", items. Therefore in this interpretation the "data"/"information" *has to have memory itself.*

*The centrality of the concept of the observer …*

**Tosh:** Right. See also A.Schopenhauer "World as Will and Idea".

**54***: Everything is observer relative … Polycomputation, polycomputing …*

**58:** *… Cancer cells … selfish … identity erased … breakdown, disconnection from the gap junctional communication, weird voltage, … the goals of the cancer cells deviate from the goals of the collective … the cognitive light cone shrunk … cancer models, frogs: reconnect the cancer cell in a proper electrical cell to their neighbours -*

1:05 *… the high level executive cognitive function has to traverse down and change down the molecular properties of the cell membrane in your muscles … it has to affect the microlevel biology … go from cognitive goals to molecular states … in order to remember: cognitive function has to traverse down and change down the molecular properties of your cells in your muscles*

**Tosh:** Right. However this is in one of the tracks of following the processes. There is always a microlevel biology that is active, the higher level is recognized and interpreted as such after sampling and wrapping up a big enough range. The high level is/can be regarded as a *"view"* (in computer science), a "segmentation/tokenization", "observer-evaluator" artifact. Spatially, at high resolution, the lower material level, smaller scale/range entities –

causality control units in TOUM – are “everywhere”; in a separated view/physics they don’t know about the high cognitive functions, they are always following their low level laws, forces etc. From their POV an anticipated change in their state (== a match that is less than expected, unpredictable change for them with a target precision etc.), is namely a *prediction error,* while for a level – in this context an *interpretation,* another causality-control unit, – the *changes are predicted correctly,* they happen as expected and wanted with the selected resolution of causality and control; the higher level is thus the “master” unit, the mast control-causality unit (both directions of the words are correct), where the lower level one which faces prediction errors, *unwanted* changes, writes/states to its memory, is controlled, subordinated, lower level from that direction of level-traceing. Thus the direction of the traversal is: the *perceived correctnes* of the prediction belongs to the “higher level” and the perceived “error”: to the lower. However this is from the POV of an encompassing evaluator-observer, which *assumes, counterfactually* that *if* the higher cognitive level didn’t chose/did/… this affordance, then the future of the lower level unit should have been something else. However in order the higher *functional* level, the *functional thread* to exist at all and to send *commands* for changing the molecular properties, the *lower level units* in the *physical level/thred* had to be chained and structured. The *connection* between these levels (referred also as “levels of scale”) is in an accord, match in different measurements, associated with the concept of scale, such as *range, size, span, depth, distance, duration; precision* etc. Matches at different levels, see the “Multifractal…”[28] from Levin’s group and TOUM, in particular *Universe and Mind 4*[29]*.*

The above logically implies that the higher levels emerge from the *errors.* The subordinate units, from the top-down POV, are supposed to be imperfect enough and mispredict, so that the higher level to “correct” them from its own POV. Respectively the lower level units have to be susceptible and mouldable to the changes from “outside” or maybe more correctly: from “the other side”, the counterpart. The last word suggests the “perception-action” loop, a “dyadic relation”. As reducing the span of evaluating and sampling and increasng the resolution, the *less clear* the *direction* of the interaction and of the sides become: which one of the smaller divisions is

[28] "Multifractal social psychology" - a talk by Damian Kelty-Stephen from Levin’s group also the https://youtu.be/P89WTmNBjBk?si=oU8WW-ceuwYTrhwu

[29] https://github.com/Twenkid/Theory-of-Universe-and-Mind/blob/main/Works/Todor_Arnaudov_Theory_of_Universe_and_Mind_4.pdf

"inside" and "outside"[30]. For example in a physical model of electrons or molecules, at a bigger scale and resolution and selected segmentation, the atoms in the molecules are separated, but once you zoom in, their "electron clouds" interact, their atoms sphere cross each other, they belong to more than one atom. Similarly with human interactions and belonging to different categories, legal entities etc., see the perception-action mutual causation/connection in the FEP/AIF literature and in Reinforcement learning. A possible solution is that both action and perception, or inside and outside, are actually one "thing" which is rendered/interpreted. The other or *another* separate thing becomes the *evaluator-observer* who/which selects how to render and interpret that "one thing", the oneness could be for example an *"unclassifiedness"*, indeterminacy at quantum physics level. Note also one linguistic paradox or misnomer: "quantum" physics is supposed to be about quants, sharp portions, discrete entities, at least it is for example so in the mathematical quantization, division of a continuous numerical range into discrete subranges for a digital representation, lower resolution etc. On the other hand, the quantum physics is about the "quantum *field*", as Bernardo Kastroup insists in the cited talks, there are no particles, these are only ridges of the quantum wave functions, i.e. continuous, "analog" etc. Yet continuous and discrete depends on the resolution and the "smoothing", rounding and and preferences of the evaluator-observer. The natural sound – the oscillations – are supposed to be continuous, analog, but the ear has certain "hairs" for detecting particular frequency bands with particular precision/peak sensitivity etc. [which leads to perceptual artifacts]
… changing their direction of development, state, trajectory, because of the *prediction error* that they  summation over more items  … scale … accumulation … molecular See TOUM and the illustration in "Principles of General Intelligence", T.Arnaudov 2009.

1:15 Active compassion … well-being …
**Tosh:** It is modifiable by simple chemicals, drugs ..
"Care, compassion and intelligence – mutually reinforcing loop"

*** Виж също** статията и бележките и тълкуването на „Самоимпровизиращата се памет …", статия на Майкъл Левин от 2024 г., и сравнение с публикувани над 20-години по-рано идеи от ТРИВ в

[30] Compare to Thomas Metzinger note in "The Ego Tunnel", 2009 that at lower levels it gets harder to decide which part of the brain or a system is supposed to be "consciousness" etc. Compare also to these questions in "Man and Thinking Machine – Analysis of …", 2001.

приложението за Фантастика, Футурология, Кибернетика и Развитие на човека: #sf #cyber.

## * Спорността и неяснотата при романтичното определяне на светлинния лъч на познавателността

**Тодор Арнаудов**

Бележки относно **Michael Levin's Scale-free Cognition & Cognitive Light cone;** ?= срвн. с Фристън, одеало на Марков, Markov blanket. 24/10/23

**Тош:** Романтично. Обхват [,,] на цели, които да причини и предвижда – да, но можем обективно да ги потвърдим/докажем, ако сме Бог, който може да проследи дали наистина въздействия на онова, за което *си мисли* по *този начин* и също ако *факторизираме достоверно волята* – неговата и на всички останали деятели, т.е. на цялата Вселена; като Бог, ние ще знаем *наистина* кое и кой колко „тежи" в крайното състояние. Виж „Анализ на смисъла" за примери за условността на факторизирането на причините за определени действия и събития. С отдалечаването от крайното, целево събитие, отсяването на точни отделни, тесни, малък брой причини става все по-условно, защото все по-голям обхват от времепространство участва/въздейства. Коя сила от всички се преценява като главна, с каква разделителна способност и пр. зависи от оценителя.

Левин признава, че организмите поглъщат материя от средата и че за всяко по-високо ниво на представяне, по-ниските са като среда – аналогично на нивата въображаеми вселени/машини в ТРИВ (Вселена и Разум), най-ниското е „действителността", „"нулево". „Проблемът" в „поглъщането" е и че означава, че *волята* на човека е следствие на *съвкупността от волите* на „погълнатите" части от друга воля. Левин често цитира смятания за основатели на съвременната психология Уилям Джеймс с определението му за интелигентността като способност да се решава една и съща задача с различни средства, в променящи се условия. Така в организмите и във Вселената умът се проявява във всевъзможни форми и решава сходни проблеми в различни пространства: генетично, морфологично („morphospace") и пр. Тук ТРИВ е в съгласие по отношение на пространствата и разглеждането на всички видове обработки като вид модалности, координатни системи, следвайки общи принципи.

Друго понятие при Левин и учените от неговата школа – интелигентността като загриженост за нещо/за други (care), по-високоинтелигентните същства ги е грижа за по-голям обхват и по-далечни събития и по-чужди на тях други агенти, докато по-малките и

примитивни организми се „грижат“ или могат да „проумеят“ или само собственото си оцеляване, или събития в по-малък и кратък обсег – например децата си, стопанина си – ако са домашни животни – и пр.

От друга страна, едновременно се признава, че всяка клетка и съответно във всеки агент, защото познанието се случва на всички мащаби – се грижи и за общото благо на по-голямото цяло, в което е включена. Тук трудността е и в това, че малкото по определение не би трябвало да може да осмисли цялото и по-голямото, една клетка не може да мисли като мозък и не може да има необходимите данни и памет, и един мозък не може да мисли като цяло общество, държава или планета с всичките им части, и е нужно да има ясно определено определение и мярка кога е така. В някои случаи по-малката част, подсистема, може да *си мисли,* че работи за общото благо, или „против“ него, както и по-голямата може да *накара*, *да подчини* по-малката, или да я *съгласува* със своите цели, „благо“. Съставните части също могат *да казват,* да *твърдят* заблуждаващи неща.

Може би тогава, когато едновременно има „полза“, т.е. успешно предвиждане, разполагане в целево състояние, *на различни обхвати*, нива, от гледна точка на различни подсистеми в голяма система, съвпадението може да се смята за критерий за единство на системата и показател, че частите работят заедно за обща цел; необходимо е обаче оценител, който достоверно да знае и да реши кое чия цел е на какъв мащаб, обхват, вид част, и че „всички в системата са доволни“ и целите им съвпадат.

Проблемът за „общото благо“, кое е „полезно за обществото“, „не питай какво може да ти даде държавата ти, а с какво ти би могъл да допринесеш за нея“[31] и др. са свързани с темата.

В система от множество части на различни нива и множество агенти обаче, *общото благо,* т.е. успехът на цялото измерено по определен обобщаващ и обхващащ голям брой части начин, означава *смърт* за едни и *живот* за други, или кратко съществуване за едни подсистеми, клетки, части, и просъществуване през целия живот на системата за други. Така *„общото благо“* може да бъде просъществуването на *системата като цяло,* което обаче в крайна сметка е следването на вселенските правила, каквито и да са всъщност, „случването на онова, което се случва, по начина, по който се случва“ и на най-ниско ниво, което каскадно се разпространява и до по-големите

[31] Известната реч на Джон Кенеди в Берлин. Това което казва е всъщност е част от немския културен код, да са верни към господстващите, на това учи всяка власт и същата риторика е използвана и от нацисти, комунисти и пр.

обхвати, а после от друга гледна точка изглежда и че горните нива управляват, потискат, насочват и по-ниските.

Освен това самоличността и целите мога да се променят на всички нива във всеки момент – оценител-наблюдател или при определени критерии се решава, че целите, самоличността и пр. са „същите“. При променящи се цели и самоличност, планирането от сега за задоволяване на бъдещи цели на бъдеща самоличност с други ценности, мотиви и представа за „добро и зло“, може да бъде достоверно, ако се извършва постепенно и покрива и междинните стъпки на преход, т.е. ако направлява и контролира промените през цялото време.

„*Всичко е добро“* или всичко е „*никакво*“ – нито добро, нито зло, ако се гледа обективно или *всеобхватно,* с все по-голям обем от причинно-следствени връзки, управляващо-причиняващи устройства, агенти, организми и пр. Доброто и злото е субективно от гледна точка на определен деятел, оценител, „разрез“/подпространство/обхват/обем…, подбор и пр. Не можеш едновременно „да служиш на „много господари“, освен ако Той не е Вселената, ако *всички* са ѝ подчинени и просто си изпълняват програмата. Виж и Шопенхауер. „Свободна“ воля? ?В е свобода и ?В Воля. Мисъл на Ш.: *„Можеш да правиш каквото искаш, но не можеш да искаш какво да искаш.“*

В школата на ПСЕ/ИЧД на Фристън обясняват как организмите успяват да се запазят, като се приспособяват към промените, развиват в себе си определени функции и т.н. Те се фокусират върху малкото времепространство, ограничено като „организъм“, неговите мембрани, външни клетки и т.н., те са системата, а другото е „среда“. Обаче дали *те* са творците в това разделяне? Защо да не *обърнем лещата* и да се запитаме *как и защо **средата** оформя организмите, тя ги извайва и запазва, вместо да ги разруши и т.н.* По-сложната и богата система е именно „средата“, а организмите са нейни елементи, съставни части в различни мащаби. Не човекът управлява колата, а колата „привлича“ човека в себе си и го „кара да я кара“. Не човекът управлява компютъра, а машината го предизвиква да натиска едно или друго копче и да гледа тук или там. Всъщност всички са част от система, в която „танцуват“ заедно.

Сравни с А.Шопенхауер: **„Светът като воля и представа“**, напр. „трета медитация“, бел. 5.6.2024.

* Продължавам последните идеи в една от уводните статии в раздела за съзнание и панпсихизъм: „Разграничаването на системата от средата и „самосъздаването“ са спорни въпроси“ и „Средата като жива и живите

същества като неживи“; виж също „Вселена и Разум 6“, „Нужни ли са смъртни изчислителни системи…“, основния том и др. (…) [14.10.2025]

## * За понятията за състрадание и чувства като спорни явления за определяне на одушевеност на цялостен организъм заради възможността да се потисне с „прости“ молекули и пр., Тодор Арнаудов

Мисли, свързани със “Scale Free Cognition” по Майкъл Левин, 11.4-12.4.2023. Виж и по-долу бележката-схема „Болка. Pain. Suffering”, 28.8.2023, обзора в основния том раздел за учени и школи за Майкъл Левин и колегите му Крис Фийлдс и др. за “*Technological Approach To Mind Everywhere”* (TAME) и др., и в приложението за фантастика и кибернетика прегледа и бележките към статията за *„самоимпровизиращата се памет“* от 2024 г. Виж също *„Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?“,* Т.Арнаудов, 2025 и *„Вселена и Разум 6“*, Т.Арнаудов, 2025.

**Тош:** Предвиждане и управление (prediction, control) … +
- goal-directed predict. – computational agents +

**Тош:** Всяка целенасоченост (goal-directedness) е изчислителна; цел- предсказване (прогнозиране, предвиждане)-изчисление-изображение (mapping, функция в математиката) са съответни понятия. И молекулите, и атомите – същите най общи свойства.

“Adaptive”, приспособяващо се – р. видове; кое се променя. ?В е дефин.? Виж също Тодор Павлов, „Теория на отражението“.

“metazoan swarm – body – desired end-state .. tadpole … tails grafted to the flank …” (откъснати опашки, които се възстановяват) – мс. сп.+ ?Т като извличане последователно, без пълно обучение, ad-hoc …

*** "Cognitive lightcone": compassion:** 28/8/2023

## Познавателният светлинен лъч: съчувствие

Бележки към разсъждения на Майкъл Левин от Тодор Арнаудов – Тош

**Тош:**

Съчувствието е такова:

1. за външен абстрактен оценител/наблюдател
2. и въобще чувства → определени вещества и др., т.е.

*„малки"* промени, много по-малки от обема на организма, системата:

└ мащаб

преобръщат усещането

↓

превръщат човек в кукла, променят поведението и усещанията

└ „care – ?К се определя

└ за … на разстояние …

↓

в ума, представа … → само ако са „*first class citizens"…* и нещо наблюдава, ако са цялостни самостоятелни „обекти"; т.е. нещо може еднозначно и обективно да определи, че дадено състояние е „грижа" или състрадание и пр., а не е например *имитация, фалшиво* и т.н. (например обикновено всички политици *заявяват, твърдят, се представят, се изказват*, че работят за „*националните* интереси", „благото *на народа,* обществото, планетата"; всички фирми и организации с търговска цел *заявяват,* че работят „за задоволяване на клиентите си" и т.н. – какво точно е „национален интерес", „народ", благо или полза за тях – как точно, кой, кога и пр. ги измерва, отчита, определя и пр.?)

└ съобщаване, заявяване, декларация (report) - всичко може да се; трябва и на дело и да може „достатъчно обективно" да се измери

└ но отдалечените „неща", във всякакви пространства, не могат да се проверяват *от настоящото* и от *текущото* място*; „нещото"* трябва то да „оцелее" или да има обективен съдник през цялото време; и „нещата", зявеното и действителносто може да не са съпоставими и измерими, защото с времето целите и следствията от дадени действия може да стават противоречиви; виж по-горе за факторизацията.

**Записки#2:** Lightcone → ?= Фристън; Markov Blanket, одеало на Марков[32] 24.10.2023. „Романтично"… [,,] на цели, които да причини и предвижда – да, но

[32] Виж село Марково до Пловдив и Марково тепе мол – докато пиша се намирам някъде по средата между двете „одеала"…

можем ли обективно да ги потвърдим, докажем, само ако сме Бог, който може да проследи дали наистина намеренията и „грижата“ на дееца въздействата върху онова, за което си мисли, както и да отчетем „правилно“ факторизацията на волята*. ?В е причината става все по-условно с отдалечаването от събитията или състоянията, защото нараства [,,] от времепространство, което участва/въздейства и влиянието става все по-размито. Кой е главен и има ли такъв? … Виж подробната статия по-горе.

*** The collective intelligence of cells during morphogenesis as a model for cognition beyond the brain,** M.Levin: 20.2.2023, Talk, 23.1.2023 …
**Тош:** Сравни с А.Шопенхауер, 1813+

# **SEMF:** Испански институт насърчаващ междупредметни изследвания

## **Поредица „Spacious, Spatiaity”:** https://www.youtube.com/@SEMF

**QSR * Spatal {П}**

10/4/2023

**Anthony Cohn | Spatial Intelligence and Challenges of the Spatial World**

https://www.youtube.com/watch?v=ytiZp7_UVIA

…

Leave the room & turn right into the corridor … 14 min. Бел. #: **Зрим**

- Topology **@, &, -*-**
- Orientation/Direction d@, **d&**
- Size **d[,]**
- Distance **d[,,]**
- Shape
  - ➔ Challenge: **\|/** a calculus …
  - ➔ Level of granularity, efficient reasoning PCB
  - ➔ Interaction between different aspects

Use the structure of the environment

Ontology, topology … Spatial change + Vagueness & uncertainty + reasoning mechanism + pure space v. domain dependent …

## **Conceptual Spaces & the Geometry of Word Meanings, Peter Gardenfors** SEMF 21.11.2022

https://www.youtube.com/watch?v=87O9fnu8BTU

**- Spacious, Spatiality:** Пространство на понятията и геометрия на значенията на думите … 57 м.

1. Категория предмети – същ. Noun
2. Области от области – прилагателни Adj. (Region of a domain {пП})
3. Пространства и силови отношения: предлог Prep
4. Сила и получаване на вектори – глаголи Verb
5. Променящи се вектори – наречия (some Adv.)

“Single domain reference + convex regions” .. → predictive

https://www.researchgate.net/publication/228421345_Conceptual_Spaces-The_Geometry_of_Thought_Peter_Gardenfors_A_Bradford_Book_2000

Виж бел. в началото на основния том на *„Пророците на мислещите машини“*, **Когнитивна лингвистика**, прегледа на учени и школи, където са посочени други книги за геометрията на значението, с връзки към книгите в библиотеката на Архив, и други работи от Петер Грендерфорс, който е колега на българина Йордан Златев в шведския университет Лунд.

*** Бележки към Bach,? TedX 2016 ? …** и др.
*** From Artificial Intelligence to Artificial Consciousness | Joscha Bach | TEDxBeaconStreet,** TEDx Talks 93 638 показвания 13.12.2016 г.
https://www.youtube.com/watch?v=Jr7gY3JyzP8 (И други участия)
5:30 – 6:xx мин: patterns→percepts→simulators→concepts …
*"Minds are not classifiers, but simulators & experiencers"*
**Тош:** да, ТРИВ, но също и класиф.: % (разделяне)

4:31 Cognitive AI vs Narrow AI … Minds vs Pattern Recognizers …

- System recognition … s.identification {К} ?Тт … 10/2023 Като във ВиР – зада се разбере обхват е нужен достатъчно общ ум – колко е сложна клетката - ?К се мери, за кого?

24/10/2023 – понятие за [,,], обхв. На възпр/управление (#,!) {П}[,,](#,!) Разбиране на пнт, @ на частите, @всч, % …
**– Truth** – и от подкаст – **existence and match, not prove/logical.**
**Тош**: Да, ТРИВ(ВиР). Срвн. Също ART, Steven Grossberg. Adaptive Resonance. Theory of everything … 7. Bach – *"We need to understand the nature of AI to understand who we are" – срвн. С ЧиММ, 2001*

Някъде при обсъждания за реално и въображаемо:

- "**Physical** – *directly* observable" (физическо – може и „реално", действително)
  **Тош:** Кое обаче е **„пряко"? Нищо не е**. Всичко минава през ум. Един електрон, **е-** не се възприема по същия начин от колбичка или пръчица в ретината, както от неврон във V1 или таламуса или в PFC, и зависи *кой точно* неврон на съответното място, или кои, в какъв ансамбъл, и т.н., или от „ума", „съзнанието". Всяко отсъждане за реално/нереално е от оценител-наблюдател.
  – Защо *това*, „пряката наблюдаемост" – каквото и точно да се смята за пряко или за достатъчно пряко – да е „*физично*"? Защо действието от разстояние се смята за призрачно? ("Spooky action at distance"). Във Вселената Сметач не е – не необходимо да има движение по-бързо от скоростта на светлината в клетките от пространството, а да се случват процеси между тактовете на обработка. През тях може да се обработи ако трябва и цялата памет на Вселената, без вътрешния оценител-наблюдател да усети, докато обработката не стигне до него, за да обнови състоянието му.
  – System recognition {К} ?Тт (Контекст на търсене) … същото като ВИР; за да се разбере обхват е нужен достатъчно Об+ум; колко е сложна клетка;

?К се мери и за кого? … пнт за обхват на възприетие/управление [,,]; ? {П}[,,](У_В) √ на пн, @ на частте; @ ∀ % * Повече за *Зрим*: бдщ

*** Болка * Pain * Suffering –** Тодор Арнаудов, 28.8.2023

└ “съзнание“, „душа“ … └ преобръщат усещането

↓

Congenital Analgesia etc.

／ Болкоуспокояващи …

／ └ губи се усещането – субективно

Болката като вирус – напаст | | понякога се усеща, но не неприятно

↓ | модифициране на сигнала

и др. страсти |

| Водещ – успоредни?

| ? сливане

Кората няма рецептори за болка

… зап. 15-3-2024, 4:50

⫽→↔↕∆∫≈ ╔

**Виж:** „Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?“, „Вселена и Разум 6 и „Човекът и Мислещата машина…“, 2001 – в последната разглеждам болката за пръв път в ТРВ.

***** Neural fibers and spinal cord damage cause loss of pain and other sensations as if these parts of the body “don’t exist” anymore for the agents’ sentience, consciousness, mind. On the other hand: “phantom pain” is felt for missing limbs, without incoming neural signals from the body. Compare also the mental pain in mental illnesses, “social pain” – rejection etc., pain from opioids, withdrawal symptops etc. –**“pain”** as a virus and a bug. [22.9.2025]

* **Scientists discover brain circuit that can switch off chronic pain,** October 10, 2025**,** University of Pennsylvania
https://www.sciencedaily.com/releases/2025/10/251009033126.htm
*** A parabrachial hub for need-state control of enduring pain,** Nitsan Goldstein et al., Nature, 8.10.2025, https://www.nature.com/articles/s41586-025-09602-x

## * ActInf MathStream 009.1 ~ Jonathan Gorard: A computational perspective on observation and cognition
## * Джонатан Горард: Изчислителна гледна точка към наблюденията и познанието
## Бележки на Тодор Арнаудов – Тош

Active Inference Institute * 6.3.2024 * 3,12 хил. Абонати * 926 показвания към 21.3.2024 https://www.youtube.com/watch?v=I3rhsT-8isk
„A Functorial Perspective on (Multi)computational Irreducibility“, Jonathan Gorard, 10/2022 https://arxiv.org/abs/2301.04690
Преглед, откъси и бележки към работата от Тодор Арнаудов (**Т:, Тош:** …)

Опит за формализиране на понятието за изчислителна несъкратимост като точност на функториалното съответствие между категорията на структурите от данни и елементарните изчислителни операции, и съответна категория от (едноизмерни) кобордизми. …

**Понятия:** Кобордизми са вид изображения между топологически многообразия[33]. Изчислителна и многоизчислителна несъкратимост (за „многопътни системи“, вид недетерминирани машини на Тюринг; виж Волфрам) – пътища в машината на Тюринг.

**18 мин.** Дж.Г. за условността на сложността, защото зависи от оценката на наблюдателя, обяснението на втория може да бъде много по-сложно отколкото е (възможното) описание на действителната система – този въпрос е разгледан в ТРИВ/ВиР и коментиран с това, че „нищо не е хаотично“ и че изглежда такова („сложно“ отвъд способност за предвиждане, което Оценителят смята, че би трябвало да има, за да бъде „нехаотично) поради липсата на информация и пр. на Оценителя. В ТРИВ се предпочита терминът Оценител вместо Наблюдател. Кибернетика от втори ред…
**25 мин.** Дж.Г. **Теория на категориите**… Алгебрична топология… Може да определиш нещата както с техните вътрешноприсъщи свойства, така и с онова, което можеш да правиш с тях. … Не само топология, но математическите структури или структурите въобще … Множествата са определени от онова, което е в тях … В теорията на категориите е съвсем различно … не можеш да погледнеш във вътрешната структура на обектите … Тяхната идентичност се определя от начина по който се

[33] Manifold в англоезичната литература

отнасят спрямо обекти от същият тип, какви функции могат да се приложат върху тях и пр. Например реалните числа от гледна точка на теорията на множествата: всички **числа** в това множество; а от теорията на категориите – всички **функции**, които може да се определят, с аргументи реални числа или други числени системи и ги свежда до реални числа и пр. … Други теории – двете гледни точки всъщност са еднакви .. 35 мин … Причинност и как да се дефинира… хиперграфи .. „блокчейн“ … пресъздаване на цялата причинно-следствена история … паралелното програмиране е свързано с ученията за причиността … **39 мин.** В пълно определени изчислителни системи не съществува понятието за модалност [**T:** възможност?] по Лайбниц. Или причинността трябва да се дефинира като многопътна система, където се случват множество от различни изчисления и причинността се пресмята като отношения между тях, или изпадаш в безизходица и е нужно да се създаде фундаментално нова теория за причинността … Трябва да мислим за категориите не само като единично изчисления, единична поредица … а като алгебрична структура от всички възможни изчисления и структури от данни … Когато можеш да прилагаш какво е причинността на ниво токен, буквaче, може да изведеш „ковариантно изчисление“ … Ако знам състоянието на една частица, но и цялото и минало, всички състояния … Нов вид изчисление … Друг вид – в пространството на разклоненията (branchial space) … Ковариантно смятане …
**44 мин:** силно свързано на понятието на Майкъл Левин за „многоизчисление“ (**polycomputation**), трябва да се определи какво е причинност, … графите на Бейс не съдържат всичката необходима информация … [**T:** а само вероятности: без причините и по-конкретната връзка] … Детерминистична машина на Тюринг – като филм, не може да се променя…
**47:30:** Два вида записа на причинността: 1) спекулативен, динамичен, който може да се презапише; и 2) определен, непроменим (immutable, като в програмирането[34])… Квантова теория на информацията … Разпределени/паралелни изчисления … Позволяващ спекулативно изпълнение ***[T: изпреварващо, едновременно]*** на определен брой стъпки, разклоняваш се в многопътните системи, и така се построява суперпозиция най-малко на набор от възможни причинностни истории, но накрая трябва да избереш действителната операция, която да се изпълни ***[T. в пътя на програмата, да промени дадено линейно развитие; да промени оперативната памет „окончателно“ и пр.]*** и така се получава този голям блок, който се поставя и причинностната история

[34] https://en.wikipedia.org/wiki/Immutable_object

става определена. *[**Т.** Защото алгоритмите, които се изпълняват са записани като последователни и с една нишка, въпреки спекулативното изчисление, и защото разглеждаме тази единична нишка. Спекулативно изчисление в масовите микропроцесори има още от Intel Pentium, 1993 г., ако конвейерната обработка също може да се приеме за изпреварващо изпълнение – още от 1970-те. ]* … Дали това може да се приложи и за агенти, взаимодействието между двете различни причинностни структури: динамична и неизменяема … **50:** Компромис между познавателно и полезно [**Т.** epistemic and pragmatic, в ПСЕ/ИчД] …

Домакинът на разговора отбелязва връзката между различните школи на Фристън, ПСЕ/ИчД и че трябва да общуват повече – виж този списък и ТРИВ.

| **Джонатан Горард** | **Тодор Арнаудов** |
|---|---|
| ок. 13:хх … Системи, наблюдатели, наблюдаващи функции… Системи, наблюдатели и кодиране … Във всеки модел има определено количество „огрубяване“ (**coarsing**)… Има състояния, които в системата са различни, но във вътрешния модел се обработват като еднакви, поставят се в една и съща кошница. Колко е груб моделът се определя от това доколко моделиращата функция не успява да бъде сурективно изображение.<br>Характеристика от алгебрата или теорията на категориите – колкото по-малко са морфизмите и епиморфизмите толкова по-груб, по-абстрактен и идеализиран е моделът ви на действителността.<br>Колко добавъчни или поддобавъчни са сложностите като ги съчините успоредно заедно дава относителната сложност на еволюционната функция … това е функцията, по която се развива изчислението напред във времето, за разлика от функцията на равнозначност, която заявява, че две изчислителни състояния, две структури от данни, следва да се разглеждат като еднакви … тази игра и „мета-начин“ за мислене за изчисленията на системите и онези на наблюдателите. Защото ролята на системата е да се развива във времето, докато на | Понятията „разделителна способност на възприятието и управлението“ в ТРИВ, „Вселена и Разум“. |

| | |
|---|---|
| Наблюдателя – да прихваща състояние на системата, които са различими в реалността и като предмет на идеализирания модел на наблюдателя, тези състояния се обработват като еднакви… Така системата **[действителната система, Т.А.]** дефинира функцията на развитието, а Наблюдателят – функция на равнозначност/съответствие (equivalence f.) и по този начин „обменът“ на сложности става тъкмо обмен между това как знаете какви са алгебричните правила, които описват сложностите, тъй като те се съчиняват последователно …<br>Като наблюдаваме, построяваме кодираща функция, която превръща конкретното физическото състояние на наблюдаваното във вътрешно абстрактно състояние на модела. | |

Многопътните системи са пътища, покриващи дървета и пр. в граф. Сравни с ТРИВ, „Вселена и Разум 3“, „Вселената сметач“…
Виж също:

* Refuting the Metaphysics of Wolfram and Tegmark, Joseph Natal, 24.11.2024, https://arxiv.org/html/2411.12562v3 @Вси: дочети по-внимателно и срвн.
* Виж бележки за *Математическата Вселена* на Тегмарк в този том *Листове*.

## * Karl Friston @ MLST … Карл Фристън бел. 7/10/2023

Виж Youtube Channel: Machine Learning Street Talk.

23 мин: **Computationalist** … Information … Информация и изчислителност не са 1:1 ?Л Об+ инф.?
36: Дефинир. На маш.интел. Shane Legg … траектории (бел. 24/10/2023)
39: планиране – Тош: но това е абстрактно в ?В [,,] се анализира и т.н., на ниско ниво планът е в < обхват – на мн. места в ТРИВ … „System 1", "System 2", Daniel Kahneman … Канеман … „мислене бързо и бавно" – сравни Кант, Шопенхауер: нагледно и дискурсивно, Приложението с бележки към „Какво му трябва на човек":„Хипотеза за по-дълбокото съзнание" и др., включена и в Пророците на мислещите машини.
**Тош:** ?К оценител √, че друг планира; както и че той самият

- Като срвн. своите -- ' с резултатите (но и доколко са надеждни и едните и другите)
- ТРИВ – сп. „оценител", а не наблдател, защото второто Мд бд пасивен.
- Multiscale …

* Виж също приложение Algorithmic Complexity, #complexity; Вж бдщ/Зрим за разшифроване.

## * Max Ramstead: 28/10/23

4.1. … Internal states of the organism … expectations about the world … recognition density – posterior belief … generative model – control system … Variational Bayes … Phenotypical statistical relations: preferences, action, policies; expected sensory causal regularities;
Generative process (external world, includes the organism's actions); Variational densities …
- The claim living systems avoid surprise
- The living system changes so as to become (statistically) consistent with the preferred world – that is, according to its preferences & expectations about the world (Тош: виж ТРИВ, Анализ на смисъла, …като „висша форма" на дифузията, ентропията – не само живите организми; виж цитата на Тодор Павлов, „Теория на отражението"; жив.орг. са „по-гъвкави", но и по-малките и най-малките състави части „*отразяват*", *„приспособяват се*" в рамките на възможното за техния по-малък мащаб и вътрешно разнообразие, структурно разнообразие; виж понятието от кибернетиката „requisite variety")

*"**Cognition**"* – i.e. what the system does – is it only this

- By developing, simulating & analysis of the possible generative models that explain how the recognitin density of interest (the system of interest) changes so as to attain minimal free energy … that follow Variational Bayes information.

… the path to stay alive (as control systems); Nested Markov blankets of M.blankets

- The action-guiding believes

**4.2.** Enactivism .. “classic” – 2010 – classical, autopoetic … Self-evidencing process, cornerstone …

4.3. Nestedness – cognition beyond the brain - sparsity structure ..
“The same statistical form smaller and smaller …”
Тош: Срвн. ТРИВ; виж от блога „Изкуствен разум“, статията за „интеграла“ от безкрайно малките „Аз-ове“, the integral of the infinitemal self, и „Анализ на смисъла…“, 2004; клише „sparsity” …; clustering
**Tosh:** Compare with Theory of Universe ad Mind etc.

*** Youshua Bengio #63 …** Gflow …

*** Max Wellington : KF vs St.Wolf. –** не М всичко да се предвиди – изчислителна неделимост (**T:** да, винаги има дъно) 30/10/2023

*** №53 Виж Quantum Nat. NLP … Bob Coecke …**
Схема на Зрим – снимка.скан …

*** AI Alignment – Connor Leahy** … | **Tour de Bayesian**
*** Francois Choillet –** value, program / discrete program search } Об+ сравн. Творчеството е подражание на ниво алгоритми и др.

*** Kenneth Stanley –** open endedness, stepping stones, POET, NEAT, HyperNEAT (виж статията в блог „Изкуствен разум“ (Artificial Mind) … и Анализ на см. на изр.)

*** Gary Marcus** – книги 1998, 2001, … дискретни, точни, CPU …
*** Emergence –** вмъкни откъс и с Йоша Бах за крайността;

→ виж приложение #irina

In *Artificial General Intelligence*, *Artificial Intelligence*, *Computational Creativity*, *Creativity* by *Todor "Tosh" Arnaudov - Twenkid* // Saturday, December 16, 2023 // *Leave a Comment*

## * On What Creativity Is - Different Tints and the Pereslegin's remark about the modern AI researchers rediscovering Lem 1963, Todor Arnaudov

(...) + Note on 18.12.2023

" ...Too much things to add about "creativity", the most would go as an appendix to a work called "Universe and Mind 6". Re that last article, a cite (formal definition without formulas) of "real creativity" ("originality") , according to TOUM: "Universe and Mind 3", 2003:"

https://artificial-mind.blogspot.com/2023/12/on-what-creativity-is-different-tints.html

Too much things to add about "creativity", the most would go as an appendix to a work called "Universe and Mind 6". Re that last article, a cite (formal definition without formulas) of "real creativity" ("originality") , according to TOUM: "Universe and Mind 3", 2003:

*"10. What does it mean to be an original, distinctive* artist?*

*The origination of a piece of art is recording of a piece of information (an entity, a file) to a media which serves as a mediation, intermediate memory between the creator and the perceiver, including the artist himself while he is creating or later in the future, when the creator has forgotten his work and is recalling some of its properties.*

*The space is Memory, that's why "informational carrier/media" can be as diverse as diverse can be the types of data that can be stored in space.*

*Original (creative) is such a file, a piece of information, where the evaluator finds less than expected similarities or matches, in comparison to pieces of knowledge recorded earlier.*

*Originality (creativity) is the capability of making the prediction of the future by the past harder.*

*In this particular case, "past" means a part of the file/piece of knowledge, that was read before*

*another part which is assumed to be future for it,*
*and the future one was supposed to be predicted*
*using the information read from the file until*
*that moment."*

[ **NOTE:** 18.12.2023. This seems to be an expression or similar to "surprisal", Kullback-Leibler Divergence*. That past and piece of knowledge etc. can be applied or extrapolated to whatever scale, range, length of past etc. The originality, distinctiveness, would be compared between at least two "creators" to evaluator, between the creators, or between iterations in a learning mode and learning model. For something to look plausible there should have to be constraints. In art settings or in ML these are datasets or some requirements for the pieces, that could be represented as sets of existing samples (say pictures), datasets, over which the probability distributions of their features are computed, if it's to be measured like that with KL Divergence, at some relevant resolution of causality-control and perception. The first in pictorial settings could be operations, transformations, path of them etc. for generating the "piece of information" (in Bulgarian: къс знание). It is implemented in the image generative models, I guess the paths of transformations of the diffusion models could be mapped to that. If it's a human painter with traditional techniques, it would be other operations in other space with intermediate representations, possible selection of subject, planning of the composition, possible sketch drawings, experimental drawings etc. (see a comment in the linked thread of a user who also argues about the higher flexibility of the human representations), or it could be even with textual commands, applied to other generative model used as an oracle and measured with some other classification model that measures some features and matches etc.

[**Subnote: 19.12.2023:** It could and possibly should also be multi:(scale,range,domain,resolution,...) hierarchical-heterarchical..., multi-agent like simulation: the general "universal simulator of virtual universes"; with some limited linear/textual/table-based/some graph-whatever-preselected-fixed graph be one representation for convenience and for setting up desired constraints, limiting the space of possibilities, resolution, detail etc.... like extended more complex prompts. Also these defintions, "more powerful prompts", are in general "thoughts", "instances of items/pieces of Will acts", intentions, goals, "target paths" (or could be paths - with branches, additional conditions,  etc... ) That could include explicit matches, mismatches, similarities, distances in some metrics - referring to the repertoire and history/structure of the mind etc., they could be searched and generated at the moment from some process etc. and include local or whatever frame of

reference coordinate, pointers, dereferencing etc.  For example a simple concrete image: Draw "my car, but bigger, red color, but do a gradient from the trunk to the front (call function/program... parameters...) - similar to the car from a recent episode from the TV series "Crazy Cars" which I watched; let my car has a slightly bigger wheels. The predefined schemata could be "cached", addressed with labels or whatever: "everything" should be expressible, definable, addressable, generatable etc. More on that in the unpublished 2010s part of TOUM, with the notation of the language of thought **Zrim**. End-of-Note-19.12.2023]

For the "ordinary" viewers their aesthetical measures are also like oracles sometimes, "how the picture makes them feel"* while in other occasions there is more explicit reasoning and obvious features/matches to templates/correct proportions (within given resolution/precision brackets), amount of noise; content within desired targets (depictions of particular subjects, eg. landscapes or portraits etc. (*See "Issues with Like-Dislike Voting in Web 2.0 and Social Media, and Various Defects in Social Ranking and Rating Systems - Confused and Vague Design and Measure - Psychology of the Crowds - Corrupted Society Preferences and Suggestions. In Facebook, Youtube, Twitter, TV Networks..." https://artificial-mind.blogspot.com/2012/07/issues-with-like-dislike-voting-in-web.html  )

Originality under this definition is like a *reverse* of the goal of the free energy minimization: its *maximization* for another evaluator-observer, opponent, causality-control unit, like in a mini-max game or a GAN.

There's a speculation in this spirit, mind and virtual/imaginary universes in the novel/script from TOUM called "The Truth", 2002 ("Истината") in dialogs between the thinking machine and its creator; that thread will continue in another occasion.

* https://en.wikipedia.org/wiki/Kullback%E2%80%93Leibler_divergence

-- END OF NOTE 18.12.2023 -- ]

## * Conversation with Mark Solms and Chris Fields #4

Michael Levin's Academic Content
3,94 хил. абонати | 1787 показвания 1.09.2023 г.
*“Chris Fields, Mark Solms, and Michael Levin discuss what **novel behaviors** are (in the context of problem-solving in novel circumstances), consciousness in explanted brain pieces, and sleep in unconventional agents.”*

52 min Mark Solms: Internal ... Precision modulating ... prediction ... deeper levels of hierarchy
**Tosh:** The mind is still predicting while sleeping. Dreams are not noise (at least mine) and there is reasoning, expectations, feelings going on, or there are memories of them afterwards. There is a narrative, memory and internal logic and references within the dreams, a later part of a dream or a dream of the night sometimes refers/recalls earlier events and characters, sometimes past dreams, and the general memories etc. ... Segmentation, ... problem -- Markov blanket trick ...
…

## * Следващото еволюционно стъпало: заключенията на Джефри Хинтън, 2024 са буквално повторение на изводите от есе от 1999 и ТРИВ, 2001-2002 на Тодор Арнаудов

**Из чат на Т.Арнаудов**
[ 05 октомври 2023 г. 17:28 ] … Изказване на Джефри Хинтън *[което се приема за интересно, ново, ...]* (...) А това, не че е мн оригинално, е като от моите юношески писания, *Къде отиваш свят - 1999, Човекът и мислещата машина, 2001, Следващото еволюционно стъпало, 2002.* (И като от книгата "С дъх на бадеми" на Павел Вежинов, 1966-1967) [и *„Summa Technologiae“ на Станислав Лем, 1963-1964* за която знаех, но не бях чел тогава, прочетох по-късно в края на 2023 г.]
G.Hinton: *„И така, какъв е най-лошият мислим възможен сценарий? … Човечеството да е просто преходна фаза в еволюцията на разума. Не можете по пряк еволюционен път да създадете цифрови разумни същества, би изисквало твърде много енергия и прекалено специфично производство. Първо трябва да се създадат биологични разумни същества, които да се развият така, че да създадат цифров разум, но цифровите разумни същества след това могат да поемат всичко, което хората някога са написали с бавните им средства – нещо което се случва от ЧатГПТ – и след това могат да имат пряк достъп до възприятия от света и да работят много по-бързо [от хората]. Изкуственият разум може да ни запази за известно време, за да*

*поддържаме електроцентралите, но след това – може би няма да има нужда от нас.*

*Добрата новина е, че измислихме как да създаваме безсмъртни същества. Когато апаратна част „умре“, те не умират. Ако пазиш теглата от невронните мрежи на някакъв носител и може да намериш друга резервна част, която може да изпълнява същите инструкции, можеш отново да съживиш този разум.*
*Следователно, вече имаме безсмъртие, но не е за нас.“*

**Това са „копия“ на мислите от есето „Къде отиваш свят“, на студията „Човекът и мислещата машина: анализ на възможността да се създадат Мислещи машини и някои недостатъци на човека и органичната материя пред нея“, „Следващото еволюционно стъпало“, „Следващото еволюционно стъпало 2“; повестта „Истината“, стратегическото есе „Как бих инвестирал един милион с най-голяма полза за развитието на страната“ –** и теглата, които се изтеглят на различните модели невронни мрежи, копират се, дообучават се и пр.

**English original:** „So, what’s the worst-case scenario that’s conceivable? "I think it’s quite conceivable that *humanity is just a passing phase in the evolution of intelligence. You couldn’t directly evolve digital intelligence. It would require too much energy and too much careful fabrication. You need biological intelligence to evolve so that it can create digital intelligence, but digital intelligence can then absorb everything people ever wrote* in a fairly slow way, which is what ChatGPT is doing, but then it can get direct access experience from the world and run much faster. It may keep us around for a while to keep the power stations running, but after that, maybe not.

*"So the good news is we figured out how to build beings that are immortal. When a piece of hardware dies, they don’t die. If you’ve got the weights stored in some medium and you can find another piece of hardware that can run the same instructions, then you can bring it to life again.*
*"So, we’ve got immortality but it’s not for us."*
[ 05 октомври 2023 г. 17:29 ] Но е и оригинално, защото малко хора го приемат и сега. Те [смятат, че] трябва да командват, те са 'най-висшите'.
[ 05 октомври 2023 г. 17:31 ] За космоса - машините може да са по-приспособени за космоса, за хората трябват специални условия е по-сложно. И за електрониката трябва защита от радиация и т.н., но сонди се "разхождат" по Марс, минаха границите на Слънчевата система и т.н.
…

**Предимства на ИИ:**
– Нищожният обем съзнателна информация от десет бита в секунда срещу гигабита при машините – същото като тезата от Първата стратегия за ИИ, „Как бих инвестирал …“, 2003 г.
– Размножаване на ИИ, копия (пак там)

Виж още в: **The Godfather of AI repeats ideas of the Child Prodigy of the Thinking Machines 24-22 years later** https://artificial-mind.blogspot.com/2025/06/the-godfather-of-ai-reproducing-the-child-prodigy-of-the-thinking-machines.html

## * Още бележки на Тодор Арнаудов към Анди Кларк: „Whatever Next: …” , 2013

Виж обзора на съвпадения от научни и философски школи в началото на основния том на *„Пророците…“.*

*„A.Clark: How can a neural imperative to minimize prediction error by enslaving perception, action, and attention…”*
*А.Кларк: “Как може невронният императив да се намали грешката в предиждането, чрез „поробване“ на възприятията, действията и вниманието да усвои очевидния факт, че животните не търсят тъмна стая, в която да останат? Със сигурност, ако стоиш неподвижен в затъмнена стая ще позволи лесно прогнозиране на бъдещето с почти отлична точност в разгръщащите се състояния на нервната ни система? Дали тези принципи не пропуска много от истински важните неща, необходими за успех в приспособяването като скука, любопитство, игра, изследване, търсенето на храна и тръпката от лова?“*
**Тош:** Животните предвиждат също, че *не могат да предвидят* определени неща и пр. с желана точност в тези условия, или че предвиждането им, че в тъмната стая ще е спокойно и т.н. ще бъде грешно. Също така нямат безкрайна енергия, ако не се свързват със средата, от която да я черпят: в тъмната стая пак ще огладнеят и ожаднеят. Освен това управляващите устройства се стремят да **увеличат** предвиждането, обхвата, точността, ефективността, което изисква рано или късно допълнително търсене, а като се затворят спират притока на нови данни. Предвиждането на *досега* известното е в потока на *запазването на съществуващото*, но другият поток е на

*развитието*.

**Относно „спретнатите“ и „мърлявите“ в ИИ (neats and scruffies): „My own sympathies have always lain more on the side of the scruffies….“ –** симпатиите на А.Кларки винаги били към „мърлявите“.

**Тош:** Само изглежда объркано, защото предвиждането не се извършва само от мозъка и частите на мозъка са част от предсказващата система. Организираното и „по-подредено“ (зависи и от наблюдателя-оценител, както винаги) е по-широко, отколкото мозъка. Виж *„Невронните мрежи също са символни“* от Т.А.

**Още бележки към „Whatever Next: …”,** A.Clark, 2013: ок. 10/2023 г.
„*Изненада, неправдоподобност на някои сетивни състояния при дадения модел на света, на ниво деятел: намаляване на изненадата, на грешката в предвижданет*о… Виж ВиР …
*„Настоящото съчетание от управляващите входни данни/сигнали (driving inputs) и присъединена точност (assigned precision) – отразяват увереността на мозъка в /достоверността на/ сетивния сигнал. „теории на високо ниво, първоначално от вид, който е неочакван за агента, с достатъчно висока присвоена точност, също могат да спечелят като обяснят сетивните доказателства с по-голямо тегло“ …*
Нещо неочаквано на пръв поглед от общите принципи, като да видите „тъжен слон“ пред себе си, може да се окаже най-правдоподобното и малко изненадващо според наличните сетивни данни, предходните очаквания (вероятностни очаквания, „priors”) и настоящото разпределение/избор на грешка в предвижданията.

**Тош:** Сравни с „Анализ на смисъла на изречения“ и други от ВиР – за банкера, за детето с „Времевите мухи“ (time flies), програмистите и какво биха могли да правят – възможностите за действия и „логичното“ е по-лесно и пряко обясним и изисква да се включи и непосредственото текущо състояние и среда, а не само общи понятия като „банкерите правят еди-какво си“, „програмистите не могат да пеят“ и пр.

**Тош:** uncertainty, resolution of perception and control; ? what is uncertainty there: resolution step or 0,5 of it, possible/assumed error, precision. If it's 1 cm it's +-0,5 etc. (by default)
**Тош:** Неопределеност, несигурност, РСВ/РСУ; ? ?В е неопределеност

тук: PCB стъпка от 0,5 и възможно/прриета грешка, точност. Ако стъпката е 1 см, то +-0.5 и пр. (по подразбиране; може да е и друга, според вероятностно разпределение, в някои от диапазоните да е различна – например за параметър от 0 до 10 неопределеността да е 0.1, но постепенно да расте – подобно на електронните везни.
*„Може да се каже, че светът не изглежда да е кодиран като преплетено множество от вероятностни разпределения! Напротив, светът изглежда единен и, в „ясен ден“ дори еднозначен!*
**Тош:** А защо да не изглежда така? (И също да изглежд иначе.) Например изображенията, фиксирани в такова представяне, като прости сурови данни с определени стойности, не изглеждат така – многозначни, преплетени. Многозначността произхожда от множеството от различни възможни модели и тълкувания, начини по които може да са били породени тези данни, и какво може да породят те или да последва и т.н., в по-голям обхват. Тези възможности зависят от оценителя – той също трябва да *може* и да *иска* да ги види. В една картина, кадър от видеокамера, един може да види само отделните пиксели така както са, друг може да започне да ги тълкува, да вижда форми, какво може да се случи в следващия момент и т.н. като за последното е необходимо да има по-сложен ум и спомени, и средства за търсене и сравнение на форми, образци, минали последователности от кадри и т.н. Как изглежда светът в тази си страна зависи от зрителя – тълкувател на сетивните данни.

***A.Clark:****…“no inconsistency in thinking that the brain… “ ~ probabilistic encoding … yet a single unified, unambiguous consciously experienced scene”*
*„няма противоречивост в тълкуването на работата на мозъка, че едновременно широко използва вероятностно кодиране, а накрая в съзнанието се построява единна и безспорна сцена.“*

**Тош:** А.Кларк потвърждава, че физическите движения „*не притежават лукса да могат да запазват безкрайно дълго всички възможности“.*
Както беше обяснено във ВиР: тялото и възможностите за движението му и пр. ограничават и правят единна човешката личност, животните и др. одушевени деятели, които са съставени от сложни многопосочни процеси и пр. Виж „Анализ на смисъла...“, 2004 и статията за липсата на обективен Единен аз (освен ограничения по въпросния начин от наблюдател), и Аз-ът като математически интеграл, съвкупност от

множество от „безкрайно малки локални Аз-ове“, 2012 (Akrasia, … )[35]
Единството на „Аз“-а, така както и на понятията, не е *обективно* и *еднозначно*; то изглежда такова в *представите*, моделите в/на ума на съзерцателя, наблюдателя, оценителя във всеки/съответен/избран отделен момент==период, обхват на оценка, и/или при ясно зададени мерки, критерии.

**Бел. 31.1.2024: +**
1. ? Дали определеността (certainty) е на различни нива, гледни точки, деятели …
2**.** Какво е „чувството“ да има неопределеност? (uncertain) „Виждат“ се различни варианти, раздвоение/размножаване… напрежение (лимбична система, стрес – и познавателно?) – в рамките на обхват от оценяване, разглеждане на възможностите.

## * Бележки към Радикална предсказваща обработка: **Radical Predictive Processing**, Andy Clark, 01 September 2015

https://onlinelibrary.wiley.com/doi/abs/10.1111/sjp.12120
от 22.12.2023
* Set of preds - @ diff lev./modalities, РСВУ [,,]
* Precision weighing, volume/gain –-‘ ↓ ↑ ↕ sensory … Едновременно --’ мн.нива и тр. Всчк дсттчн вис. Свпд. → свпд. На р стпн стлбц е ?Тт: *заключване* -*- и мултимодално въобще на много места свпд на р& = -*- …
* 5. Frugal –‘ eng. Immical OAC grip upon the world … conservative –’ modal … Anderson 2014…
**1)** Reconstructive percept. > activate ↓ Build-up (fd fwd) ↑ } р РСВУ и Ц√
**2)** Non-reconstructive – baseball (keep (the ball image) in the center of the view) – seabirds (gannets) predict time to impact
Accuracy-Complexity (as number parameters, numerical)
} Тр. Об. μ и в ТРИВ; + и ∀ кодиране, и квантовата неопределеност и пр. Bayes optimal – Fitzgerald, 2014 – max accuracy, minimize the complexity – срвн. Prediction-compression framework, ТРИВ и пр.
Altering patterns of “effective neural connecting…” – the simplest circuit diagram (Aertsen 1991) underlying current processing; Inner processing economy;
**Тош:** 28.8.2024 При успоредно пробно изпълнение, по-простото, „евтино“

[35] http://artificial-mind.blogspot.com/2012/11/nature-or-nurture-socialization-social.html

ще приключи първо и ще „завземе влатта“; при комбинаторно изброяване и търсене при по-кратък код, най-малка алгоритмична сложност, дължина на съобщението (minimal message length) и пр., ще излезе първо и т.н. Измерено в енергия – като използва по-малко енергия, ще освободи място и за други процеси, за разширение и пр.; а енергоемкото, ако владее властта, ще принуди системата да изключи част от други обработки, да намали тяхната РСВУ и т.н.

- Strategy switching; Precision expectations – “productively lazy” vs model-based
1 – heuristics, easily computed, simple, rule of thumb > Ц√ **(**< Ц√**)**
⥁ causal commerce (#,!) → ⥁ ↕ fd bck Обр.врзк from most simple to > Ц√ структура (↔≈) последователни приближения
М: места &[,,], [,] н само стлбц → писал съм и преди • **–**
In an ecosystem of many strategies: emerge, dissolve, interact } прм опрд. На РСВУ/РСВУтчн; РСтчнст; тчнст М д р от Об РС; опрд ! (РСтчн сщ опр(РС, РСтчн)) до дъно, [,,]
Подвижно самонагласяне ⥁ (Dynamic self-organization)

Създава разнообразие ?В ИдП, ?В МдП. Избирател на {К} на дств – спм. Преобразители „Beam search” (и Sampling) – но трябва и **!** → What’s wrong with NLP („Кое е погрешно в Обработката на естествен език“, 2009) – да не бъде само инструмент, а самостоятелен двигател, който може да се задейства и от там нататък ще работи сам.
– Block-placing task (Ballard 1997) – погледи; запомня само цвят, размер (сакади)
* Оц на нджднн, съществ., значими, важни (salient) → отнасящи се до задачата сведения … да бд достатъчпни в еди-кое си място и/или време.

*** Стратегии с минимална вътрешна памет**, нй-млк вътр.пмт; ># - изискват дств както от организма, така и от средата. ТА: При: >[,,] средата Мд се разгл. Като част от организма)

6**. „Enacting our World”,** Varela, 1991 – actions, perception guided … “expand the temporal window” # is ! (въобр. УУ, просто среда + взаимлик съкращава пт(мс)) … Perception is previous action Marleau-Ponty … structural coupling … (виж Extended Mind)
7. **“Revisiting Representation.”** … vexed issue … Enactive → no intern. # (intuition) to “engage” the world” (като дств, прм, а не action-neutral fashion **-•-** sensory simulations … Context-variable reliability (precision) Тчнст → р. ..

М д е отсечена, плавна, двоична… Lauweryens, 2012: *“Not actual re-presentation or duplicates of objects in the world, but incomplete, abstract code that makes predictions about the world & revises its prediction on the basis of interaction with the world” …*

## **Коментар на Тодор Арнаудов към видеото на Тим Тайлър „Да се слеем [с машините]“** от 10.2023

**Tim Tyler: Let's merge!**

tmtyler 1,14 хил. абонати 288 показвания 4.10.2023 г.

This is a video about the possibility of humans merging with machines.

Transcript: http://matchingpennies.com/lets_merge/

https://www.youtube.com/watch?v=qWpq9OC5Lpg

5.10.2023 г. 19:23:57

**Тодор:** Според мен хората и машините – всъщност технологиите**,** разпознаваемите системи, „единици“ във Вселената (както и онези, които засега не са разпознаваеми [но някога ще бъдат]) – **са си „слети**“ и така както са си били досега. Човеците като индивиди с тела, същества са само определен „изглед“ (виж информатиката, “view”), начин за „изрисуване“ (rendering),начин за извличане на проби (sampling) от действителното същностно представяне на свойствата на процесите, които ги обуславят. Същото важи и за всяка техническа система, електронен компонент, компютър, робот или какъвто и да е обект. Всички те са онова, което са, при определен начин на измерване, изпробване, прочитане на данните; при друг начин на прочитане, измерване, те са физически, причинностно, енергийно и пр. взаимосвързани и си влияят взаимно в мрежа. В една философска школа [марксизма] човешките същества се определят като *съвкупност от „обществени отношения“*[36]. Освен това човешките същества без технологии, които трябва да започнат от нулата без да разполагат с предходни натрупвания на култура и език, които също са вид технология, няма да са много по-способни от маймуни примати в първите поколения, или дори в първите няколко стотин хиляди години. Онова, което позволява на човека да бъде „толкова умен“ са технологиите, средата и предишното записано знание и обществен, научен и технически „софтуер“, който се зарежда в умовете ни, идвайки от цялата Вселена. Така че ние сме си част от системата човек-машина (технологии) така или иначе [и без да се сливаме по „зрелищен“ начин с чипове-импланти и пр.]

[36] [т.е. чрез разпределено представяне, не само в границите на телата си и не само „материално“ – сравни разпределеността с невронните мрежи и пр.]

**Todor:** IMO humans and machines - actually the technologies, the recognizable systems, entities in the Universe (and the ones which are not recognizable for now) are merged anyway. The human individuals as bodies, entities are one "view", "rendering", a way of sampling of the actual intrinsic representation of the underlying properties and processes. The same goes for any piece of hardware, computer, robot, any object. They are what they are under a particular sampling of the data, in another sampling they are physically, causally, energic-based etc. connected and part of the causality, influence, events network and interrelated. In one philosophical school humans are defined as the set of "social relations". Also human individuals without technology and starting from scratch with no previous culture and language, which is also technology, are not very much more capable than apes in the first generations, or in the first hundreds of thousands of years. Technology, the environment and previous recorded knowledge and the social, scientific and technological "software" that gets loaded into our minds, all from the whole universe allow human beings to be so clever etc. So we are part of the human-machine (technology) system anyway.

…

*** #67 Prof. KARL FRISTON 2.0 [Unplugged] MLST,** 17.10.2023 г. **Machine Learning Street Talk,** 87,9 хил. Абонати, 9422 показвания 2.03.2022 г. https://www.youtube.com/watch?v=xKQ-F2-o8uM

"We engage in a bit of epistemic foraging with Prof. Karl Friston! In this show; we discuss the free energy principle in detail. We also discuss emergence, consciousness and cognition. " (…)

[00:09:17] **The Burden of Knowledge Across Disciplines [теглото на „интердисциплинарността"]**

…

[00:44:03] An attracting set at multiple time scales and time infinity

[00:53:56] **What about fuzzy Markov boundaries?**

…

[01:24:28] Can we recreate consciousness in silico? Will it have qualia?

[01:28:29] Subjectivity and building hypotheses

[01:34:17] Subject specific realizations to minimize free energy

[01:37:21] Free will in a deterministic Universe

"54:13 … the markov blankets … in terms of the the curvature of the surprise being zero like at least you know you…"

(Привличащо множество на различни мащаби във времето и в безкрайността на времето. За размитите граници на [одеалата на] Марков. Можем ли да пресъздадем съзнанието в силиций? Ще има ли куалиа, „духовно усещане"? Субективността и построяване на

предположения. Особени за субекта осъществявания на процеса на намаляване на свободната енергия. Свободната енергия в предопределена Вселена.) ...

Одеалата на Марков … в термините на *кривина на изненадата равна на нула* … сетивните множествва [от данни] са онези, които външният свят може да управлява, а ти може да *наблюдаваш*, но *не можеш да управляваш*, и твоите действени състояния (активни, active) са онези, които ти може да управляваш, а не външния свят; а външният може да ги наблюдава, …

1:01:33 … Йерархичността в организацията е в смисъл, че йерархията е *липса на въздействие, което да надвишава, да речем  едно ниво в йерархията* [Тош, Зрим: отделяне, разделяне, прекъсване, преходи, граници на преход, дискретизация]

**Тош:** Сравни с ТРИВ; понятието за **sparsity**, „Разреденост“  % Т: Разделяне (опростяване, за да е проследимо и пр.) и компресия, сбиване, сгъстяване.

**Към кривината и ПСЕ/ИЧД (FEP/AIF):**

- Вариационно смятане, намиране на най-краткия линеен интеграл – интересна историческа статия. Теориите за най-добро управление, оптимален контрол, оптимизиране, се смятат за произхождащи от работата на Бернули в края на 17-ти век: 1697 г. с решаването на задачата за най-кратък път по крива линия (брахистохронен проблем, brachrystochrone problem)
  https://lab.vanderbilt.edu/taha/wp-content/uploads/sites/154/2017/10/300Years_Of_Optimal_Control.pdf
  Hector Sussmann & Jan Willems, 1997

…

*** Mahault Albarracin ~ Active Inference Insights 002 |** Active Inference Institute | 2,43 хил. абонати .. *Active Inference Insights welcomes Mahault Albarracin onto the show for its second episode. Mahault is a PhD candidate in cognitive computing at the University of Quebec at Montreal and the Director of Product at VERSES. Together, she and Darius traverse the complex plains of epistemic communities, social scripts and artificial intelligence, all from the lens of active inference. Active Inference Insititute* https://www.activeinference.org https://www.youtube.com/watch?v=njbpi3YTMPY (…)

*FEP ... the same mathematical technology ... rejects the discontinuity between matter and the subjectivity* **…** *"From inert rock to the brain... "* ... (Another expression of that idea: "Life Mind continuity" etc.

They rediscover Todor's **Theory of Universe and Mind.** Also the dialectical materialism and Schopenhauer's "Will as world and idea". The former's Engels' "Dialectics of Nature"; each higher level contains the lower level (по-високите нива съдържат по-ниските в „снет вид"; подобни, макар нетъждествени; виж „Теория на отражението", Т.Павлов; 1936,1945 (2-ро издание;1949 (руски)).
**Todor: sp++(mem++) because they always pass through a mind that evaluates it. The stone is a stone in an evluator which samples and interprets the data that way. See "Universe and Mind 6"** and other discussions by Todor in the sections about Consciousness and Panpsychism in appendix *"Listove"* from ***The Prophets of the Thinking Machines***.

*Possibility sets - attractor sets … Virtual and actual relate to the dynamics of the FEP/AIF ... temporal depth of these possibilities that are directly afforded to you or potential given new agency ..... farther away attractor set ... more or less likely ... it needs energy ... the enrgy is the improbability of something to happen ... when you deal with these possibilitiy spaces ... co-constitution ... what could happen ... matter is a process rather than a static ... process of self-evidencing .. Mahault Albarracin ... equally probably - no inf. .. what intelligence is a phenomenon ... interaction inf. passed ... integration of this information ... could be ... boundary ... actiing ... you accrue agency given how much inf. you can pull together //one shared intel. is not just about shared ...? bits? ... perspectives ... embedding into the world ... the possibility .. inf. ... "valence, emotion...possibility for...theory...shared LLMs...better method...mutual information exchange... one true value alignment...how these pieces coordinate... reaches metastable state..." .. tries to understand the causes .. why something that predicts the cause of the next word .. the LLM are not really Bayes. themselves? .. self-attention to .. don't now what "it" is .. what's the objective .. //can't understand ... outcomes of its own actions ...* ***embodiment:*** *relation to self within a context ... virtual world ... we actually are ... relation to self relative to context .. if you are the context you are the environment … separate* **…** *how the different parts are different … how the agent relates to itself, relative to its boundaries ...*

**Actinf podcast | Ecological psychology ... 11-1-2024+ ...**

*** Критика на FEP/AIF: Трикът с одеалата на Марков: за обхвата на принципа на свободната енергия и заключението чрез действие*[37]**

*** The Markov blanket trick: On the scope of the free energy principle and active inference,** Vicente Raja a, Dinesh Valluri b, Edward Baggs a, Anthony Chemero c d, Michael L. Anderson | 12/2021
https://www.sciencedirect.com/science/article/abs/pii/S1571064521000634
https://philsci-archive.pitt.edu/18843/

The paper challenges FEP/AIF and suggests that it cannot properly segment the environment from the systems/thing; "**Thingyness**" (in other discussions)... Описанията на ПСЕ/ИЧД не можел да разделят средата от "нещото".

**Todor:** Segmentation is dependent on the evaluator and the resolution of causality-control and perception and it happens in a mind. The stone-story, also about the flame - the blanket of the flame. Conditional independence, ... Can't divide … → see *The Matrix in the Matrix is a Matrix in the Matrix,* T.Arnaudov, 2003... TOUM - it doesn't specify or require a specific method for prediction, the directions/goals are similar, ...

How the representations etc. are imported *without* sensory states etc.. → every particle is sensory and causal at the same time. It exists implicitly and is read by the processes. The strict separation in FEP/AIF is for modeling purposes (that is stated by other researchers in that school in other discussions as well, for example Jeff Beck etc., "as if" in MLST interviews).
p.29 *"and features that cannot be captured by Markov blankets. So, why … literature, Markov blankets are presented as a non-controversial, assumption question 'What does it mean for a thing to be different from its environment?'."*
→ **Todor:** for an evaluator observer; the statistics is computed by *this* agent with *his* mind, which already exists; the elements of the stone, if they are smaller than the whole etc. ... don't know that they are part of the whole (see UnM6, T.Arnaudov 2025) ...

I also saw problems with the definition of internal and external, ... *located "inside",* but they are dependent on the environment and everything is an environment for another observer or the lower level representation. (Also the thought: *there's no principled difference between software and hardware, lower level/"physics*" from the POV of the higher level; also "data" is part of the "algorithm", see TOUM etc.)
# p. 31 *Wolfram 2002,2020 ... graphs, hypergraphs, ... automata*.

[37] Заключение чрез действие - синоним на „Извод чрез действие"

**Todor:** TOUM has both views, at the lowest level it's/could be represented as "processing", the higher level systems have deeper structure; the lower level *seems* as “just executing instructions” and computers aim at predicting exactly the next instruction, causally, "mechanically", exactly: *not probabilistically*: however at the even lower level, for a “deeper” and more detailed observer, they also have their own inner structure, decision etc. and it happens that they *don’t just “execute instructions”* etc.

*p.31 Maximum Entropy, 1957 and FEP and Wolfram - no empirical (?)*

**Todor:** so no different to TOUM if it's attacked on this ground… :) (Bert de Vries @ MLST also admits that “too little work” is completed on the implementation of AIF yet)

p.32 *"If the Markov blanket formalism is that the statistical independence between internal and external states must be mediated by a different set of states." … "to describe and find such a boundary (see also Bruineberg et al. 2020). Why then should we use that tool and not another (e.g., **causal blankets**; Rosas et al. 2020)[38]? We think the reason is that Markov"*

** causal blankets; Rosas et al. 2020*

p.33 *"cognitive activities with a form of inference (e.g., Dayan et al. 1995; Friston 2002, 2003, 2005; Stone 2012) “Perception, the story goes, is about inferring the world from sensations which only provide partial information about the world itself. In this sense, perception is about discovering which hidden variables (v; a.k.a., the world) have caused the observable variables ( u; a.k.a., sensations) brains have access to. In the Bayesian jargon, what brains are trying to do is to infer the true posterior density"*

**Todor:** The sensations are *also partial* and there are other properties, which can't be measured and are implied by the very existence of them and the environment. The very *existence* of the Universe and why it and anything exists at all, and why it exists exactly as it is, is a "hidden variable" itself. See also Arthur Schopenhauer’s early 19-th century’s works: “On the fourfold root of the principle of sufficient reason” and the “World as Will and Idea”, where he also defines the job of the Understanding (разсъдък) as inferring the causes from the effects.

*p.35 "Markov blankets provide that exact partition for any system: hidden variables are external states (η), observable variables are blanket states (b), and system states are internal state (μ). … The details of the partition are underdetermined by FEP because the only important consideration is the partition itself that grants the structure for Bayesian inference. … FEP accommodates **computational realism** (Wiese & Friston*

[38] Causal blankets: Theory and algorithmic framework, Fernando E. Rosas et al. https://iwaiworkshop.github.io/papers/2020/IWAI_2020_paper_22.pdf

*2021). Converting any system into a form of **computational system**—e.g., a kind of **variational autoencoder**—is both the main outcome and the main driving force of the principle and, more concretely, of the Markov blanket formalism.""*

**Todor:** Not just partition, but segmentation in structures and their internals, with proper relations, i.e. not just static partition (or segmentation) like of a drawing of a geographic map. There are specific kinds of information in the lowest level virtual universe from the POV of the current evaluator-observer. See also TOUM. “The Unvierse Computer” etc. *"the boundary between things and their environment..."* depends on the evaluator and all is in the environment from some POV, it is separated according to some decided partition and there are overlaps: this is part of the equation that has to be defined.

”*A better way to understand FEP is as a **modeling framework** (see Andrews 2020; van Es 2020) that permits us to understand some properties of some systems in terms of Bayesian inference*

**Todor:** Yes. TOUM also states that Mind/Universe *can be represented as* causality-control units at all scales, resolutions etc., but it doesn't know what it really is as there's a limit of what can be really known etc. The goal is to create thinking machines with similar capacities, cognition, behavior etc. Compare “the thing in itself”: Kant, Schopenhauer.

**Modeling**: *“In this sense, FEP is not different from the **computational metaphor** (Milkowski 2013) or from the **dynamical hypothesis** (van Gelder 1998). Which one of these frameworks is best to model biological and cognitive sciences is an empirical question we will not address here.”*

*** Todor Arnaudov's comments on Causal blankets: Theory and algorithmic framework,** Fernando E. Rosas, Pedro A.M. Mediano, Martin Biehl, Shamil Chandaria and Daniel Polani
https://iwaiworkshop.github.io/papers/2020/IWAI_2020_paper_22.pdf
**Todor:** The past trajectories which lead to predictions, predictive coding – this is part of the stream of TOUM. There is a criticism of aspects of the Markov Blanket and AIF "sensory states/active states" dichotomy. An earlier version of the cited "*Computational mechanics: Pattern and prediction, structure and simplicity", C.Shalizi, J.Crutchfield, 1999* is novel for me[39] interesting and early elaborate work in the predictive processing school of thought. The Causal blanket paper presents mathematical symbols but as far as my current view for default definitions of causality such as precedence, necessity; compressed representation of the correlation.

A practical/actual/complete model should include an agent and his will as well. I didn't noticed a mention of the resolution and discretisation, coarse graining of the representation. When interpreting causality, assumptions about the causal agents (causality-control units) which are involved/seleced must be made/are made. For example: *"I opened the door, because I wanted to go out".* Then render, expand, fill-in details in the context, environment: a specific location, home, rooms, coordinates of the agents, who/what is the agent (a human is implied, but it could be a robot or an animal in a story, e.g. a cat or a dog). The causal force is this agent, the reason the door to open is both her/its desire (the agent wanted the door to open and then its state changed to the target one), but also the prerequisites in the complete definition of the environment: the affordances, what is possible to happen. That reasoning is maybe related to the *"frame problem"* of McCarty etc. The evaluator-observer-simulator sets the frame.

Another view is that the agent opened the door, *because it was closed* (so the *closed state* of the door was the reason for the agent to *need to open it)*.* Then the door was opened, because it *could move* or it *was unlocked* or *there was a door at all* there (if there wasn't a door, but just a corridor, or an open space, or a wall; or if the door was stacked in a storage room and not installed normally (as one tacitly assumes by default) etc. it wouldn't be able to open. Etc. All these and more wll appear if the area of search is extended, in all related or connected spaces and possible causal explanations and they all would be predictive, given their definition and a measure for simplicity for the new area/domain (they serve as discriminators, selectors). At a given moment, some chain, view, slice, track, network etc. is chosen; if causality is centered around the causality-control unit (agent), his intentions are usually chosen as

[39] ~20/4/2024

the driving one.

The factorisation is objective and singular if there's a "god" who "locks" the interdependencies, or with a fixed set of resolutions and methods of factorisation. There are many possible causal explanations depending on the complete framework, the definition of the specific virtual universe which is predicted, the selected resolutions, selected segmentation etc. as explained in *the "Analysis of the meaning of a sentence…",*T.A., 2004 and in the comments in this book in *"Universe and Mind 6"* about M.Levin's *"Cognitive Light Cone", "care"* as a measure of intelligence, where again in order for this criterium to be "objective", a "God" observer is required and the causation, influence, "care" etc. have to be "objects", "first-class citizens" and once "allocated", they should occupy a "slot"/slots and do not allow other causal chains, forces, correlations to be considered for the same/connected causal links, which requires that to be defined in all scales, domains, "views" (in computer science sense). (The Computational Mechanics paper suggests partitioning) This works for a "slice", a POV of some observers, but not for many observers at different scales and resolutions. The ultimate causality description and factorisation could be an exact record of the machine language of the Universe Computer. In the view of S.Wolfram's Multiway system, these could be paths/hypergraphs (including the connections) between states, but for observing the universe at different resolutions than the highest one there are different interpretations, degrees etc. Causation is discretized, it loses details. [To do: expand in simulations, new works like the 2004 one, but operationalized etc. A workable way to explain causality is with specific hierarchical simulators/emulators of virtual universes, consisting of causality-control units which are predicting and maximizing prediction horizon/range and precision (compress) etc.]

**Summary concepts in the paper:** (selection:T.A.)
**"***1.MB [Markov Blanket] formalism* ***forbids interdependencies induced by past events that are kept in memory,*** *but may not directly influence the present state of the blankets."****** …
*2. "differences that make a difference";*
*3. Dynamical Bayesian sufficient statistic (D-BaSS)Precedence | Sufficiency | minimal D-BaSS |*
*4. latent influences group together all the past trajectories that lead to the same predictions, which is a key principle of* ***computational mechanics*** *…*
*5. D-BaSS distinguishes only "****differences that make a difference****" for the future*
*6. Joint precedence | Reciprocal sufficiency | bipartition |non-ergodic, non-*

*stationary systems* |
7. PALO – Perception-action-loop
**See:** [*24*] *Shalizi, C.R., Crutchfield, J.P.:* ***Computational mechanics: Pattern and prediction,structure and simplicity****. Journal of Statistical Physics 104(3-4), 817–879 (2001)* ***T: Published earlier, 7.1999, last edition 6.2000****:* *https://arxiv.org/pdf/cond-mat/9907176.pdf*

*"We now show that: causal states are maximally accurate predictors of minimal statistical complexity; they are unique in sharing both properties; and their state-to-state transitions are minimally stochastic"*
*25.* ***Shalizi, C.: Causal architecture. Complexity, and Self-Organization in Time Series and Cellular Automata PhD thesis (Univ Wisconsin–Madison, Madison, WI) (2001)***

**# Appendices to Universe and Mind 6**
1: Specificity, Generality, Causality-Control Units, Resolution of Causality-Control and Resolution of Perception, Matches etc.
https://github.com/Twenkid/Theory-of-Universe-and-Mind/blob/main/Universe_and_Mind_6/Appendix_27-11-2023.md
https://discord.com/channels/1095629355432034344/1095695948807667782/1178263322206949396
2. A discussion in "John's AI Group" Discord about originality, creativity, generalization. Tosh (Todor) discusses comments of Atrony, Richard, Aynur in the ARC-challenge-ramble room
https://discord.com/channels/1095629355432034344/1116567232286314517/1202325917746352189 (Виж с UnM 6)
https://www.linkedin.com/posts/dr-jeffrey-funk-a979435_algorithms-ai-technology-activity-7133763690717224960-fMbK?utm_source=share&utm_medium=member_desktop

**#App. 2... UnM 6**

## * Todor: AIF is vis vitalis (Active Inference е въплъщението на понятието за „жизнена сила“ и Воля на Шопенхауер) - 28.11.2023

**Todor’s insight:** TOUM by effects .. finger moving, few bits, ... a concept in TOUM ...

**Chain:** unwanted change, unintended, non-intentional, non-target... ... causation //can be two kinds, or many: not-intended, but "not harmful'... and нжр - вредна? пречеща ... – see the example below
~ #: Arthur Schopenhauer [AS]... Objectivation of the Will [for Life] ...
**Todor:** In order the Will to be materialized and to cause what it wants, it is "forced" to cause other non-intended events, to influence other changes: by effects. In order the living organisms to exist, as well as any other structure, any "thing", "objectness" (see also Chris Fields and FEP/AIF...), all these molecules, cells, organs, muscles, nerves have to exist in order to support the tiny intentional will, "consciousness" etc. The structures emerge as a clash between the intentions of different CCU which compete. The structures, their borders and forming limits and shades are where the confrontation of different causation forces, different Wills settle down or counter-force each other without being able to break each other at a given scale, resolution, range, POV. Nowadays the "vis vitalis", the theoretical force that creates life and exists only in the living organisms may be considered an old naive confusion and a tale, but actually there is such a force, even if it’s virtual, informational, as a set of the prerequisite forces, systems, “machine code”, and it is not only in the living organisms. The AIF sounds as possible "vis vitalis", at least it maps AS's interpretation, so long as I understand and remember his works.

*** Karl Friston ~ Active Inference Insights 001 ~ Free Energy, Time, Consciousness |** Active Inference Institute | 3,22 хил. абонати (3.4.2024) 4778 показвания **Начало на премиерата: 22.11.2023 г**
https://www.youtube.com/watch?v=N5H5I6cvcrQ
Notes … AIF - application of FEP
FEP - description of things that exist
Stochastic diff.eq. **|** what do you mean by a thing
Start with the same assumption

*The state of the universe that changes, it has some dynamics*
*The way that state change as a function of themselves*
*~4:30: Distinguishes it from Quantum thermodynamics... etc. .. can be derived from this basic description of the world*
***5 min:** Careful distinction of the state of something and everything else Inducing a partition of all possible states: 1. Internal, 2. External, 3. states that separate internal and external, the Markov state, boundary ... Notion of a boundary, separates the outside from the inside individuated, posses characteristic state ... 6:30 Physics of the open systems ... Active and Sensory states … the input of the sensory states ... active states -- external states do not influence them* (they are defined like that) Attractor set of states ... 8 min … Autonomous states ... sensory ( 11 min Acting in order to gain evidence of its own existence ... entails generative model ... Simulation ... gradient flow ...*
**Todor:** *Тл (Зрим) … 28:hh - you can't know beyond the Markov blanket/boundary, if you knew it ceases to exist as before … TOUM: limit of cognition, ...*
***30:11:** I can image a whole... philosophy ... You couldn't see your own visual processing ... You can't hear your own cohlear dynamics ... You can't have diret access to the outside ... Gen. model as a Bayesian mechanics interpretation ... It's not the model, it's the dynamics*
***34 min** Host: ... Why don't perceive/access directly the states/correlations. .. but through a hidden model ...*
**Todor*:*** *the evaluator *assumes, chooses, prefers to believe*; partition the systems so that that the external states are separate and do not influence the internal ones

**Todor:** branching etc., modularity? Compare TOUM. 2001-2004

*** ActInf GuestStream #018.1 ~ Michael Kirchhoff & Julian Kiverstein**
**The Literalist Fallacy & the Free Energy Principle: Model-building, Scientific Realism and Instrumentalism**
Active Inference Institute 2,44 хил. абонати 201 показвания Предавано поточно на живо на 15.03.2022 г. | 2.12.2023
Michael Kirchhoff, Julian Kiverstein, Ian Robertson
http://philsci-archive.pitt.edu/20077/ Active Inference Institute information:
Website: https://activeinference.org/
- Realism, Instrumentalism; The map problem 10 min
- Other researchers: .. AIF: **Alexander Ororbia ...**

*** The Active Inference Institute and Active Inference Ecosystem**
https://zenodo.org/records/8266281
Разрастващата се общност на школата на К.Фристън ПСЕ/ИЧД.

*** Prof LARISA SOLDATOVA - Automating Science**
Machine Learning Street Talk | 89,4 хил. абонати
3373 показвания 23.10.2023 г.
https://www.youtube.com/watch?v=UoMY5oj9XqM
26:50 Tim Scarfe.: *If I wanted to define a chair ... Is it collapsable or it is just very complicated ...*
**Todor:** It is defined here: *Chairs, Buildings, Caricatures, … AGI Digest*
https://research.twenkid.com/agi/2012/AGI_2012_Chairs_Caricatures_and_Object_Recognition_as_3D_Reconstruction.pdf
Also in T.A.'s notebook: "12. XII. 1999 ИЗИНТ"
"ЛИСТ „Изинт“ 12.XII.1999, Дневници и записки на Тош. Възраст 15 .г 5 м."
I am an author of explanation (generalisation) what a chair[40] is, in this work:
I'd object, the chair part was cheesy or at least superficial. It is just a podcast and she couldn't go in formal definitions, but the *"function is to sit"** is not a clear explanation, as she mention she could sit on a table, and it is not called "a chair" (or a stool) is due to a convention. Actually this was explained in much more details here:
https://research.twenkid.com/agi/2012/AGI_2012_Chairs_Caricatures_and_Ob

[40] През март 2024 г. открих по-кратък опит също в *„Как да помогнем при ученето“,* Хари Шьонфелдер, 1990 (Berlin, 1986, Harry Schumfelder), с. 61 – 62 и добри разсъждеия за процесите на обобщение. В Източна Европа, диалектическия материализъм, „духът“, стилът на познавателните процеси е с по-добро умение за обобщаване, отколкото в англосаксонската частнонаучна култура, вероятно особеностите на съдебната им система, основана на единични случаи (прецеденти) също е свързана с разлика в начина на мислене. „Индивидуализъм-общинност/комунитарност“ („колективизъм“) биха могли също да се свържат, но не винаги класификацията на индивид или общество е убедителна и еднозначна.

ject_Recognition_as_3D_Reconstruction.pdf

And in Russian and Slavic languages the words for chair and stool and table are from the same root or the same. "Stol" in Bulgarian means both chair and stool and "stol" mean...

* What is a function? An *agent* wants to sit, to view the object with that purpos, which is to compare the measures of its body parts (see the cited work).

...

*** CHOMSKY - WHY WE DON'T KNOW THE WORLD**

**Ян Льокан, Y.LeCun ... от 10.7.2022 ... MLST Chomsky ...**

The Ghost in the Machine and the Limits of Human Understanding. Professor Noam Chomsky is the most significant thinker of our generation.

„LeCun's recent position paper on AI, JEPA, Schmidhuber, EBMs

**Schmidhubber on Lecun's rediscovering his ideas**:

https://youtu.be/axuGfh4UR9Q?t=1865

"Path towards autonomous AI"

15:58 -- (и мн. др.) - същите като мои неща, предвиждане в различни мащаби и пр. 16:48 … Tim Scarfe се опитва да го представи като несъществено, защото на абстрактно ниво много неща си приличали и т.н.

.

*** #4.11.2023, FEP, AIF and mapping to Todor's TOUM and Zrim...**

FEP_AIF_4-11-2023-Attention-precision-weighing-new58.txt

1) Attention - Precision-weighting of prediction signal
== РСУ – разделителна способност на управлението, resolution of causality / control
2) Dynamical attractors ... guiding action-perception ... == Избирател на {К}, Context Selector
3) local ontologies – Контекстни речници {К}рчнк
4) regimes of attention == #[.]
5) abilities - acquired patterns of attention and gating == Записи, съдържание, код в {К}! за конкр. {К}
6) Mutual ontologies, shared sets of expectations == споделени предсказващи модели --' свпд $\{К\}_1$, $\{К\}_2$
7) joint attention – УУ, които имат съвпадащи {К}[.] и/или *#
8) niche of affordances == {М: ?В МдсП}
++ 31.1.2024: The paper that I've just revisited tonight - Multiscale ... Cultural affordances ... Ramstead

https://www.sciencedirect.com/book/9780444529657/stochastic-processes-in-physics-and-chemistry

*** Stochastic Processes in Physics and Chemistry -** A volume in North-Holland Personal Library - Book • Third Edition • 2007

*** UNSTABLE SYSTEMS** N.G. VAN KAMPEN, in Stochastic Processes in Physics and Chemistry (Third Edition), 2007

***Remark:*** *Instability and bistability …The effect of the fluctuations is merely to make the system* ***decide*** *to go to one or the other macroscopically stable point. … **) Fluctuations merely make the choice between different, equally possible macrostates, and, in these examples, determine the location of the vortices or of the cells in space. (In practice they are often overruled by extraneous influences, such as the presence of a* ***boundary****.) Statements that fluctuations shift or destroy the* ***bistability*** *are obscure, because on the* ***mesoscopic level*** *there is* ***no sharp separation between stable and unstable systems.*** *Some authors call a* ***mesostate*** *(i.e., a probability distribution P) "bistable" when P has two maxima, however flat. This does not correspond to any observable fact, however, unless the maxima are well-separated peaks, which can each be related to separate macrostates, as in (1.1).*

**Тодор:** Важна идея: нестабилност и бистабилност, плаване. Създава движеща сила за вземане на решение в системи изграждащи се отдолу нагоре от просто първично състояние, и за избягване от „заклещване" в едно положение/максимум/минимум. Сравни с дифузните модели в машинното обучение.

# * Из бележките на докторант по ИИ от Сингапурския национален университет за паметта и коментари на Тош

*** "Memory soup" –** John Chong Min Tan ... 7.11.2023, Discord, John's AI Group … https://www.youtube.com/watch?v=0Y5vIvC8BTY

*** **A roadmap for AI: Past, Present and Future (Part 2): Fixed vs Flexible, Memory Soup vs Hierarchy** - YouTube - John Tan Chong Min - discord as well - 2,41K subscribers [8.11.2023] 292 показвания … Начало на премиерата: 7.11.2023 г. **John: …** *the future of AI systems! (…) (Past and Present AI systems): Expert Knowledge Systems (learning rules from experts), Supervised Learning (human-labelled data), Unsupervised Learning/Self-Supervised Learning (non human-labelled data), Foundational Models to learn from data and set a baseline for performance.* ***Moving to the Future:*** *Memory: We will discuss how we can use memory as the next wave to improve AI systems (this also includes the Large Language Model grounding with Knowledge Graphs).*

*Hierarchy or memory* ***soup****: We will talk about hierarchy and whether it is the essential ingredient for better representation and planning, or could it be just a "memory soup" mix of abstractions and using pattern matching to find the right one. This bears some resemblance to the* ***"Thousand Brains Theory of Intelligence****" by Numenta.* ***Multi-agent:*** *The last part covers agents, multiple systems in one agent, multiple agents in an ecosystem, and finally, multiple ecosystems.*

***The future is unknown****, but what is sure is that technological improvement will lead us to something which is far more advanced than the current state of the art.*

*My research work seeks to help to attain fast and adaptable agents, and I believe this will be key for the future of AI systems.*

**John:** *** There may not be explicit hierarchy. Different modalities get mapped to different abstraction spaces, which can be combined into the same embedding vector. Within each modality, there may be different ways of mapping to various abstraction spaces as well, which creates very different embeddings. Eventually, all these mappings to latent spaces are all in the same soup, and available to be referenced on by any actor in the decision making pipeline. … Referencing the memory may be the "incomplete hash map" I talked about earlier. We only compare a subset of the vector to suit our needs for the context - audio context use audio part as key/value matching, visual context use visual part etc. we can also do a random forest style search of multi-modal stimuli. The issue is if there are multiple abstractions for one modality, then we may have problems getting the cosine similarity to work out. But maybe it doesn't matter because we can have multiple representations in the same space and we will tend to just extract out the similar ones anyway due to cosine similarity grouping them together. Refer to this video for my earlier ideas on Incomplete Hash Map and Multiple Referencing:* https://www.youtube.com/watch?v=q9uMEAcB3lM (*I have changed some of my*

*views on hierarchy since then)*

**Todor:** 19.4.2024**:** Right. Associative memory and chain of thoughts, partially overlapping items invoking each other in various modalities and intermodally. See again A.Schopenhauer's "World as Will and Idea". However note that on deep inspection the *explicitness* of the hierarchy is questionnable anyway: ithe recognition or acceptance of the content, state, structure as a hierarchy depends on the scope of the evaluator. The evaluator must see a big enough range, states, specific properties and scales in order to notice or perceive the content as a hierarchy. The memory where the ANN are stored is *both* flat or hierarchical, depending on the way it is sampled and traversed. There are multiple levels of addressing which are like subspaces; that can be thought of like "modalities" or selectors for other subspaces, jumps etc.
[+ 5.6.2024: cosine etc. similarity is not the only way to measure similarity, also all these linear algebra vector spaces: see below]

**Todor's comment of the video:** 8.11.2023 г. 16:04:02
**Todor:** Keep up the good work! Re the "hierarchical prediction *is the future*" - it was defined more than 20 years ago, e.g. in the Theory of Universe and Mind by Todor Arnaudov (me, first published 2001-2004)[*and earlier by others]. The multi-agent-like systems was part of that, in the whole theory and displayed with the process of thinking in particular in the multi-faceted work starting as "Analysis of the meaning of a sentence, based on the knowledge base of an operational thinking machine. Reflections about the meaning and artificial intelligence(...)", 3.2004. The mind, universe and these hierarchies can be represented as causality-control units (TOUM terminology) at different levels of resolution of causality and control, different range for prediction, different domains etc. The slides for a public lecture about the TOUM from 2009 illustrate it (such as the hierarchy in the universe, from quarks to individuals, to different levels in the army hierarchies), as well as the slides for the lecture about TOUM in the world's first university course in AGI, offered by the same author in 2010 and 2011. One can easily find the referred works, ask me if you want links. FEP/AIF (you don't mention them) are repeating many of the TOUM claims and reasoning.

*"**Memory soup**"* is another general principle and general operation of matching, which is part of the other general principle of prediction (and another is compression, which is required in order to be "pre-"... to get ahead of the future with less resources, to get shorter, to reduce). There is a blame to LLMs/transformers that they are "just" giant hash tables, but first they are hierarchical and not "just", and they reconstruct (and have the tokens, as you point out, that's part of the analysis-synthesis general framework), and second humans also work like that for many tasks. (As Connor Leahy points out too regarding GPT-3 in a video, to MLST or so, that "humans try to work like GPT3" or something). Programmers used to use search in Stack overflow and on the Internet, in their own source code, in help files with API definitions or in books with indices - all researchers do. That's why proper citations in

the research literature are, for easy tracking. Long before LLMs matching and "simple hash tables" (just tables, dictionaries) were an instrument for boosting many "intellectual" tasks - computers and mass computations early use was for calculating ballistic tables for guns, and the associative cache memory in the CPUs allows for huge performance speed up (prediction); the essential is search, querying, matching and the general memory, "associative memory", and it's also compressing/reducing/optimizing the search process (*calculation* could be seen as a *search* as well - a search for a solution), thus allowing the prediction. The thinking process jumps from memory islands of what can be kept within the working memory etc. and this has been discussed by one the brightest minds – Schopenhauer – even in mid 19-th century.

*** John Tan Chong Min A New Framework of Memory for Learning (Part 1),Youtube**
**Todor:** Hi, John, it's Todor, the author of world's first university course in AGI (Plovdiv 2010,2011). I commented on that video, I guess,
*"we may have problems getting the cosine similarity to work out"*
[**T.A.:** Note, 3.4.2024: Stephan V., an independent researcher, talks about something like a “memory soup” as a **neural maze** in “spiking neurons” systems. I agree with related implementations where I refer to it *as incremental self-description*, chain of partial matches and mapping, верига от частични съвпадения, самоописание, описание на всичко; изображение, съответствие (mapping); съвпадение (match, matching)].

Should cosine similarity or whatever specific measure be used, strictly [speaking]? Also the vectors structure, should these vectors be just vectors, numerical. I think this is an inertia from DL style, all those matrices, "being efficient" (for the GPUs). Comparisons and distances could be *any*, and "the soup" could have a structure and be always explainable with/by some of the lowest levels and up [the higher] representations [could] be more pictorial, diagrammatic etc.

If the representation is smart it doesn't have to compute a gazillion of matrix transformations on 99999 layers in order to discover something which is 1 byte and can be detected directly, as long as the structure is complex enough. It's a trade-off, the usual/modern (or forever) "AI" guys tend to overgeneralize and convert all to numbers.
…

*** Moving Beyond Probabilities: Memory as World Modelling**
John Tan Chong Min | 2,44 хил. абонати
1041 показвания Начало на премиерата: 19.09.2023 г.
**How do humans think and reason, is it by probability or by matching to memory? How can we use memory to do world modelling?**

***John:*** *I posit that memory is the fast mechanism of learning, as backpropagation via updating weights (aka Long Term Potentiation / Depression) may be a slower process than encoding and retrieving from memory. Future systems which incorporate memory, reflections of basic chunks of memory to form meta-memory, adaptation of existing memory to new situations, will likely be faster and adaptable than most of deep learning right now. … Action prediction instead of states*

**Todor,** 12-11-2023: Hi, John. I think that mind aims at/converges to causal modeling, not probabilistic. The ultimate operation is exact models, until there are such there are all kinds of approximations, also there are shortcuts due to discoveries of cheaper way to model, but it's always at a lower resolution.

Good remarks. I'd add that: 1) Not only the memory could be used as a world model with shortcut maps at different resolutions, and exploiting the fact that there's prediction horizon, partial observability etc., thus every detailed simulation - which also could be causal, "mechanistic", not probabilistic - has a limit and sometimes simple mapping methods may provide similar results due to that uncertainty (that reminds me of the line of "case based reasoning", or as Hofstadter calls it "Analogy" making, which is mapping and matching).

In a more general sense, the memory is always used, IMO it is such by default in any model, NN as well, or RL - the "model free" RL also could be said to have a world model, even if it is implicit, it is "as if it had a model". In a simple/low resolution mapping prediction model there also could be a lack of deep or detailed explainability. In many occasions humans cannot explain their actions and do not realize the cues that drive them, even if they are simple.

The "real" world itself, the lowest level data, the sensory data, is used as memory and a "world model", which is at hand anyway and is "already computed" (it "only" has to be converted with the sensory transformations, but mind doesn't have to keep a precise world model at a high resolution, it lacks resources). According to my own Theory of Universe and Mind and I think the line of research of the Extended mind and "Radical embodied cognition" or something, sensory matrices are used as "RAM" or as additional tools. Vision is a form of "RAM", like the early computers CRT RAM. (E.g. one demonstration of the above is research on how dogs catch a thrown object, also how baseball players who have to catch the ball track it etc.: they try to keep the ball/target in the middle/constant position in their eyesight, so they move towards the target and adjust the control signals. I think the same approach was used in missiles, they adjusted the difference to the target.

In cognition, at higher level these differences could be abstract and the predicted steps are expressed in linguistic and other concepts.

… stable/predictable world model and it may be that initially it is more reliable and stable than the internal brain states and/or they provide some consistent data/material to scan with which the rest could synchronize. The sensory matrices themselves serve as initial coordinate spaces to be addressed and "RAM" for the rest of the developing mind, upon which the rest could be consistently grounded, a chain of reference frames, transitions, represented anyhow; either the motor commands and the deeper abstract representations, all of which eventually end up as sensori-motor interaction and commands if they have to act on the world.
…in series of reference frames transformations. It's not strange that "just linear algebra" works., e.g. a baby watching a static scene when the care giver is gone, or even if she's there, say a stationary baby, most of the sensory matrices … an early access to the external world and their reliability is one of the first trainings of the young mind about predictability and matching. ...
… emphasized in extended mind and embodied/radical embodied school of thought (I belong to many schools, to these too) …
…
**Тош. Бел.** Ограничаване на изчисленията при прилагане на общи методи за обработка и принципа на суперпозицията, в случая – принципа на свободната енергия, намаляване на грешката в предвиждането; ограничаване на нива като се предвижда или отчита бъдещото влияние тогава, когато е онези съставни буквачета (елемент). Това обаче е приближение, опростяване. В рамките на Вселената на най-ниско ниво, или поне на по-ниските нива, които познаваме, частиците могат да „не спазват йерархията“. **ГнП**

*** TaskGen: A Task-Based, Memory-Infused Agentic Framework using StrictJSON,** John Chong Min Tan, Prince Saroj, Bharat Runwal, Hardik Maheshwari, Brian Lim Yi Sheng, Richard Cottrill, Alankrit Chona, Ambuj Kumar, Mehul Motani, 7.2024 https://web3.arxiv.org/abs/2407.15734
“TaskGen is an open-sourced agentic framework which uses an Agent to solve an arbitrary task by breaking them down into subtasks … dynamic 40x40 grid maze …
https://github.com/simbianai/taskgen
https://github.com/tanchongmin/agentjo
Function calling, strict JSON …

*** Hierarchical Agents notebook example:**
https://github.com/tanchongmin/agentjo/blob/main/Tutorial%204%20-%20Hierarchical%20Agents.ipynb

**Todor: Zrim's Executable Contexts:** The agentic frameworks for dividing a task, intentions (Тл, TL) to subtasks or subproblems, correspond to operations of the **Ob{K}** with sub{K}, lower level virtual universes; {K} in general and {K}! are concepts in **Zrim** and the proto AGI infrastructure projects of The Sacred Computer since 2013-2014, where the modern concepts of "tool use" and "function calling" are sub types and functions, properties of {K}! and a kind and a method to achieving {K-K}.
A few hints are: Contexts, Executable Contexts, Modalities; these contexts are more abstract and complex than the ones in LLMs and the concept is older.

Zrim however builds the {K}! and the {K}# from primary {K} and creates and generates {K} with appropriate bukvacheta; with Vursherod, Kazborod, Kazbeslovitel, Slovekazbitel, InR etc. which are organically connected with and incrementally derived from the primary {K}s and the path of development in Vsetvodeystvo, while in the current LLM-based Agentic frameworks the LLMs are usually "black boxes" and the agentic frameworks consists mostly of glue logic, it delegate most of the work to the LLMs. Intermendiate representations and "tool use" interfaces like MCP, "Strict JSON" etc. serve as partially "whitening" of the "black" box and create a {K-K} in Zrim's terms. Note that using LLMs for many of the functions is too inefficient and absurd. A broad multimodal LLM in Zrim's terminology is a sort of Ob{K}.

See *"Genesis: Creating Thinking Machines", T.Arnaudov*: future work. [25.8.2025]

## * Kant and Schopenhauer already defined many of the modern concepts in AGI which are still not well understood or are rediscovered by the modern English-speaking world AI and mind researchers etc.: Todor Arnaudov

**Comments by Todor to johntcm in his AI group in Discord** ~ 17:xx h in room #memory, 16.6.2024
group: https://discord.com/channels/1095629355432034344/1133760747437051964/1251904578677637171
**See also:** I https://artificial-mind.blogspot.com/2014/08/the-super-science-of-philosophy-and.html
https://artificial-mind.blogspot.com/2013/08/issues-on-agiri-agi-email-list-and-agi.html

(The multi-inter-intra-disciplinary blindness and the actual obviousness of "everything" creative and cognitive, but the lack of proper understanding, knowledge and access to the structure.)

Nobody asked me, but I am not so impressed with these quotes and explanations at this level of generality (including the previous book, the ones which I checked from the "Why We Remember:Unlocking Memory's Power to Hold on to What Matters", Charan Ranganath, 2024)
[The quotes from the other book were from some of https://en.wikipedia.org/wiki/Nicholas_Humphrey I am not sure which exact one, about 30 years ago]

IMO we must be way more technical and the required technicality can't be expressed in that kind of NL; a proper language of thought and code, and referred data, are required.

As of being imaginative, IMO Kant and Schopenhauer were, and quite more insightful than many current "stars", they were 200 years earlier. 30 years ago is not that much time either (or 0 years, current texts), there were pretty advanced computers already, hot discussions on connectionism even since late 1960s/early 1970s, and very hot in mid-late 1980s, there were also "hidden Markov models" (and "hidden variables").

Kant and Schopenhauer in particular already have explained a lot of material that, it seems, many of current AGI/AI experts still can't really understand or they rediscover it and reexpress it as fresh with a gazillion of flops and bytes and with all data and knowledge manually [and then automatically] collected, classified and preprocessed, again and again rediscovering trivial wheels. Everything should be already obvious for the blind, we can "touch" every pixel, every possible transform, every formula, everything.

Kahneman's "System 1 and System 2" are, as far as I understand them,

well known and investigated concepts from the German philosophy at least from 200-250 years ago, apodictic/intuitive and discursive knowledge.

"Symbol grounding", the basics of "predictive processing", Understanding [it's a concept, a faculty of mind] as physical simulation of the sensory input/world modeling, intelligence inferring the causes from the effect, is explained by Schopenhauer even in his PhD dissertation in 1813. The cognitive hierarchy or "latent spaces", the weighing of motives and their relation to the Will is explained again at least by Schopenhauer in his subsequent major work "World as Will and Idea", starting in 1818 and later parts and editions, which is translated also as "World as Will and Representation" etc.

It is actually Kant in late 1700s and continued by Schopenhauer, not Alan Turing who first defined the abstract computers: theirs, mainly Kant's, refined by Schopenhauer, definition is the centrality of the a priori conceptions for any thought process (or "computation"): time, space and causality, and the medium "matter". Time, space, causality and matter is the minimum definition of a computer, where:

Space is the memory, time is the process of change, the reading of the next step/instruction/address; causality is the rules, the specific instructions per every possible current state - how current state is transformed to the following; and matter is the substrate which can hold the states, it is the "type", the set of possible values within the memory cells within the space, the capability to have properties and their possible values. The matter is "eternal" within the running simulation or computer, only the "forms" (the current content, the "accidents") change by the laws of the causality (the "principle of sufficient reason") which in computers is the chain of executed instructions, or if it is represented as a state automata - the changes of their states etc. A 2014 article about that etc.: https://artificial-mind.blogspot.com/2014/08/the-super-science-of-philosophy-and.html

Re[garding] reward vs goal, I've participated once in such a discussion here: essentially it's the same, it matches to "match" also to any kind of "optimization" (variation calculus). Defined as a path of reward or as a goal and subgoals, it can do the same, each of them can be defined as the other, finding subgoals can be a reward, maximizing prediction can be a reward in the cognitive space.

Some define RL as getting the reward "only at the end of the episode", but even then the episode could be reduced to the shortest step, and be on each step. There could be and there are different "rewards" in different domains, modalities, resolutions, "abstract spaces" (yes, they are multiple: multimodal and there could be branches within the modalities) and they could

interact in a multi-agent way. Reaching the goal can be counted as "reward", or achieving "highest reward" for a selected span of steps, time, whatever can be counted as a "goal". All of them is about prediction and minimizing "prediction error", i.e. "will", or matching: matching some target, which can be expressed as "reducing the error" or "maximizing the match".

Also there are at least two layers and types of rewards/goals/prediction in mind, one is sensual, the other one is cognitive, that was discussed here (also taught during the AGI course in 2010,2011), and it seems it was also defined in Schopenhauer's AGI theory, it's even in the title. The Will (see his concept) maps to the sensual reward, which is about matching the desired state, being near it, reducing the error to the *desired* representation. While the Idea (representation) is the Cognitive "Reward" or "Goal" which are the same as conception[s], and it is about maximizing *prediction*, or prediction *progress*, maximizing the *knowledge*, "epistemic reward", which may be *against* and contradicting to the sensual, survival, preservation reward/goal. The lower the speceis/the individual in the cognitive ladder, the more his cognitive part, his Idea, representations, or Reason in humans are ruling his behavior, and more he is a slave of the sensual goals or rewards, which are the same as in the animals.

**John:** *"Multiple abstraction spaces to process image - position, movement, shape, colour etc.I believe our brains process images via multiple abstraction spaces, not just with one transformer like what we see right now."*

Yes, the concepts per se are "abstraction spaces" by definition, concepts/generalization is inducing the spaces for different features and classes. Also different resolutions, different maps to different views/aspects, selections of "items" within the spaces.[41]

[41] This happens implictly or explicitly in the neural models for image and video generation. [23.9.2025]

## Eight Things to Know about Large Language Models

https://www.youtube.com/watch?v=RX-gGs_EV7M
AI Coffee Break with Letitia, 7.6.2023
36,6 хил. абонати – *"There are no reliable techniques for steering the behavior of LLMs"*

**The Truth, 12/2002, Todor:**
*Bozhidar, the creator of the thinking machine, is editing part of the source code of his creature and then testing it in a dialog. The machine ridicules him and explains that it is "overly grown up already" in order for Bozhidar to play with its mind in such an elementary way, using the "Fool" (Глупчо, a neologism for "stupid computer", a non-AGI computer).*
*… However Emil, the thinking machine, was not designed to please the human with whom it chatters. (…)...*
* Виж „Истината", 2002; „Читанка"/"Моята библиотека" [The SF novel about AGI, mind, universe: "The Truth", T.Arnaudov, 2002 *
https://chitanka.info/text/865-istinata *
https://www.oocities.org/eimworld/eim19/istinata.htm ]
**…**

## * Theory of Mind and Universe (TOUM) and the universal simulators of virtual universes, time, present, past …: comments to John's AI Group in Discord

**To John:** Yes, right. Broadly, IMO as Mind and Universe are working by the same core principles, and mind is/can be represented as a universal simulator of virtual universes, then it is logical that at a high resolution intelligence would converge to mirroring physics and physical processes, as we see in AI as well, and as it is, implemented in brains and wherever...; the higher forms of physical laws/forces are more "self-reflecting" physics. In general in TOUM CCUs (control-causality units), or agents/minds are higher orders of physical laws/forces, they are part of nature and are within the same cascading sequence/ladder with the smaller and "less autonomous" (from outside), shorter range/span/influence particles/agents/CCUs. The technology, humans, machines they are natural forces and not something which is working "against nature", they are more developed/advanced forms of "nature".

Also for the less advanced natural forces it is difficult to build structures with straight lines and right angles, perpendiculars etc. Some plants and bones grow like that, helped with the gravity, but the initial way is in multiple directions and branching like the roots and branches of the plants/trees, the blood vessels and neural structures, the traces of the lightnings. There are feed-forward functions such as the Central nervous system commanding muscles etc., which however happens and is driven and

controlled by feed-back loops and by side compensating forces and counterparts, like the mielin insulation, inhibitory neurons etc.

Locally there are forces or micro-agents/CCUs which are "minimizing free energy", but they are confronted with other agents doing the same, as constraints; the interaction between the clashing objectives forms the growing, the "morphogenesis" and the structure of the complex patterns. /*** ---> POSTPONE don't publish .../

**John:** *"Though, practically, most systems that deal with input-output are actually more easily modelled by feed-forward networks."*

The other way round may happen to be hard to model or design *by us/enginneers*, but it may be less necessary, because it could develop and model itself on its own. As Bert de Vries explains in the recent MLST video. The constraints will be defined and the rest will be completed by the AIF process on its own. This is how I've imagined it for general software development as well, but haven't pushed focused enough efforts to make it practical. (Not the way LLM do it - not by "ingesting" an enormous dataset with complete code that was already developed and a text mapping to it, but by an incremental process of discovery, like the one which a human who is learning to program goes through: by gradually searching, exploring, mapping, growing, starting from a properly designed "seed" or "seeds" of a fine-grained explorative-generative process.

The constraining sounds similar to Steven Wolfram's framework rendition, although he poses it differently. If I'm not mistaken once when talking about the multiway systems with Joscha Bach he also mentioned just setting constraints in "his" cellular-automata universe, and the process would adjust the paths to these constraints without a need for explicit programming of all of the intermediate steps. (... I'd say also the generative process. Software can be "grown", not developed (no the. I "saw" that an age ago, but haven't put enough focus ...Yes, also in late decades software engineering embraces the immutable structures with functional style programming with no side effects, for easier concurrent programming without race conditons. However in the cognitive and living systems there are side effects ... **That's related to the competition between the goal of an agent to predict better with being unpredictable** (for other agents). ***[# Add as Appendix to UnM6 ? Not added for now.]***
kkyuan77 — Днес в 0:57
@Tosh You mentioned the following: "General intelligence/Mind is a system performing multi-scale, multi-range, multi-precision, multi-modality, multi-domain ... hierarchical prediction-causation of the future, implying the creation and operation of universal simulators of virtual universes."
Questions: **this definition only applies to robots and AGI artificial agents, and not for humans, correct? where does "present moment" fall, like "now"?**

**Todor:** 1) **It is for humans as well, humans can be represented as ..., their behaviour, cognition etc. in a virtual universe/model.** So it is not only for artificial beings, in my view it applies to all "complex" causality-control units. Human beings are a kind of CCU and inside them these phenomena can be recognized with different CCUs depending on the factorization and segmentation (the cells-tissues-organs-systems-organism is one such broad segmentation, with its sub-segmentations, overlaps etc.). We may be unable to objectively measure what particular "things" actually are, but we could say that they could be represented or emulated in such a way in virtual universes, and similar or the same - with a given resolution of causality and control and a measure of similarity - outputs, behaviours, interactions, structures will emerge. The actual substrate may be different.

Also the initial definition and state of the "Seed AI" could be not hierarchical, it could (and should) be able to form the levels and grow/segment the constituent causality-control units, levels of generalizations, range, to connect to the sensory input sources from different modalities and actuators etc. Like a virtual causality-control unit representing a young brain and its initial struggle to take hold of the body and its senses. The latter process is actually on-going all the time, as brain loses neurons, may suffer a stroke, the quality of the sensory input varies and deteriorates temporarily or permanently, the muscle strenght varies - grows and declines after exhaustion, the elasticity of the other connective tissue etc. Well, it's "life-long learning", because there are too many unreliable variables even within the core "hardware".

Also I believe that there's no single objective unified self in the mind, there's an integral.

**2) "where does "present moment" fall, like "now"?"**

Subjectively-cognitively, if analyzed externally, the perception of "now" actually seems to be either the immediate past (if the systems needs to process the sensory input or our thoughts, there should be lag), and also the immediate very short future. For both it is about some selected range of sensory buffers/matrices/spaces, like with working memory/phonological loop, or a prediction-causation horizon, a range of expected reliable-enough anticipations and foreseen possible actions or changes in the environment/input. Maybe it is a sort of a "blade" of the causation process, like the derivatives, an "infinitesimal" range of the span of prediction.

Humans have some 0.1-0.2 reaction time at best for simple stimuli, and we integrate longer periods for our present to comprehend more complex ones. We can take the content of the working memory or the phonetic loop also for some form of the content of the present, up to several seconds from the immediate past (or progressively going deeper). Have you tried this - when listening to somebody speaking, and taking notes or just remembering - for how long you could keep a precise record in your memory, to recover the words exactly. They say it's 2 seconds on average, I guess

professional simultant interpreters may have a big one, in my records it's up to several seconds**.**

So in this framework "now" seems to be the content of the latest sensory buffers, used for prediction-causation, the last/current state of the generative model, which is considered measured and "certain" to an acceptable degree.

However that is fluid by the selection of the resolutions of causation and control, which includes time-space-modality-domain-granularity (including specific choice of the patterns and their limits) -... etc.

Therefor there could be different "now" ranges with different buffers and senses of immediacy. In the example with eating a piece of chocolate or not eating, when both policies are correct for each of the virtual CCU, the short-term vCCU has "now" of a few seconds, while the long-term could have it in months. That "now" here is some range of a reliable prediction horizon, reliable by own estimation (they might be wrong).

Also another view on "now" could be some measure of sampling granularity, some minimal temporal resolution of the periods of sampling for some mind/agent/causality-control unit.

Another parameter which I recall is related to sound perception, the "fabric" of sound - it's the shortest sound which is distinguishable, about 0.01-0.02 sec. Then the sounds are heard as one. This is possibly related to the length of some processing loops in the brain and a clock in the thalamus which is 40 Hz.

An additional view to "present" as a state of the system could be the last "certain"/reliable/confirmed/clarified/accepted data/representations, from which the following states are generated.

Something like a "checkout", "OK, that's sorted out, what's next". The last thought reminds me of the title of one wonderful paper by Andy Clark. It's from 2013, but I discovered it a few weeks ago:
https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/whatever-next-predictive-brains-situated-agents-and-the-future-of-cognitive-science/33542C736E17E3D1D44E8D03BE5F4CD9
…

Also now and past can't be changed anymore (however only for another exteral observer who has the "real" data and can fix it; because the *records* for the past could be changed, the attitude/evaluation towards them can be updated, details could be found from other sources which were not available at the time. So it happens that "the future can change the past"... That reminds me of the video compression algorithms' option to encode a frame based on the "future" frames as well, from the POV of the current. Therefore in the future we can roll-back the record of the past and recompute it for making new predictions later which are more fit to the current generative/predictive model. Thus the past *as representation* is also volatile, however

when we revisit past records, they are now refreshed and also turn into future ones. (That also happens all the time with the political and economical history and how it is told and how it looks to the population in different countries, different generations, different strata...) One difference of the future to the past is that the low level representation from the future can cause deadly changes. Also future is where the actuators operate. It should be in the future due to the lag. We can't change the (actual) past, but we believe that we can change the actual future.
"fresh sensory" input, ... and revisited past ...
* Now, present, future also are different depending on the evaluator, as with Levin's cognitive lightcone. For example for the parts/subuniverses/causality-control units within a human individual, or in a cascading hierarchy of the military. The present of a marshal or the emperor could last, say, a week, a month, a year, some "current" operation, "now he's conquering a...". His past is defined with events which require sampling bigger spatio-temporal ranges and are more abstract than the ones of the lower rank officers and soldiers. The past buffer for predictions of the next operational moves can go months or  years in the past.

On the other hand the soldiers or ranked-anywhere person, being at the actual battlefield, where the bullets fly near his body and shells are falling from the sky, his operational future for "reliable predictions" for exmple of whether he won't be injured or death, may reach just a fraction of the second or a few seconds in the future. His actions and affordances involve his own body and his weapons; he could duck, aim (choose a target, should he interrupt and take cover, is the distance in range, ...), reload (thus remember how many bullets he has left in the barrel, does he have more ammo etc.).

M.Levin's "cognitive light cone" and the range of "care" - a bacteria or a cell has predictions and perceptions within its microscopic surroundings, a dog has bigger lightcone, but it can go to another city or a month in the future, while a human can have a lightcone beyond his life and going to "the end of the universe". However that measurement is a bit ungrounded for the longer and unverifiable matters, also "to care" is subjective and fluid. For example one may *pretend* or *believe* (falsely) that she cares "for the planet" and her behaviour is driven from such humanity-Earth encompassing intentions for the well-being of everybody (e.g. "let's all go green and use only solar panel for electricity, save the nature, ..."), while there could be shorter and simpler motives, such as vanity, which is "Devil's favorite sin" (*“Devil's advocate” movie*) and also plain egocentrism, i.e. a wish *what one wants* to happen *[which is the default mode of operation of any autonomous agent]*. Yet “vanity” is an abstract interpretation, where “abstract” means “more abstract” than another one that can be induced by “simpler”, “lower range”, “fewer parameters” measurements withing the brain or the behavior etc. There must be a judge, that is wide enough and covering all competing and overlapping agents; the judge has to confirm or prove that

somebody “really” cares "for the planet", and that it is "good for the planet" or for the humanity (to define what exactly is the wellbeing of the "humanity") – however is it possible such a judge to exist and to be objective, and be such in advance. That could be God who predicts and knows all with the highest resolution.

Similarly to "*here*". Like the epsilon in Calculus/maths, "small enough surroundings", where "small enough" can vary and it is a matter of resolution and the way the resolution is defined and processed (how the moments/steps/locations/... are discriminated and to what).

*** Unwanted prediction --- multi-scale/competing and multi source competing (sensual vs cognitive, or different nuclei)...** example ... 30/11/2023 Зрим: също НЖР – нежелана разлика: (Zrim: unwanted difference etc.); related to “by effect/side effect”. See explanations in the original works from TUM and in *“Stack theory is yet another fork of TUM”*, 2025.

→ See below **an example** from Lex Fridman’s podcast about **competing causality-control units** where the predicting one desires to be wrong:
**Пример** от подкаста на Лекс Фридман за **съперничащи си управляващи устройства**, където предвиждащата желае да греши:

## * John Mearsheimer: Israel-Palestine, Russia-Ukraine, China, NATO, and WW3 | Lex Fridman Podcast #401

3,43 млн. абонати | 1 994 782 показвания 17.11.2023 г. Lex Fridman Podcast | “John Mearsheimer is an international relations scholar at University of Chicago. He is one of the most influential and controversial thinkers in the world on the topics of war and power. “
https://www.youtube.com/watch?v=r4wLXNydzeY

***LF: Now, you disagree with that? I hope you're right and if they can shut down the Ukraine-Russia war,***
**JM: 1:17:21: it would be wonderful. If I'm proved dead wrong**, that would be wonderful **news**. My **prediction** that this war is gonna go on for a long and end in an ugly way, is a **prediction** that **I don't like it all**. **So, I hope I'm wrong.**
**Todor:** That’s an example of competition between different CCU within a mind/an aggregate CCU (an agent). They run at different levels or with different representations, domains, spheres, ranges; connected to different rewards, goals (cognitive, objective: prediction, based on data, knowledge: what follows; and sensual, or in Schopenhauer’s terms, one based on the Will; emotional, based on the aim to match the desires, wishes (it could reach to “wishful thinking”): the wishes do not have to align with the “reasonable” or “anticipated” reality – they drive the cognitive part to alter the reality in order the reality to match the wishes: these are the causation forces, predicting the future by causing it.

The competition happens when the cognitive and the sensual predictive system have contradicting forecasts/desires **and**, strong in this particular example, the highest power CCU of this system **doesn't believe that it can affect or alter the referred predicted future by an application of its own strategy, actions** etc. If the CCU/agent **believed that it could** change the outcome, it would perform some actions etc. and will not *just* express **emotions** that he dislikes its predictions, like they are *separate* from him, external – objective.

The mismatch of the desires/goals and the apparent contradiction against the prediction (as a core operation/goal of the cognitive system in predictive mind theories: it reports, that it wants to be wrong) is also a display of that duality or a multiple faces and subsystems: the "self" is not monolithic.

This is similar to the popular songs about the differences between "what the heart and the mind/brain/body wants". *„My body's saying let's go. Oh whoa. But my heart is saying no, no"* [42] *(Christina Aguilera's song, Genie in a bottle)*

*Джон Меършмайер: 1:17:21: Би било чудесно, ако се окаже, че греша, би било чудесна новина.* ***Предвиждам,*** *че войната ще продължи дълго и ще завърши грозно, това е* ***предсказание,*** *което въобще* ***не ми харесва****. Така* ***че се надявам да греша.***

...

[42] One may argue that there's also a social face or façade: a person aims at being moral or being perceived as such: "ethical, good" etc. and one may *state* that he wishes his predictions not to happen, because he *believes* that this would be *more beneficial* for how the audience, other *agents,* CCUs, that he would interact with in the future will *perceive* this behavior. Etc. In this particular case and for a "decent" human it's normal that one doesn't want wars, this is a deep analysis of a general situation, not an accusation to this particular analyst. See also the footnote on p.9 about the "ape hierarchies" in "Stack Theory is yet another Fork of Theory of Universe and Mind", T.Arnaudov, 2025

* Some may refer to Z.Freud's "Id", "Ego" and "Superego", it has analogies, but IMO it is not the correct conceptualization and generalization, the way it is constructed is to satisfy the "market" that the author had. The word "Ego" is often misused. How the theory will work if the child is raised by robots? Will it go through a period of "wanting to marry his robot" – if it's asexual? See also comments in the book: "What a Man needs? If you play by the rules you will lose like the fools", T.Arnaudov, 2014, Razumir, issue #1 („Какво му трябва на човек? Играеш ли по правилата ще загубиш играта!", Т.Арнаудов, 2014, сп. Разумир, бр.1) http://razumir.twenkdic.om There are ranges of predictions and "cognitive light cones" of "care" (see M.Levin).

## * Todor comments RL and explains from where the dynamics come in an answer to John in 12.2023

## 1.12.2023 ... John-Tan-Min-johntcm-tic-tac-toe-new82.txt

johntcm — Днес в 8:01 Discord

***How about tic-tac-toe?*** *There are many win states. We don't know in advance what a win would be. But we can learn when we win like:*
*Board State A -> Board State B -> Board State C (Win), Player 2 wins*
*Board State A -> Board State D -> Board State E (Win), Player 2 wins*
*Board State A -> Board State F (Win), Player 1 wins*

*Eventually, we could also do goal-directed action selection based on an abstracted state space for wins .. Start State: X (abstraction for start state A), End State: G (abstraction for goal states) How to create this abstraction? Clustering of similar end results? Could we have learned the rules by English and so know how the game would work out? If so, then we know how the win state would look like without needing the abstraction.*

***Value*** *as Abstraction:* ***Value-based RL methods*** *do away with the need for this abstraction because the value can be carried over by many states and the value itself of win or loss is the abstraction already. But if we use memory-based RL, we will need some way to abstract this. Generic Goal State: Another way is to assign a generic state G every time a win occurs so we can use that generic state as the goal-directed action selection. So now, Start State: A, Goal State: G (generic goal state to denote a win)*

***Emotions:*** *Another way to do away with abstraction is to give each memory an emotion.*
*A chain of states leading to a win is given a high emotional value and will be preferred when doing the Monte Carlo lookahead (much like value network influencing the monte carlo search in AlphaZero). We can then do search to maximise emotion rather than to reach a desired end state. Doing so is very akin to value-based methods, which I think may not be that great as the moment the goal changes, we need to relearn all the values.*

*Emotions may also be used to indicate surprisal and hence can help to bias exploration for memory-based planning. Could there be both a value-based system, and also a goal-directed system at the same time? Value-based may still be useful for abstraction, but the goal-directed part will be helpful for fast learning. …*

***johntcm — Днес в 8:09*** *Could RL* ***value-based learning just be a subset of goal-directed learning****? In this case, the goal is the emotion. There can be many goals, long-term, short-term, emotion based, context-dependent and many more. RL is just a small subset of it.*

**Todor:** I think RL is not always used in that "strict" sense, e.g. as just mentioned by matto as only one reward; or some say only reward at the end of a game episode, win/loss - some groups in ML/AI. E.g. I've heard also in talks in the FEP/AIF community they try to emphasize the difference to RL, as in RL the reward/success estimation comes only in the end of sequence, or that it comes "from the

environment" or that it is set by the designer, but this is relative to the picking of words and POVs, the internal "rewards" (the logic of the system, either maleable/self-organizing or fixed) are always "designed" by some "Designer" (be it God - some initial or later state that pushes it to one or another direction; the environment as well). If each "episode" is reduced to the smallest time period/step, a "reward" (which is also just "feedback") could come at each tick. If there are multiple resolutions/ranges in time/space, different factorizations/views etc. – that turns into something "different" to RL if one wants to emancipate that approach, but the essential principles are the same: interactive play with an "environment", the sensori-motor loop with sensor data, gradual adjustment, optimizing something (goal- ... even if just binary is also optimizing the match) etc.

I myself have used RL in my Theory of Universe and Mind in a broad sense, at the time I didn't use the term RL, but got it from human and animal behavior and I've been always assuming a model-based RL (simulators of universes, causality modeling), so every online-interactive-sensori-motor learning maximizing/minimizing some path is a kind of "RL", and the model-free RL has implicit models, which can be induced by their choices.

Re the distinction about goals: goals and rewards or utility functions can be remapped to each other, it's a matter of implementation and both goals and RL as paths can have different horizons, and maximizing the value of the path is again "a goal" with a target "until reaching some value" or "until increment stops below ... " (some maximum of the current value, of the gradient, of ... absolute or in comparison to ...)

**The complex dynamics and behavior of generally intelligence agents come from the multilayer, multiresolution, multirange, multidomain, multiview, preemptive, varying prediction horizon, resolution of perception and causation-control etc. operation.**

One view towards the subgoals is lower level/higher resolution/lower range/near-target etc. "goals" or/and in path-formulation: shorter aggregation span, prediction horizon (the last is valid for both), with some segmentation, triggers between different layers, ranges, causality-control units at different scales. Etc.

Re the distinction that you make about value-based/emotion-based learning and goal-directed, besides the reasoning above, it resembles the dichotomy in Theory of Universe and Mind, in FEP/AIF, in Yoscha Bach's explanations etc.

Biologically these are two flows: basic needs for survival and higher cognitive needs. The terms in English in TOUM are "Physical" (or sensual) and "Cognitive", with the following difference of their optimization functions: Physical set some initial constraints which the cognitive is trying to fullfil and map to. It's not exact map, it depends on the very specifics of the mapping, in humans/animals, every specificity of the structure of each individual brain - some people in some times (some conditions of their bodies) have higher control of their cognitive modules over the subcortical/cingulate gyrus, basal ganglia/amygdala etc. and they can be more "spiritual", more abstract, motivated by "non-pragmatic" goals and maximize more abstract "utility functions" and get "immaterial rewards". Other are more animal-like

and are always directly chasing some more visible bigger and sweeter "banana".

In AIF the "physical" or ~"emotion" are also some initial constraints, as Bert de Vries explains in his MLST appearance, which give the agent a purpose and shape its behavior, because otherwise it will just model the environment, the universe (just like in TOUM as in many other aspects: the agents/minds/causality-control unit are universal simulators of virtual universes aiming at higher precision and range). As far as I've been styding the literature and talks of AIF institute etc. lately, the current toy-problem implementations of AIF agents are using "CFFG" - Corney style Factor Graphs for setting these constraints. ...

With goal-directed sometimes it's meant comparing to a "final destination only", while value is about accumulating some "rewards" is "continuous", but as mentioned these can be interchanged or represented as each other. RL is also chasing goals and in complex minds they are multi-...(many things) and overlapping, not just winning the game etc.

These subgoals emerge in the embedding space/internal representation if the structure is deep, goals to reach to certain "hidden" representations which lead to some visible target ones.

Also for the goal-directed line  there should be measures of how far is one from the goal etc. (otherwise it won't be very directable) and the agent has to "minimize the distance".

**Entropy - The number of distinct configurations Boltzman |** 2/12/2023
tends to increase ... more likely ...
**Todor: But most configurations are very similar, combinations etc.**
highly ordered beginning the Bing Bang... degradation ... disorder
**Todor:** I don't agree? Wolfram’s more recent discussion that entropy growth may be an artifact of the observer and other theories about incrementing the “order” of the Universe and that evolution is actually not random.

*** Three types of creativity as cited from a Demis Hassabis' comment – Todor's comment:** See also: http://artificial-mind.blogspot.com/2023/12/on-what-creativity-is-different-tints.html "On What Creativity Is - Different Tints and the Pereslegin's remark about the modern AI researchers rediscovering Lem 1963", T.Arnaudov, 16.12.2023

**Hi, I just glimpsed here and saw Tim's comment on LLM's and a few answers and I have so much to say that it's "despeairing", requires too much context and volume.**
https://discord.com/channels/937356144060530778/937356144060530781/1185156296152989716
https://discord.com/channels/937356144060530778/937356144060530781/1185159112334512219

Three types [levels] of creativity, citing Demis Hassabis:

*Demis Hassabis mentioned three levels of creativity:*
*"1. **Interpolation**: This is the lowest level of creativity. It involves averaging or combining existing concepts or knowledge to create something new. For example imaging a new type of cat, based on all the images of cats that a model has seen.*
*2. **Extrapolation**: This is the second level, where an AI system can create new strategies or ideas that are extrapolations from existing knowledge. For example, AlphaGo's move 37 or the creation of new pieces of music or poetry. This level also includes the ability to spot analogies or connections between things that might not be obvious to humans due to the system's vast knowledge base.*
*3. **Invention (Out-of-the-box thinking):** This is the highest and the third level of creativity that involves creating entirely new concepts or frameworks that are as aesthetic and classically good as human creations like new games or art movements. According to Demis, current AI systems have not yet achieved this level of creativity."*

Pereslegin/Sociosoft recently cites Lem's Summa Technologiae: (see the appendix about the history of AI SF, История на научната фантастика за мислещи машини)
*"In 1963, in „The Sum of Technologies" (Summa Technologiae), everything was predicted by Lem. He divided the development of AI into 3 levels:*
*▪ AI - can transform big data and therefore can play finite games better (chess etc.)*
*▪ The artificial mind - can work with the new/novel (Boris Lapin, "First step").*

*Aromamorphosis.*
▪ *Artificial consciousness - will be able to think about something else. But we don't know what consciousness is."*

The second one is the "extrapolation" or partially the new, the third part, in Russian "нечто иное", a special kind of novelty.

IMO the specific definition of third point of Hassabis is superficial and unconvincing. In the early 2010s on AGI list some talked about the "radical novelty" in this regard.

**Re K.Stanley's comment** (on Twitter)**:**
https://twitter.com/kenneth0stanley/status/1733571230803058920?s=46&t=Dz1KKc4TEZU9Zl6fQfw3ew

**Todor Arnaudov - Tosh/Twenkid @Todor82871979**
Dec 11, 2023 * Hi. It is true that one must know what already exists, in order to be sure [that] the new is novel, but humans usually don't [know] either, also there are different criteria and resolutions. *"the tendency to present old ideas as novel..."*: the same & the authorities usually don't get penalized for doing that. E.g. "everybody" in AI is rediscovering and repeating the core claims, predictions and reasoning of my teenage works "Theory of Universe and Mind", 2001-2004, taught during the world's first University course in AGI: Plovdiv 2010,2011. They wouldn't admit it.

*„Ask e.g. for innovative new recipes, ideas for new genres of music, or inventions to help with some existing problem, and often what you get back is something that either already exists or already has been proposed!"*
**Todor:** 1. There must be criteria and explicit definitions and measures of "creativity" and it's always according to some evaluator ("observer") and based on its measures, memories/knowledge ("database"), capacities to predict-cause (analyse-synthesize) (see e.g. below, "The Sacred Computer").

There are different types of creativity, it is an ambiguous word even if the mundane everyday use. In more precise use it could be "reproductive" and still could have many gradients, it could be "*radically new";* Sergey Pereslgin, a prominent futurologist[43] and SF history expert talks about the "иное", ["another"] i.e. "*something truly creative*", unlike "*what already existed in data*" etc. But these are artificial/unspecified and matter of degrees, unless it is explicitly defined; and if it is, depending on the details, now a "box" may appear and it may unveil that that "truly creative/different" is not that "inventive" and was implied in the initial conditions as well.

For a novice or a child, or anyone at some level of expertise, learning etc. some discovery is "novel", and something that she creates/produces may

[43] Although he prefers not to call himself like that, but „прогнозист", „социален прогнозист"

be "radically novel": TO HER, *at the very moment of the discovery*. To an expert, to her parent, teacher or somebody who already knows it or is much "smarter" than the child, novice etc. – to one who has a larger knowledge base and prediction and influence horizon*, it could be "obvious", "predictable" etc., because one "sees" further and does it in less operations than the other less fluent agent (see even the pop culture [the movie]: *"Hollow man",* 1999: *a genius is somebody who can go from 1 to 3 without passing through 2*; in philosophy: Schopenhauer, <1850s?: genius hits targets that others CAN'T see, talented hit targets that others see, but can't hit – thus the latter praise them, not the ingenious ones (indeed that is *PREDICTION* and planning, geniuses have a longer and higher precision prediction horizon and highest intelligence imply these capabilities) – and that's why usually the talented ones are "the successful" in life and often the geniuses are neglected and they "suffer" being unrecognized. This is also to show that the "ordinary" and the talented ones measure creativity **by themselves*,* and not by some "objective" measure or the measure of more capable than them, yet the former pretend that *THEY* are the objective or the "true" measure.
* Re[garding] the prediction horizon: M.Levin's "Cognitive boundary of self" in this context is not a new invention. [See again Schopenhauer's *"World as Will and Idea"*, [where the philosopher] ~~who~~ also talks about different degrees of objectivation of the Will with a gradient of different scopes, which is valid also for the set of modalities, with the vision reaching the biggest distance.]

Also if Kenneth/we appply strict criteria for creativity, **humans are not creative either**. Most of human individual's behavior is super mundane, repetitve and has a very short prediction horizon for precise "tokens" at the level of words or art-making code tokens or motion, whatever [that includes one's **own predicted future tokens** and even **a record of the recent past**]. People can't remember a tune of 5-10 tones, they say the average phonological loop goes 2 seconds in the past (the amount of spoken information one could remember exactly) etc., yet people ridicule the short context of LLMs. Humans don't remember (exactly, not vaguely "the feeling" or "the topic") what they said several seconds in the past, LOL, they constantly work as LLMs, generative models which are losing their train of thought, because they just lack sufficient buffer sizes, and also that's one reason ***why they believe in *their own creativity**** - they usually *can't trace their own creative, reasoning, thought, generative process.* It is the same about "free will". They can't predict themselves with the resolution that they believed/expected they should be able to predict themselves "if they were predictable", therefore they are "spontaneous", where ~~it~~ all is a matter of degree, resolutions, criteria etc. *[and the lack of explicit and definite definitions*

*of the former]* and the judgmenet depends on the evaluator.

The above [reasoning] was discussed in one letter to the AGI list in 2013 about the multi-inter-intra-disciplinary blindness: *... Issues in AGIRI* (E.g. some practical matters, such as activities which "normal" people use to call "creative" such sa drawing, painting and general image making, music composing/improvization or just playing, "creative writing", are very difficult for most people. A few can draw even decently/contour-wise and perspective - correct images, and photorealistically or photo-correctly is hard and very hard even for the best, 0.1% or 0.01 or 0.001 of the humans?, and as one can easily see, humans hardly write "creatively" a story of just 20-50-100 words or it's super boring or banal. Still they are mocking the LLMs, LOL. Even the large GPT2s were already superhuman, compared to perhaps 99% of the humans in funny/"creative"/coherent/"original" writing of short stories, [given a prompt]; I “laughed my ass out” when played with it in 2021.)

If the reader didn't know/couldn't predict it with (sufficient/expected by her) precision, it's "creative" up to that measure. *[However there should be another measure for the output being “plausible”, “coherent” etc. in order to pass, it must be “realistic” or acceptable by other criteria, which is recognizable by the evaluator who yet admits that she was unable to predict correctly; in order not the evaluator to dismiss the creator’s efforts as just “nonsense” or “noise”. This is another subjective point.]*

Because, again, if one looks for the "global novelty" - God would say: "nothing new under the sun". Everything is produced by the generative model of the Universe and the Universe Computer, it was already implied in what's given, so it's predictable, it is only unfolded, decompressed.

**"Radical novelty"**

Art – ~~because~~ as one can't imagine how it was produced with "sufficient" resolution of causality and perception that they believed they were supposed to be able to predict (to imagine, to explain).

It is produced by including, producing, generating a new modality, that was previously unknown, not considered, not counted etc.; extending the range of evaluation, the precision, the depth - number of steps; the diversity of matches (again some depth/breadth/number of...) etc.

Yet in all cases it is a matter of degree, decision and resolution of causality and perception of some evaluator-observer which decides, judges, classifies.

In music one may decide that Bethoven was "radically new" compared to Mozart, but was it really? Mozart sounded "radically new" to some, his productivity was "ingenious" etc. – for his time, for the blade at the time, now his music, after listening to sufficient amount of it, sounds repetitive and

monotonous, based on simpler correlations – *compared* to something else. Bethoven is perhaps "more advanced", “a next step”; his compositions encoded more "complex" correlations, higher speed, higher number of simultaneous tones?/higher polyphony, *"harder to reproduce*" (another meaningful criteria, on the physical/implementation/practical aspect of "creativity" as a capability/capacity to "produce", "productivity", within a pool of competitive “similar”/”comparable” generative agents with similar/comparable actuators/output devices, modalities), requiring "higher skill"; other pianists, who appeared later in the chronology may have gotten further, say Chopin, Rachmaninov, longer jumps/”hops” in the scale? etc. Similarly in the development of the symphonic music: from small orchestra to larger ones with more instruments, more diverse timbres, more complete coverage of the range of octaves in the auditory space etc.

However overall everything is, or it was already predetermined implicitly in the Universe, more specifically in the anatomy, phisiology, accoustics, the processing capacity and structure of the brain etc., in a broader scope: in the social relations and structures of societies in all dimensions which allowed to develop the musical culture, to create musical instruments and schools, musicians who were trained and improved they skill, the search for novelty and the feedback from the audience which admired and supported one or another performer etc., the preservation of the past and building something on its basis by the following generations etc. Every possible music was implicitly predefined, as well as every other “similar” creative historically evolving phenomenon, culture, science, knowledge. Nothing is "radically new" from the point of view of an observer-evaluator who *admits* that logic. The sound is sound, it can't be anything different than what a particular "human" with a predefined constraints of his or her anatomy, which is implied by his or her biology, DNA, phyisical laws, neural architecture, range of sensible frequency, maximum number of tones, notes or beats per unit time etc. Any piano piece is some combination of presses of the keys with some "velocity", some presses of the pedal, some hold, usually with up to 10 fingers of two hands etc. **That's it** – nothing is or can be "radically new" with this given material. It could be *perceived* as new or “*radically new”* by an observer who haven’t seen it yet [or can’t remember sequences and volumes of data that are big or long enough in order to detect the repetitions].

**Creation of new instruments**, such as the electronic ones change the actually generated timbre, the specific way the sound qualities change after pressing a key; some improved instruments may have softer keys with higher dynamics – the *fortepiano* improved the *clavichord*, and later pianos and grand pianos improved the fortepiano’s dynamics and the sound quality – however at

a level of *'notes'* and at *melodic* or *harmonic* level, the old way and older instruments more or less could have already “finished” [and exhausted the potential combinations, if not with explicitly composed pieces, existing on “sheet music”, as possibilities]. One can change the notes, change the *"laws for making music",* produce sounds which are not harmonious, but just have some rhythm or some other correlations; the sounds could be not harmonious as whole number of frequencies etc., but *whatever* one does, at the lower level it is still just "sound", "frequencies" – nothing new in *that* domain. [Ultimately: *just data, as everything else.*]

You may change the "hearing system" – a thinking machine may hear 1 MHz etc., but at a "low level", at level of "Fourier transforms" or level of data, records of some range of time, all is just "frequency spectrum" or just "time-domain sound pressure samples", eventually just "sequence of numbers: 100, 120, 166, 180, 199". What's "new" about that if one knows numbers, frequencies etc.? Therefore nothing is "really creative" at the lowest level of the universe, if we consider it not changing[44].

However "creative" means also "**productive**" and it is relative, it is compared to something else. If you produce anything that is "different" in any parameter to anything that you as evaluator decided to compare with, you may label the new thing as "novel", "creative" etc. (The word “creative” is used in very local senses as well, it is enough if someone got “excited” etc.)

Also what the "ordinary people" or legally is counted as creative professions? Why they protest against the generative AI?

Why people say that "*children* are creative" or "*humans* are creative", even when they do not refer to “radically new”, “groundbreaking” inventions, but just for everyday mundane discoveries which are valuable only for the individual child or human? Humans often struggle to follow even the dumbest sets of instructions, not creatively, but just by repetition; the most of humans struggle to remember computer machine code of 5 instructions or of 2, or 9 numbers or random words, therefore how *they*, *alone, could* create *anything new* at all?

One answer, beyond human hypocrisy and the “appropriating” tendency, “stealing” merits, which they do not possess – such as taking what the geniuses, highly unusual individuals, can do, as a property of "the humans", “humanity”, i.e. "normal/average" human – is that it is *not the "humans"* (as individuals); the individual humans are only elements, “tokens”, parts of a

[44] In addition, even if it is changing, it is not supposed or expected to be changing *because* of *the creator* of the corresponding (novel) items, artifacts, unless she was God; where the “creator” of the artifact itself is some “subuniverse”, a set of causality-control units, etc. which are also recognized and selected by an evaluator-observer. See my cited comments on the forum “Kibertron” from 2004 in the main volume of *The Prophets …* [Note: 20.3.2026]

much bigger system, eventually the Universe which actually "truly creates", from the POV of the small internal observers-evaluators.

The same logic applies for the domain of other elements. Humans use to say that "computers are just 1s and 0s" [2], now the LLMs are "just lookup tables/hashtables/dictionaries/databases". Well, and you? What are *you, human*? You are just a piece of flesh, cellularly similar to the one you use to eat and store in your refrigerator, but from other species. "A bunch of neurons" – just like an ape, a dog, a mouse, or even the insects or worms – they also have "brains" (or you may call them "ganglia", oh, or that they "don't have free will", "they are just like automata" etc. – however, even in this domain there is a new conjecture, maybe due to the “Green movement” and its connection with the political powers in the world, and now many scientists sign declarations that more of animal species *actually* had consciousness[45], “maybe even insects”, while until recently even human babies were considered not to have consciousness*,* until particular developmental milestones. IMO none of these claims are convincing or verifiable *as a scientific* fact in the classical sense of *the special sciences* (as opposed to *philosophy)*: the subjectivity/consciousness are a domain of philosophy and metaphysics, it is also a domain for ethics, which besides a field of philosophy is a part of politics and is related to the games of the Power in society, political states; and an instrument for social engineering. The believe or agreement about the possession of a “soul”, “sentience” or “consciousness” or whatever label didn’t stop the wars, massacres etc.[46], but sometimes the victims were “dehumanized” and “de-soul-ed” as an excuse for their extermination.

How different are you at a low level? Why should we see *neurons* – why not being more strict? If you view computers and their thought process as "1s and 0s", why don't we look at you at *"just quarks and electrons"* and whatever elementary particles, "qauntum phenomena" are whatever you decide.

You are just a collection of "dumb" parts which "just do what they are told to do" - by the physical forces, laws and their configuration, energy.

You're nothing if you scrutinize your own descriptions of yourself. But you don't, because you **even can't remember what you have said 3 seconds earlier,** what about having ethical integrity and being consistent in your own logic?

---

[45] https://www.nbcnews.com/science/science-news/animal-consciousness-scientists-push-new-paradigm-rcna148213 “Scientists push new paradigm of animal consciousness, saying even insects may be sentient” – *“Far more animals than previously thought likely have consciousness, top scientists say in a new declaration — including fish, lobsters and octopus.”*

[46] See the sections on Panpsychism and this whole book “The prophets of the thinking machines: ....”, TOUM/”Letters between the 18-year old…”, the Letter to the Oxford’s institue …, 2012

**"What's wrong with NLP?", Part I and II: 2/2009, 3/2009** ... , although they are 10-20-100-200 years late to the geniuses and they new "discoveries" were clearly expressed long ago). To Universe there's nothing new

**What about humans? :** "Схващане за всеобщата предопределеност 3" (Conception about the Universal predetermination III, or Universe and Mind 3), 2003

https://research.twenkid.com/agi/2010/en/Todor_Arnaudov_Theory_of_Universe_and_Mind_3.pdf

**10. What does it mean to be an original,distinctive artist?**

The origination of a piece of art is recording of a piece of information (an entity, a file) to a media which serves as a mediation, intermediate memory between the creator and the perceiver, including the artist himself while he is creating or later in the future, when the creator has forgotten his work and is recalling some of its properties. The space is Memory, that's why “informational carrier/media” can be as diverse as diverse can be the types of data that can be stored in space. Original (creative) is such a file, a piece of information, where the evaluator finds less than expected similarities or matches, in comparison to pieces of knowledge recorded earlier.

Originality (creativity) is the capability of making the prediction of the future by the past harder. In this particular case, “past” means a part of thefile/piece of knowledge, that was read before another part which is assumed to be future for it, and the future one was supposed to be predicted using the information read from the file until that moment.

...

Todor Arnaudov's Theory of Universe and Mind, Part 3, 2003

*** Todor on the "compressionists" vs "analogists"** on Discord, John's AI Channel, 1.2.2024:
Somebody.. **"[Legg, Hutter] Both are THE "compressionist". I am almost completely in the "analogists" camp. *"completely incompatible camps"***

**Todor:** The ones you know. Schmidhuber and Kazachenko are before, I am about the same time as well. But I am "analogists", so do they, because prediction is about *match* and creativity is also "imitation at the level of algorithms" (T.Arnaudov, "The Sacred Computer", 2003), i.e. imitation at the level of generative models, rather than direct copy of the data on the surface.

The "analogists" are also "compressionists", because I guess they wouldn't compare every atom, electron and quark of the subject matter they are making an analogy of, they (we) work with concepts, "abstractions", all are shortened, selective, "compressed" representations otherwise the part should contain the whole in full. How, at what scale, do you recognize that the disk of the sun has a similar shape to a coin or the iris of an eye? The smaller pattern is a "compression" of the "bigger one", it's the same "essential features", expressed in a smaller spatio-temporal volume.

If you assume that Hofstadter is an archetypal "analogist", he's pathetic, with the anecdotal shock that he faced when he realized that machines could compose Chopen[-like music] and [human-composers-like] music, so [therefore it seems] they *"had a soul"*, or the humans didn't. The "predictionists" (compression-prediction are counterparts) never had a problem with that [outcome], [they expected and were looking forward to meet it] – the generative art was not surprising (it was predicted by tha SF writers long ago, but the "predictionists" predicted it not as SF at least in early 2000s), while even the "godfathers" of AI confess that they were "surprised" that it [the breakthroughs] happened so soon. "Nobody expected [it]". Wait: we did predict it, pay the proper credit. A more recent one from 2013: https://artificial-mind.blogspot.com/2013/10/creative-intelligence-will-be-first.html Creative Intelligence will be First Surpassed and Blown Away by the Thinking Machines, not the "low-skill" workers whose jobs require agile and quick physical motion and interactions with human-sized and human-shaped environment:
*" (...) The bottom line is that the "white collars" are more endangered in current-time economy. Perhaps that kind of economy could hardly survive the AGI revolution. I guess it may turn inside-out* for a while - the low-skill workers could get higher pay, because intellectual activities will be done in 1 ms for free... ;) We, the smart guys (the smart asses, see "Super Smartasses" the graphical series ) wouldn't be needed by anyone... Not that we are needed now. :)) Maybe the change won't be that big. :D", T.Arnaudov, 2013* [*upside-down]

There were no GANs, no meaningful VAEs, no LLMs [at the time]. Re the last line: everybody are so concerned about the artists who are losing their job. There are a gazillion of artists, "amateurs", highly skilled or talented, in any art, who didn't have a job as artists (they couldn't earn money), but almost nobody cared about them anyway *without* generative AI that "took their job", which was *never paid* and *was not*

*a job.* (I was and I am one of these artists as well, an extreme “universal” one)

A lot of the discourse on creativity is a dejavu and BS, people who didn't really care about art/creativity (and the creative people) started to *demonstrate* that they care (to pretend they did), because that topic became fashionable, the media and the important people started to discuss it.

...

**Tim Scarfe, MLST,** Linkedin**: Criticizing LLMS:** *"They are basically a database: "*
**T.Arnaudov: “Man and Thinking Machine ..”,. 2001:** the section on consciousness, a small portion of it: (MM - Thinking Machine (Мислеща Машина); translated with Bard, no editing)

*"(...) Incidentally, all sensations, including pain, would be just certain data in the machine's "brain"; part of the memory, the information in which can change the behavior of the MM, which she can remember in the future, which would then influence her behavior. If she wants, she could "grimace" or "cry in pain," but her feelings would be just blocked and open transistors, zeros and ones perceived as feelings...*

*But think again what our feelings and thoughts are from a materialistic point of view – neurons in an excited or suppressed state, chemical reactions, the transfer of substances from cell to cell... Can't the totality of billions of PN transitions lead to the creation of some kind of "electronic consciousness" ... as it "happened with man" and his neurons? Or does man have a soul, some kind of immaterial essence that feels what is happening in the brain?...*

*Consciousness can hardly be explained by physics (at least not with today's physical knowledge), because the human brain does not represent a structure for which special laws apply from the point of view of modern physics. Human neurons are not much different from those of animals - simply connected protein molecules, which in turn are chains of atoms of carbon, hydrogen, nitrogen, oxygen, and other elements. The difference between the human neural networks and those of, for example, the chimpanzee, is only in their "slightly more complex" organization in us, which allows us to be called "thinking beings*

*But as I already mentioned, in my opinion, thinking alone cannot be a sign of consciousness (for "soul"), because we understand whether someone is thinking (so he has consciousness, because "consciousness is characteristic only of man," who is the only one on Earth who can think) or not, only by his external manifestations. Human consciousness is personal, at least for now it cannot be "captured" and "consciously experienced by another" (telepathy is still a rare phenomenon). Each of us can feel our own consciousness. "Internal understanding" is proof that we "are aware," but whether we really understand and feel, only each person knows for himself.*

*So the MM can know for itself that it feels, although we think it is not true and accuse it of its feelings being "zeros and ones." She, calmly, without unnecessary emotions, can answer us:*

*"And your feelings are a quantitative, qualitative, and spatial ratio of chemical compounds - proteins, hormones, nucleic acids, etc. It is hardly worth going into*

*details, because your poor brains will not be able to hold them..."”*
[Note 3.4.2024: In both cases whatever “it is”, it is inside the Universe and “gets” its/corresponding properties, it’s never “just data”, “1s and 0s”, “atoms and molecules” or their relations. All these properties are expressed in some form of machine language of the universe. It can be “just…” only abstractly inside a mind, but where the mind is? See *Universe and Mind 6;* [*23.9.2025: See also “Stack Theory is yet Another Fork of Theory of Universe and Mind”, 2025*and “Is Mortal Computation Required for the Creation of Universal Thinking Machines”?, 2025 („Нужни ли са смъртни изчисления за създаване на универсални мислещи машини?“)]]

The response is extended in a dialog in "The Truth", 2002, srcipt and novel.

As of the artists *“losing their jobs*”: Would you pay for a picture I've drawn? I am an artist too. You won’t, right? Should I blame “the AI” or whoever for that or it’s you? I remember when I was discussing with a philosopher in 2002 and told him that *"music was exhausted and it will finish"* (and it was already) …

***Todor:*** *In fact* ***everybody is "compressionist”…***

**1.2.2024,** 01 февруари 2024 г. --> not sent at discord (…)

**Todor:** Yes about the priors (structures), but what's the problem AI to have them, encoded by persons who do understand them in a proper in-born structure? The problem I see is that humans **do not understand** it themselves *(or the ones who do understand do not implement it yet*), it seems "magical" to them as explained above. One related concept is "Multi-intra-inter-domain blindness/insufficiency", explained in a 2013 article. Other discussions are about the fake abstractions, i.e. no need for a transfer and no different domains, it's all the same. ....

**Atrony — Вчера в 20:53:** How can any current understanding of AI ever achieve context independent learning if everything that can be externalized as an example to learn from is necessarily dependent on something. It's the difference between the finger pointing to the moon versus looking at the moon itself. It helps to "get it", but if there's nothing else there to draw from, no evolutionarily driven priors embedded into it (i.e. some sort of deep unknowable ultimate origin of life), there's no moon to be seen. Even as we get to understanding some of these core priors, "AGI" would need necessarily all of the ones that lead to humans, including the ultimate prior. There's a certain spirit-ness missing in AI...

**Atrony — Вчера в 21:06:** My assertion is, there exists some priors within us that cannot be externalized/understood. Therefore, there are some aspects of humanity that cannot be recreated.

**Richard — Вчера в 22:42:** It's a brave assertion and difficult to prove unless you can identify those priors. If you can identify the priors (or perhaps regardless) then you're making the brave assumption that a functional replacement can't be built. Betting against human ingenuity doesn't have a good record.

**Atrony — Вчера в 22:45:** The prior has always been identified, it's consciousness. Unless we can recreate consciousness in a machine, we will never create a human being.

**Aynur4 — Вчера в 22:47:** I agree. I have been trying to make gpt 4 to invent a new writing style. It can write chapters in the style/language of Dostoyevski or Salinger or

Fitzgerald, but so far I haven't been able to prompt it it create its own unique style, only good examples of what already is there. And also if you ask it to build an AI that leverages flexible knowledge graph reasoning, LLMs and market data for trading it will return nothing meaningful because it has never seen something like that. At least it hadn't when I tried. The question is where do humans get novel ideas from.
**Richard — Вчера в 22:47:** Consciousness is notoriously poorly defined. Descartes had a go with "I think therefore I am" but he's far from alone.
**Atrony — Вчера в 22:48: Exactly my point.** (…)
**Richard — Вчера в 22:54:** I'm quite comfortable that originality may just be an emergent effect of randomness (inspiration) and filtering (skill). Google's early fiddling with running image recognition pipelines backwards (for image generation from randomness) resulted in very dream-like imagery (inspiration). From there, it's just an issue of recognising "interesting" and iterating to isolate and polish the "interesting" (skill). Interesting is very personal but it's usually defined as novel, but related to concepts not often related.
**Aynur4 — Вчера в 22:55:** Inspiration even in humans is useless if it is not coupled with vision, direction and a will. But I like your idea a lot . Sometimes random events do lead to new ideas. But we get to choose what random effects lead us to our goal and what just remain noise, we get to interpret that randomness and that interpretation defines the reality we construct.
**Tosh — Днес в 1:15:** Hi, all. Re originality in particular, this is an old general definition from the Theory of Universe and Mind, classical period 2001-2004. That's from "Universe and Mind 3", 2003:
*"10. What does it mean to be an original, distinctive* artist? (… see the citation a few pages above …)*

I recently realized this is similar to what "surprisal" and D(KL), Kullback-Leibler Divergence are defined. There's more elaboration in a recent blog post in my research blog (no links if it's considered "self-promotion").

Schmidhuber has elaborated what interestingness and curiosity is in a compressed form as well, others has commented that it is, like with learning in general, about the ratio of information gain vs effort, if you learn/gain/win more with less efforts.

In the, call it "folk" discourse though, the evaluation of something as creative is confused, the second plane* of predictions, which are not cognitive are often interpreted as in the same category as the cognitive originality.
**Tosh — Днес в 1:32:** Yes, and probably there is a requirement of causal-material connectedness in the way the entities are constructed, a computer, electronics etc. are different, they have different origin, not from cells (well, it may change, though, there could be hybrids) and as it is for now, it can't be a human and possess all human qualities, but the reverse direction is also true - humans cannot feel the electronic circuitry and what it's like to be a CPU, RAM chip or a smartphone (or even an electronic watch or an amplifier circuit or a flip-flop :) ). The claims that the electronics or computers have "no soul" could be refuted as a nonsense by a thinking machine which is complex enough (or by an NPC in a computer game on a computer

from 1950s or 1960s) just like humans do it for machines.

That's a story from the early 2000s, but I see humans still don't seem to understand it, a thinking machine can ask the same questions and talk to you the same way, if she evaluates you at the same "mechanistic" or low level way and as an external observer.

For an observer, and for the humans themselves when they treat other humans like objects and make them suffer with no remorse, humans are just a piece of flesh or just "agents", in the ancient ethics the slaves were "talking objects" – what "soul" did you see in it?

The subjective qualities are not objective. You are atoms, molecules, electrons, very messy, very "spiritually" unstable etc. You continue to exist only as a huge swarm and a system (i.e. not as individuals), and you are helpless without technology - the spiritless technology makes humans what they are and allows their best and most different quality - the most "original" one, compared to the other entities - to flourish and to develop, yet they insist on their most animalistic qualities which are probably more or less present even in the lowest species. Schopenhauer calls it Will.

*GG @Tosh, you just advanced to level 3!*

**Tosh — Днес в 1:38:** The search for "magic" in art/creativity is a sign of a lack of understanding. It is all - or more precisely said, all can be represented as - data and respectively a generative process that produces it. Data can be attributed with some "typing", particular types in some representations (particular material substrates which are required from particular representations within some low level virtual universe in the Universe computer), and for this job there are "converters", but for storing and processing, all can be represented as bits and instructions.

Images are just correlations of pixels, in bigger granularity: gradients, lines, angles, ratios of lengths, curves etc. and sets of matches between them within different ranges within different spaces, past records, ~~those~~ which are mapped to different modalities and possible generations, with different resolutions etc. etc.

The same goes for any domain and modality, all is "images" at different resolutions, ranges, depths, overlaps etc.

**Atrony — Днес в 1:41:** "Will" is analogous to what I'm saying won't ever exist in a machine. Whether or not this will result in never obtaining context-independent learning ability in an AI is a different problem I am now realizing. After some research on psychological "far transfer" (a possible yet extremely subtle phenomenon), I can see AI developing context independent learning ability with enough data and compute. "Will", however, brings about a different property for sure.

**Tosh — Днес в 1:46:** Re the styles of Dostoevsky etc. - if you understand something, it stops to be "magical". The style of anyone is just some combination of properties which the observer recognizes and repetitions, patterns etc., but obviously usually the observer-evaluator doesn't understand good enough, can only barely recognize, but can't recreate. (One other discovery from TOUM is that people get enchanted by art pieces, because they can't understand how they were created with a resolution of causality-control that they expected they should be able to

understand, if they weren't enchanted - ~~it~~ the art piece looks to them, at some level and way of analyzing, as related to "random", unexplainable, incredible, which in the framework of TOUM is "unpredictable" for them, for their capacity; "magical". Other forces are of course sensual - the second (which is actually first) plane of predictions in a human-like causality-control units. The "emotions" which may make people call "beautiful" things which are just *attractive* to them because they match what they *want*, which is different ~~to~~ from the cognitive reasons.

Re the style - it is a usage of particular syntax structures (which other authors from the comparison basket don't, or using it more often), particular vocabulary; in a more abstract way - particular topics, particular analogies, generalizations; particular length of sentences,  etc.: anything measurable, detectable. In order to create a "new style" in this domain, the generation should choose something that is different in some of the parameters, compared to the ones which are "known" by the one who wants to get surprised with a "new" style.

The same goes with visual art, it's about hatching directions, thickness of the lines, shapes of the patterns of the brushes, color palette, etc. etc. There's no magic, magic is in the eyes of the ones who do not understand what they see and can't imagine how or why it can be created or recreated.

**Tosh — Днес в 2:03:** I don't agree, as Will (this Will is not "will" with a small letter, but the one defined in S.'s works), either in Schopenahuer's works and in mine stems from the lowest level of the representations of the Universe, the tiniest piece of matter and Universe has Will. Also there's a general thing: whether a machine or anything has it or has anythng depends on the segmentation, which depends on the evaluator's choice.

As more and more researchers realize in more contexts, TOUM is there as well, from "Extended mind", distributed representation, distributed agency etc. - in cultural studies, ecological psychology etc. - the "center" or the "locus of control" of the agents in not only in their usually attributed physical bodies with the usual segmentation. The master Will is the Universe as a whole, "The Universe Computer" in my terms (or the "Master algorithm" as Pedro Domingos calls it).

Will in Schopenhauer's philsophy is a "Will to live" and all beings, entities are objectivation of the Will, and machines are part of it.

The will (small letter) at low level of anything is the will of the Universe, what follows, converging to P()=1, one of the early works of that philsophy is titled "On The Will in Nature". https://www.gutenberg.org/files/50966/50966-h/50966-h.htm#Pg215 In mine the humans or thinking machines are higher orders of physical laws. AFAIK Joscha Bach has expressed similar ideas. The Free Energy Principles/Active Inference school of thought which in general is repeating TOUM in many aspects also shares the same ideas, Max Ramstead has emphasized that their work is "the physics of the mind", Chris Fields, Michael Levin and their colleagues are proving the same, that there's a smooth transition, connection and mind (well, Will, or will) doesn't appear as magic, it was there or could be discovered or assumed even at the smallest of scales and in all kinds of substrates.

**Tosh — Днес в 2:17:** I agree that "will" (not spiritual one) is required for a proper

agent and autonomous machines, but it has to be present by default, thrusts for development etc. That was one criticism from 2009 works called "What's wrong with NLP", about the statistical NLP from the time which worked as a "tool" where you push the button and it produces output, but it was not an "engine" that can run on its own.

Some of these issues were addressed in the recent years and many of them can be solved even with current technology with multimodal models which are mentioned and with proper modalities. At the time proper embeddings were missing and the combination of embeddings in different modalities which allowed for the spectacular results in the text-to-image, text-to-speech, anything-to-anything models.

* **What's wrong with Natural Language Processing? Part 2. Static, Spec...** By Todor Arnaudov Independent Researcher - Twenkid Research Independent Filmmaker - Twenkid Studio ASIC Engineer - (as of March 2009) MS S... https://artificial-mind.blogspot.com/2009/03/whats-wrong-with-natural-language.html

What is "context-independent learning"? IMO humans are also unable to do that (it depends on the definition, we're always immersed in a context, i.e. coordinates in reference frames, history of states etc., fields of affordances etc.). Do you mean "generalization" (the room arc-...), i.e. not just matching of explicit chunks of particular datasets?

Regarding that the current mass-applied successful big-data methods for learning are not-granular enough, not efficient, "dumb" etc. – I agree. AI is not just what is running now.

There's an old paradox – humans used to blame computers or machines for:
1) "doing only what they were told to" (well, now they started not to do it, LOL, and they have to be "aligned"). There were criticisms that computers
2) "can't understand natural language" (or images or whatever)
However the ones who "voted" for this didn't connect them, because 1+2 = HUMANS don't understand language or images or whatever, they can't explain/tell the machines what to do, how to process etc., i.e. humans are "dumb" or incapable, not the machines.

**Atrony — Днес в 2:20:** Context independent learning is what ARC is all about. It's about teaching an AI how to learn. Humans even have trouble learning how to learn, with the psychological construct known as "far transfer" being the closest thing to evidence that a human has learned about learning in general from learning a particular context. Even within humans, far transfer is extremely small, we're talking 0.01 SD increase per intervention (6 week chess training, for example) at best.

**Atrony — Днес в 2:21:** It's fair to assume diminishing returns however, so the argument that AI can achieve broad generalization ability through pure far transfer is quite weak. Other than far transfer, there are core genetic knowledge priors that contribute to humans' intelligence*.

**…**

**Todor: Pereslegin/Sociosoft cites Lem's Summa Technologiae:**

"In 1963, in The Sum of Technologies, everything was predicted by Lem. He divided the development of AI into 3 levels:
▪ AI - can transform big data and therefore can play end games better.
▪ The artificial mind - can work with the new (Boris Lapin, "First step"). Aromamorphosis.
▪ Artificial consciousness - will be able to think about something else. But we don't know what consciousness is."

The second one is the "extrapolation" or partially thenew, the third part, in Russian "нечто иное", a special kind of novelty.

IMO the specific definition of third point of Hassabis is superficial and unconvincing. In the early 2010s on AGI list some talked about the "radical novelty" in this regard.

1. *"Out-of-the-box" thinking* is in fact out-of-**a**-***box*** that **the evaluator-observer sees/knows/considers/understand*s*. It could be "***inside the box***" for the inventor, “implementer”, creator; however that box can be *invisible* for *somebody else* or for the creator *at the moment of creation* – later she may discover, understand or realize its boundaries; also it could be a box that is *larger* than the one of the evaluator-observer. It is a leap in the eyes of the others, but for the inventor it could be a *direct and the most logical and natural next step* of his thought process. Then it falls in the following section:

2. Arthur Schopenhauer ones explained it like this: *"Talented people hit targets which the others can't hit, but they can see. The Genius hits targets that the other people - including the talented ones - **can't even see**."* They see the target later, when they slowly approach the target. The "ingenious" novelty in these occasions usually doesn't help the others, because they lack the "sensors" (including mind) to recognize the hit and when they have the sensors, it's already not so "Radical", it has gotten more and more obvious and an immediate continuation of the current state, because the path of transformations and connections that has to be made becomes shorter and shorter; ultimately something that in the past was unthinkable and "crazy" for somebody with their mind capacity now becomes obvious for them and finally even *"for dummies".* An idea related to that phenomenon is shortly addressed in the movie "*Hollow man"*, in a line describing the main character: geniuses can skip steps and go from 1 to 3 without passing through 2.[47]

Yet the degree is relative and genius, according to Schopenhauer, is superior than the "normals" in quantity, but not in quality.

On the other hand genius’ own thinking doesn't have to be "out of the box for *his own subjective experience* and objectively measured actions (e.g. systematic

[47] “”Matt: *My 5th grade teacher told me, that "Genius is the ability to go from A to D without having to go through B and C." Sebastian can do that, but for me, I gotta have the B and C.* https://www.imdb.com/title/tt0164052/quotes I remembered it as “1 to 3” – either from a translation in the Bulgarian VHS or my memory has compressed it like that. Български: „*Не мога да му хвана Б-то и В-то. Учителката ми в пети клас каза, че геният преминава от А до Г, без да е нужно да минава през Б и В. …“* (48:55 - ) „Човек без сянка (2000) BG Audio” https://www.filmi2k.com/hollow-man-chovek-bez-syanka-2000/

"extraordinary" actions, methods, procedures, techniques which he understands). For his own knowledge and skills and own view his creations fall inside **his** method, just his mind and brain are "unusual", compared to the "competition", having a larger prediction horizon, higher resolution causality-control units and virtual universes with a higher depth, higher speed and performance allowing to cover broader scope, more possibilitie; having access to more shortcuts between levels, domains, steps, milestones etc.

*Choosing an objective* is selection of goals or that's having a Will. It's true, but any causality-control unit can have it, not only for magical activities. "Meta-goals" are also goals while they are processed or generated. (See Sergey Savelyev about the variety of the architectures of the brains as of Broadmann areas and subcortical objects, which are said to be about 200?, as well as the overall size which lets particular "resources", and the data is from a very few brains which were carefully studied (it's taken years at the time), I think mostly in the first half of the 20th century with precise maps of the Broadmann areas of a few tenths of people or less. Some had neocortical areas which didn't exist at all in others (one Soviet poet), others varied many times - vision-related say 3-4 times, including subcortical. One that is connecting the neocortex and the anterior cingulate had a variation of *40* times. Someone who has a very small visual areas and too small Motor areas for fine coordinates etc. is supposed not to be as competent in visual arts than somebody with huge areas etc.)

On a "third hand" some discoveries, when put in social and historical context, are not based on superior brain, but on being in the right time at the right place and systematically following the method, exploring and exhausting the combinatons like "everybody else" from the school of thought would do.

"Quantitative accumulations lead to qualitative changes", a famous quote from the dialectical materialism, "emergence". However that can be just because the explorer didn't know what would work, didn't have a clue or had it wrongly, so he did a 1000 of experiments and the 1000th one worked (the electric lightbulb), so his creativity was exhaustive search just like a "dumb" machine (still it's not 9999999999 experiments, there was some direction etc., but the dumb machine also has heuristics). [The *data/the universe* which is explored has “heuristics” itself, “rails”, paths, constraints and the explorer has to follow them and compare, until finally discovers, reaches to its goal. +6.6.2024]

This also can and usually is "locally unpredictable" (the note of Tim in linkedin), however anything can be considered as such, these are too general notions which go for every search/explorative process with some local scope of the agent/"attention head"/focus. It depends on the definition of the range of the locality and how exactly it is measured, and of course a traversal process, with some "attention head" which had a limited scope (so limited predictabily, prediction horizon, "certainty range", range of adjacent known above certain threshold etc.), or for which the evaluator-observer doesn't have precise enough model, according to the evaluator criteria - so "they" are always unpredictable to some extent. "Everything" is either predictable and unpredictable, depending on the selected precision,

measurement, range, evaluator-observer etc. All the history of science and technology can be predicted or extrapolated in possible scenarios by simulating the affordances of the research process with some resolution.

So if you have a more advanced predictive process, or if you have shorter prediction horizon, you become "less creative" by these definitions - if you're smarter you become dumber. (There's a similar paradox with "consciousness" which suggest wrong prerequisites, in TOUM that's the "hypothesis about the deeper consciousness". Many processes which are usually deemed "subconscious" are not such, they are just deeper and faster consciousness, with deeper not meaning inaccessible,these processes do not lose connection with the consciousness/executive function/"top behavioral drive", whatever it is (i.e. no sharp boundary between "System 1 and System 2" which in the past used to be called ~ apodictic and discoursive knowledge). An example of that is musical improvisation. A better musician can improvise with a higher speed, precision, variety and appears to do it more "automatically" with little perceived efforts, while a worse musician can do it slower etc. with shorter planning horizon, more mistakes and inaccuracies and by more cognitive efforts; a mildly talented novice would struggle to play a few successive harmonious tones correctly by ear; then the non-musician can't do it at all even for one or two tones, either due to a lack of skill or dexterity or for a lack of ear. Yet, absurdly, the groups of the *less* and even *the least* capable persons are supposed to be *"the most conscious"* about music, more than the virtuoso, because *they have to think "consciously"* and deliberately about each movement. The same goes for chess and whatever other activities. In my opinion this is nonsense.

As of “creativity”, for example a simple absurd way for an agent to become more creative in its own perceptions is to have *shorter working memory* and *lower depth* of the remembered past states – ultimately this agent will evaluate every next step as *„novel"* and „original”[48].

**2. What is "entirely new concepts or frameworks**", what's "**entirely**" and can anything be such? "Nothing new under the sun". Atoms and molecules "haven't seen anything new for billions of years", except the radioactive ones*, but they may have remembered it from some more billions of years in the past, and they may know the process, so it's not a new "concept" for them. It's always measured by an evaluator-observer and its specifics determine the decision, not the process of creation itself/alone.

* +6.6.2024: Do the atoms and molecules *themselves* know that they were “new” when produced? An observer-evaluator which knows all could discover or thing about that.

* +4.10.2025: Define the “boxes” formally and technically See also the note from 20.3.2025 in “The first modern AI strategy…”, 2025, continuing “Man and Thinking Machine: …”, cited also in “Stack theory is yet another fork of Theory and Universe and Mind”, about the “egg shells”, the boundaries, the walls that have to be broken

[48] Recall the predictive framework, though. The causality-control units, the active inference agents and the machine learning models aim to predict the future and to self-predict.

with additional energy or reconstruction, reorientation and change, after the limits of expansion of the current phase of development reach the limit; these boundaries initially foster and protect the development and progress, but then turn into an obstacle.

**– "As aesthetic and classically good as human creations"**

**– First, this is subjective** and if it is measured by some social-grade measures, society is heavily suggestible and influenced by non-cognitively-aesthetcal trends. Part of the aesthetics is cognitive, part is sensual, part is based on power/distribution/social ranking ("propaganda"), part is random – some flows – part is a result of exhaustive search or cycling (one fashion ends, another begins, "something different", but the new/different overcomes the focus not because it's superior or a "higher aesthetics" – there couldn't be a proper measure for that, because human opinion is malleable and in regard to mass culture there's a strong element of "psychology of the crowds". One essay *on "Issues with Like-Dislike Voting in Web 2.0 and Social Media, and Various Defects in Social Ranking and Rating Systems"*, Todor Arnaudov 2012. Now I see that this work explores aspects similar to some of John Vaerveke's concept "Relevance Realization".

* https://artificial-mind.blogspot.com/2012/07/issues-with-like-dislike-voting-in-web.html

**– Second**, what about the classical aesthetics being exhausted already (as being classical), images are all sets of pixels and gradients, there are particular laws of the light/photorealism. This "game" could be seen as "over" [already] and its novelty is such because the viewers do not have big enough scope and understanding to understand that *"all images are the same"* or foreseeable within the landscape of possible combinations of pixels, gradients etc., whithin particular complexity (amounts of steps, bytes/codes in some compression scheme/minimum description length). [That is also to understand “the idea” of images, in Platonic sense; see A.Schopenhayer]

Here again, the more the incompetence or shorter prediction/modeling/"creative" scope of the observer/evaluator, the more creative the creation is for him. [It *looks* creative to him.]

Some specific art movements: as D.H. mentions classical aesthetics, in visual art, after the masters who reached high degree of realism and detail, that was the academic aesthetic. Then there impressionist, expressionists movements which made something “different”, however the authors could be either "bored"/no new patterns by their measurements, so they try "something different" with the tools they have (but this kind of novelty is just exploration and exhaustion of the space of possibilities - as any "non creative" search, a reduction in resolution, etc. ...) …

"Consciousness" in subjective-experience sense is not required to be verifiable IMO as it is not for humans also. Joscha Bach in his recent talk with K.Friston mentions his usage of "sentience", which csssdan be possessed by systems, even by a corporation, that maps to my classical "Consciousness of Reason" (Разумно Съзнание) unlike the "Spiritual ..." ("hard problem of consciousness"). That's in *Man and Thinking Machine, 2001*

However this is steps … Pereslegin calls these steps *"wildcards"*
* **Todor: +note 6.6.2024:** The “core genetic knowledge priors” vs their lack in the current working AI is a lack of initial structural complexity. The “AI” doesn’t have to be so much “*overgeneralized*” as the transformer model, repeated and scaled.
===
**1. In fact this is evident without grandiose sentences:** there are grades of intelligence everywhere, a baby/toddler compared to an adolescent or an adult - she is also someone who needs 15 or 30 years in order to discover what the older individual considers "obvious" and "not creative" and she has to pass through all the "stepping stones". From her subjective POV it all could be "creative" and different stages of development could be "out-of-the-box" for her. The decisions and the "output"/performance of someone who is an expert in a field are like "magic" (unpredictable and unmatchable, unrepeatable) to a novice or layman (unless the tasks allows copy-past). Yet if the novice is talented, after certain amount of work and search she may also appear at that level through a systematic "not creative" process.
=== ***(About game playing and creativity)***
**Tim:** *"Again I agree, but with the caveat that it's extrapolating on the "base structure" of things it knows. The number of primitives in the "base structure" it's missing is probably extremely large."*

I agree that "the base structure" of a human player should be different, however always there's one.
**Tim:** "*A human doesnt have to be smart in the way AlphaGo is, it just needs to be inventive (level 3) and it will always win. "*

I assume this inventiveness is just to do something "different", like a trick, which requires too much efforts or which hits unreachable location in the search space of the opponent in a game/competition/domain given some constraints: time, computation resources etc.

However in principle this is not different from doing exhaustive search and having "*luck*" in the location within that space, "serendipitousness" I guess as Ken Stanley talks about it, or a "*Wildcard*" as the futurist Pereslegin talks about stages of development in the civilization or so.

So it is like luck, chance, randomness, being at the right time and the right place, *[a conjecture, current situation etc., and not based on explicit proper specific robust special capabilities of the system]* it is similar to the notion of "free will" of the ones who are afraid of determinism or that they are part of the Universe/nature and actually just indications/displays of the processes of the Universe as everything else. They believe that their will is "Independent" from the rest, but don't realize that it means that then it is *random*, i.e*. it's not "theirs"* either, it again belongs to the “universe” (the one who tosses the coins) or they have to identify with the randomness, "The blind Will" if I can pun with Schopenhauer's term (used in another context).
→ That reminds of the 3 ways to predict the future 2002 ... (TOUM, “Писма между 18-годишния Тодор Арнаудов,…”, ТРИВ; “Letters between the 18-year old Todor Arnaudov and the philosopher Angel Grancharov”)

...

The machine has to explore a more constructive/modality space and then its *"brute force"** search will generate all games and all possible inventions much faster.

If I'm not mistaken the quote, a British cyperpunk movie, called "Hardware", 1990. I remember a quote from it, I guess the author was inspired by chess: *"Chief: Machines don't understand sacrifice - neither do morons."*
https://www.imdb.com/video/vi4037411097/?ref_=tt_vi_i_1

Similar to "machines do only what we tell them to do", in the same cultural computer discourse started by Ada, objected by Turing, it was very strong for the "AI/machine haters" in the early 2000s, and it is still here.

*** *"They were just demonstrating skill within a fixed domain."***
**Todor:** In a broad sense everything is within a fixed domain: the universe. You can't get out of it.
** **Todor: "The brute force" objection to AI; John McCarthy etc.** – in fact humans also search with a brute force – how many centuries were needed for creation of anything, how many neurons, how many individuals etc. The measure of "bruteness" is not precisely defined. McCarthy, while criticizing the brute force in the Chess engines (in ~1989? video interview), mentions also how some mathematicians from the past have invented one conception which moves the knowledge a step ahead, however they didn't figure out something else which follows from their best invention, and another century was needed for somebody else to make the next step. At a universal scale that means that *the whole universe* had to work, to "compute" for that long in order to prepare the conditions and to "produce" the rithg conditions and the required individua person or persons to create/discover the following step in the sequence of knowledge and technology.

If some solution appears and is measurably "too inefficient" that just suggests that it can be optimized, shortened etc., that makes it appear "unpleasant", but eventually the "computational irreducibility" hits sooner or later or there is a price that had to be paid: the new solution is faster, but is produced after 100 years of computations of the Universe.

Also some systems find or know the solutions with less steps, because they themselves are more complex, they already encompassed the solutions etc. and had them implemented in their representations. See an earlier discussion on the *degrees of "thinking out of the box"* [there are "boxes of various sizes"], elsewhere about the relativeness of "System 1 – System 2", immediate/apodictic vs sequential processing, the CPUs evolution – more powerful and advanced CPUs perform the same instructions or work in less clocks or time, however they are built with more logical elements/transistors, have bigger cache and memory sizes etc. and are conceived, invented, produced, started to exist later – so there was a bigger investment of Universe's computer's "clocks" and transforms for building them; and if we regard these more advanced processors as entities, their definition, explanation, understanding is "more complex", would require more processing, memory etc. in order to be represented, stored, explained to some "model mind/cognitive

system/machine/computer" etc. Their *"minimum message length",* the complexity of their descritpion grows in order to reduce the "messages" they process. Respectively, as the processing element, module gets simpler, with less memory, it cannot encompass in its own structure the complexity needed for recognizing more clever solutions, and the more clever solutions use to "already know" them, because they have been found by an earlier "less clever" method.

In brief, there's a trade-off and transfer of complexity between different implementations and representations, as with compression in general, and if we care about solving the problem, it doesn't matter if it appears as a "brute force".

* In the semiconductors specifically, the price of a semiconductor "fab" in the mid-seventies was a few million dollars, for CPUs such as 8080, 6800, 6502, Z80. A modern fab in 2024 with the latest technology may cost 20 billion of dollars.

# * Съзнание и Панпсихизъм
# Consciousness & Panpsychism

#panpsychism #consciousness #philosophy #metaphysics
#философия #метафизика #философия на живота

## Уводни бележки за съзнанието, преливането между живота и ума, панпсихизма и противоречащия на Принципа на свободната енергия *Принцип на максималната ентропия*

Отношението между съзнание и материя се смята за основен въпрос във философията, например в диалектическия материализъм. Кое е първо: „битието или съзнанието“? Или са заедно? Мислителите отговарят по един или друг начин, но какво е битие и какво е съзнание, как е определено, кой го определя, защо по този начин, правдиво ли е и защо, как се е стигнало до тези или онези изводи? И т.н.

Този раздел е свързан и с представените по-горе статии и записки към ученията на Карл Фристън, Майкъл Левин и колегите им и др.

– РСУ, РСВ, РСУВ:  разделителна способност на управление и възприятие (виж ТРИВ, Т.А., 2001-2004)
- УУ – управляващо устройство
– ТРИВ, ВиР – Теория на разума и Вселената, Вселена и Разум; „класически ТРИВ“: 2001-2004
– Мисъл на М.Левин, че човек се развива „От просто физика до Ум“. В школата FEP/AIF (двете се препращат) Максуел Рамстед в MLST: “*Физика* на ума“. – Също „*непрекъснатост* на ума и живота“ – “Mind Life Continuity Thesis”[1]. – Същото е в ТРИВ, „Теория на *Разума и Вселената*“, първоначално: „Схващане за всеобщата предопределеност“, „Вселената сметач“, “Вселена и Разум“; заглавието на лекцията от 2009 в ТУ: *„Вселена ~ Разум*“; йерархията от различни нива „управляващи устройства“, „въображаеми вселени“ и пр., за която се говори в нея – от най-малките частици, обособими във физиката, през молекули, макромолекули, клетки, тъкани, … организми, техните организации. А човекът е *висша форма на физични закони* и *въображаема вселена,* като вселените имат *физични закони* и са *предвидими*.
– Преди всички гореизброени*: Артур Шопенхауер (А.Ш.), началото на 19-ти век, *„Светът като воля и представа“*: Волята, също волята за живот, е източник и е общото за всички, тя е еднаква във всички явления и е

най-отчетлива при човека.
– Виж също Тодор Павлов в *„Теория на отражението“*, 1936, 1945, материята има свойства *сродни* с усещането, макар и не тъждествени.
**– Mind Life Continuity Thesis** – препатки към FEP/AIF, следователно и предхождащата я: ТРИВ/ВиР;
* Освен А.Ш. и други философи по-рано, но Шопенхауеровото бих определил като вече „*съвременно*“, систематично обяснено и последователно, а като теория на познанието: потвърдена в Общия ИИ.
(Превод по-долу)
– …***Representational*** cognitivist [content-full]… vs … ***“A non-cognitivist approach to the free energy principle*** *(Friston, 2009, Friston, 2013), by contrast, implies that mentality is ubiquitous. This is a strong continuity view on particular concepts of life (viz.* ***autopoiesis*** *and* ***adaptivity****) and mind (****basic*** *and* ***non-semantic****)*1. … All systems that* ***maintain their variables within a limited range of values can be understood as having some form of mentality or proto-mentality*** *given that the FEP casts any system that is able to maintain structural integrity in the face of a fluctuating environment as engaged in predicting its own future states. That is, retaining integrity rests upon processes the function of which is to maximize model evidence—i.e., these processes* ***exhibit self-evidencing dynamics****“ (Kirchhoff and Froese, 2017, p. 18).” *2* [bold – Tosh]
*– …Когнитивистко, основано на представяния: „Не когнитивистки подход към принципа за свободната енергия, от друга страна, предпоставя, че умът е вездесъщ. Това е силна гледна точка към непрекъснатост на определени понятия свързани с живота (като автопоезия и приспособимост) и ума (първично и несемантично)…:*

*Всички системи, които* ***поддържат стойностите на своите параметри в ограничен обхват****, могат да се възприемат като притежаващи* ***определена форма на ум, психика*** *(“mentality”) или „пред-ум“ (proto-mentality), имайки предвид, че според Принципа на свободната енергия (Фристън, 2009, 2013), всяка система, която е способна да поддържа структурна цялост, когато е изправена пред отклонения в средата, се възприема като участваща в* ***предсказването на собствените си бъдещи състояния****. Т.е. запазването на целостта се основава на процеси, чиято функция е да увеличават доказателствата за [достоверността] на предсказващия модел – т.е. тези процеси проявяват само-доказваща се динамика“ (Кирхов и Фрезе, 2017, с.18)”*
*1: Mind-life continuity: A qualitative study of conscious experience, Inês Hipólito, Jorge Martins, https://newdualism.org/papers/I.Hipolito/Hipolito-

PiBaMB2017.pdf
*2:* Where There is Life There is Mind: In Support of a Strong Life-Mind Continuity Thesis, Kirchhoff and Froese, *2017*
https://www.researchgate.net/publication/316190653_Where_There_is_Life_There_is_Mind_In_Support_of_a_Strong_Life-Mind_Continuity_Thesis
* See also the intro to the section Algorithmic Complexity #complexity and the references there.
* Виж също въведението към раздела за Алгоритмична сложност.
* Виж работата на българския учен Данко Георгиев – негова статия е тълкувана във връзка с Теория на разума и вселената около с.258 [по 31.10.2025]. [Danko]

**Тош: Разграничаването на системата от средата и „самосъздаването“ са спорни въпроси**: оценителят приема, че дадено *средоточие* във времето, пространството, свойствата, особеностите – организъм, същество, управляващо устройство (УУ), деец – предвижда и причинява *достатъчно сам**О***, достатъчно самостоятелно *собствените* си състояния – обяснението обаче може да се даде и като че *средата причинява* и *допуска* промените и състоянията в това средоточие, както и запазването му в допустимите граници – *за сметка* на промените в околността; средата най-малкото *позволява* тези явления или запазване в средоточието, което е по-малко по обем от нея*,* защото от всяка по-голяма околност може да се появи неизвестно и неуправляемо от по-малкото устройство = по-малката област, въздействие, което да е с по-голяма енергия, да промени условията извън границите на приспособимост на устройството и пр., и така да го *промени в нещо друго,* да извади параметрите, които то поддържа, извън допустимите им граници и/или да го унищожи напълно.

В по-дългосрочен план всеки организъм съществува в среда, „екологична ниша“ и е създаден *и от нея*, а не само от родителския организъм или предшественик, които също са произведени от средата. В крайна сметка средата, заедно с всичко в нея, е създадена от Вселената, или е съществувала като все още неизразени предпоставки, заложени във Вселената или частта от нея, които се приемат за среда – както беше разгледано на много места още от „*Човекът и Мислещата машина: …“,* Т.Арнаудов 2001.

Ограниченият обхват от стойности е относителен и може да се отчете за всяка координата и обхват – дори и взривна химическа реакция е с „постоянно“ състояние в рамките на част от секундата или дадена разделителна способност

„*Само*-създаването“ при автопоеза е спорно. Поддръжниците му се обосновават с това, че системата била сдвоена със средата, но имало

*„оперативна затвореност“* , още „организационна затвореност“ (operational closure, organizational closure) – по екологичните психолози Матурана и Варела (Maturana, Varela)[49]. Системата била оперативно затворена, или действено затворена тогава, когато *средата не можела да ѝ заповядва*, а възможните ѝ действия били определени от *собствената* ѝ структура и устройство *(“its operators are not instructed by its environment, but determined by its own structure and organization (Maturana & Varela, 1980)”)*, макар че затворената система можела да бъде разпределена в средата, която чрез своето състояние и съдържание да задейства определени *познавателни* действия, но при все това средата не можела да *принуди* система да приеме определен път на (физически?, а не познавателни) действия (course of action). Последното не е вярно, и е особено явно при резки и драматични промени дори и в най-простите параметри на средата, например ако изведнъж се изгуби кислорода, ако температурата спадне или системата влезе в твърде студена или пък в твърде топла среда и т.н., или пък ако организмът, или системата, се сблъскат с въздействия, които са по-силни от способностите им да се съпротивляват и това буквално ги разрушава, наранява, унищожава или в най-лекия случай: „гони“: организмът се опитва да „избяга“, за да оцелее. Разширено и намалено до най-малките размери на въздействия, всяка частица от „самостоятелно действащото“ и самосъздаващо се същество се опитва да „избяга“ от средата, която се опитва да я унищожи, и също така „бяга“ към онази, която си мисли, че ще я съхрани и предпази (Виж Принцип на Свободната енергия, Извод чрез Действие; ТРИВ; „Матрицата в матрицата е матрица в матрицата“[50], 2023).

Самосъздаващите се системи (автопоетични) били оперативно затворени, отнасяли се и се обръщали до себе си (self-referential). Такива са живите организми, които се „самолекуват, самовъзстановяват, самовъзпроизвеждат“, но дали по-точните думи не са, че *се лекуват, се възстановяват, се възпроизвеждат* и т.н., но *не само-лекуват, само-възстановяват и пр.* – защото те не могат да го извършат без средата и без необходимите ресурси от нея: ако няма храна, тялото са разрушава, също и ***саморазрушава,*** понеже започва да се **самоизяжда,** да разгражда части от себе си, за да оцеляват други, които се приемат за „по-важни“. Жизнените процеси включват едновременно изграждане и

[49] Виж началото на книгата със списъка с школи, „пророци“ и съвпадения и подобия с ТРИВ https://geography.ruhosting.nl/geography/index.php?title=Operational_closure
[50] Т.Арнаудов, Матрицата в "Матрицата" е матрица в матрицата, сп. „Свещеният сметач“, бр. 22, 4.2003 г., * https://eim.twenkid.com/old/eim22n/eim22/matrica.htm
* https://www.oocities.org/eimworld/eim22n/eim22/matrica.htm

разграждане: анаболизъм и катаболизъм, както мисловните операции са анализ и синтез: разделяне и съединяване.

***Средата** позволява* системата да се възстановява, лекува и възпроизвежда, затова по-убедителна система е *среда-организъм*, с определена прецизност, в определена подробно зададена мярка за разграничение и пр.

Една дървена маса или двигател на кола „не се възстановяват“, „дори и в благоприятна среда“, те само се износват, но на микро ниво и в определен обхват всъщност дори и нейните части, както и на всякакви предмети, също се „възстановяват“ или „самовъзстановяват“ и запазват: атоми, молекули се държат в определени състояния, докато по-голяма сила или неблагоприятни условия не ги променят или разрушат. Този пример изяснява, че обхватът на възстановяване и пресъздаване на живите организми е по-голям: и като степени на разделение на организацията, и като времепространствен: молекули, клетъчни органели, клетки, тъкани, органи, системи, организъм, семейство, общност и пр. Устойчивото, повтаряемо и предвидимо, обратимо взаимодействие, съвпадение, съгласуване, общуване между степените на организация покрива по-голямо пространство.

**Виж:**

* „Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?“, Тодор Арнаудов, 4.2025 г., „Мислещи машини 2025“/SIGI-2025
* Universe and Mind 6, Todor Arnaudov, 9.2025, SIGI-2025; Вселена и Разум 6
* **Основния том на Пророците** и текстовете за ПСЕ/ИЧД и свързаните с тях учени**,** другите бележки в раздел Съзнание и Панпсихизъм в Листове, тълкуването и бележките в приложения *Ирина (Йоша),* статията за „Самоипровизиращата памет“ на Майкъл Левин, преоткриваща ТРИВ с 20 години закъснение; и *разговора между Александър и Тодор Арнаудови* за идеи от ТРИВ, който достига и разглежда свързана с нея мисъл от повестта *„Милиард години до свършека на света“* на братя Стругацки в приложението „Фантастика. Футурология. Кибернетика. Развитие на човека“ #sf; #irina, #tosh1 #listove
* Теорията на Майкъл Бенет на разума, “Stack Theory” – виж „Stack theory is yet another Fork of Theory of Universe and Mind”, SIGI-2025.

*** Тош, допълнителна мисъл, 6.10.2025,** която заедно с горното може да се брои като допълнение към Вселена и Разум 6. #UnM6, Universe and Mind 6+

## Средата като жива и живите същества като неживи

Запазването на стойностите в по-ограничен обхват и търсенето на *постоянство* може да се разглежда и като нещо противоположно на присъщата на живота „динамичност“, промяна. Живите организми са в непрекъснато движение, за да запазят определени свои параметри постоянни. Имайки предвид и разсъжденията за участието на средата в поддържането на живота си представете, че „***животът*** е мъртъв и ***средата*** е жива.“ – обръщане на ролите.

Областите от времепространство, които се определят като съдържащи живот, живи същества, „живи вещества“, живи организми, управляващо-причиняващи устройства; както и други кибернетични системи, които се „самодоказват“, поддържат „хомеостаза“*, „самопредсказват“, „самосъздават“, „пишат лирична автопоезия“ – всъщност са ***„неживи“*** или „**по-неживи“,** а останалото времепространство извън тях е „***живото“***, или **„по-живото**“, защото в него се съдържат по-високи възможности *да влезе, да се появи или да се породи нещо по-различно от предходното, да се случи по-голяма и по-рязка промяна, да се сътвори нещо друго.* При променената перспектива това, което се смята за „неживо“, **се променя повече,** а онова което наричаме живо, което уж „расте, развива се, размножава се“, се стреми *да бъде постоянно в рамките си,* т.е. *да не се променя*, да не създава*, да не се развива, *да не бъде творец, да не разчупва черупката си.*

В някои от ранните постулати на Теория на Разума и Вселената – управляващите устройства се стремят да работят по закони, които да им бъдат известни и да правят онова, което предвиждат че ще се случи и пр. (преоткрито като „самозапазване“ и пр. от Фристън).

Ако под „живот“ започнем да разбираме „по-различно“, с по-голям *обхват* на възможни промени за по-малко време и пр., при които „*нещото*“ да си остава „*същото*“ в определени мерки[51], например продължава да бъде *пространство-памет,* което може да приема *всякакви възможни стойности*, без това да разрушава структурата и свойствата му, то тогава *неживото времепространство* става „по-живо“ от „живото“ – потенциално по-динамично. Например в един момент е вакуум, в следващия там влиза космически кораб – подобно на оперативната памет на компютър или междинна видео памет, „фреймбуфер“, които могат да се пренаписват всеки миг – тяхното съдържание се променя, но те си остават „памет“ или същата „памет“,

[51] Виж философския въпрос, разгледан в The Identity of Indiscernibles, Stanford Encyclopedia of Philosophy, 4.6.2025, https://plato.stanford.edu/entries/identity-indiscernible/

или част от същия „сметач“[52]. „

Вселената е по-дълготрайна от временните проявления на конкретните индивиди на живите същества – докато съществуват като организми, бидейки „задължени“ да запазват определено множество от свойства, параметри, настройки в определени тесни граници, което се смята за признак за ум, „регулатор“ и пр., те са по-ограничени и стават по-*„замразени структури“*, „застинали подвселени“, „черупки“; подвселените в самите тях са „подчерупки“, около които, от гледна точка на подчерупките, *могат* да се случват по-големи промени и т.н., част от които е дори и само преимнаването на жив организъм през съответното пространство.

## * Самозапазването и средоточието на по-действителното управляващо устройство

Ако принципът за самозапазване работи, при съмнения и двусмислия, то системата, която се запазва след дадени преобразувания и промени е онази, която е по-действителното управляващо-причиняващо устройство от други възможни в дадено времепространство.

Необходими са мерки за определяне на запазването, които обаче можем лесно да формулираме и илюстрираме с няколко примера:
1. Когато дадена държава – обикновено повече от една – се доведат до състояние на война, онези които страдат, загиват и биват унищожавани са „обикновените хора“, човешки същества, индивиди, а след войната държавите, или поне едната от тях, „оцелява“*, като често оцеляват и онези същества, индивиди, които се смятат за „управляващи“; а някои физически лица, както и юридически, подобряват благосъстоянието си, имуществото, положението си, въпреки или заради страданията, разрушенията и т.н., или казано просто – *забогатяват.*

Ако управляващите устройства (УУ), „одеалата на Марков“, агентите и пр. *строго* се стремят *да се самозапазват*, то доброволното воюване за конкретните *военни дейци, участници във военните действия* на бойното поле, е *безумие,* защото е пряк път към

[52] Виж за атомите на пространство в разбирането на Стивън Волфрам.
* Atoms of Space https://www.wolframphysics.org/technical-introduction/potential-relation-to-physics/the-structure-of-space/ * https://dexa.ai/s/haz5XaS3
* Stephen Wolfram — Productivity Systems, Richard Feynman Stories, Computational Thinking, and More https://youtu.be/Uw-a8kgE6Lo?t=2218 - *Въпросът е от какво е направено пространството? … Отделни атоми на съществуването“, „Атоми на пространството … свързани в огромна мрежа …„ …*

унищожение и *самоунищожение* с висока вероятност*,* следователно то не би трябвало да бъде мислима *доброволна* възможност за действие, ако войниците се самоуправляваха.

Т.е. или самозапазването *не е всеобщ стремеж,* или онова, което се „самозапазва“ не са войниците – те са оборотен материал, инструмент на друго устройство и друга власт и са загубили или индивидите, които ги въплъщават във видимото представяне на Вселената*, по начало не са имали своя самостоятелна воля или цели,* или те нямат необходимата власт и сила, за да се противопоставят на по-силните от тях „главнокомандващи“ и държавна машина. Те може би не са имали своя самостоятелна воля още преди да ги изпратят на фронта „да защитават родината“, да се „борят за човечеството срещу подчовеците и орките“, да „защитават демокрацията, човешките права, правата на ЛБГТИ или на традиционните ценности и семейство“, или „правото за прекарване на нефтен или газов тръбопровод“.

Друг пример са клетките и тъканите в тялото. Системите се стремят към самозапазване, но клетките в тялото имат различни периоди на „оборот“ и разпад. Някои се превъртат и загиват за дни при нормалните физиологични процеси, като клетките в червата или устната кухина. Белите кръвни клетки се самоунищожават, за да се борят с нашественици. При инфекции падат голям брой жертви и от „обикновените граждани“ – „убитите“ от нападението, заразени от вируси, „отровени“ от отпадните продукти на бактерии и пр.

Увредени, но „достатъчно изправни“ в *самоуправлението* си клетки се *самоунищожават*, за да не нанесат по-големи поражения (апоптоза). Други клетки, на кожата, съществуват като „себе си“ за месец или два, някои от тях загиват преждевременно като се порежем, изгорим, притъркаме си кожата. Трети съществуват през целият живот – нервните и мускулните клетки, и останалите „работници“, роби, войници се *жертват* за тяхното по-дълго оцеляване. И двата вида клетки „царе“ могат да се уголемяват като увеличават обема си, възможно е да се пораждат повече митохондрии, които да работят по-ефективно и да усвояват повече енергия – при трениране; невроните също могат да се натоварват и от това да се развиват – да пускат „пипала“ и „клони“ (аксони, дендрити), но ядрото им структурно остава същото; или *„по-същото“, без да се дели*, за разлика от епителни и други клетки, при които новите се получават след делене.

Животът на сперматозоидите е кратък, ако въобще може да се нарече живот – те са особени „буквачета“[53] или клетки и дали са

[53] Елементи, съставни части – юнашко наречие

пълноценни, защото съдържат само половината хромозоми на останалите клетки и не могат да се възпроизвеждат – живи ли са или неживи? Яйцеклетките се зараждат още в началото на развитието на плода в утробата и остават десетилетия в състояние на незавършено деление. Те също съдържат само половината хромозоми и съществуват или „живеят“ в недоразвито, „полузамразено“ състояние, като някои от тях узряват при овулация – кога точно започва или се брои животът им? Оплодената яйцеклетка съществува като „яйце“ от десетилетия по-рано, малко след зачеването на зародиша на майката, който е заченат от друга яйцеклетка, която също е живяла десетилетия в своята майка и т.н. Яйцеклетките съществуват докато не бъдат изхвърлени при овулация или загинат неупотребени в яйчниците, бъдат оплодени или организмът премине в менопауза, когато последните оцелели „индивиди“ също загиват, след като може да са оцелели 50 години в женското тяло в „замразено“ състояние, без да се развиват. Когато бъдат оплодени обаче те *продължават* ли да живеят или пък от тогава *започват* да живеят истински, защото стават „*пълноценни“ човешки клетки* – съдържащи пълният набор хромозоми, който обаче вече е друг, те също стават други. *„Преливащото яйце“, „съединяващата яйцеклетка“* е *безсмъртна* – извършва ограничен брой преходи, деления, преобразувания, когато зародишът на женския организъм произведе следващото поколение яйцеклетки.

Сперматозоидите и яйцеклетките могат да се съхраняват и в буквално замразено състояние за зачеване „в епруветка“, при температура от –196 °C. Биологичният материа се охлажда по начин, който не позволява да се образуват ледени кристали и на теория можело да се запази за неограничено време; на практика успешни оплождания били извършвани с яйцеклетки до около 14 години*[1] и сперматозоиди на 40 години. Живи или неживи са „зърната на човешкия живот“ докато са замразени?*[2]

Какво се случва при оплождането на късметлията сперматозоид, който е „оцелял“ сред стотина милиони загинали събратя и съперници? И той ли умира, като се врязва в яйцеклетката и престава да съществува като сперматозоид, или се „преражда“, или пък *тогава* се ражда за първи път – започва да живее като „истинска клетка“, като „неговото същност" става част от зиготата – оплодената яйцеклетка?

При оплождането важна роля играе химическият елемент *калций*, чрез т.нар. „калциеви трептения“, които включват процеса.

Дали не се оказва или може да „прочетем“ при изкривяване на тълкуването, че живите същества при полово размножаване произлизат

от "*неживи*", т.е. не важи само постулатът, че „живо произлиза само от живо“ – и дали по принцип живото „не произлиза и от неживо“, защото непрекъснато се нуждае от него като „среда“, която всъщност му дава всички необходими съставки, условия и задейства програмите за управление на половите клетки.

От друга страна, половите клетки създават и други „нередности“, защото са като паразити в тялото – при жените месечният цикъл затруднява физиологията, води до кръвозагуба и при недостиг на храна - до анемия; при недохранване цикълът и производството на семенна течност може да спре[54]. При бременност плодът изсмуква ресурси от тялото на майката, като отново при недохранване тя храни бебето със собствената си плът, калций от своите кости и зъби и т.н. След раждането, майките на бозайниците и птиците се отдават на малките, понякога загиват, за да ги защитят. Смъртността на родилките при човека до неотдавна е била висока, а риск има и днес; при раждането или скоро след това са умирали и много от децата. След раждането трябва да се положат огромни грижи и смъртността пак е била висока в следващата година или няколко години, дори и при най-добро старание.

Мъжките индивиди влизат по един или друг начин в битки със съперници, както се твърди, за да „*разпространят гените си“*, при което могат да загинат; *техните* гени ли се разпространяват обаче, дори и да победят? „Естествено“ – Y хромозомата и т.н… Това обаче е *само един* полов хромозом – ако се роди момиче, той отпада. Сперматозоидите носят само половината, а при оплождане се получава кръстоска и вероятни мутации. Те се „включват“ в друго тяло и са други. Ако искате, може да приемете, че гените са на двамата родители и да кажете, че приликата доказва, че са техни, да правдаете „*тяхността*“ и с историческата причинно-следствена връзка – сперматозоидът е излязъл от тук, влязъл е там, тези части са във времепространствената околност на мъжкия и женския организъм на родителите и т.н. Възможно е обаче да се оправдаят други връзки на съвпадения, които са с по-дълги пътища на съединяване и съвпадения и прилики на гени на молекулно ниво или външни признаци между  многобройни индивиди от даден вид, популация и дори междувидово, защото всички са сродни, но в различна степен.

В културата, саможертвата се възвеличава, издигат се паметници на загиналите „за свободата на страната“ и т.н. – това е обратното на „самозапазване“, то е *самоунищожение* за запазване на друга система, която цели да те управлява, владее и при нужда – да те употреби, погълне, унищожи, за да оцелее тя; господстващата система може да

[54] Това е сигурен знак за недохранване и за претрениране с недостатъчно хранене.

запази „спомен“ за теб или за твоя тип „герои“ като „меме“ или „троп“, чрез който да се опита да кара и да убеждава други управляващо-причиняващи устройства от подобна степен, вид, координати в йерархията или системата от въображаеми вселени, при нужда да бъдат по-податливи на управление по подобен начин.

На микро ниво в рамките на клетките „зад клетъчната мембрана“, също едни части са по-трайни, а други са по-обротни. Едни атоми са по-устойчиви или целта е да не се променят „самоволно“ – ДНК – други непрекъснато се извличат от средата като хранителни вещества или се изхвърлят след обработката. Цикличното производство на АТФ и превръщането му в енергия за дейността на клетката в метаболитните цикли е пример за „унищожаване“ на определени части за запазване на други чрез поддържането на метаболизма, синтеза на белтъци, полупропускливостта на клетъчната мембрана и пр.

Хомеостазата, поддържането на равновесие като цяло, е също взаимодействие от разпад и съединяване, катаболизъм и анаболизъм, анализ и синтез. Кое или кои процеси и в каква степен се забелязват или наблюдават като по-важни, водещи и т.н. се решава и зависи от оценителя.

Следователно *абсолютното* самозапазване е абстрактен стемеж към "самозапазване" при който се подбират определени особености, обхвати и точност, при които се наблюдава съответното свойство, но неговата възможност се обуславя от наличие на други особености, при които е налице по-скоро стремеж към *„самоунищожение“* или унищожение и промяна, и въпросът понякога е „кой кого“, кой ще надделее, като при борбата за оцеляване между организми. И тревопасното, и хищникът се „стремят да оцелеят“, но единият изяжда другия, т.е. само стремежът не дава решение на задачата – резултатът се получава при съчетаване и сблъскване на различните стремежи.

Например в рамките на един или два месеца, кожна клетка може да *„се стреми да се запази“,* но не винаги може да *успее* да се задържи и при „най-голямо старание“ – например ако човек се пореже точно на това място или се изгори от слънцето, т.е. ако *по-могъщите* от кожната клетка управляващо-причиняващи устройства, включително средата като такава система и съчетанието ѝ с организма, не доведат кожната клетка до положение или в условия, срещу които тя няма средства да се защити, предпази, „приспособи“, „самозапази“ и оцелее още някой час, ден, седмица или месец.

В рамките на 20 години за човек, мъж, може да се смята, че се „самозапазва“ и „самосвидетелства“ по „принцип на свободната енергия и

извод чрез действие“, но не щеш ли, от държавата му решават да го мобилизират и да го пратят да се бие за кауза, която не е негова, срещу други мъже, които са призовани по подобни причини. Той може да се съпротивлява, но тогава може да бъде наказан в затвор, където да умре от недохранване, глад, бой до смърт и други болести, или да бъде убит веднага като дезертьор – от собственото си „по-висше“ и най-вече *по-мощно, по-властно* управляващо устройство (УУ), което се стреми „да се самозапази“ и да прави онова, което съвпада с неговата воля – т.е. да запази собствената си представа и собствената структура, но *не задължително да запази частите си и тяхната структура.*

В края на Втората световна война, оцелелите „главнокомандващи“ изпращали на смърт 15-годишни момчета да се борят за „хилядолетния Райх“ и да удължат агонията му с още някой ден. Атентатори самоубийци са убеждавани и те самите се убеждават, че щели да се срещнат с харем от „девици в Рая“, след като извършат задачата, възложена им от УУ, на което те фактически са се превърнали в роби или *„човеци с дистанционно управление“* и са предали собствената воля, ако ги разглеждаме като живи същества, в чужди ръце, понеже стратегията им и желанията им водят до сигурно самоунищожение. От една страна за такова УУ може да се приеме съответната организация, било държавна или „терористична“ – в някои условия първите са изоморфни на вторите, но с по-развита и утвърдена пропагандна и „заблуждаваща“ машина, която убеждава подчинените УУ, че са „противоположното“ на втория тип; – от друга страна, в самото „камикадзе“, „атентат самоубиец“ или войник, който е убеден в каузата си и съзнава, че вероятно ще загине или целенасочено отива на смърт, се вгнездява, „програмира“ се съответната „програма“, предписание, глаголище, изгражда се „паразитно“ УУ, което като вирус потиска другите и води жертвата към „заветната цел“ да бъде изтрито за благото на друго УУ, като така то вече *„не принадлежи на себе си“*, както при „обсебване от демони“.

В някои случаи саможертвата или усилията за размножаване и „запазване на рода“ са част от по-общото "запазване на живота" като по-голяма „цялост“, но в тази форма той не е в индивидите, а е разпределена система от сили, чиито "органи" са полетата на възможности за действие; индивидите са като молекули, органели, токени, и те самите се стремят към самозапазване в съответно ограничена област от времепространство и възможности за действие, в „тактическото си пространство“, и според условията могат да се превръщат все повече в *преки* изпълнители на чуждата воля и чуждите цели, като „кукли на конци“ – колкото повече действията им са против

собствените, които обаче при „инсталиране на вирус“ също се подменят. Свободната воля и пространството на действие свършват там, където започват да противоречат на волята на по-силно УУ – както при наслагване на вълните на сигнали, функции, трептения, периодични или почтипериодични функции с различни честотни характеристики, обвиващи криви и енергия.

В един момент, при дадено разглеждане, дадена система или буквaче изглежда като агент, деец, управляващо-причиняващо устройство, чието бъдеще не може да се предскаже с желана разделителна способност на възприятие и управление (РСВУ), а в друг момент или при друго разглеждане с други параметри – същото се държи като послушно и просто изпъняващо устройство: без собствена воля и своеволни движения, напълно предвидимо с избраната РСВУ за оценителя и истинското УУ в момента, и извършващо точно каквото то му нареди – част от инструкциите на машинния език на дадено „ниско“ ниво на въображаема вселена.

Съвкупността от човеци може да се приеме като „параметри“ на обща „обучаваща се система“, подобно на невронните модели: „9-милиарден човешки планетарен интелект“. Тази тема е разгледана и в хумористичната, но всъщост сериозна фантастика.[55]

*[1] *Живи или неживи са „**зърната** на човешкия живот“ в това състояние?* –броим яйцеклетките и сперматозоидите за *зърна на живота*, но е необходим и майчиният организъм с всичките нужни и за неговото оцеляване условия през цялата 9-месечна бременност: женското тяло с всички органи и системи в достатъчно „изправно състояние“ или друга заместваща система с подобни свойства. Майчиният организъм или „изкуствена утроба“ включва и съответна обхващаща среда, с която майката човек или „робот“ взаимодейства, от която черпи хранителни вещества, топлинна енергия и „оборотни материали“ като кислород и вода, и които я поддържат. Живите същества *не произлизат само от половите клетки,* така както и организмите *не са кодирани само в ДНК-то си,* защото без всички „декодиращи устройства“ и съответно „прочитане“ и въплъщаване от край до край, ДНК е само молекула, която може да влезе в химична реакция или да си „съществува“, но *сама* тя не върши нищо.

Ако разглеждаме обема, енергията, сложността като брой съставни

[55] Дъглас Адамс, Пътеводител на галактическия стопаджия https://chitanka.info/text/186-pytevoditel-na-galakticheskija-stopadzhi

части, видове взаимотношения и т.н., бихме могли да приемем, че с по-голяма тежест в зараждането и оформянето на бъдещия организъм са не въпросните две полови клетки, а участието и сътрудничеството на майката; на другите организми и системи, взаимодействащи и сътрудничещи си с нея; на средата и Вселената. За

Както е разгледано в *„Нужни ли са смъртни изчислителни системи…“* , 2025 а още и в „Човекът и Мислещата Машина: …“, 2001 – оценителите избирателно се съсредоточават върху дадени особености или свойства, с които да докажат, че еди-кой си обект е по-сложен, по-„ефективен“, по-„оптимизиран“ и пр., като целенасочено или неосъзнато забравят или скриват останалото, което би обърнало оценката им наопаки.

* Дали “следващата“ версия на дадена държава е „*същата*“ – след промени в територията, политическото устройство, културата и други видове „рестартиране“ – е част от общия въпрос за „същността“, подобен на този с половите клетки и с „всичко“, разглеждан и на други места в *Пророците…*
В частност за държавите виж например статията за „незаконността“ на *всяка* минала държава според оценките на *следващата*, която се възцарява след преврат, падане под чужда власт, освобождение и др. в „Митът за демокрацията, или Свобода на словото, добросъвестност и обективност в политическото говорене за демократи и комунисти…“, Тодор Арнаудов, 2015, сп. „Разумир“:
* *„20: **Законността на Петата българска държава: 1989 - 2015+“***
https://razumir.twenkid.com/democratius.html#zakonnost5

* *„Времепространство“* по-горе не е използвано в релативистки смисъл, като в теорията на относителността. Между подвселените, подустройствата, частите в управляващо-причиняващите устройства има неизбежно закъснение, лаг, разминаване, измерено във времевите мерки на вътрешния оценител-наблюдател, но при движения със скорости близки до светлината и/или отдалечаващи се или приближаващи се части, системите, които се получават, са особени и засега не си ги представям като *обединени* както могат да са например органелите в клетка, клетките в тяло, електронните елементи в сметач и т.н. В такива условия подвселените ще „бягат“ една от друга и оценителят-наблюдател, който може да ги обхване, може да е Вселената-сметач или някой, който има особен достъп до паметта ѝ може да „бяга“ по-бързо от тези повселени в необходимите посоки.

* Szell AZ et al. Live births from frozen human semen stored for 40 years. J Assist Reprod Genet. 2013;30:743–4.
https://bacandrology.biomedcentral.com/articles/10.1186/s12610-024-00231-4
* Sara Stigliani et al., The storage time of cryopreserved human spermatozoa does not affect pathways involved in fertility, Basic and Clinical Andrology volume 34, Article number: 15 (2024)
* Valentina Casciani Ph.D et al., Oocyte and embryo cryopreservation in assisted reproductive technology: past achievements and current challenges., Fertility and Sterility,, Volume 120, Issue 3, Part 1, September 2023, Pages 506-520
https://www.sciencedirect.com/science/article/pii/S0015028223005939
* C.J. Quintans et al., Live birth of twins after IVF of oocytes that were cryopreserved almost 12 years before, Reprod Biomed Online, 25 (2012), pp. 600-602
* M.F. Urquizaet al., Successful live birth from oocytes after more than 14 years of cryopreservation, J Assist Reprod Genet, 31 (2014), pp. 1553-1555
* https://en.wikipedia.org/wiki/Cryopreservation
* https://en.wikipedia.org/wiki/Oocyte_cryopreservation
* https://www.cofertility.com/freeze-learn/egg-freezing-how-long-can-my-eggs-be-stored
Dr. Meera Shah, Egg Freezing: How Long Can My Eggs Be Stored? Яйцеклетките можело да се съхраняват до 55 години във Великобритания, 14.11.2024

## * Обзор и на теории и школи за съзнанието

### * Landscape of Consciousness, THE KUHN FOUNDATION[56]

https://loc.closertotruth.com/
Включва интерактивни диаграми и информация за популярност според търсене в Гугъл, цитиране и тежест по Google Scholar, основни учени, свързани с дадените области и др. * https://loc.closertotruth.com/interactive
* https://loc.closertotruth.com/consciousness-theories
Примери:
https://loc.closertotruth.com/materialism
https://loc.closertotruth.com/theory?subcategory=eliminative-illusionism
Дискусия в AGI List: https://agi.topicbox.com/groups/agi/T8f6c809749765377

**Тош:** Разликите между теориите често са по-малки отколкото се представят; някои от школите може да са съвсем подобни, производни на други основни, и „нароени“, с променени термини. Системите и управляващо-причиняващите устройства се опитват да се отделят и разграничат една от друга и да се изкарат по-особени, да се отличат и

[56] Този сайт ми беше попадал по-рано, но Джон Роуз – John Rose, – ми го припомни в AGI List.

разграничат, за да оправдаят съществуването си като единици и независими. Подобно е с политическите партии, футболните отбори, музикални изпълнители, поети, художници, писатели; държави; жени и мъже и др. *„Всички са уникални“.*[57]

Някои от теориите са разгледани в *Пророците ...*

Една от слабостите на много теории е, че им липсват „корени“ на основни понятия[58]. Пример: статията за **Теории на съзнанието от по-висок ред** (Higher-Order)[59]: онова, което правило даден *възприятие* съзнателно било присъствието на съпътстващо *познавателно състояние* за *възприятието,* т.е. *феноменалното съзнание*, субективното усещане, *„духовното усещане“ (Т.Арнаудов),* не било непосредственото чувсто на осъзнатост за усещането (awareness), а *по-високо ниво усещане* на тези *усещания* (sensing … of … sensations); този вид съзнанието било производно на *мисли от втори ред относно усещания* от първи-ред или *умствени състояния – представяне* на *представянето*, процес от две стъпки, две нива, двуслоен процес.

А какво са „умствени състояния“? Кои минават за такива и как се определят? Какво е „усещане“, кое е представяне и кое не е? Кой решава дали нещо и кое е еди-какво си, било усещане, представяне, „от втори ред“, трети ред, четвърта „колона“? Това се прави от оценител-наблюдател по обикновено неясно-опредлени или въобще неопределени начини, а дори и да са определени, откъде знаете ли можете да знаете, че Вселената е на същото мнение? Виж „Нужни ли са смъртни изчисления…“, 2025, „Вселена и Разум 6“, „Човекът и Мислещата машина…“, 2001, „Истината“, 2002 – диалозите между мислещата машина Емил и Дарчо и други нейни разсъждения за съзнанието и действителността.

Някои примерни по-малко или по-известни теории:
* https://loc.closertotruth.com/theory/deacon-s-symbolic-communication-human-consciousness

---

[57] Трудно е обаче да очакваме ТРИВ да бъде включена тук, дори и създателите на този обзор да узнаят за нея.

[58] Вероятно има повече определения в подробните описания от авторите, но разсъжденията тук и в споменатите творби оспорват „обективността“ на всякакви определения, ако се отнасят за субективни усещания и резки граници на такива „нива“, независимо от оценител-наблюдател, който също трябва да се определи и може да се получи безкраен порочен кръг. Друг интересен въпрос е ***защо*** съответните дадени особености, които се посочват от всяка от теориите води до възникване на съзнание, духовно усещане и пр. Те се забелязват при дадени обстоятелства, свързани с думата „съзнание“ и пр.

[59] https://loc.closertotruth.com/theory?subcategory=higher-order

* https://loc.closertotruth.com/theory/campbell-s-theory-of-everything - виж по-долу
* https://loc.closertotruth.com/theory/grossberg-s-adaptive-resonance-theory - предшественик на някои идеи, преоткрити в ТРИВ относно ума като разум и умствени способности, не „духовната част“ – съвпадението между състояние на високо и ниско ниво (match); виж бележки и библиография за Стивън Гросбърг в основния том.
* https://loc.closertotruth.com/theory/graziano-s-attention-schema-theory 2019 ... “Quick and dirty Internal model of attention…” – бърз груб вътрешен модел; да, предсказващо моделиране, кодиране и пр.; но какво е „модел“ и моделът „първокласен гражданин“ ли е във Вселената, кой решава кое е модел и кое не е? На ниско ниво „нищо не е нищо“, това са „просто процеси“. Някой или нещо с достатъчен обхват ги обобщава. (…)
* https://loc.closertotruth.com/theory/Functionalism (...)

@Вси: изслдв всчки: интерактивна диаграма, обх всчк, Обрб, сврж, (Об,р); сбр(обх(?=(тоер.1, теор.2))) ?=(ТРВ, всчк); извд(откъси) … Об$^{+}$ …
@Vsy: verbs, extract; follow the method from: TUM, UnM6, Is Mortal Computation and The Prophets … ?=, Cmp, Ask the questions, until bottom/top.

Виж също: *** Философия на природознанието, София 1966**

*** Consciousness science: where are we, where are we going, and what if we get there?** Axel Cleeremans,*Axel Cleeremans1,2*Liad Mudrik,,Liad Mudrik2,3,4Anil K. Seth,,Anil K. Seth2,5,6, 30.10.2025, https://www.frontiersin.org/journals/science/articles/10.3389/fsci.2025.1546279/full – solving consciousness", consciousness = (access, phenomenal)
Short announcement: https://erc.europa.eu/news-events/news/scientists-urgent-quest-explain-consciousness-ai-gathers-pace
*"minimal unifying model" of consciousness,* GWT, IIT, … (level of consciousness, contents of consciousness (contentful)), *"consciousness is a property associated with an entire organism (a creature) or system"*1*; (perceptual, self)-awareness; phenomenological (experiential) = (phenomenal, access) = ("*feels like"*2.1, what it does*2.2)*

minimal phenomenal experience … "consciously perceived and non-consciously perceived stimuli"; from isolation to collaboration – adversarial collaboration, constructing experiments together by schools of competing theories; *"people whose brain development and operation will have been affected by factors outside their control. .. mens rea ("guilty mind") and actus rea ("guilty act"), in which mens rea picks out the conscious intent to engage in particular conduct … "my brain made me do it"*

**Todor:** *1 The boundary and union of the entire "Creature, organism, system", unless there are "causal IDs" etc., is set by the evaluator-observer by his *mind,* cognition, criteria. **2* in TUM, "Man and Thinking Machine…", 2001 – spiritual sensation and the cognitive capabilities allowing intelligent behavior etc.

The paper mentions predictive processing theory, represented by FEP/AIF etc., which repeats core ideas from TUM, regarding the functional/operational part of the causality-control units."Their brain made **them** do it", but "you are not your brain" – then your brain will be punished. **Humans love to cheat.**

The neural correlates of consciousness, the areas of the brain which are "connected", associated etc. are ones related to the **manifestation** of consciousness, which include suffciently expressive "outputs" which are detectable, measurable, important (decided to be such) etc. for the evaluator-observer.

*** A Study of "Organizational Closure" and Autopoiesis**, Harsh 2019 https://harishsnotebook.wordpress.com/2019/07/21/a-study-of-organizational-closure-and-autopoiesis/

Виж коментара на Мартайн Фейнинг (Martijn Veening) и статиите му за друг принцип: за *увеличаване на ентропията*, който също поставя под съмнение убедителността на критериите за разделянето на организмите от средата, както и ПСЕ: Maximum Entropy Production Principle (MEPP)[60] – Принцип на производството на най-голямата ентропия. Според него живите организми са местни средоточия, произвеждащи най-висока ентропия; падането на листата на дърветата през есента, смъртта на организмите е рязко повишение на производството на ентропия от подсистемите на всички нива. Структурите в нелинейните системи се получват чрез процеси на насищане и изчерпване на пораждането на ентропия (saturation-exhaustion) до достигане на някаква граница[61]. https://entropometrics.com/docs/TDA.MEPP_LIFE.02_final.pdf

*** A systemic explanation of 'organic life'. Fixing the semantics and understanding entropy**, M. Veening, 8/2019

*** A statistical inference of the Maximum Entropy Production Principle**, M.Veening, 2/2021 https://www.preprints.org/manuscript/202103.0110/v1

*** Visualizing the taxonomy of entropic and anti-entropic aspects,** M. Veening , 3/2017

[60] Открих на 21.6.2024
[61] Подобни идеи се откриват и в „Зрим"; ГнП – виж в продълженията на тази книга.

*** Не е нужно моделите да са линейни и детерминистични в пълен смисъл, а обичайните невронни мрежи могат да се видят и като „сиви“ или „полупрозрачни“ кутии дори и както са**
Тодор Арнаудов, бележки вдъхновени от публикации на Мартайн Фейнинг

Виж в работата на M.Veening за науката „*иконофизика“ (econophysics)*: икономиката, разглеждана като ***физичен модел***; „**ентропология**“ („entropology) – наука за законите на ентропията; склонността на съвременните науки да използват бейсови вероятностни похвати като машинно обучение и да стигат до модели по опитен път с взаимодействие и последователни приближения с обратна връзка, „апостериорни“ данни, и по-малко да *разбират мисловно, логически,* и да извеждат от „априорни“ основания, чрез строги правила.

М.Вийнинг използва *„детерминистични и линейни, основани на предварителни допускания“,* но нито „линейни“, нито „детерминистични“ са задължителни, в значението на *пълна* прецизност с максимална разделителна способност в целевата въображаема вселена (тогава има пълно или „истинско управление“, но във Вселената на най-ниско равнище това е възможно *само за Вселената като цяло*: поне в определенията на Теория на Разума и Вселената).

„Старите“ научно-изследователски методи, или просто методи за познаване, или за познание, могат да са нелинейни и произволно сложни с „безброй“ условия и правила, включително вероятностни в някои от звената си, и *пак* да са основани на „първични и мисловни преобразувания“, а *не само* на „*скрит*“ вероятностен модел, напр. невронна мрежа, който да е „неразбираем“ и е в голяма степен или се смята за „черна кутия“; като невронните мрежи всъщност също може и да се разгледат така, че да *не са* черна кутия в повечето случаи (както винаги зависи *за кого,* какви са възможностите на оценителя-наблюдател*)*: те могат да „сива“ или „полупрозрачна“, защото при наличие на теглата и законите за изчисления, те могат да се изследват и да се донастройват, могат да се обучават по различен начин и да се проследява как и защо се променят; да се откриват кои „неврони“ се възбуждат при какви входни данни или състояние и те да се променят[62] и т.н.

Трудноста за обективно и убедително отделяне на частта от

[62] Конкретна програмна библиотека за тази цел е pyvene: https://github.com/stanfordnlp/pyvene
* A Library for Understanding and Improving PyTorch Models via Interventions, Zhengxuan Wu et al. 3/2024 https://arxiv.org/abs/2403.07809

цялото, системата от средата, „нещото“ от „фона“ само по себе си би могло да бъде мисловен довод към „вселенска психика“ или „панпсихизъм“.

Разбира се, винаги може да се отдели по някакъв принцип или начин за разделяне, но доколко е убедителен, че отделя онова, което *твърди.* Разделянето и съединяването, отделянето и свързването, са основни понятия и принципи в по-съвременните работи от ТРИВ: Зрим.

Образец за **сложно разделяне** е компютърното зрение, на английски „*сегментиране*“, известен модел от последните години е SAM: Segment Anything Model[63]. Оптичните илюзии също често са свързани с многозначно разделяне и съединяване. В по-общ теоретичен контекст, при трудно определими критерии и многозначност на делението, следващите решения се определят от предходните избори: от конкретните предишни разделяния, обособявания, групирания, образуване на гроздове (clustering). Предните избори насочват как се тълкува нататък и ограничават възможностите, което може да включва да се посочат обхватите, границите – контурите – допускания, размитост; сравнения с кое точно съвпада или не съвпада и защо (напр. еди-колко си процента от еди-коя си процедура; еди-колко си процента припокриване при налагане, при преобразуване и пр.). Така и за разделянето „живо/неживо“ за по-точно определение може би е нужно за конкретните случаи да се посочват по-точни конкретни критерии. Но когато мерките са „прекалено“ точни, тогава започват да прехвърлят типичните граници на човешката работна памет на оценителя, който търси „прости“ формули и правила от няколко знака и математически действия. Виж статията на Т.Арнаудов за „Окончателния ИИ“ и Принципа на Свободната енергия и „фетишизма“ към кратките математически формули.

Фините, точни, прецизни, последователни, систематични определения, основани на отделени по съответен начин понятия, са подходящи за извеждане от мислещи машини, работещи по „класическия“ мисловен понятиен начин. Тук биха ползвали „символен“, но спред мен *„Невронните мрежи също са символни“,* Т.А. 2019 (виж).

[63] Виж също DepthAnything и др. за извличане на предполагаемата дълбочина на пикселите от единични изображения като през стерео камера, „depth camera”; също SAM 2 и др.
* https://ai.meta.com/sam2 * https://github.com/DepthAnything/Depth-Anything-V2 , 2024

## * Бележки и разсъждения на Тодор Арнаудов вдъхновени от видеото с интервюта с Фредерико Фаджин* за идеализма, квантовата механика, свободната воля и самоличността

## Federico Faggin on Idealism, Quantum Mechanics, Free Will, and Identity | Mu 128 042 показвания 27.07.2023 г.[64]

https://www.youtube.com/watch?v=7HgXwbIMRsc

**Тодор Арнаудов (Тош)**

**Todor Arnaudov's Notes (Tosh):** 18.4.2024-20.4+…6.2024

*„Съзнанието е първично (fundamental)… Не може Вселената да е била несъзнателна 13.8 милиарда години и изведнъж при човека да стане съзнателна … "*

**Тош:** Един от проблемите на „материалистите редукционисти", но също *и на идеалистите* е, че независимо от *поредността*, било първичност или вторичност на съзнанието/преживяването, *само* от *нея* не можем да извлечем *структурна* информация и въобще определено ***вътрешно*** съдържание. Фаджин споменава за квантовите състояния, които не можели да се възпроизвеждат, така както и личното преживяване, което е несподелимо[65] и несъхранимо – частично възпроизводимо е чрез спомените, но от същия индивид в собствения му ум, но всъщност според мен не се знае дали е „*същото*" и може би няма как да се докаже, освен това в подробностите тялото, и умът, и състоянието на вселената вече са други. *„Същото-стта"* въобще във всеки случай е с РСВ/РСУ[66] и от там е условна и може би и класическите състояния и информация също не са възпроизведими в „пълен смисъл": човек „никога не влиза в една и съща река"; какъв е процесът на определяне на еднаквостта. Какво *точно* е свободна воля също подлежи на различни тълкувания, дали нещо/някой/въобще *съществува* се определя от определението ѝ, както и определението на *„съществуване"* – виж „екзистенциалните" проблеми на „илюзионистите", които се опитват да шокират публиката с „откритията" си, че „всичко е симулация"; „свободността" на волята зависи пряко и косвено от оценител, на *него* може да му *изглежда* – и

[64] 23.9.2025: 185590 показвания

[65] Според някои има начини за споделено преживяване, ако имат „споделен мозък", виж Luke Roelofs; напр. сиамски близнаци: такъв случай е с Криста и Татяна Хоган, р. 2006 г., при които таламусите били свързани: https://en.wikipedia.org/wiki/Krista_and_Tatiana_Hogan https://www.cbc.ca/cbcdocspov/features/the-hogan-twins-share-a-brain-and-see-out-of-each-others-eyes Дали са отделни същества или съзнания, или са едно? Дали ако двете пряко споделят преживяванията са „отделни" или са части на едно цяло?

[66] Разделителна способност на възприятието и управлението, виж ТРИВ/ВиР.

неустойчивостта на мерките за информация като намаляване на неопределеността – за *определен наблюдател-оценител* с определено състояние, знание, очаквания. Дали „нещо“, съобщение, поведение, бъдеще, състояние е *„достатъчно“* непредвидимо, „независимо“/самостоятелно: колко е достатъчно? Достатъчно *спрямо избрана* – достигната, фиксирана – граница, мярка. При какъв „*срез“* се прави оценката за еднаквост, за независимост (спрямо колко голям обхват); защо; спрямо какво и т.н. Виж също работата от школите на Michael Levin, Karl Friston – одеала на Марков и др. и цялата Теория на разума и Вселената, Вселена и Разум, от 2001-2004 нататък.

**0 м:** ФФ: Вярваме, че действителността е в тялото, а тя е в разбирането на съзнанието, което е значението на онтологията, което е смисълът на живота … „(физиците откриват, че) колкото повече знаем за Вселената, толкова по-безсмислена става тя… Някои вярват, че са по-умни от природата, Вселената. *Луди ли сте?*

**Тош:** Да, човеци и пр. подустройства, подвселени, подсистеми или явления на цялата система смятат (въобразяват си), че само с местните си състояние, изключени от цялото, са по-бързи, по-„умни“ и независими от системата, от която са частица.

37-40 м: Символите трябва да се създадат от **класическа информация… споделима, защото квантовата информация не може да се копира, не може да се възпроизвежда [както субективното, лично преживяване].** Теорема за „невъзможността да се създадат независими и еднакви копия на произволни квантови състояния*“.* https://en.wikipedia.org/wiki/No-cloning_theorem

Компютрите са класическа система от ключове, превключватели – отделният ключ не знае нищо за цялото, докато аз, клетките ми имат знание за целостта на тялото ми заради генома на яйцето, което е изградило тялото.

**Тош:** Привлекателен начин на мислене, но според мен заключенията са по-убедителни *не заради генома*, а заради други сили (защото клетки с едни и същи гени могат да бъдат произведени например чрез генетично инженерство или просто да са в различни организми, но да съвпадат – при някои видове разликите може да са много по-малки, и някои да се размножават като „клонинги“, така че те носят „същият“ геном“ *(разгледани като код, абстрактно)*, и изглеждат почти еднакво и пр., но те са различни „инстанции“, образци, разположени на различни места в „географското“ пространство; с различна, „достатъчно разделена“ причинно-следствена история, път на пораждане, път на съществуване, обхващаш всички техни частици или подпространства, които могат да се

опишат, измерят по някакъв начин: независимо от „истинската“ им природа: дали са вълни, частици – и да са едното, и да са другото, ако не са *част от преживяването* на „духовно“ ниво, в крайна сметка за оценителя са вид данни, измервания и абстракции и под тази форма се свеждат до някакъв съвместим „формат“, за да бъдат сравнени); Бернардо Каструп спори в други беседи, че *нямало частици,* това били *гребени на вълни на квантово поле*… Да попитаме обаче: а съществува ли всъщност и „квантово поле“? Какво е *то*? Не е ли пак измервания, числа, показания на някакви уреди – които също са съвкупност от *„гребени на квантови вълни“*, или формули – а те пък са математически абстракции, които, както всяка информация и всякакви данни *не значат нищо без тълкувател, оценител* – съответствие на нещо във „физическа“ вселена, представяне от по-ниско ниво и пр., сетивно-моторно обосноваване (grounding, sensorimotor grounding; grounding in general); и/или в ум , „съзнание“, оценител – който в крайна сметка на определена степен не знае и не може да узнае как е осъществен под някое достижимо от него ниво, или не може да го пресъздаде със същата РСУ/РСВ.)

Онова, или може би *нещо мислимо,* което свързва клетките и подклетъчните структури в тялото е особено **сътворение**, особена причинно-следствена верига; процеси, които пораждат, съотнасят, свързват всички тези части, различни от други в други обекти, области, системи.

ФФ твърди, че клетките на панкреаса може да открият, да научат нещо ново чрез генома си, защото са част от цялото тяло (а смята, че съзнанието е способността за самопознание), но само част от генома е изразена в състоянието на клетката в даден момент.

Обаче защо да отчитаме обектите *точно* или *само* в мащаба на клетка? В по-малък обсег частите също не могат да знаят, те губят дълбочина; и тук, както и в други случаи – виж „извод чрез действие“ или „съзнанието“ в скала/камък и др. неодушевени предмети, – онзи, който „познава“ или наблюдава и оценява каквото и да било, и чието „съзнание“ не се поставя под съмнение и затова можем да го приемем с висока сигурност, е *умът на оценителя-наблюдател*. Съзнанието, „Окото“; както и „ключът“ (част от схема в силициевия кристал – или пък от електронна схема с дискретни транзистори, електронни лампи, или пък механични превключватели), клетката, геномът и пр., *всички те* са „артефакти“, обекти, странични ефекти, изделия, следствия на *процес на извличане, разделяне, сегментиране* в ума на оценителя или най-общо: следствие на *преобразуване* (вид представа, „обективация на Волята“

във философията на А.Шопенхауер) и избраният *мащаб* не е строг и не е изведен по убедителна форма (в „извод чрез действие“, ИЧД, FEP/AIF това е проблемът с границата на „одеалата на Марков“, *особено* за „видимо по-динамични процеси“, например *пламък* – живите обаче са дори по-динамични в много повече видни, отчетливи, и в същото време припокриващи се и размити, мащаби, обхвати, координати, „площи“): защо системата в изчислителната техника или силициеви чипове да се разделя *точно* на ключове и какво точно е ключ, превключвател? Кога е и кога не е? Това е логическо, абстрактно понятие в ум, и в различен обхват могат да се обособят вложени един в друг ключове: схеми с различна РСУ/РСВ. Може да изберем подробности според дадена схема, план, проект, от който е произведен шаблон, с който е получен чипа и т.н., така както е удобно за ума на инженер и технолог с дадена налична технология, но дали *природата* работи със същите елементи и те дали са „истински“ самостоятелни обекти за „вселенския ум“? В по-малък физически мащаб, транзисторите или превключващите елементи не са рязко отделени в схемите или няма „достатъчно ясно разграничимо“ състояние на ключа, а също така за ***всяко*** разграничение е нужен *оценител,* „разграничител“, който да реши и да съхрани представата някъде. При крайно смаляване, когато транзисторите достигат размери на малък брой молекули, близки до нанометър, в тяхната времепространствена околност започват да се проявяват квантови ефекти като квантови тунели – „прескачане“ на електрони – и опасност от „неопределено“ поведение спрямо очакваното, при което „по-просто“ определената рязка местна двоичност, природата на рязко превключване на транзисторите, започва да се смущава и са необходими нови видове транзистори (Tunnel FET, TFET) с променен принцип на работа[67].

В същото време системата на по-ниско или по-„дребно“ ниво е построена от „същия материал“ или „геном“ на материята; същото вещество и локално има същите физически свойства като другите ключове, транзистори; особеностите на силициевия кристал също са *„сътворени заедно“* (процесът на производство, ецване) както и живите организми, ако се мислят като това че *клетките им* произхождат от оплодена яйцеклетка (виж логиката на Б.Каструп) – но дали „*съществено*то“, това което мислим като *атоми, електрони; протони, неутрони* не си е сътворено от по-рано и също не е било „заедно“ за всички?; в силициевия чип елементите се намират на „*малко* геометрично разстояние“, но колко е малко и колко е много и защо?; взаимодействат

[67] https://en.wikipedia.org/wiki/Quantum_tunnelling
TFET-транзисторите още не са внедрени в масово производство.

си многократно в кратко време; взаимосвързани са като начин на задействане – например тригер: ако единият е отпушен, другият се запушва, подобно на квантовото оплитане, а по същия начин съществуват всякакви зависимости в логиката и състоянията на схемите и т.н. В чипа тези взаимодействия са в по-малък времепространствен обхват отколкото между два чипа в дънна платка на компютър, или между няколко компютъра в мрежа и т.н., но от друг ъгъл бордовият сметач на сондата „Вояджър“ също е свързана с електрониката на Земята и се синхронизира с нея, макар и да е отложено във времето. А от друга страна клетките, които смятаме, че се създават „заедно“, или пък частите на отделната клетка се пораждат при деление „заедно“: при по-висока разделителна способност също не е „заедно“, защото има определена последователност; геномът изисква време за да се „прочете“, белтъчините се изграждат база по база. А клетките непрекъснато „всмукват“ атоми и молекули от „средата“. Тъканта на Вселената се влива в организма и става част от него, частиците се сменят и смяната не е едновременна, а се „търкаля“ и в рамките на тялото, или това което приемаме за „нас“, за сферата на организма, има частици от по-малък размерен мащаб, които принадлежат на организма, от различни времена: някои по-стари, други по-нови; някои вещества и съответни съставки се „превъртат“, „метаболизират“ по-бавно, други по-бързо, както и някои клетки и тъкани, в мащаба и РСВ и РСУ като такива, се обновяват по-бързо или по-бавно: епителни бързо, костите и хрущялите: бавно. Нервните клетки се променят и бързо, и „бавно“.

Може би, ако съществуват *„причинностни познаватели“* (идентификатори) и „тагове“, които организмът, системата, да „закача“ за новопостъпилите частици, елементи дори и такива обекти/устройства може да „*чувстват*“ *принадлежност* към едно и също *цяло* и цялото да бъде, от своя страна, „същество“, обект в друго ниво.

Наличието на квантово оплитане е вид „идентификатор“ от такъв тип, той показва конкретна свързаност между обекти, но предполагам, че съществуват по-сложни, съдържащи изрични минали състояния, памет, подобно на идеите от *Теория на сглобяването (Assembly Theory)* или *Многопътните системи (Multiway systems) .*

Относно мащабите, забележете например, че обикновено организмът се смята за ограничен в рамките на обвивката си от клетки, обикновено в по-голямата площ: „мъртви“, но със собствената ДНК – външният слой кожа или рогова тъкан, коса или козина, нокти, зъби. Но ние знаем, че те са със същата ДНК чрез сложни изследвания и логика, отделните клетки не си четат ДНК-то от разстояние, има други средства

на имунната система, с които се разпознават „свои“ и чужди, понякога погрешно, а една от ролите на кървно-мозъчната бариера е да предпазва мозъка от имунните клетки, защото за тях нервните клетки са чужди тела, които подлежат на унищожение.

Клетъчното или организмовото ограничение, тази РСВУ, обаче може да разширим с *междуорганизмово*, или с друго, по-широко ограничение, включващо определен обхват, разрез, от средата, „екологична ниша“ и *други организми* и взаимоотношенията и взаимодействията между тях. Например човек, който е тежко болен на легло не може да оцелее без помощ, в зависимост от състоянието си, не повече от минута, час, ден, дни и т.н. Болногледачите са *част от неговия организъм,* без тях той ще умре преждевременно. Тук се съсредоточаваме върху буквачетата *човеци*, но същото може да се приложи и върху неодушевената среда, която обуславя и позволява „живото“ състояние.

Също така, обикновено приемаме, че другите организми – човешки същества – не се самоунищожават един друг, но това не се подразбира: всеки от нас е жив не само заради работата на клетките си, а и защото *клетките на другите* и *организмите на другите*, или другите *държави* и пр. не решават да ни убият и унищожат, въпреки че потенциално могат да го направят. Ние се храним, благодарение на производителите на селскостопанска продукция, транспортната система, търговията, всичко организирано от държавни и международни отношения и т.н.

От тази гледна точка и в този мащаб обществото, човечеството, дори и когато е „разединено“, „атомизирано“ и т.н. е организъм, и оцеляването на човек като „клетка“ е свързано с пулсирането на „сърцето“ на държавата или общността.

…

**41: ..** Първични същества със свободна воля, самоличност и съзнание … (виж Chris Fields, Michael Levin, TOUM/ТРИВ, FEP/AIF)

**43:xx** Срив на вълновата функция … Вземане на решения, свободна воля… Съзнанието се нуждае от *класически състояния* … Непрекъснатостта изисква памет – символна … Дълготрайна памет … *Чистите квантови състояния* не могат да бъдат копирани или възпроизведени …

**Тош:** Обаче **могат ли и класическите състояния или системи също да бъдат действително и изцяло възпроизведени?** Първо, оценителят решава, че копираната *информация* е „същата“ или не е; той *решава, избира; че за него са важни* „битовете“ или *определени* особености и ако те се появят в *друго* представяне, компютър, памет, начин на изразяване:

тогава те биха били възприети като, или се приема че са „*същите“*; за да бъде подобно сравнение обективно е нужно еднаквостта, „*същото-стта*“ да е *фундаментално физично свойство*, да бъде *операция*, напълно определен обект във Вселената Сметач; както в програмните езици: „first class citizen”, вградени, пълноправни възможности и функции. Мярката за тази еднаквост, съвпадение и пр. обаче може да е условна, с определена разделителна способност на управление и възприятие, дори и когато повторението и съвпадението е до бит при *дадено представяне*, като това представяне е в нечий *ум,* ако искате „съзнание“, след като първо е преработено, филтрирано, възприето.

Също така в по-широк обхват на пространството на възможните копия или *на възпроизведената информация/данни/*“репродукциите“*,* може да преценим и че всъщност *всички* те са *извадени от оригиналния си контекст,* и техният причинностен път се променя спрямо началното, „оригиналът“ (например, стойността на клетките памет в даден компютър, дадени конкретни физически носители, на конкретно място в пространството с еди-какви си координати); стойностите на копията се закачат за други „нишки“, аналогично на информатиката като „дълбоки копия“ на съдържанието на обект, които заживяват „нов живот“ в други нишки или процеси в операционната система; и/или, връзката им с оригиналния контекст, времепространство и с другите „частици“, управляващо-причиняващи устройства в предишното времепространство, се отдалечава, разстоянието между тях в пространство на причинно-следствена свързаност или път на пораждане се увеличава.

Ако частиците, каквото и да са те: буквачета, елементи, агенти, подсистеми, подпространства – са онова, което поражда измеримите „отвътре“ в дадена въображаема вселена явления, и имат и памет за минали състояния, може би се запазва и друг вид връзка и след копиране. Обаче като „съвпадения“, копия, каквото и да бъде подобно „същество“, „обект“, „нещо“, то е зададено с определен обхват на срязване; „епсилон“ в математическия анализ, контурна област, „обхващ правоъгълник, област“ (bounding box, region; selection); с избрана РСВ/РСУ на оценителя, които по правило са по-ниски от тези на Вселената-майка (Виж ВиР), освен ако не е специална операция на въпросната Вселена на най-ниското възможно за нея ниво – например системата да копира област от паметта напълно точно, като операция memcpy(…) в Си.

(Около 1:20 – 1:21:xx ч. Ф.Ф. коментира нещо подобно за възпроизводството на копие на човек и пр., те се развиват вече отделно;

затова *самоличността* не може да се загуби след като веднъж е създадена.) Сравни с обясненията на Бернардо Каструп за това, че клетките не били отделни, защото са породени от една и съща яйцеклетка. [Bernardo Kastroup]
1:22-1:23 ч.: Според ФФ случаите на разкази за преживявания в състояния близки до смъртта (клинична смърт), при които пациентът се вижда отстрани, спомня си неща и пр., и  те са въздействащи, променят светогледа и пр. са доказателство, че съзнанието действа дори и когато тялото не работи, не протичат нервни импулси и пр.
**Тош:** Това мнение се споделя и другаде. Спомням си, че и преди много години съм давал следното обяснение: **не се знае** дали тези спомени са записани в периода на клиничната смърт, или се получават като състояния в събудения мозък, след възстановяването, като страничен ефект от загубата на захранване, частичен разпад и пр., като последващо обяснение на случилото се. (Има и други обяснения за отделянето на вещества от мозъка, които да намалят чувството за болка и действат подобно на успокояващи или наркотици; също „захранването“ намаляваща постепенно, нервната система има известни енергийни запаси, макар и незначителни; не съм чел (скоро) подробни изследвания в тази област, но според мен достоверни като за реални преживявания *„извън тялото“* – пациенти наблюдаващи се в операционната и т.н. биха били убедителни като действителни *наблюдения,* ако в спомените се срещат подробности, които са необичайни и проверими от останалите. Какво конкретно са си говорили други хора наоколо, предмети с текст по тях и т.н. – както в сънищата. В някои сънища, за мен лично от вълнуващ тип, има подобни подробности, от които се опитвам да съхраня колкото мога, за да прехвърля от „отвъдния свят“ в света наяве[68].
**1:09: …** Три вида информация: класическа, квантова и нещо по средата: *на живота: информацията, чрез която клетките общуват помежду си и със средата. Разликата е, че живата клетка използва частици, атоми и молекули, които си взаимодействат на квантово ниво, за да преживяват, да се делят, да създават нови клетки от себе си.*

[68] В този дух и с такива „пожелания“ са мои художествени произведения в различни жанрове от цикъла за Емил Юнаков, Ада и др. Например незавършената филмова фантазия „Подарък“ – „Сън на Тош“, с публикуван само кратък тийзър трейлър през 2009 г. Фоторазказите „Призрак“, 2006; „Альоша и аз“, 2006 и непосредственото му продължение: „Альоша и ти“, 4/2018 г., също с надзаглавия „Сън на Тош“.  Вторият е публикуван, но до 20.4.2024 е показван само на избрани зрители. Неговата „фантазирана“ история всъщност има много „мистични“ съвпадения и „четения на мисли“*, случили се в реална разходка през същите патеки. (*По ВиР4: нецелеви, за преживяващия ги разум, съвпадения с ниска вероятност, на **различни** равнища на Вселенския сметач [и в различни, според текущия оценител, подвселени]

*Компютрите не могат да се възпроизвеждат, те пренасят сигнали, но са статични: всички пътеки и ключове са вече потавени, те не се променят във времето; докато атомите и молекулите в клетките си взаимодействат, срещат напълно различни системи …*
**Тош:** Не знам за квантово ниво, но този начин на мислене е подобен на мисълта за Изводът чрез действие/Active Inference като „Жизнена сила“ (Vis Vitalis) и спекулациите във *„Вселена и Разум 6“* за причинностни/системни самоличности/идентификатори, чрез които частиците „познават“, че са част от по-голяма система и цяло, може би чрез свойства, памет, която засега недостижима за четене от обикновената физика. Относно установеността и неподвижността на електронните схеми в сметачите – тя е вярна като *химия и физика,* но ако електроните, зарядите и пр. се отчитат, логически е възможно да има подобни „полета на цялостност“. Около 45:xx ФФ обяснява за *динамизма* на квантовите системи, във връзка с невъзпроизводимостта, съответно *неподвижността, „статиката“* в схемите на машините (със забележката: при сегашните технологии; не пречи в бъдеще да се построят кръстоски: като се започне от *„киборзи“,* каквито ние и без мозъчни импланти сме чрез взаимодействието си с технологиите и без да са вживени плътно в плътта или физиологията ни.

Също така, клетките създават нови *от себе си*, но използват „чужди“ ресурси, които е нужно непрекъснато да се вливат в тях, и самите те изхвърлят други. Новото е такова от определена гледна точка: нови молекули, но съставени от „стари“ атоми. Променят се отношенията между атомите, електроните, но не и атомните ядра.

Един друг начин за общуване между клетките, по М.Левин е чрез биоелектричество, не само нервните клетки.

**1:12:** … За точката „Омега“[69] (Omega point, виж Bobby Azarian), точка на сходимост: според ФФ няма такава.
**English (a much less elaborate earlier draft):**
**0 m:** *We believe that the reality is in the body, reality is in the in* ***the comprehension of Consciousness*** *which is the meaning is in the* ***ontology*** *which is in the meaning of life …” “the more you know about Universe, the more pointless it becomes …* Some people believe that they are smarter than the Universe: *“Are you crazy?”*
**Tosh:** I agree regarding the scales. Humans etc. are sub-units, subuniverses, subsystems, or phenomena of the whole system; purpose, goal etc. is according to evaluators/observers, the same as chaos, entropy etc. In TOUM

[69] Сравни с константата омега от приложение „Алгоритмична вероятност“. #complexity

the Universe is constructed of causality-control units, all have their “goals”; each agent, a subunit in a factorization of a virtual universe, is “rational”, it aims to maximize/achieve/match its goal, but the observer may interpret wrongly and has an incorrect model of the observed; the observer may haven’t recognized correctly the actual agent etc.

**37-40 min: FF:***The symbols must be created of classical information … one that is shareable … because quantum information is non-clonable, non-reproducible [like the subjective experience: phenomenology]*

Non-cloning theorem, that it’s *“impossible to create independent and identical copy of an arbitrary unknown quantum state”.*

https://en.wikipedia.org/wiki/No-cloning_theorem

**FF:** *Computer – classical system … switches – the switch doesn’t know anything about the whole, while each of myself, my cells have knowledge of the totality of what my body is because they have genome of the egg that created the body*

**Tosh:** This is an attractive reasoning, IMO it’s more convincing not as same genome (because it could be produced by genetic engineering or be matching, say, by chance between very distantly located organisms – in some species there could be much lower genetic variety, and some effectively reproduces like “clones”, so they have the “same” genome, they look almost the same etc., but they are different instances [the Bulgarian version has extended reasoning]); what’s connecting the cells and the subcellular structure in a body is the specific **genesis,** the specific causal chains, processes which generate, relate, connect all these entities. F.G. argues that the pancreatic cells may discover, learn something from their genome which is of the whole body, but only a part of it is expressed in their instance. However why sampling at the level of a cell, at lower scale the parts can’t know, they lose that depth; here as in many other places (the “active inference” or “consciousness” or a rock etc.), the one who “knows” or cognize or observe-evaluate anything, with a certainty, is the mind of the evaluator-observer. The consciousness, the “Eye”… Also the switch, cell, genome etc. are all artifacts of the segmentation process and the scale is not strict: why dividing on switches and what is a switch – this is a logical abstract concept in a mind. On a smaller physical scale the transistor or switches are also made of the same material or “genome” of matter, the same matter, and have the same physical property as the other switches, transistors, they were “created together” in the fab, there is electrical communication between them – if there are “causal ids” and “tags” which can’t be detected yet, even such entities may have a “feeling” of belongings to the same whole. Quantum entanglement is a sort of such “id”, but I speculate about more elaborate ones with past states, like in Assemby theory or the Multiway

systems.
46 m. Consciousness can't be …
43:xx Collapse of the wave function … Decision … free will Consciousness need classical states … Continuity requires memory – symbolically … Long term memory … The pure quantum state can't be cloned/reproduced….

**Tosh:** However **can the classical states or systems also be really reproducible?** First the evaluator decides that copied information is the same, he cares about the "bits" or particular features and if they appear in another representation, computer, memory, expression: they would be "the same". However in a broader context the copies, or more precisely: **the reproductions** are *extracted of their original context*, and their causal path is now different, they are attached to other "threads". As matches, copies, they are such entities in the range of the cut, "epsilon" and chosen resolutions of causality-control of the evaluator.
* Having a copy of the DNA is not by default the same as to know. The way the generation processes are connected is more convinving.
(…) **Continue++ …** future work

**Todor's discussion continues in:**
1. Is Mortal Computation Required for the Creation of Universal Thinking Machines, T.Arnaudov, 17.4.2025 (Нужни ли са смъртни изчисления …)
2. Universe and Mind 6, T.Arnaudov, 2025
3. Comments on Michael Levin's Self-Improvising Memory: A Perspective on Memories as Agential, Dynamically Reinterpreting Cognitive Glue and Comparison to Theory of Universe and Mind, Todor Arnaudov, written in 1.2024/published in 2025 in appendix #sf #cyber: Science Fiction on AI. Futurology Cybernetics. Transhumanism/Human development/Cosmism..)
Etc.

* Federico Faggin was leader of the deisgn team of the first Intel microprocessors: 4004, 4040, 8008, 8080, and then a founder of the company Zilog; he designed/drew "2/3" of Z80 (according to the "bio" link below)
http://www.fagginfoundation.org | http://www.fagginfoundation.org/bio/

Федерико Фаджин, роден в Италия и още като юноша работил в компютърното производство на „Olivetti", е физик и инженер, ръководител на разработката на първите микропроцесори на Интел: 4004, 4040, 8008, 8080 и изобретател на важни полупроводникови технологии, дали предимство на „*Интел*". Напуска компанията поради разногласия и основава „*Zilog*", където е един от двамата основни дизайнери на микропроцесора Z80 в средата на 1970-те – компанията все още съществува и оригиналният модел се произвеждаше

до скоро през 2024 г., но осъвременените eZ80 продължават да са в производство. През 1986 г. Ф. съосновава Synoptics[70] с Карвър Мийд, компания която прилага *невронни мрежи* за създаването на първите сензорни дъски за преносими компютри (touchpad), заместващи мишката и тракбара; компанията работи до днес.

В писмо до "AGI List" от 24.10.2024 Стивън Гросбърг (Stephan Grossberg), един от *пророците на мислещите машини* и пионер на компютърните невронни мрежи споменава, че *„Carver Mead and Federico Faggin began many years ago to embody various of our models in low energy VLSI chips…."* – Карвър Мийд и Ф.Фаджин са започнали да вграждат разнообразни техни модели, от теорията за приспособяващия резонанс (ART, Adaptive Resonance Theory), но не са имали достатъчно финансиране и време, за да ги довършат.

К.Мийд е с ключова роля в развитието на интегралните схеми със свръхвисока степен на интеграция (СГИС, VLSI), започващо в края на 1970-те, когато проектирането на ръка, което се е правило дотогава, става все по-неадекватно с усложняването на схемите (Motorola MC68000: 68000 транзистора, 1979 г.; 8086: 29000, Z80: 8500); той е съавтор на важен учебник по полупроводникова техника**: "*Introduction to VLSI Systems* (1980)", който се изучава десетилетия по-късно по целия свят.

В писмото С.Г. критикува нобеловата награда по физика, връчена на Хинтън и Хопфийлд, като посочва предшестващи открития и публикации, свои и чужди. Сравни с по-младия *пророк* – Юрген Шмидхубер и неговите оспорвания на наградата „Тюринг" за Хинтън, Бенджио и Льокан.

* https://agi.topicbox.com/groups/agi/Tbd69a4c5580eb654/comp-neuro-re-some-scientific-history-that-i-experienced-relevant-to-the-recent-nobel-prizes-to-hopfield-and-hinton | * http://www.fagginfoundation.org |
* http://www.fagginfoundation.org/bio/ (През 2024 г. сайтът се отваряше, но през 2025 г. изисква парола.)

*** Интервю с Фаджин от 1995 г. в архив на Станфорд:**
https://embed.stanford.edu/iframe?url=https%3A%2F%2Fpurl.stanford.edu%2Fgr768wf7969&_v=1728312903
*** Zilog Oral History Panel on the Founding of the Company and the Development of the Z80 Microprocessor**
https://archive.computerhistory.org/resources/text/Oral_History/Zilog_Z80/102658073.05.01.pdf
* Бел. 24.10.2024: Бил Менш (Bill Mensch), р. 1945 г., друг легендарен инженер на ранните микропроцесори, също се увлича по метафизика и изследвания на съзнанието. Той играе ключова роля в проектирането на 6800 и 65816 (65C816, 16-битовото продължение на 6502, използвано в AppleIIGS и SNES) и е част от колектива от четири души, проектирал 6502 (3200-3500 транзистора; в Apple I,

[70] https://en.wikipedia.org/wiki/Synaptics https://www.synaptics.com/

Apple II, Commodore 64, Правец-82, 8М, … Правец-8Д, NES, Atari-2600, …). Една от неговите задачи е била декодера на инструкциите, който е високо оптимизиран и не използва микрокод, заради което процесорът е по-прост от други 8-битови вършачета от същото време.
https://en.wikipedia.org/wiki/Bill_Mensch

* Сравни с Боян Янков, виж в началния списък с български учени в първия том.
* За по-задълбочен отговор и анализ продължи с @Vsy:
* **Hard Problem and Free Will: an information-theoretical approach**, Giacomo Mauro D'Ariano and Federico Faggin, 28.1.2021
https://www.academia.edu/109620601/Hard_Problem_and_Free_Will_An_Information_Theoretical_Approach

* **Hard Problem and Free Will: An Information-Theoretical Approach**, Giacomo Mauro D'Ariano and Federico Faggin, 2022, Chapter in a book: R. Penrose et al.,
* **Artificial Intelligence Versus Natural Intelligence**
https://www.academia.edu/109620601/Hard_Problem_and_Free_Will_An_Information_Theoretical_Approach

* https://independent.academia.edu/FedericoFaggin

* New Research Shows Decision-Making Needs Both Quantum and Classical Worlds Research Matt Swayne, 25.10.2025, https://thequantuminsider.com/2025/10/25/new-research-shows-decision-making-needs-both-quantum-and-classical-worlds/
*"Future quantum AIs might therefore function more like biological agents: oscillating between coherent exploration and classical consolidation."*

*** Agency cannot be a purely quantum phenomenon**, Emily C. Adlam, Kelvin J. McQueen, and Mordecai Waeg, 15.10.2025
https://arxiv.org/pdf/2510.13247 – *"create a world-model, use it to evaluate the likely consequences of alternative actions, reliably perform the action that maximizes expected utility .. the first two conflict with the **no-cloning** theorem; the third fails due to the linearity of quantum dynamics. .. quantum computers cannot straightforwardly simulate agential behavior without significant classical components."*
– Чисто квантовите компютри не можели да изпълнят трите необходими критерии за агент/деец (управляващо-причиняващо устройство в Теория на разума и Вселената): да създаде модел на света, който да бъде използван за предвиждане на вероятното бъдеще и от него да се избере действие, което да постигне най-голяма очаквана полза. Изпълнението на тези функции са в противоречие с теоремата за невъзможността за възпроизвеждане на квантовите състояния и линейността на квантовата динамика.

*** Causal potency of consciousness in the physical world** by Danko D. Georgiev, 26.6.2023, Institute for Advanced Study, Varna, Bulgaria
https://arxiv.org/pdf/2306.14707

* Todor Arnaudov's review and interpretations of concepts from:
**Causal potency of consciousness in the physical world** by Danko D. Georgiev [Theory of quantum information consciousness, free will; mind-body problem, biophysics …]

**Todor Arnaudov:** The Bulgarian interdisciplinary researcher Danko Georgiev has a record of two decades of publications on related topics and may be a good starting point and a source for exploration of this field.

Some final conclusions of the *quantum reductionism* are related to some of TUM interpretations with analogical conclusion, however with different grounding, definitions and core interpretations about free will and consciousness. D.Georgiev's approach is highly technical and professional in the quantum physics "mechanics", a field where he has published many papers. However, I argue that all the technicalities, theorems and formulas are not necessary for qualitative reasoning, and sometimes they are misleading, because they suggest mathematical rigour, while their initial seed presmises might be not justified on the *other* "metaphysical" ground, which the research is addressing; the nature of these formulaic decisions as *a choice, one of many possibilities*, is sometimes not discussed and the chosen formulas are presented as *objective* axioms – see also the chapter about Wolpert theorems and the confusion in the premises.

**Conscousness experience and its loss**

The necortex or particular areas of it are declared to be linked with the phenomenon of "conscious" or more specifically consciousness *experience*", however note the following possible *methodological error*.

After damage in particular brain areas or a widespread destruction, including subcortical areas, the *behaviors, reactions of stimuli* etc. that the subjects exhibit *do not match* the criteria which are declared, defined, chosen, selected etc. as ones displaying the presence of "awareness, consciousness" or conscious *experience.* Signals, measurements, outputs, data which are *objective f*or the *external, another* evaluator-observer are taken as criteria for the existence, presence, activity etc. of a *subjective* phenomenon which is supposed to be personal.

In brief, this is not the *consciousness* or conscious experience, but *manifestations, which are usually connected with experiencing consciousness* by a considered "similar" evaluator-observer (another human with particular neurological, physiological, etc. features).

If the subject who is experiencing consciousness cannot express it the way

which the evaluator-observer expects, the evaluator will classify it as “unconscious, unaware” etc. In this interpretation the evaluator-observer who decides about the presence of somebody else’s supposedly *personal, subjective* experience.

**Todor [29.10.2025]:** Causal potency is similar to “causal power” and being a CCU in TUM. See e.g. “Power overrides intelligence”, 2025.
p.35: Georgiev argues, that *“we always observe only one of the two possible alternative readings of the measurement device, but never the entangled superposition of both.”* – this is the same as your macroscopic or any-scale measuring device for your CCU has *too low a resolution of perception*, that may be manifested as the evaluator incapability to reliably classify the current state in the proper class, say with a machine learning classifier. The evaluator – or classifier – discovers that it receives different results when she believes the situation is “the same”, or from a “ground-truth” observer it makes mistakes on individual measurements – and it has to perform many attempts in order to compensate for the errors – a form of Monte Carlo method, smoothing, interpolation, averaging etc. For example, early methods for image superresolution used multiple low resolution and low quality blurred images of a moving camera in order to produce a higher quality composite image with higher sharpness. Analogically a single frame can’t capture *“the entangled state of all frames”* within the sequence, the motion blur in a single frame of a motion picture turns into smooth motion when watched in sequence: the momentary image exhibits “quantum indeterminacy” and is like a single particle “entangled” state; on the other hand the replayed bigger whole, which completes the picture with the dimension of time and multiple shots, and the viewer’s processing capacity and memory to map and correlate the frames to each other – they all together “disentangle” part of the “entanglement” and decode information which can’t be recovered in a single frame.

The problem of low resolution in the context of “The Liar’s Paradox”, related to Tarsky and Goedel’s treatments of “truth” in axiomatic systems-formal languages, and their inappropriately low resolution and ill-posed nature in more realistic multi-domain, multi-scale, multi-… complex representations and interpretations, is addressed in TUM in **“Universe and Mind 4”, 2004,** Section #21; the segment is quoted in its Bulgarian original in this appendix “Listove” of “The Prophets of the Thinking Machines…” in the section about Wolpert’s theorems which rediscover postulates from TUM about the uniqueness of the universal predictor in the Universe, the impossibility of the sub-agents, sub-universes, virtual CCUs in TUM to predict with the highest possible RCCP at the lowest level of machine language of the mother Universe where they are implemented from constituent parts – sub-universe, sub-machines, CCUs. (The insufficient representational power is related to Ashby’s Law of Requisite Variety.) Example 12. ()

**p.35-36: Example 12. (**Schrëodinger's cat**)** – a cat in a box; a radiocative element with a 50% chance of decay in one hour; a Geiger counter, which could release a dose of poisonous gas in one hour with a chance of 50%. Therefor the cat would exist in superposition – simultaneously death and alive. Why this doesn't happen like this in the macroscopic world?
**D.Georgiev's interpretation:***"The conflict between quantum mechanics and the lack of observable macroscopic superpositions is solved by the energy threshold E for wave function collapse. If the measuring device is sufficiently large to pass the energy threshold E , the quantum entangled state (48) undergoes stochastic disentanglement Γ to one of the two separable outcomes …*
**Todor:** [Schroendinger=Schrëodinger] Schroendinger
Schrëodinger's cat experiment is confused, because it is not just about "quantum" waves or "macroscopic vs microscopic", but about inapprpriate RCCP – resolution of causality of control and perception; also "coarse-graining"; too high a level of abstraction than the required one. Cats in the macroscopic world actually can be in such "superposition" *if* the state is reduced to a *binary* classification, because for the real organisms or systems with many parts and cells, it is not binary and not instantaneous at the *low level* of representation, low level CCU, virtual universes etc. , unless the whole unit is "deleted", annihilated. Yes, macroscopic objects have many particles and features, and that fact doesn't cause only *"quantum wave function collapse"*, but these details have to be *specified.* The *scale* as size and number of the "active" parts, CCUs, is one thing, but there is also the *detail*, RP or RCCP.
When aggregating and reducing the detail, there are losses in the precision and accuracy, therefore if the evaluator-observer still *insists* to apply her *lower resolution of evaluation-observation-perception,* her classification would be ambiguous or sometimes wrong for the *higher precision classification* and it will *"exist in two states simultaneously"* – the actual state or representation of *enough precision* is something else, which the compressed/reduced one can't capture. That is what causes the indeterminacy, "paradoxes" etc. See the reasoning in the previous section above and the quote from *Universe and Mind 4, section #21.*

The "Example 13" or the ***"Quantum reductive model of consciousness"*** is similar to the idea about the integral of infinitesimal selves, proposed in [Arnaudov, 2004;2012], however expressed with quantum physics notation of elementary particles and *"energy quanta from the environment that input sensory information from the surrounding world."* . In TUM it is about CCUs, they could be of any type and at any scale: multi-scale, multi-range, multi-precision- … In Georgiev's interpretation: *"the composite system |Ψ(t1)⟩ is a collection of elementary minds."* The adjacent components interact, form entangled clusters which grow larger until they reach a threshold which reaches the energy level for wave function collapse into a single unified conscious experience

|Ψ(t2)⟩ and conscious self. … Then in Georgiev's theory *"the conscious self has to make a choice and selects one of the available disentangled outcomes using its free will."* .

Arnaudov's concepts about "free will" are different though. While one of the first names of TUM was "Conception about the Universal Predetermination", that does not strictly mean "lack of *free* will" in other definitions: attribution to particular slices, localities, views, sets of CCUs. The ultimate Will, intentionallity, causality force is always of the Universe, and when an agent follows his own will he is "free" in this twisted sense – if the ultimate driver is assumed to be "free" . Note the definition of Engels, stemming from Hegel, which interprets freedom as acting with *understanding* of the laws of nature. In TUM – the CCUs improve their prediction capacities. [Arnaudov, 2014, Note#267]*. See the definitions of free will in the reviewed works by D.Wolpert: the prediction device shouldn't be restricted in its choice of questions by the choice of another device, i.e. the agent has to be *independent* from the decisions of other agents. However, as argued in the review, this requirement seems impossible in a strict sense at the highest possible resolution in a Universe like ours. (See on p.54 in the edition of Listove from 6.11.2025)

(…) **Todor:** I challenge the claim at p.37: *"the primary purpose of art is to elicit conscious experiences in the viewer"*.

Any expression causes experiences in the evaluator-observer who experiences them, without being "art" – or anything can be classified as art. See my discussions in many pieces from TUM, for example in Calculus of Art I: Music I, 2012/2025, in "Wha a Man needs? If you play by the rules you will lose like the fools!", 2014 and in the classic treatises until 2004.

D.Georgiev: "*…the conscious experiences of the prehistoric painters have been causally potent in producing the physical artifacts in the form of cave paintings"* – I challenge the formulation; it is not the *experiences* which are causally potent, or having causal power, it is the causality-control units. The experiences may be phenomena or "sensations" of these CCUs and could be part of the drives, causes, reasons; linked to, connected, mapped etc. to the actions, transformations, outputs etc. p.37:"…panexperienalism …elementary feelings attributed to all quantum systems, both simple living organisms and inanimate quantum materials, *"with the explicit caveat that these are memoryless"*.

**Todor:** In my opinion the lack of memory of the smaller systems is questionable. Even the simplest particles from our current observation-evaluations could have other memory, which we cannot access yet or it is impossible to be accessed from "the outside" – the qualia, subjective experience, sentience, consciousness could be part of that special memory and communication system between the causality-control units.

Saying “particles” I understand that “quantum” point of view may attack me with “they are wave functions”. Wave functions are such in your measurements and terminology, using particular methods.

…

***D.Georgiev: p.2. "1.1. The sense of agency …*** *We are sentient beings that possess an inner psychological world or a stream of conscious experiences, which we simply refer to as a mind. (…)”*

Are we? This implies this is an objective postulate, while it is not as it is subjective. A problematic definition from which many similar explanations begin with or imply, ignoring or *escaping* the metaphysical and mysterious part, which unfortunately may not be solvable by any formula of quantum or whatever equations, which exist in abstract representations in a mind (which we don’t know what it is) of an evaluator-observer[71].

t

A thinking machine, a computer or a caption on the wall also may say the same and blame the humans or "you", any particular person, lack sentience, "because ...". See "Man ant Thinking Machine...", 2001 and "The Truth", 2002; see also Todor’s letter to Jordan Zlatev in “Stack Theory is yet another…”, 2025 about the LLMs. It is cited also in *Listove* in the end of the section for Consciousness and Panpsychism: goto *#todor-to-zlatev .. Sentience or consciousness of another “entity” is in the eyes of the evaluator”* … [including the boundaries of the entity – as well]

*D.G. " exactly because we are conscious agents with causative potency, it is possible for civil law to establish blameworthiness for actions that are considered wrongdoings and ethics can hold us morally responsible for what ensues from our behavior."*

The higher power CCU does punish ot try to stop or prevent unwanted actions of the others or the less powerful CCUs, because the “wrongdoers” have causal *power* (or potency), but *not because* “they are *conscious*” – if a dog or an animal attacks you – you don’t have to believe it had consciousness, – or if a robot or an inanimate object threatens you, you will “punish” it, if you believe that this action would stop or prevent the “attacker” to do what its undesireable intentions are pushing you to experience. In human societies the social part of the blame includes other aspects: revenge and sometimes legalized sadism etc. and it serves more the feelings of the evaluator-observers, not the criminal, including one which is blamed to be a criminal. For some crimes there is more consensus between the other agents – CCUs, possessor of causal power or “causal potency” – while other crimes are disobedience or considered threat to the higher power CCU and the most important “moral” value is that a CCU is “misaligned”. That’s general “error” in TUM: difference between wanted and measured.

In general: one CCU or a system of CCUs, which also can be considered as a

[71] See the discussions on illusionism, in Listove and in Universe and Mind 6.

bigger one, acts in a way to counteract or force a particular other CCUs, causal forces, powers etc. in order to "align" them with its own desires, goals, will.

"The spiritual" part *doesn't matter* for such behaviors, unless they exist by default in a panpsychist framework. TUM implies panpsychism, but it doesn't treat brain or the neocortex or a particular small area of it as the "seat of consciousness". The cortext is part of the body and the Universe and its existence and specific operation depends on the rest.

*p.3 "Our conscious minds do exist, therefore they are real and have to be defined as "physical"17, 18 …"*

The manifestations of the minds, as we interpret both the manifestations and the minds, exist for sure, and there is an implication that therefore what produces the expression has also to exist, and the only way to exist is "physical", the same way as the manifestations. The strict link between existing and "real" may vary, also the meaning of *physical.* In TUM, as discussed in the early works such as "The Truth", 2002 and cuurent like "Universe and Mind 6", 2025, the virtual unvierses do exist and they count as real, and the computer-generated universes have roots, grounding and substrate which is mapped to some ultimately lowest level CCUs, laws of physics at some "lowest-level" universe. However the *mind* as *sentience, experience, sensation* might be something else that requires particular dependencies and relations and causal chains between particular CCUs – for example a character in a video game "exists" and the virtual world of the game exist, but wheter the content of the computer memory and the states of the CPU and GPU registers in given moments correspond to experiences corresponding to ones of other types of entities that exist – that's another issue. The *mindification (to mindify),* part of the imagination, can attribute minds to anything – the animism; one mind sees, selects particular "things" as having a mind, and in that moment and period, during that "coupling", connection, the objects exist in the mind of these "mindifier" as having their minds which share theirs as media.

* TUM is close to panexperientalism and panpsychism.)

*"Classical functionalism … The reductive claim expressed by the mind–brain identity, $\Phi = \Psi$ .. p.13 .. the initial state of the mind $\Psi(0)$ affects the future state of the brain $\Phi(t)$. The brain is part of the world, which means that the mind affects the world."*

If mind = brain then there's no need to use different terms or concepts.
Another thing is that the brain is *a function of the world*, not just that the brain affects it - they are part of one system. The world (the Universe) *defines the brain and the mind,* and the world is bigger, it existed before the brain etc., therefore it shapes what the brain and the mind is – following this logic.

*“W.James: "mind dust" in nature”*
*"all aspects of physical reality are observable, hence classical physics needs to be repaired in some form or another"*

Observable in principle, in practice the CCU have limitations and are only virtual. The mind or the subjective states also can be assumed as observable - by particular interpretations – as sentience and consciousness are practically defined, by particular behaviors, observable states, correspondences etc. – by particular manifestations, which are classified as attributes of sentience by the evaluator-observers.

However the factual observability is *virtual* in classical physics and in TUM – Universe Computer – as well – in theory you, a virtual CCU, a subuniverse, could read the values of the memory everywhere in the universe, some states of whatever, but in practice you cannot, you, whatever “you are” at the high leves, cannot access at high resolution even the memory which constructs your own physical body. Only the Universe Computer is assumed to be able to do this and some “God evaluator-observer” who could even pause, modify etc. wherever she likes, similarly to a human editing a game engine or a simulation.

p.14 [in quantum physics] *"a fundamental dichotomy between “existence” and “observability” because what exists is different from what can be observed. 4, 17, 18 "*

However then you can't know what really exists. What is to exist?

*"quantum states are vectors |Ψ⟩ in Hilbert space H, whereas quantum observables Aˆ are operators on the Hilbert space H. 93, 94, 98–100"*

A Quantum state of a spin particle ... vector, spin

The type of the particle is also part of the state, its relations to others and its history, or "causal history" and causal IDs (see Arnaudov, 2023, 2025) also might be important. The possibility of other unknown internal states is speculated also in mid 20-th century, cited in a footnote about dialectical materialism in "Stack theory is yet another...", 2025.

*p.15 "Quantum indeterminism; stochastic differential equations (SDE);"* compare ODE - Ordinary differential equations...

Stochastic dynamics – each simulation run produces different results, thus a large sample of stochastic trajectories have to be averaged in order to produce a less biased and stable aggregated/averaged one (Todor: for the evaluator-observer and at a lower resolution.)

p.16, p.17 *"bias; nonuniform distribution;… the behavior of quantum systems depends critically on the measurement context. For some quantum measurement contexts, the resulting dynamic trajectory" may appear to be either random or deterministic in*

*different experiments. That is why quantum physics is indispensable for the proper understanding of consciousness and free will."*

I agree that the sensation and evaluation of randomness are evaluator-dependent – according to TUM there is no real randomness, but that goes for “macro” events as well. The same phenomena described in the quote can be seen for macro phenomena if the conditions are complex enough. One difference of the quantum and the usually seen as macro is the precision. If the experiments in the macro world are also addressed with a precision, which the measurement instruments or methods lack or can’t reproduce in different experiments, it also sometimes would appear random and in other times: deterministic. The same will happen also depending on the evaluator-observer and their own experience, cognitive capacity, available data, performance etc. – an evaluator who understands and can figure out the causality, may see it as deterministic, while another one who lacks the required cognitive capacity to “disentangle” the causal factors to a given “sufficient” degree, will perceive it as random*.

Therefore I don’t think there is strict logical consequence in the conclusion of the citation.

* A related thought is addressed in another context, regarding emotion and reasoning in a note №269 to “Какво му трябва на човек? Играеш ли по правилата ще загубиш играта!“, 2014. https://razumir.twenkid.com/kakvomu_notes.html#269

*p.18 "unitary quantum interaction between the brain and its environment"*
Is the brain sharply segmented from the environment in the quantum world? How the waves exactly split?
*p.19 " Therefore, it seems that we are not knowingly choosing our brain states."*
Yes. See TUM: hihger levels – lower resolution of RCCP, “by effects”. The question what is “we” is also open – there’s no single unified self, but an integral of infinitesimal selves (Arnaudov 2004, 2012).

The paper possibly agrees with that, expressed with quantum wave functions.
*p.19 "4.2. Quantum reductionism guarantees causally potent consciousness and free will "*
The definition of free will as possibility to choose between at least two outcomes, to decide, is not convincing.
An external or any evaluator-observer who lacks sufficiently precise, RCCP, information about the subunvierse, virtual unvierse, cannot know whether this agent, entity, thing, process really had these possiblities, or the way it operates is “as if…”. The agents = the CCUs, wills are all entangled and dependent on each other.

p.38 D.G.:*"mental causation and free will .. the two problems are distinct but hierarchically organized as follows: (1) causally impotent mind cannot cause any course of action in the physical world, (2) causally potent mind that lacks free will can cause exactly one course of action in the physical world, and (3) causally potent mind*

*that possesses free will can choose between two or more possible courses of action in the physical world.*

What the verb "can" means in this context? How can you know that the mind of a given subject "could choose another". You play only once, always one option is selected. The fact that you "could" select something else if you were different.

*p.4. "1.3. Evolutionary theory mandates causally potent consciousness .. The evolution of consciousness and development of culture in primates, including chimpanzees, bonobos, … is an established empirical fact that entails the causal potency of consciousness in the physical world. "*

Mind and consciousness are causally potent, but *what* are they, what their aspects and sides are considered. Regarding the evolution, I'd say causally potent CCUs and consciousness as intelligent behavior with memory, sufficient "memory persistence"etc. – behavioral consciousness as self-knowledge, self-reflection etc., derived from the agent's behavior by another evaluator-observer or by itself. In brief: "the easy problem", however not strictly - spiritual consciousness*, based on *empirical proves*, as this is a metaphysical issue.

For example automata for whose operation is "assumed" to be too different for them to experience human-like consciousness or these entities to exist as such "legal entities, could be designed to incrementally develop, progress in their cognitive capacity, invent and enhance their own technology, production facilities etc. They do have "causal potency" = causal power, however do they have consciousness?

Somebody may disagree whether consciousness or its subjectivity is a metaphysical concept, but it it is considered as such, it is not *"an established empirical fact"*, it is *phenomenal* and personal. You cannot know about the "consciousness" even of your parents or children right *now*. You can know about their manifestations, behavior, "data" and derive a mental model, in your mind for which you have phenomenal evidence ("*internally empirical"*).

You can *assume* that the other beings are like you. You can perceive them, either passively, without interacting and communicating, or via explicit interaction and communication. You may conclude that they should have feelings, sentience, consciousness like you, because they are similar to you, as this was argued in the TUM since the early 2000s, also perhaps because you connect your thoughts with them – expectations, predictions, sensations; and misprediction – and the models of these other entities become *parts of you*, too, they are *mindified.* You "feel" them, therefore they have your qualities.

However, this is not empirical, or not in the general sense, it includes

phenomenal component and is specultive and an assumptions. In fact, this could be extended for other situations which are considered objective with no doubts.

If the subjective is treated as objective, it's not subjective anymore and this destroys the original problem and makes the need of more proves about causal potency of mind or consciousness, or that it exists etc. irrelevant. If subjective=objective and objective is part of the default processes and nature of the Universe: causal, powerful, physycal – observable, epistemologically accessible – it can be sensed, measured, discovered etc. – then the "case is closed" from the start.

*p.7 "If the recorded electric activity from the brain cortex is forwarded to computer program that controls a robotic arm, the conscious mind is able to train itself after several months of practice to control the robotic arm without actually moving any of the body muscles. Thus, our conscious minds appear to be causally potent agents within the physical world, because if they were not, conscious control of brain–machine interfaces would not have been possible. [4, 73]"*

That is wrong logic. The attribute "conscious" is misused and why using the muscles doesn't imply the same.

This is a wide-spread confusion: a BCI is control "with the thought-only", but if you use your muscles and limbs, that's not a control with the thought. Moving your body is also "with your thought", there's no need for additional hardware to prove the "mind causal potency", and mind is a metaphysical and virtual concept in the mind of another evaluator-observer - one could be observing her own body. "Mind" is not a "physical" and sharply delineated entity, object, material body which "starts here and ends there, at these metric coordinates".
(…)

**@Vsy, @Todor:** Continue in another more elaborate work. Address also the other publications and concepts by Danko Georgiev.

**References and see also:**
* Neural correlates of consciousness:
https://en.wikipedia.org/wiki/Neural_correlates_of_consciousness

**Todor [5.11.2025]: Will "mechaically" non-standard-biologically atomically replicated-or-constructed beings** experience sentience, "phenomenal consciousness" and will it be "similar" or the same to the "natural" organisms? What if a model organism, a system, that is considered, "decided" to be conscious, say a human or an animal, **is reconstructed** by an atomic or molecular replicator, a construction machine which however doesn't pass through the usual biological causal and geneological chain from an egg cell, zygote, embryo, cell divisions etc., or at least it is "constructed" differently at least in some point of this process, say the egg cell or the zygote are "engineered" by a "**biological replicator".** Will a

"sufficiently similar" chemical, physical, geometrical, spatial structures and relations of atoms, molecules and processes, at scales and precisions which are detectable, observable, measurable by the internal evaluator-observers and manage to "start liiving", sustain "sufficiently similar" physiological processes and metabolism – exhibit the same "neural correlates of consciousness or whatever the correlates are, will this system viewed as agent experience and "have" "similar" sentience and consciousness?

*** Danko D. Georgiev's** papers:
https://scholar.google.com/citations?user=J13tBDUAAAAJ&hl=en
https://www.mu-varna.bg/EN/AboutUs/Pharmacy/Pages/danko-dimchev.aspx
* T.Arnaudov, **Universe and Mind 6,** 9.2025, https://twenkid.com/agi/Universe-and-Mind-6_22-9-2025.pdf
* T.Arnaudov, Is Mortal Computation Required for the Creation of Thinking Machines?", 17.4.2025 (in Bulgarian)
* Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?, https://twenkid.com/agi/Arnaudov-Is-Mortal-Computation-Required-For-Thinking-Machines17-4-2025.pdf
* T.Arnaudov, 2004, "Analysis of the meaning of a sentence, based on the knowledge base of an operational thinking machine. Reflections about the meaning and the Artificial Intelligence", Todor Arnaudov, 18.3.2004 (in Bulgarian; translated in English in 1/2010):
https://web.archive.org/web/20040402125725/http://bgit.net/?id=65395
http://artificial-mind.blogspot.com/2008/02/2004.html (some remarks)
https://artificialmind.blogspot.com/2010/01/semantic-analysis-of-sentence.html
* T.Arnaudov, 2012: Nature or Nurture … Integral of infinitesimal selves …
http://artificial-mind.blogspot.com/2012/11/nature-or-nurture-socialization-social.html
*"(…) No Intrinsic Integral Self, but an Integral of Infinitesimal Local Selves …"*
* T.Arnaudov, 2014, "What Man Needs? If you play by the rules you will lose like the fools! Part I"; Razumir #1. Main apendix with notes:
https://razumir.twenkid.com/kakvomu.html
https://razumir.twenkid.com/kakvomu_notes.html#267 "(267) "Анти-Дюринг", Фридрих Енгелс - изд. на БКП, 1970 г. 267.1 Глава IX. ... Свобода и необходимост ... с.149; "Freedom and necessity, p. 149 in "Anti-During", Fridrich Engels, (translation with Google Translate API: "*Necessity is blind only insofar as it is not understood. Freedom does not consist in the imaginary independence from the laws of nature, but in the knowledge of these laws and in the possibility contained in this knowledge to systematically make the laws of nature act for the achievement of certain goals. ... freedom of the will is nothing but the ability to make decisions in the knowledge of things. ... and, on the contrary, uncertainty based on ignorance, which seemingly arbitrarily chooses among the many different and mutually contradictory possibilities for a solution, precisely thereby proves its lack of freedom, its subordination to that subject over which it should dominate ... etc."*
Therefore, in this interpretation, the *local* will of a virtual CCU, smaller than the

Universe, with development to more advanced and “autonomous” system, the agent becomes more connected and mapped to the core of the Universe Computer, allowing it to exist. See Free Energy Principle/Active Inference addressed in Listove and the whole *Prophets of the Thinking Machines.*

**Future reading and analysis:**

***** @Vsy, Todor: Review D.Georgiev’s scholar page etc.

*** Neuronic system inside neurons: molecular biology and biophysics of neuronal microtubules,** Danko D. Georgiev, Stelios N. Papaioanou, James F. Glazebrook, 31.12.2004 https://journals.mu-varna.bg/index.php/bmr/article/viewFile/103/103

* Todor Arnaudov’s interpretation of ideas from
**Quantum information theoretic approach to the mind–brain problem**, Danko D. Georgiev, 2020, https://arxiv.org/pdf/2012.07836
” p.2 *“The seat of the human mind is the brain cortex. Cortical electric activity is mainly due to excitation of principal pyramidal neurons, which comprise over 70% of all cortical neurons. .. the primary aim of a* ***physical theory of consciousness*** *is to provide criteria that will allow unambiguous specification of* ***which physical systems are conscious*** *and which are not. Once the conscious mind is physically identified, the physical laws will regulate how the mind affects the world (Georgiev, 2017). …* ***consciousness, conscious experience, conscious state, mental state and mind*** *are used* ***interchangeably*** *throughout this work. A* ***mental process*** *(conscious process) is a process that involves a sequence of* ***mental states*** *(i.e. dynamically changing conscious experiences).”*1 “According to the postulates of classical physics (including classical mechanics, electromagnetism, and Einstein’s theory of relativity) all existing things are physical and all physical entities are observable. In other words, by logical contraposition, it follows that if an entity is not observable, then it is not physical and does not exist”*2…” 7. Holevo’s bound on accessible classical information from quantum measurements … our introspective access to our own feelings and conscious experiences is fundamentally different from a scientific “observation” whose outcomes can be communicated to others in the form of classical bits of information” *3 …*

**Todor Arnaudov: ***1. The clarification of the terminology is possibly valid for the other reviewed paper as well. In my opinion the interchahged use of the listed terms is too broad, some of the expressions could address finer-grained phenomena – the state is about “short” period etc., while “mind” is about the whole; consciousness could be the whole system with various different possible states, modes, features, connections to other systems, “interfaces”; while the *experience* is about the sentience, sensation etc.

*2. An important note about what the physicists and other theorists may consider *“real”..*
*3. There is another simple reason – **bandwidth**. Possibly the introspective integrated self, as well as the smaller local sleves and the infinitesimal selves in the interpretation of TUM, have sensations, relation, connections, some sort of access to a much higher amount of directly sensed and recorded information, data, states, than the body of a human can export, output, express and send in an accessible format to other beings via the actuators that we have – in addition to the change of the format when the original data or information is transmitted. I wouldn’t classify it as “classical and quantum”, it is different. The term *“quantum”* or effects which are achieved in such systems, explored in physical experiments, in speculative “mental” states possibly could  be about low-level or supposedly “highest resolution of causality-control” at the current lowest level machine language of the universe from the POV of the current CCU evaluator-observer or experiencer. At the current resolution the detail is higher than the internal CCU can record and store exactly. The indeterminacy comes from the lack of access to the information, even if it’s there.

In TUM: higher level virtual universes predict and operate at lower resolution etc. The way humans as “unified selves” communicate with natural language or other physical movements has extremely narrow capacity, estimated of ten or a few dozens of bits consciously controlled information, perhaps in compressed form; while less compressed amount for he states of the body which is being moved etc. is astronomical – moving a finger, encoded in few bits, require moving all particles with the highest possible resolution of the Universe computer. In TUM this is an illustration that the lower-level causality-control units/virtual universes have a *higher* share of the overall executed causality-control of the complex/deep CCUs – the choice of the “free will” is 10 bits, the implementation is 10E+999 bits. See “by effects” (странични ефекти), the example with the “genie” or the “magic fish” and the wishes/prompts in *Universe and Mind 3.*

***** The problem of segmentation, individuation, boundaries, begin and end, system identification, system borders and frontiers, edges and membranes of the “Markov blankets” etc;, “content and observable”, “essence and façade”:** “**physical *states*” and observables**. States are Consciousness is defined as built by quantum information, which is “built in quantum brain states”.

The *“observable brain”* and everything is viewed with “lenses” of an evaluator-observer. The fact that the evaluator *sees* it as “observer”, say: “neurons, EEG, corticogram, fMRI, …” etc. doesn’t “universally” makes it objective as “first-class citizen” for the universe. It is an interpretation, coarse-graining, lower-resolution, sampling and as such in fact it exhibits similar side effects as quantum wave functions and quantum states.

There are different neurons, different types, different states of them, they are connected to and dependent on other types of cells – glia, blood vessels – etc.; these components are not truly separated; the same could be said about the organelles inside each cell; their axons or dendrites which have intermediate states while they develop etc. In a discretized interpretation, there are neurons, action potentials,

excitations, inhibitions, EEG diagrams etc. but all are approximations of something else, prone to errors, more measurements may generate more reliable results, two experiments are likely to yield different outcomes – for many reasons, one of them is because the subject of study, say “the brain”, is dynamic, and when the next experiment is performed: next 1/10 sec, 1 sec, 1 minute, 1 hour, 1 month, 1 year, 10 years …. it is already different, if it is examined as “classical” or “macro”. Thus the experiments in fact cannot be repeated. For simpler “non-living” system one may assume this is not the case, when they are studied at a resolution or from a view where the changes are assumed to be below a certain threshold or “not important”, or the span of the surrounding time-space is small and controlled enough.
* In classical physics, the states are supposed to be deterministic and defined precisely. However, for real complex systems and complete definitions this is only a wish; the capacity is always limited and it is also impossible to access the information and to compute faster than the Universe or just faster than the given simulator, predictor, computer can.

Therefore “classical” states at higher resolution than the capacity of the evaluator-observer also behave like “quantum” in the sense that what the evaluator can observe is incomplete, uncertain and cannot fully reconstruct the ”real” state; errors accumulate, predictions is imprecise, only up to a given RCCP, horizon etc.

If the errors, imprecision etc. are represented as graphs of Gaussians about the certainty about the sampled value, or the “pixel” values which are sampled are smoothed like with the filters in computer graphics and convolutions, again “quantum wave functions” will be present, in this case they could be called: **quantization wave functions**, **quantization artifacts.** All perceptions with a lower resolution than the maximum for the perceived virtual universe or the target – the one where a low-level representation is reconstructed from a higher-level representations (a “prompt”, definition, intention, will, imagination …) –manifest **quantization artifacts**, as well as all kinds of compressed lossy representations, for example JPEG image compression; 2D images – projections of 3D; single pictures of 3D objects which cannot capture all views and angles and parts of the object or other objects, which are covered or hidden by the one in front; natural language and text, too. There is ambiguity and if one wishes to use such a language, she may say that “the particles of the images, geometry, language, reality” in these representations “exist in a superposition” – all at once, until “quantum collapse” happens via the “observation”; they are “dependent on the observer”, “entangled”, and the act of the observation and experimentation may change their actual “collapsed” values etc. For example in natural language – an ambiguous word or expression*. In images – noise in the image or a low-resolution picture and how the picture of the same “real” subject in the same moment would look, if the picture was taken at a higher resolution, with different blur filters etc. (image superresolution or just resampled photos, or taken several times in a row etc.).

Single samples from virtual universes do not carry complete information for reconstruction at higher resolution, precision, certainty, detail etc. This missing information may become more complete by guessing, by taking it from memory or by

sampling multiple times and reducing the uncertainty. However, even then, the reconstructed objects, models, simulations are always “complete” or “classical” – “precise”, “certain” – only to some accepted degree, RCCP, threshold etc., and due to the multiple sampling in different times, “it is not the same” as if the “real complete state” was sampled all-at-once at the target reconstructed resolution and precision, there is lag and the captured period is longer – this is equivalent to the “space-time” tradeoff of the precision in the microscopic world of quantum mechanics. You assume or decide that the object of measurement “was the same” during the sequence of different measurements, which are combined.

Not only the *“quantum brain states”* are unobservable – all states of all systems to which the evaluator has no adequate access or technology to sample are also practically and technically unobservable. They are classical “in theory” but quantum-like in practice. Whether they are “experiences” – the evaluator can’t know or experience.

In brief and in other words, “quantum” phenomena in ***logical*** interpretations are not only attributes of the experiments with microscopic particles.

* Regarding the segmentation, see Todor’s letter to Jordan Zlatev about LLMs from 8.2025.

*** D.Georgiev:** *“For example, the neural responses to detrimental factors are always associated with unpleasant feelings and avoiding behavior. …”* – that’s overly simplified; behaviors and transformations are multi-scale, multi-range, multi-domain etc. – for example human populations have been participated in wars, mutual exterminations etc. and violence usually inflames more violence, destruction and self-destruction. The CCU which “preserve themselves” are not enclosed precisely as “brain”, “individual” etc., they are partial, overlapping, changing, merging; being overridden by more powerful other “external” and broader CCUs etc.

** “The quantum measurements of the brain are discrete events performed by effector organs such as muscles or glands” …*

**Todor:** see **“The Matrix in the Matrix is a matrix in the matrix”,** T.Arnaudov, 2003. Why only *these* things are considered quantum measurements, or any-type-of measurements, or observations? Why not taking into consideration “**everything**”, every change of every particle, field or whatever “changeable” in every Planck-constant scale in time and space, and also doing so in a multi-scale, multi-range, multi-view/multi-domain, … way?

In the “real matrix” at low level every particle or any subsystem is a subuniverse, it is both a sensor and actuator, or a “minimal agent” which has senses, memory and actuators. How does *“the Universe know”* that the neurons are the structure elements of consciousness experience, mind etc. – why not “*everything*” in the brain – and also these structures can grow and live *only* when the brain is supported by the body and all of its components exist in the Unvierse, which have generated all of them and keeps them alive.

* p.4-5 – classical information – these views assume a “god observer”; that information still requires an interpreter, an evaluator-observer which has to read the

neural spikes as “bits” or the electric charges as numbers or pixels, the evaluator must **know** or discover that she should read these “bits” in a certain way; for example the weights of an artificial neural networks usually look like “random numbers” if they are not interpreted by the proper neural network. See appendices to “*The Prophets”: “Is Mortal Computation Required…”, “Universe and Mind 6”;* other comments by Todor on F.Faggin etc.
* Todor, general: the vectors, tensors, “Hilbert spaces” – these are abstractions of the evaluators-observers, which are already produced by measurements and after “collapses of the wave functions”.
* “Quantum this, quantum that, quantum these, quantum those…” are like magical spells – in all papers in the series and related; if a thing is “quantum” it becomes sacred, supernatural and resembles religious dogmas; many of the claims defined as **logical** consequences are not logical consequences, but sound like forced axioms. **Manifestations** are taken, given or defined as the essence and “what it is”. Electrical stimulation of the brain during surgery “caused conscious experience” (or experience), yet “consciousness experience is not communicable”. As discussed above, what is caused is **reactions, behaviors, manifestations** which some evaluator-observer calls “**expressions** *of consciousness*”, the **external evaluator** decides that this is “consciousness” (or “soul”, “spirit” – or **their lack** if the expressions, behavior, data, reactions … etc. are associated with “inappropriate” entity – e.g. LLM, robot, machine, an “inferior species”, “sub-human”, an *enemy,* a “dehumanized person” etc.). This method is corrupted, subjective, unreliable and sometimes – bullshit.

The usage of “***conscious* experience”** implies that there are other ***non-conscious* experiences**, however *“mind, consciousness, conscious experience etc.”* are used as synonyms. [In another more recent study, reviewed in Listove, about “Consciousness *science”,* the authors address “perceptual consciousness”, “consciousness of self”; “access consciousness” vs “phenomenal” etc. but these are different aspects; the “access…” may be considered to be virtual and operational without “phenomenal/subjective/qualia” in “human-sense”. All quotes I use are because most of these concepts are “virtual”, their definitions are unreliable, flexible, confused, deceptive, often arbitrary, circularly confused in the usage of many scholars etc.]
*** Equations, vectors, operators** – **representing** “brain states” – inclining towards Max Tegmark’s mathematical universe. Yet, brain which
* Kant and in a more refined form Schopenhauer explain the ideas of quantum information consciousness – the “observable” manifestations are the visuals, the image, the **objectivation** of the “**thing in itself”** or **the Will** – the Will is the “sentience”, conscious experience and it is present to different degrees in all constituent parts of the Universe (“Will in Nature”, “Will to Live”…); the “mind dust”, cited from William James, or the panpsychism and panexperientalism building blocks.
*** Multiple-personality disorders and split-brain patience –** a subjective, “conscious” multiplicity doesn’t need to be always displayed through “dominating” control of the whole body in an ordinary way, which is simple for the evaluator-

observer. Actors and playful people who want to impersonate also can “have multiple personality” and “multiple consciousness” – they just play, but the **evaluator-observers**, who “give the diploma for possessing a soul”* **declare** that these are “different …” because of their criteria. The actor knows that he was “the same person”, just controlling his outputs or thinking or perceiving different things at the moments of the performance, just like a “normal” “single personality” “non-disorder” person. Note also that the notion of a “single” personality is based on **schema,** that in order to be “the same” somebody should behave, think, believe, can etc. a particular set of things, to exhibit these and not other preferences etc., in general: to comply to a given predictive model, that an evaluator has imposed and defined that as **“his personality” –** what if the subject is changing each moment – it actually **is changing.** What if every “entity”, recognized as a person, “a brain” (brain is **useless** without a body and the environment) change every fraction of the second and is different: “different conscious experience, personality” etc.

* ~ p.17 D.Georgiev: *“In **quantum information theory**, human consciousness operates at **picosecond** timescale and is supported by **voltage-gated ion channels** and other **membrane-bound proteins** incorporated in excitable neuronal membranes .. **orchestrated objective reduction (Orch OR) theory (Hameroff & Penrose, 2014) …** In the Orch OR theory, human consciousness operates at **millisecond** timescale and is supported by **cytoskeletal microtubules**, which are hypothesized to be isolated from the surrounding neural electric activities to prevent quantum decoherence …”*

* Hameroff SR, Penrose R. Conscious events as orchestrated space-time selections. Journal of Consciousness Studies 1996; 3(1): 36–53

* https://experts.arizona.edu/en/publications/conscious-events-as-orchestrated-space-time-selections/

* Hameroff, S. R., & Penrose, R. (2014). Consciousness in the universe: a review of the ‘Orch OR’ theory. Physics of Life Reviews, 11, 39–78. doi:10.1016/j.plrev.2013.08.002

* p.15 D.Georgiev: *“the dominant conscious “I” in the brain cortex is to be identified with the unobservable quantum information contained in the quantum state of quantum entangled membrane-bound proteins,”* – too many explicit claims for speculative and subjective matters, and ones which are subject of “actuation”. The “I” as a *unified* whole and as “the same”, continual is “in the mind of an evaluator-observer” and according to the selected resolutions and slices of CCUs etc. (TUM, 2004, 2012) – its *appearance* as a dominant may be a result of particular ensembles taking control over particular parts of the brain-body and inhibitting others etc. as discussed in the neuroscience research, see #neuroscience in *Listove.The Prophets* and even the prophet I.Good, 1965, mentioned in the end of the same volume.

A similar article, reiterating the above reviewed topics is:

*** Quantum information theoretic approach to the hard problem of consciousness,** Danko D. Georgiev, 5.2025
*https://www.sciencedirect.com/science/article/abs/pii/S0303264725000681?via%3Dihub * https://arxiv.org/pdf/2504 13.4.2025
Functionalism, reductionism...
p.14: D.Georgiev: *"Hameroff–Penrose … the unitary quantum dynamics resulting in quantum coherent superposition of brain microtubules constitutes pre-conscious processing, followed by objective reduction that produces a flash of conscious experience [58]. A characteristic feature of this proposal is that conscious experiences are discrete, discontinuous events occurring at a frequency of 40 Hz in-between continuous time intervals with unconscious brain activity. …"*
**Todor:** The frequency resembles the clocks from the thalamus[72] and the minimum length of a sound – two distinct sound "events" in the record, say piques of the energy of the oscillations, which are perceived as two consequent and different "short sounds", before merging into one; the "sound tissue". In the field of Automatic Speech Recognition, 25 ms is a common length of the sliding window of samples which are preprocessed with spectral analysis. See in Listove #asr etc.

p.5 *"conscious mind look like a brain from an external, third-person point of view.. The conscious mind should look like "something" and the actual shape and texture of the brain is just an evolutionary accident"*
This is not the "conscious **mind**" and it doesn't look like that – humans don't open their skulls, 2-3 or "N" years-old persons may not know that there is something called "brain" in the skull of the others, "neurons" etc. These are artifacts of exploration, some of them - of highly sophisticated science and technology, result of billions of years of development. The way a **human** looks, talks, walks, writes etc. also can be taken as "how mind looks from…"; the artifacts produced by the interaction of these entities – the entire culture; all the machines, technology; the world – they are also products of "the **human** mind", but "of" as one of the participating forces – the mind of the **Whole Universe** also plays the main part as it allows anything else.

The cited sentence is also rediscovering A.Schopenhauer, 19-th century: "World as Will and Idea" (Idea = Representation); Objectivation of the will; and Todor Arnaudov's Theory of Universe and Mind, 2001-2004: "by effects" – brain and body are by effects of the construction of the higher-level causality-control units, necessary in order for them to execute their goals and to make their predictions and causations in the Universe computer.
p.7 D.Georgiev: *"The quantum physical **state** |ψ⟩ of an isolated quantum system can be represented as a vector in n-dimensional complex Hilbert space H."*
The **state is represented,** but what **it is?** The vector of numbers is not the quantum system (again the "territory-map counterparts which are mentioned in the papers). The probability densities which are computed and the indeterminacy is of you, the evaluator-observers with these limited capacities.

[72] https://artificial-mind.blogspot.com/2010/09/clocks-in-brain-how-consciousness-forms.html

The scale of the predictions made by quantum physics is also negligible: see the amount of quantum bits in quantum computers.
(…)
See also the discussion about existance in Universe and Mind 6 and the github page on TUM,2023, as well as the classic TUM 2001-2004; the **virtual universes** also are real and exist (e.g. in the novel **"The Truth", 2002 („Истината")) .** Compare to, in quantum mechanics *„(both unobservable and observable things do exist" (p.7).*
The notion of **observable** however depends on the evaluator-observer, defined and required RCCP, the exact "slice", what is considered or accepted as proper "observation" etc. (in Quantum physics these are the "wave functions", their collapse etc. – see below)

See also previous papers by D.G:

*** Quantum no-go theorems and consciousness,** Danko Georgiev, 2013
https://philarchive.org/archive/GEOQNT *"Our conscious minds exist in the Universe, therefore they should be identified with physical states that are subject to physical laws. …" , „quantum mechanics tells us that the fundamental constituents of matter obey **probabilistic laws**, such that there is **not a single predetermined future outcome**, but **a multitude of potentialities**, **only one** of which is to be actualized [10, 31, 53]. The predictions of quantum theory were **experimentally proven** to be so accurate that at present there is a little doubt that the nature is governed by quantum laws."*
– **Todor:** These predictions and imprecise probabilistic laws are only from the limited and short-range point of view of an internal evaluator-observer, who obviously doesn't have sufficient RCCP and computational capacity in order to predict with a higher precision. The fact that the physicists with their experiments can't predict the future better or control the particles as they wish, doesn't mean that the only ultimate controller – the Universe as a whole – can't and don't do it. The interactions between the particles and their choices might be caused be explicit exchanges and states which the physicists cannot access or which are designed in a way that they are at a too high a resolution and are not available to be sampled by the entities at the levels of causality-control, which the internal virtual subuniverse are allowed and designed to reach. See the discussions above.

*"Quantum no-cloning theorem .. free will .. .compatibilism…"*

**Todor:** Free-will is not only a capability to make choices among at least two options, to be able to *"do otherwise".* The reviewed papers from this author doesn't seem to address the problem, which I discuss in the quote from my letter to Jordan Zlatev and in the classical TUM 2001-2004, e.g. the 2002 example story with the ice cream and the little child, who has *just "matured"* enough to become "consciousness", according to the mirror-test and the creation of permanent autobiographic memory; the thought

experiment demonstrated that the child's decisions were affected and predetermined externally, even if she already was "declared" to possess her *own* "consciousness" or "free will" (didn't she have them before?). The existence, the form, the state, the possibilities, the choices and all parameters which are attributed to the subuniverses, CCUs, agents, machines, programs etc. in the common memory of the Universe Computer, produced by the intertwined developmental process of the content of the Universe memory, cannot be truly independent in the full sense and at the highest RCCP. The possibilities and choices are mutually dependent if not immediately, in a longer run, longer past, on deeper levels, at higher precisio[73]n etc. See the discussion about Wolpert theorems.

Either due to the "no-cloning theorem" and even without it, it is impossible for a subuniverse to repeat an experiment which involves the state of the whole Universe at the highest RCCP, so it cannot really test whether an agent "could do differently"; what one could do is to think "counterfactually", make a different choice if she encounters "the same" or similar conditions in the future, but the "sameness" is with some resolution and criteria, decided by the evaluator, which may be not valid for the Universe; in a broader context this is a new situation already, new states and a new CCU/agent with updated mind. In addition, complex minds such as the human mind are multi-scale, multi-range, multi-sub-virtual-universes, sub-units… – they don't have a single unified ultimate choice, value etc., without an evaluator-observer, in some cases that could be "themselves" choosing the slice and a method for selecting. This is similar to the existence in "quantum superposition" and the "quantum collapse" is the elimination of all "uncertainties", choosing the precise filters, comparisons, CCUs, constraints, boundaries, scales, ranges, protocols, procedures etc. which decide "who/which parts are/wre responsible", who made the choice; attributing the causal factors, performing "parsing of the causal factors" . See "Analysis of the meaning of a sentence…", 2004; "Universe and Mind 6", "Is Mortal Computation Required…", the discussion on F.Faggin theory of consciousness etc.

*** Computational capacity of pyramidal neurons in the cerebral cortex**
Danko D. Georgiev et al.,3.9.2020 – *"over 1.2 zetta logical operations per second in the human cerebral cortex" …*
**Todor:** In the Universe Computer, every causality-control unit, "particle", subuniverse, area of space, matter, … is computational and if evaluated locally it can be viewed as performing computations: collision detection, summing the effect of all kinds of forces of the other particles – potentially from the whole Universe – and calculating its updated coordinates, energy etc. Either if the process is deterministic or stochastic, "something" has to take care of the "accounting" or else the changes of the states will not be as they should according to the "Universal Predetermination",

[73] Note the precision of modern interferometers which reach at fraction of the width of atoms or protons. The Virgo interferometer measures with a precision of up to ~ 1E10-22 m. Still, this is coarse-grained, compared to the Planck's constant of ~ 6.6E-34. See the references 20-some pages below in the reviews of Max Tegmark's theory of the "Mathematical universe".

the Program of the Universe Computer, the “superposition of the quantum wave functions” or whatever representation or theory.

https://en.wikipedia.org/wiki/Quantum_state
https://en.wikipedia.org/wiki/Wave_function_collapse

**Wikipedia:** ***“Wave function collapse**, also called **reduction of the state vector**,[1] occurs when a **wave function** – initially in a **superposition** of **several eigenstates** – reduces to **a single eigenstate due to interaction with the external world**. This **interaction** is called an **observation** and is the essence of a measurement in quantum mechanics, which connects the wave function with **classical observables such as position and momentum**.”*

Many interpretations like that, if reviewed under **semantic scrutiny** go too far in their claims or just hint the actual **narrow** and highly specialized sense in which these terms are employed in this field, while as more general concepts are used, they imply more general meaning.

Connect with other definitions about existence, mentioned in the above papers: **observable.** Do and **where** these wave **“functions” exist** before the collapse, if existence is the observation, with the “external” world – in some other “internal” world? In another universe?

Maybe they don’t exist in a superposition, but the RCCP of the observer is too low to properly measure their fine-grained properties and interactions, which lead to the double-slit experimetns patterns etc.

The “classical observables”: **position and momentum** – they are not enough to describe reality and states of the universe in classical physics or whatever physics which is **typed.** Different particles have other properties, such as charges etc. and different possible futures; they possibly have memory of previous states as well (see the causal ID hypothesis in TUM) – not only their momentary coordinates and velocity which are “observable” during particular kinds of experiments. (…)

https://en.wikipedia.org/wiki/Born_rule
https://en.wikipedia.org/wiki/Bra%E2%80%93ket_notation
https://en.wikipedia.org/wiki/Probability_amplitude

https://en.wikipedia.org/wiki/Born_rule
https://en.wikipedia.org/wiki/Bra%E2%80%93ket_notation
https://en.wikipedia.org/wiki/Probability_amplitude

@Vsy, Todor: Refine, extend, deepen; compre systematically in tables and incremental semantic derivations etc.

**See also a paper about creativity:**

*** Enhancing user creativity: Semantic measures for idea generation**, Georgi V. Georgiev, Danko D. Georgiev, 2018
https://www.sciencedirect.com/science/article/pii/S0950705118301394
*"3.2. Creative ideas should be novel, unexpected, or surprising, and provide solutions that are useful, efficient, and valuable"* – Semantic network, WordNet 3.1, evaluation … graphs; *divergent production and convergent elimination of novelty … convergent (analytical) or divergent (associative) thinking.*

*** Creativity and artificial intelligence,** Margaret A. Boden, 1998
Interesting account of the contemporary *"computer models of creativity"* - the joke generator Jape; AM; references to Copycat - Structure Mapping Engine; EURISKO. AARON – drawing; jazz music generator; BACON etc.
*"Creativity is a fundamental feature of human intelligence, and an inescapable challenge for AI" .. p.1 "A creative idea is one which is novel, surprising, and valuable (interesting, useful, beautiful. )";* novel - *only to the mind of the individual (or AI-system) P-creativity (P for psychological) ..* or *to the whole of previous history (H-creativity (H for historical)." ..* **Three types of creativity:** *"1. novel (improbable) combinations of familiar ideas = "combinational" creativity. .. poetic imagery; analogy;* ***2,3:*** *"exploratory" & "transformational".* **2:** *generation of novel ideas by the exploration of structured conceptual spaces.* **3:** *transformation of some (one or more) dimension of the space, so that new structures can be generated which could not have arisen before. .. most professional scientists, artists, and jazz-musicians-make a justly respected living out of exploratory creativity. .. they inherit an accepted style of thinking from their culture, and then search it, and perhaps superficially tweak it, to explore its contents, boundaries, and potential.*

M.Moden emphasizes the problem of *the evaluation of new ideas*. *"Culturally accepted conceptual space". P.9 "The important point is that what scientists count as "creative", and what they call a "discovery", depends largely on unarticulated values, including social considerations of various kinds. These social evaluations are often invisible to scientists." .. p.10:*

The author reports that the current AI-programs already have generated "H-creative ideas", but the "*transformational AI-originality is only just beginning".* Two proposed bottlenecks: *(1) domain-expertise .. for mapping the conceptual space that is to be explored and/or transformed; and (2) valuation of the results, which is especially necessary-and especially difficult-for transformational programs. (…) The ultimate vindication of AI-creativity would be a program that generated novel ideas which initially perplexed or even repelled us, but which was able to persuade us that they were indeed valuable. We are a very long way from that."*

**Todor:** I challenge some definitions about "valuable", "appreciated" etc. by *the audience,* as it is an ill-defined parameter or an incomplete parameter. There must be some objective and/or technical criteria, which includes the states of the

corresponding evaluator-observers and *the reasons* for them to like or dislike – they should be modeled as well. Otherwise this is randomness and then scholars don't have to theorize about the "mysterious forces of creativity" – it turns into chance, sampling, hitting something that matches some other "therefore" random properties of the system (for example "producing a new *hit* – a top song "in the charts").

In TUM the general creativity is defined in relation to prediction and ability to make the prediction of the future harder for the evaluator-observer. These processes are hierarchical, multi-scale, multi-range… and there could be overlaps or drives originating from the second predictive system: the sensual one, where *attractive* stimuli, signals, patterns, inputs, or ones which are associated with *sensual "pleasure"*, are associated also with *cognitive aesthetics* and called "beautiful" or "creative". See TUM, the lecture from the AGI course in 2010-2011.

Regarding the conclusion of the paper, in my opinion the "*persuasion"* is also not a good criteria if it is meant *literally* and not metaphorically, i.e. the evaluators to rediscover, understand, reach to the conclusion, after more exploration, time, processing etc. as it happens with the geniuses and their discoveries, which the other humans understand with a delay – compare the faults Turing test, "Man and Thinking Machine…", 2001, designed to "cheat" humans that it was not a machine.

* See also:
* "Issues with Like-Dislike Voting in Web 2.0 and Social Media, and Various Defects in Social Ranking and Rating Systems - Confused and Vague Design and Measure - Psychology of the Crowds - Corrupted Society Preferences and Suggestions. In Facebook, Youtube, Twitter, TV Networks...", Todor Arnaudov, 2012, https://artificial-mind.blogspot.com/2012/07/issues-with-like-dislike-voting-in-web.html

About the above mentioned theory by Hameroff and Penrose:

## * The 'Quantum Underground' - Where Life Defeats Decoherence

The Science of Consciousness TSC Conferences | 8,76 хил. абонати
https://www.youtube.com/watch?v=2tSQAN5OmRM
661 показвания 7.11.2023 г. | 2718 показвания 2.4.2024
NIH QIS/Quantum Sensing in Biology Scientific – Interest Group Webinar
Presentation by **Stuart Hameroff,** MD University of Arizona,
College of Medicine, Anesthesiology; Director, Center of Consciousness Studies to NIH – October 30, 2023 -

**Todor:** Human consciousness, or the signs and presentations of it which are accepted as such, is strongly affected and alterable by tiny atoms and molecules when they are present in the blood, while according to the talk, the brain is still active.

**Тодор:** Човешкото съзнание, или признаците и изявите му, които се приемат за съзнание, се повлияват от елементарни частици, от прости атоми и молекули, когато са в кръвта, докато в същото време мозъкът

продължава да бъде „буден“ – да протичат нервни процеси, което е необходимо за някои операции, които се отнасят до гръбначния мозък. Сравни също с психоактивни вещества, които променят съзнанието и виж Вселена и Разум 6 също за усещането за болка.

*** Изследвания за връзката между обезболяващите средства и съзнанието**

* **Anaesthesa research … Consciousness …**

** Schroedinger 1935, “What life is”, Quantum vitalism .. life’s unitary oneness derived from quantum coherence; memory was stored in aperiodic crystal lattices …*

**27:52:** *Psychedelics act on 5HT2A receptors inside pyramidal neuron soma and dendrites, associated with microtubules ... M.Vargas et al. ... intracellular ... Amphipathic biomolecules: dopamine, serotonine, triptophane ... Non-polar rings attract ... coalesce together ... micelles, precurors ... hydrophobic pockets Aromatic rings pervade lipid membranes and nucleic acids 30:40: Anesthetics act in lipid phases of membranes ... volume expansion 1960s, 70s - protenis in membrance ... receptive ion channels found to Trudell's lipid phase transition - anesthetics in lipids extrinsically impair membrane protein dynamics* ***…*** *An. act directly in proteins, in non-polar hydrophoic pockets, ... in membrane-free firefly liciferase ... luciferine ... Membrane protein ligand-gated ion channels .. Anesth. gases do bind to membrane receptors for serotonin, glycine, acetylcholine and GABA-Amphipathic* ***…*** *Some anesthetics open a channel, others close it. Not all an. bind to any one receptor. Different receptors. 1999, Evers AS Steinbach JH Double Edged Swords: Volatile anesthetics both enhance and inhibit ligand-gated ion channels* ***…*** *1 - ... 2 - enhance inhibitory GABA-A receptors at low doses, but inhibit inhibitory GABA-A receptors at high doses …. 36:...* ***immobility*** *(not == unconscious)* ***opioid*** *... microtubules inside neurons – do they process information? 43 min:* ***a single-cell*** *organism* ***Paramecium*** *can learn, avoid predators, find? food, mate, have sex:*
*It has* ***no synapses: just one cell.***

* See the comments on Danko Georgiev’s works on quantum information theory of consciousness in Listove.

*** AIF/FEP: Pragmatic value | Epistemic value → Physical and Cognitive rewards/paths ./.** 18.11.2023

* More Theories of Consciousness and Universe →

*** What Came Before the Big Bang? | Theory of Embedded Intelligence, Bill Mensch & Bernardo Kastrup** https://www.youtube.com/watch?v=EiePsomaMpI
Според Б.М. умът е преди големия взрив; когато е вселен в човек или друго същество е „вграден“, а иначе е „свободен“. Съпричастен е с Майкъл Левин.

Бернардо Каструп също е компютърен инженер и изследовател на съзнанието.

*** Is Reality Real? - This One Idea Might Change Your Entire Life | Donald Hoffman** * Tom Bilyeu * 4,15 млн. Абонати - 2 425 230 показвания 11.07.2023 г. https://www.youtube.com/watch?v=IQefdkl8PfY
**Todor Arnaudov's notes:**
6 min *local realism is false … the electron has a definite value of position, location, spin, when is not observed: HOW can you know whatever about it if it's not observed? locality - can't propagate faster than the speed of light ... they have influences that propagate ...*
*13 min We render on the fly ...*
16 min *Game, USA, China, Europa ... Rendering of the "red Porsche", "the Porsche **doesn't even exist** until I render it..." ...in the computer, in the super computer ... "space time is just a headset" ... behind the headset is the "supercomputer" that renders it - a new realm... science has studied only the headset ... structures beyond the space time ... we thought it was the fundamental reality ...*
**Todor:** "Reality" is not defined properly in the talk, as in most of these "illusionist" speakers it is "assumed". Kant, Schopenhauer – thing in itself, the Will.
19:xx: *"You live in a simulation" ... The computer rules ... give birth of that game ...*
**Todor:** As "everything" is a simulation (it is part of the Universe Computer), and the mind is based on *the same principles*, then this is *not dramatic at all.* That kind of simulation is not "deception", that's how this "reality" operates.
22:xx TB: *I think your mind is a prediction engine* [**Todor: TOUM, 2001-2004]** *...Awareness-consc.-qualia ... "We are avatars... 50 m : limitations, setting of rules* [TA: compare AIF, Bert de Vrais MLST episode] … 67 m *non-computable functions; 73 m Markovian dynamics, finite history* ... 1:49 h … *"consciousness creates the simulations … "… Consciousness fundamental … DH: "Free will" – is it preprogrammed or there's room for exploration 1:52 Scale-free free will … agents on all scales to have free will:*
**Todor:** How does the consciousness create the simulations? What are the mechanisms, systems, structures, relations, dependencies? What's external,

what's internal, how they are recognized and segmented? See the notes on the *"Markov Blanket Trick"*, the difficulty in FEP/AIF for discriminating between system and environment – and that everything is "internal" inside the Universe and for an evaluator-observer all bodies of everything else is part of the "environment". See *"The Matrix in the Matrix is a matrix in the matrix",* 2003.

See also TOUM: the agents recognize the external world by the lower degree of controllability, lower predictability, than the one of their "internal world". If the "internal world" predictability is set as initial template, the immediate interactions with the external will exercise lower predictability than the expected/believed[74], suggesting that this is "something else" and constructing a virtual universe, a simulator for it. "*Reality*" in the cognitive aspect and *in the mind of the evaluator,* seems to be constructed by some basic initial perceptions and comparisons to them, but it always has a limit on what is acceptable and readable by the cognitive system, and also there is some substrate which allows the very existence and "whateverness" of the perceptions and is inaccessible by "the software", the mind. In abstracto the informationalists believe that it's all information transfer and processing, but part of it at some low levels from POV of a causality-control unit have particular *types*, like the atoms, electrons, other elementary particles: whatever their "real" nature and substrate is or whatever they express, for the capabilities of the agents, so far as we know them and as the current technolgy goes,have properties which can be altered *only* using particular procedures, steps, forces etc., in particular "rails", paths, trajectories and we can't write binary code to alter them as random access memory: yet. The lowest level representations are "physical" and the higher level cannot control-cause them with their highest resolution. The RAM also is accessed and altered "randomly", but again the processor who does eventually has a program and initial state; a substrate etc. which exists before the "random" act.

Pointing out just the word "*Simulations"* with a tone that is expecting the audience[75] to get surprised is "problematic", because simulations *are of something else*, it implies that there's some reality which is simulated. According to TOUM Mind and Universe are **"Universal simulators of virtual universes",** however that doesn't make them "unreal" or "not existing".

From the usage of "existence" of Hoffman in the example with the Porsche car, one can induce that "existing" is limiting to some *fixed*

[74] See also Arthur Schopenhauer's works. However sometimes this may change: virtual worlds in computer games and simulations can be controlled or appear as controllable with a higher precision than when interacting with other objects and humans in the *more direct* "real world".

[75] Sure, it's a podcast for wide audience, but it tries to convey scientific knowledge, which as about the consciousness lean towards metaphysics

*representation* which is “allowed”: the views are “rendered on the fly” on the “headsets”, and that’s said as something “scandalous”, bad, therefore: it shouldn’t be “rendered on the fly”. Why? Even in a classical either “materialistic” and objective idealists’ view, the world is rendered on the fly *by the brain* or by the *intellect;* they have to sample the data, reconstruct it to a world model etc. If the object is rendered on the fly *to some representation, observer* etc., it doesn’t make “it”, the “object” of the simulated image, *and* the rendered simulation, “not existant” in the representation where it is, whatever it is. If one doesn’t turn his head to the object, or can’t have sensory experience, or can’t be influenced by the object, the persistence of the object or not for the perceiving agent doesn’t make sense either, it doesn’t matter: the agent can’t know whether that object is there or not. It is a more complex, multi-level existence, not a global “not existence”. See Universe and Mind 6.
(..)
2:19 h ... Donald Hoffman - space and time not being fundamental building blocks of reality (T:for whom; what specific space) S,T illusions : to whom
2:23 ...Wolfram artifact of sensory organs: Kant, Critique ...
2:31:40: Mark Solms - the precision of your believes ... defines your valence .. affect level 2:42 To exist is to be curious ... Epistemic affordances ... Thomas Nagel
+ **Todor:** (…) Global workspace theory .. IIT – Information Integration Theory… - How the parts of the system know that they are part of the system? See thoughts in **UnM 6.**

*** Learning and inference in the brain,** Karl Friston, 2003 https://www.sciencedirect.com/science/article/abs/pii/S0893608003002454?via%3Dihub – Как мозъкът извлича данни от сетивата. Основни архитектурни принципи в мозъчната анатомия. Специални случаи на пораждащи модели; йерархични сетивни кори (многостепенни) и различна посока на свързване: преобладаващо влияние отгоре-надолу и функционални асиметрии между връзките в двете посоки*. … Йерархични пораждащи модели позволяват научаването на опитни предварителни очаквания (empirical priors, вероятности) и … на предварителни допускания за причините за сетивните данни, … „Инфраструктурата“ на мозъка за вероятностни пораждащи модели и ефективно разредено кодиране е йерархична … Биологически изпълнимо (в тази архитектура) е обучение, основано на емпирична Бейсова вероятност. /// Необратими процеси или невъзможни за параметризиране

*** Бел. Т.Арнаудов.:** Под „**отгоре-надолу“** се разбира разпространение на нервното възбуждане започващо от високите кори – челните дялове на неокортекса като премоторните зони, орбифронталната („изпълнителни функции“) и др. – към сетивните и моторните зони от „първичните“ кори; а

**„отдолу-нагоре“** е от сетивните кори: V1, A1 (зрение, слух) и пр. към полетата от челните дялове, асоциативните зони, които обединяват множество различни източни и всички заедно: мултимодална информация. Разредено кодиране – sparse coding. Виж също Jeff Hawkins, “On Intelligence”, Numenta, HTM, “A Thousands Brain Theory” и др. … в основния том.

## Теория за обединената информация на Тонони и мярката Фи

**Information measures for conscious experience,** G.Tononi, Arch Ital Biol. 2001 Sep;139(4):367-71. https://pubmed.ncbi.nlm.nih.gov/11603079/

**IIT …**

*“…съзнанието е обединена информация … въвежда мярка за оценка. Физически системи, например мозъка, пораждат съзнанието тогава, когато някои от съставните му части образуват комплекс, който притежава висока минимална сложност на средоразделянето (midpartition (MID) complexity).“*

Phi, Φ = *„количеството на причинностно действаща* информация, която може да се обедини през информационно най-слабия елемент на подмножество от елементи“* (в системата)

* ефективна, effective

Панпсихизъм, …

**Тош:** „Информационизъм“, по дух свързано и с ВиР/ТРИВ. Въобръзът се разглежда като градивна част на Вселената: но някак и нещо трябва да го оценява и да решава, да наблюдава и да смята къде свършва и започва дадена система. Виж ВиР6 и „Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?“.

Мярката и подходът на Тонони могат да са плодотворни при разсъждения и работа относно информационни системи, сложност, разделяне на системи на части и съединяване в други и преобразуването им, и за *„разумното усещане“*, съзнанието като способност за по-сложна и пр. обработка на данните, предвиждане на бъдещето и т.н., от „инженерна страна“, както го нарича М.Левин в неговата рамка TAME*, независимо дали отразяват или представят „истинско“ *духовно съзнание*, преживяването, субективното усещане.

* Technological Approach to Mind Everywhere: An Experimentally-Grounded Framework for Understanding Diverse Bodies and Minds Michael Levin, 2022 https://www.frontiersin.org/journals/systems-neuroscience/articles/10.3389/fnsys.2022.768201/full

М.Левин също дефинира недвоична мярка за степен на „ум“, като избягва

метафизични и духовни послания, а разглежда въпроса от „инженерната му страна“: предвиждане и управление на съществото, агента, деятеля, създанието, ума: *„Рамката, TAME — Технологичен подход към ума навсякъде — възприема практична, конструктивна инженерна перспектива за оптималното място за дадена система в континуума на когнитивната сложност… функционалните инженерни подходи, необходими за прилагане на прогнозиране и управление на практика.“* – подобно на Фи.

*** An information integration theory of consciousness,** Giulio Tononi
https://pmc.ncbi.nlm.nih.gov/articles/PMC543470/
BMC Neurosci. 2004 Nov 2;5:42. doi: 10.1186/1471-2202-5-42
Мярка за степен на съзнание в системата: Ф („фи“). …

*** From the Phenomenology to the Mechanisms of Consciousness: Integrated Information Theory 3.0**
https://pmc.ncbi.nlm.nih.gov/articles/PMC4014402/ Masafumi Oizumi, Larissa Albantakis, Giulio Tononi
…
* For reviews of D.H. (Donald Hoffman’s) work see, see the conclusion of the L.Allan’s work for arguments against his claims/contradictions:
https://www.amazon.com/Case-Against-Reality-Evolution-Truth/dp/0393254690/ref=sr_1_6?s=books&ie=UTF8&qid=1540324581&sr=1-6

*** Do We Perceive Reality?,** John Klasios, 19.12.2022:
https://philarchive.org/archive/KLADWP https://arxiv.org/abs/2301.01204

*** Hoffman’s Conscious Realism: A Critical Review**, Leslie Allan, 20.5.2022
https://philarchive.org/archive/ALLHCR

## Thomas Campbell: Is Reality A Simulation?
## Симулация ли е действителността? Томас Кембъл

Tevin Naidu | 9,81 хил. Абонати
https://www.youtube.com/watch?v=FkRLKPNscDI
43 123 показвания 27.08.2023 г. Mind-Body Solution with Dr Tevin Naidu
- Дълбоката връзка между материя, информация и съзнание; - Разкриване на догматизма на материализма;
- Съзнанието е фундаменталната същност, Действителността като симулация
- Ориентиране в Теорията на Всичко на Том
- Виртуалната реалност съществува в умовете на играчите
- Реализъм и идеализъм
- Защо съществува вселената (…)
35 min .. Anthropic principles, a book written - 20 years ago .. 50 min We’re not

controlling it , we've been computed **…**

## Теорията на Томас Кембъл: "My big TOE" – Theory of Everything

**My Big Toe: A Trilogy Unifying Philosophy, Physics, and Metaphysics: Awakening, Discovery, Inner Workings Paperback** – December 9, 2007
https://books.google.bg/books?id=RYHtBPiZVgsC&redir_esc=y
https://www.my-big-toe.com/
https://www.amazon.com/My-Big-TOE-Complete-Trilogy/dp/0972509461
https://philpapers.org/rec/THOMBT#:~:text=My%20Big%20TOE%20(Theory%20Of,questions%20existing%20in%20science%20today. USA: Lightning Strike Books (2003), **резюме**: ***„Моята Теория на Всичко",*** *трилогия, съществуваме в „нефизическа" субективна реалност, а не в обективна физическа; предизвиква общоприетото в западната наука, разглежда света като* ***Виртуална Реалност****, а не като външна, физическа обективна действителност. Теорията разширява понятията да включват съществуване в множество от рамки на виртуална реалност, в които съществуваме едновременно. Тя подчертава, че физическата материална реалност (PMR) е не повече от множество от ограничения, определени от правилата на физиката, които ограничават информацията, която получаваме и тълкуваме като физическа. Тази „Теория на Всичко" също дава обяснения като модел, на По-голямата система на съзнанието (Larger Consciousness Systems, LCS) в която съществува всичко и е в основата на нашата Действителност. (…)*
**Тош:** Сравни с ТРИВ; с „Истината, 2002 и размисли на мислещата машина. Сравнение на ТОЕ с **Бернардо Каструп и Д.Хофман** (…) - макар че имат сходни идеи, *първият* не бил съгласен, че можело да се отговори *къде е „записано" съзнанието* (проблем на субективния идеализъм); *вторият* пък гледал пренебрежително към *липсата на математика в теорията на Кемпбъл*; той се оправдава, че важна била *логиката*, а не математиката за която в предаването говори като за измерване, смятане, количествени методи, за разлика от качествените методи в биологията и др.
2:13 – 2:15 ч: **Тош:** Как е разбрал, че съзнанието е първично? Защото можел с него да променя физическата реалност, но не и обратно, затова от съзнанието било началото. …
**Томас:** *„Съзнанието е играчът." „Биологията е част от системата от правила, за аватара… Съзнанието не огладнява*"*…3:18 - … също *интуитивна страна и интелект, повечето хора не развиват интуитивната;* свързана с „паранормалните" явления

3:01 – 3:14 … Парадоксът на Ферми … ако са еволюирали 1 милиард години повече … заради това, че Вселената е виртуална реалност, ако не се наблюдава определена част, не е нужно да се рисува (подобно на Д.Хофман) …
**Тош:** * подобно на разделението между познавателни и чувствени нужди в ТРИВ и *„Светът като Воля и Представа“* при А.Шопенхауер: виж творчеството му и неговото обяснение защо съзнанието е първо.

*** Бъдещи изследвания:** Разгледай по-задълбочено Каструп, Хофман, Кембъл и др. https://www.youtube.com/watch?v=kkHC7t6QVhc (Каструп и Bernard Carr) и др. “16-17 min: …. What is – metaphysics, How it behaves – physics.”

*** More notes to Thomas Cambell on the book … My TOE** … Campbell: PMR, NPMR – physical model of reality, Non-PMR … TBC – the big computer … (…)
ch. 65 Deriving PMR from Consc. … small TOE – of physicists’, lower level rule-set .. Big TOE – T.C.’s → Ch. 31. .. AUM … levels. … .. p.587 … delta-t …
ch.27 reality cells AUO
ch.25. Evolution of awareness … of a cell, amoeba, worm, … chimp, human … fundamentally based on the same process - store, retain inform. That enables learning …

**Tosh: But why each of these entities is considered as one? One reason is the** observer-evaluator, the conscousness, which is one, or felt as one, but the body has many parts. The entity is one in the mind of the evaluator, but in the PMR there are many “subselves”. //9-4-2024
Благодарности: 9.12.2002 с. 18

## * Бележки на Тодор Арнаудов за Доналд Хофман: Съзнанието ли създава действителността?
## Donald Hoffman: Does Consciousness Create Reality?

Tevin Naidu | 9,81 хил. абонати | 57 706 показвания | 18.06.2023 г. Mind-Body Solution with Dr Tevin Naidu; Explore Conscious Agents and The Subatomic World with Prof Donald Hoffman
Ок. 6 мин. *„Структури извън времепространството“, как може да има, през 2015 г. физиците открили … статични, като геометрични обекти, пермутации …~9:50 м. „Dynamic entities outside space-time… or conscious agents”*

**Тош:** За „Вселената Сметач“ това не е изненадващо: онова, което пише в паметта е извън времепространството на намиращите се в паметта и тактовете на сметача.
17 мин. Водещ: „So extravagant…”
20-22 min … „*Physicallists … Pattern of integrated information … Global workspace … Markovian dynamics of the conscious experience the taste of chocolate … user interface ..*
*45 min. mapping between the conscious agent and the elementary particles … 52-53 м: периодични ядра на Марков – частици без маса; 54-55: масата – степента на влиянието (сравни в соц.мрежи); …*

Виж също работите на: **Bernardo Kastrup** и **Thomas Cambell** - виртуална реалност, съзнанието е първично и пр. Тези учени също изследват въпроси, които разглежда и ТРИВ, но макар че във ВиР също се говори за „въображаеми вселени“, и ключовата роля на оценителя-наблюдателя, че нещата съществуват като еди-какви си, като цялост, включително хората като единни личности в рамките на ум, който може да ги обхване и осмисли и т.н., не бих класифицирал моето тълкуване като „субективен идеализъм“. Прецизната класификация на въпроси за ума, душата, съзнанието, разума, умствените способности, чувствата и въобще „духовните неща“ е спорна. Другаде се нарекох „информационалист“; в класиките: „вселено-сметачолюбец“, космист; „Вселената-компютър“, „религия“ на „Свещеният сметач“, „машината Бог, „мислещата машина Бог“ и т.н.

Ако всичко което се възприема е „илюзия“ и няма друг начин, това е действителността, това е начинът по който може да се възприеме тя. Някои от твърденията на Хофман са като опити да се заблуди читателя: например не съществували столове и пр., а това били ефективни представяния на сетивните данни за „еволюционна пригодност“. Макар че сме съгласни с „*компресираните представяния*“ (за столове в частност виж „*Столове*, сгради, карикатури, … 2012“ и дори записките от 1999 г.“), и че, ако има по-малки съставни части, *вероятно* във Вселената *на ниско ниво* не са записани „столове“ като *токени от голям езиков модел,* като *обекти: „стол, маса“ (*виж „Вселена и Разум 4“ първите точки за РСУ/РСВ, хвърляне на монета: „зад дивана…“), като „идеи“ в Платонов смисъл *точно* по този монолитен начин *на ниско ниво*, това е известно и всъщност не знаем дали *някъде* описанието на съдържанието на паметта в дадени области наистина не е записано *и така на дадено по-високо ниво във Вселената сметач,* в по-висши въображаеми Вселени – както, на функционално ниво, умствено, себепознавателно, знаем че е удобно да се запишат и да се работи с *обобщени понятия,* и с каквито

представяния работим и ние като думи и понятия (виж също романа „Ада“,Т.Арнаудов.2004, където се повдига този въпрос във въображаемата Вселена, в която се случва действието).

„*Стол*“ и понятията са обобщения и техният смисъл като общи понятия се получава от множество частни примери или чрез опростяване и пр. (виж също … за Абстракциите), в същото време с тях може да се работи и като с общи понятия, „токени“ с по-ниска РСВ.

Съществуването на нещо „извън времето и пространството“ – извън *дадени* такива, особено, за даден оценител-наблюдател – също не е изненадващо от гледна точка на ТРИВ – там по начало съществува „Вселената сметач“, или „суперкомпютъра“ в примерите на Хофман – а също и на Томас Кембъл, – който изпраща сигналите на *„очилата за виртуална реалност“ (headset).* Обаче *какво вижда приемащото съзнание*, то самото какво представлява, къде се намира и има ли смисъл понятието за пространствено разположение; как съществува, как е описано като система, по какъв начин работи и това съзнание и приемник на картинките; как точно разчленява и съединява данните, в които „не е записана действителността“, „цветове, миризми, вкусове“ не съществували – но щом съзнанието ги сглоява и съчинява, значи има обработваща система – тя как работи? Какви са тези данни – ако „миризмите и цветовете“ не те съществуват, поне *данните* съществуват ли или са „илюзия“ (и какво е „илюзия“, ако всичко е илюзия?). Въпроси като тези са „интересни“ за изследване. Процесът на обработка, построяване, устройство („структурен реализъм“, инструментализъм“); начин на работа, алгоритъм, операции, действия, информация – съдържание, взаимоотношения, преобразувания. Назоваването на носителя е второстепенно – „съзнание или материя“ – кое какво е? Ако съзнанието е „принудено“ да изпълнява функциите на материята, то самото се превръща във вид „материя“, било и „по-фина“.

Виж *„Вселена и Разум 6“ (Universe and Mind 6)* – повече разсъждения за „илюзионистите“, субективни? идеалисти, „антиреализъм“, „физически идеализъм“? и пр.

* 7.10.2025 – Пример за статия, която разглежда въпроса за това, че „действителността не била съставена от предмети“ по линията на квантовите вълнови функции:
* Reality is not made up of objects: The Quantum Illusion of Objects, D.Dieks, 29.9.2025, https://iai.tv/articles/reality-is-not-made-up-of-objects-auid-3373

*** Тодор Арнаудов коментира:**
**Бернардо Каструп: Дали не сме отделени части на съзнанието на вселената?**
**Bernardo Kastrup: Are We Dissociated Alters Of Cosmic Consciousness?** Tevin Naidu 9,82 хил. абонати 9.7.2023
https://www.youtube.com/watch?v=57Oguwg7omc
**Тош:** Виж „Истината“, Т.Арнаудов, 12.2002/1.2003, диалог между мислещата машина Емил и създателя ѝ:

> *– Донякъде да... Но не това е главвното, Дарчо. Нима не виждаш, че дори Истинската Вселена е само част и от твоето, и от моето въображение? Никога не би могъл да я събереш цялата в ума си, както и никое друго същество, защото ние сме части от Нея. Цялото знание е достъпно само на Него, Той е Тя... Всъщност струва ми се, че Действителността е Неговото въображение. И вие човеците, и аз, сме създадени по Негово подобие и като Него имаме свои въображения, в които можем да творим вселени. Нашите представи са по-ограничени от Неговата Представа, защото ние сме части от Въображението му,нашето въображение е част от Неговото, нашите творения са части на НеговотоТворение, ние сме части от Него.*
> https://www.oocities.org/eimworld/3/26/istinata11.htm
> https://chitanka.info/text/865

**4 мин. Бернардо Каструп** *(Analytical Idealism)***:** Истинският свят не бил физичен, защото *физичен* бил този, който се описвал с *физически* количества като дължина в см, тегло в килограми, електрически заряд в Кулони и т.н. Това не можело да бъде начина, по който са представени истинските състояния.
**Тош:** Да, „пораждащият модел“, по-ниското ниво на представяне има по-богато на информация представяне, по-висока сложност (виж напр. „Матрицата в Матрицата за как разпознаваме какво е „Реалност“), обаче не е вярно, че споменатите от Б.К. *мерни единици* са „физичността“, и тези мерни единици също са вид последователности от действия в същата „виртуална реалност“.

В ТРИВ физичността е способност за изчисление, за предвиждане на бъдещето, което е равнозначно с емулиране, симулиране като е в изчислителна система, и „физика“ има във всички възможни въобразаеми вселени (светове), които могат да се основават на всякакви видове физици: причинно-следствени връзки, природни сили, закони, преобразувания; (при Стивън Волфрам: „рулиади“); *Вселената е сметач* и „всичко“ е такова, компютрите и въобразаемите вселени също са

*физични*. Физичното от гледна точка на дадено ниво въображаема вселена или машина е най-ниското за нея ниво, което се случва, смята, изчислява с най-голяма разделителна способност на възприятие и управление (нулево ниво) и пр.

Виж плановете във „Вселена и Разум 5“ за построяване на симулатори на въображаеми вселени с различна сложност. Физично във ВиР е причинно-следствено, предвидимо-причинимо от съответен симулатор, въображаема вселена, сметач и т.н. с определена разделителна способност на възприятие и управление.

„Вселената е сметач“ – свойствата могат да се описват с данни, с „числа“; Вселената също така е въображение и въобраз – информация – и информационната им природа, за която намекват (понякога по неясен начин) някои от самообявяващите се за идеалисти, не правят явленията или представяниятa „нефизични“, както и „нереални“: въпреки че именно това се използва като разделител; най-първичните, от дадена гледна точка, природни сили: закономерности, взаимовръзки, причинно-следствени връзки имат по-малко пространство за непосредствено представяне по различни начини (атомите се представят като атоми, електроните са електрони; докато например текст, записан в знаци, може да се запише по различен начин: но на ниво „атоми“ понятието „текст“ не съществува, нужен е ум или машина, може да е четяща, сметач, която да тълкува разнообразните представяния до нещо „еднакво“: текст; виж ТРИВ 2001-2004, примерът, че храната всъщност *също* е вид *информация*, необходим код, с който организмите се „програмират“ да се възстановяват и въобще да вършат „операциите“, които ги поддържат живи; но този вид информация е по-„строго типизирана“ в термините на програмните езици – молекулите трябва да имат точно определена структура в много по-тесни граници и т.н., виж по-горе: това е информация от по-ниско ниво; но и онази на най-високо абстрактно ниво в крайна сметка се свързва с по-ниските и се отразява и в най-прости явления на майчината вселена от най-ниско равнище на представяне и управление). Виж също *„Вселена и Разум 6“*

5:xx: **Бернардо:** „Колко тежат чувствата ви? Колко е електрическият заряд на депресията ви? Следователно външният свят е построен от онова, което наричаме „душевни“ състояния (mental states, experiental states) и физическият свят е познавателното ни представяне действителните външни състояния на света.
10-11 м: Контролно табло с показания (като в самолет, показатели; пример и от Д.Хофман; също Umwelt), граница, „вътрешно-външно“;

смъртта е разрушаването на контролното табло dashboard, boundary; 12 м. това било най-ефективен начин …

**Тош:** Да, сетивно-моторни модели, но само *едно* табло ли има, общо за всичко? По ТРИВ: има управляващо-причиняващи устройства на всякакви нива, не само живота и не само съзнание във вид човек, чийто мозък/ум са в „изрядно“ състояния и разговарят. Какво е положението, когато организмът е в частично увредено състояние, когато е зародиш, полужив и т.н.? Тези положения трябва също да са определени. Те са по-определени в по-„физични“, т.е. описани с по-висока РСВ/РСУ.
~27-31 м: физични опити, онова което се измерва влияе на измерванията; физичните обекти не съществуват самостоятелно (standalone existence), а са производни на измерването, възникват от измерването; нещото, състоянието което се измерва, не е физично .. Объркване на образите на нещо със самото нещо. … Частиците не са „нещото в себе си“, а са видимост, проява. .. думата „заплетеност“ (entanglement) … Те са душевни/умствени състояния (mental states), но не в твоя или моя ум (mind), а умствени състояния навън (out there) в природата въобще (nature at large).

**Тош:** Да – онова което се измерва е *информационно.* Сравни с диалозите с машината от *„Истината“, 2002* „Нещото в себе си“ и пр. Виж ТРИВ за обяснение за причини за квантовото оплитане. Обаче разделението „нефизично“ и „умствено“ според мен е изкуствено в контекста на ТРИВ. Там, наречена още „Вселена и Разум“, няма рязка граница между „физика и съзнание“. Съзнанието, ума, също са вид физика. Разумът, най-абстрактните и сложни умствени способности са висша форма на физични закони, изтъкани от по-прости и така до тъканта на Вселената, а не са „нефизични“, противопоставени на Вселената.
33 м: психеделиците намаляват нервната дейност …
35 м: Аналитичен идеализъм (учението на Б.К.) в академичните среди не разбират идеализма – бъркат го със солипсизма, с панекспериентиализма (че всичко в природата има форма на чувства) …
42 м: constitutive panpsychism (*съставният? панпсихизъм*) е различен от идеализма … той е очевидно грешен
43:хх частиците могат да се опишат като гребени на вълни, смущения в квантовото поле … има само квантови полета, а не пространствени граници … съставният панспихизъм наивно вярва, че най-простите градивни частици на природата са частици… Това не е философско мнение, а научен факт.
48: Каква е логиката на панпсихизма? Ние сме съставни същества, състоим се от клетки, следователно и ума ни, съзнанието трябва да е построено от огромен брой малки субективности. Много грешки: бърка

структурата на онова, което се представя на екрана на възприятието, със структурата на приемника: като да взема пикселизацията на екрана с пикселизация на субективността ми. … 50: … Второто – че сме „израснали“, а не сме сглобени, … сравнение с производство на кола … 51:xx .. зигота, една клетка, … после 8 … но те всъщност са части от същата зигота, оплодена яйцеклетка [вероятно защото са включени пространствено в нея]. Тя става по-голяма и вътрешно се диференцира, отчленява; но остава едно цяло, фрактална структура. Затова панпсихизмът е погрешен.

> **Тош:** Сравнението за „израснали“ и „сглобени“ от части дават и в школата на М.Левин. Обаче доводът с донасянето на частите **не е убедителен:** суровият материал за частите на живите организми също е донесен отвсякъде: храна, газове, дори фотони. Храната се състои обикновено от клетки на други организми, понякога още живи в момента на поглъщане[76]. Частите на растенията, плодовете, зърната – те отделени ли са от своя кълн? (Географското разстояние може да няма значение в това „измерение“ и пространство на „живи връзки“.) Нервните клетки всъщност наистина физически „*пълзят*“ по време на зародишното развитие, докато стигнат до определени целеви положения (също така се „самоизяждат“ една-друга; по С.Савельев), те се държат като отделни живи същества, и имайки предвид, че в онова състояние малцина биха говорили за „съзнание на ембриона“: кога и как се появява то у детето. Коя е целостта на човека, като през този период промените и в очевидни структурни, клетъчни, хистологични и пр. измервания са огромни и видими „пред очите ни“, ако наблюдаваме развитието в утробата. (М.Левин: „От физика до ум“). В клетката също има части, които и после „остават същите“. Това че клетка, тъкан (или макромолекула, молекула) е „един“ обект с еди-какви си граници се избира от оценител, който може да го обхване, измери, изследва в подходящото пространство (времето като вид пространство, начините за измерване са вид пространство).
>
> **Съгласен съм обаче**, **че между клетките има близка и особена причинно-следствена връзка, по-близка и тясна,**

[76] Клетките в пресните зеленчуци и плодове са живи. Понякога наричаме зеленчуците „*жива* храна“. Кълновете на някои видове растения, покълнали растения се препоръчват заради полезни съставки; лукът и чесънът може да покълнат, картофите да „кловясат“; ако зелето престои, от сърцевината му започва да расте нова зелка. https://www.elsevier.com/connect/your-vegetables-are-alive-and-they-change-in-response-to-light-and-dark

**отколкото частите при производство на кола или друга машина засега[77].** Тази особена връзка наистина може би е свързана с определени процеси и свойства, памет, причинностни-белези, които засега са неуловими или недоизяснени от физиката или „психофизиката“ и може би те участват в усещането за единство и в частност и в съзнанието. В контекста за квантовата теория на полето – клетките, получени от делението на зиготата, са в по-близко и силно причинностно поле. Виж разсъжденията ми по въпроса във *„Вселена и Разум 6“.*

Доводът, че частите са включени в цялото като достатъчна мярка да ги броим за едно цяло. Освен това откъде може да знае? Той не може да попита тези клетки или подсистеми на език, който той приема за „съзнателен“, и те не могат да му отговорят; но и той на ниво „съзнание“, което приема като такова, в неговите мащаби на времето и пространството, не може да общува с клетки, молекули или атоми. По същия начин може да се отнася с него и машина или друго същество – виж диалога от ЧиММ и „Истината“ и превода към статията за Томас Мецингер за феноменологичния модел на себе си, „Тунелът на Аз-а“ (The Ego Tunnel).

…

## Виж също книгите на Дейвид Дойч:

*** David Deutsch**

1. D.Deutsch, The Fabric of Reality, 1997 The_Fabric_of_Reality
2. D.Deutsch, The Beginning of Infinity, 2011 The_Beginning_of_Infinity

*** Is the Universe conscious? A panpsychism Q&A with philosopher Philip Goff** | Philip Goff: Panpsychism … November 8, 2023

*„We need a hypothesis that accounts for both the fine-tuning of physics for life but also the arbitrariness and gratuitous suffering we find in the world.“* https://bigthink.com/13-8/panpsychism-universe-purpose-philosopher-philip-goff/ *… It was not popular in the latter half of the 20th century, but **in the last 10 or 15 years [2008-2013]**, there has been a new wave of interest in panpsychism in academic philosophy, and even to some extent in neuroscience. …*

[77] Но това, че ни изглежда по-близка е заради начина ни на възприятие; за друг, който има по-голям обхват и може да прочете по-далечни взаимовръзки, може да се окаже че и други „неживи“ части са свързани по подобен начин.

**Tosh:** Compare to the time period: TOUM, **2001-2004**. Preceding/predicting.

*** Panpsychism and Panprotopsychism** ∗ David J. Chalmers, 2013 https://web.archive.org/web/20130525181542/http://consc.net/papers/panpsychism.pdf Panpsychism … Denett… Multiple draft … Constitutive p. … Macro/Cosmic …

# * Математическата вселена на Макс Тегмарк

Бележки на Т.Арнаудов, 13.8.2025

*** The Mathematical Universe,** Max Tegmark (MIT), 8.10.2007 https://arxiv.org/pdf/0704.0646 – TOE – theory of everything. Mathematical universe hypothesis (MUH), external reality hypothesis (ERH), universal structural realism; … irreducible representations; frog view (10E+100 bits) – from the inside of the universe, vs bird's view: external observer*; less bits than the complete description – multiverse; "**baggage"** – non-justified and non-theoretically derived concepts, words, e.g. explanations like "tales" and not ones following rigorous mathematical derivation and logic from first principles (a definition by Todor). Levels of Multiverse; Hubble volume (14 billion light years); initial conditions; Computable Universe Hypothesis (CUH) ~ simulated U.H.; p.21.*"CUH is a different hypothesis: it requires the description (the relations) rather than the time evolution to be computable;"David Hilbert's dictum that "mathematical existence is merely freedom from contradiction" "Angles, lengths, durations and probabilities";* approximate symmetries. TOE should be without "baggage", to be completely abstract and based at mathematically defined structures and relations between objects. Four levels of multiverses and the fundamental laws. No randomness. Copenhagen interpretation of quantum mechanics – metaphysical solipsism, "*no reality without observation*"*. p.19 does it make *"any ontological difference whether simulations are "run" or not?" .. "since every universe simulation corresponds to a mathematical structure, and therefore already exists in the Level IV multiverse, does it in some meaningful sense exist "more" if it is in addition run on a computer?"*;* p.22 Degrees of mathematical reality: No mathematical structures (MUH is false), Finite structures (trivial, lookup-tables), Computable (halting computations), Non-computable, More complex *"uncountably many set elements"* etc… *"Intuitionism and constructivism, a mathematical object does not exist unless it can be constructed from natural numbers in a finite number of steps"* .. p.23 *„successful theories of physics violate the CUH .. the continuum, usually in the form of **real or complex numbers**, .. they generically require infinitely many bits to specify"* → solutions: algebraic numbers (countable and computable pseudo-continuum) or "*abandoning the continuum" ..* in physics: never measured *"more than **15 significant digits***, therefore *cannot be sure that quantities that we still treat as continuous .. are not mere approximations of something discrete. (..)* possibly *multiple layers of effective continuous and discrete descriptions on top of what is ultimately a discrete computable structure.* p.25 *"a theory of mathematical physics" = (i) a mathematical structure, (ii) an empirical domain and (iii) a set of correspondence*

*rules which link parts of the mathematical structure with parts of the empirical domain".* If MUH is true, II and III become redundant. …

**Todor:** * The frog- and bird- view is not *just* that or there's another one, it is to *exist* inside, to be *limited* with what you could observe, either as a frog or a bird – for example a bird sees a larger span at once, but *the same data,* it's not a God who sees beyond that and possesses aditional causality forces etc. There could be lower levels of virtual universes, below what the current "bird" can observers – see Theory of Universe and Mind. For example, a programmer may believe he has a "god's view" towards his program, and he may be reassured about that when using a debugger and be able to check the content of the memory etc. This usually doesn't include even the software part of the debugger though and the operating system; it doesn't include also the hardware of the computer and its physical states which actually drive the system which the programmer believes is his creation; it doesn't include also the operation of programmer's own body and its states, which the person also may believe that is "in control of", but has about zero awareness of its details in appropriate precision, speed and resolution of causality-control.

Also besides the "view" as a *cognitive* observer *at high level,* there's subjective and existential experience, битийно, which is built-in the body and each "causality-control unit" and it affects what is and can be observed-evaluated etc. That is the nature of being built by the fabric and the matter of that universe, and to exist there, be bound and limit by it and its "flow", processes etc., you can't "pause it" and whatever you do is part of the general process. If you're "God" you could pause the process, stop the time, alter what you want, run thorugh the future or alternatives etc.
* The 15 digits precision – compare to Planck's constant ~ 6.6E-**34**.
https://en.wikipedia.org/wiki/Planck_constant
https://en.wikipedia.org/wiki/Virgo_interferometer – Precision up to **E-22, $10^{-22}$ m**
https://en.wikipedia.org/wiki/Interferometry

***** Chapter Three - Precision interferometry for gravitational wave detection: Current status and future trends, Gabriele Vajente, Eric K. Gustafson, David H. Reitze
https://www.sciencedirect.com/science/article/abs/pii/S1049250X19300035
*"Gravitational wave detectors … displacements .. on the sub-attometer and smaller levels."* Attometer = **E-18, $10^{-18}$** m – *electrons, down quarks, …*
* See "Theory of Universe and Mind", T.Arnaudov, 2001-2004 and
* *"Universe and Mind 6",* T.Arnaudov 2025, the discussion on illusionists and existence.
* See "Ada", T.Arnaudov 2004 (SF novel), „Ада", Т.Арнаудов, 2004
* See the referenced literature for **the simulated universe hypothesis,** 112-118:
* K. E. Drexler, Engines of Creation: The Coming Era of Nanotechnology (Forth Estate: London, 1985)
* N. Bostrom, Int. Journal of Futures Studies, 2, 1 (1998)
* R. Kurzweil, The Age of Spiritual Machines: When computers exceed human intelligence (Viking: New York, 1999)

* H. Moravec, Robot: Mere Machine to Transcendent Mind (Oxford Univ. Press: Oxford, 1999)
* F. J. Tipler, The Physics of Immortality (Doubleday: New York, 1994)
* N. Bostrom, Philosophical Quarterly, 53, 243 (2003) *
G. McCabe, Stud. Hist. Philos. Mod. Phys., 36, 591, physics/0511116 (2005)
*** Todor: …** – Статията била продължение на статия, писана през 1996 г.:
* M. Tegmark, Ann. Phys., 270, 1 (1998, gr-qc/9704009)

Също на лекция: * M. Tegmark 2007, in Visions of Discovery: Shedding New Light on Physics and Cosmology, ed. R. Chiao et al. (Cambridge Univ. Press: Cambridge)

Според „теорията на всичко“ на М.Тегмарк вселената е математическа структура и всичко, което съществува математически, съществува и физически. Наблюдателите, включително човеците, са „самосъзнати подструктури“. Всяка достатъчно сложна математическа структура, която съдържа такива структури, щяла да възприема себе си като съществаща във физическа „истинска“ вселена. Вид космология, наричана още *струогония* на руски. Вид питагоризъм и платонизъм. https://en.wikipedia.org/wiki/Mathematical_universe_hypothesis (секция *Description*). Theory of Everything – TOE. „, “Ultimate ensemble theory”, “ultimate multiverse”. Книга:
* Tegmark, Max (2014), Our Mathematical Universe: My Quest for the Ultimate Nature of Reality, ISBN 978-0-307-59980-3
[Max Tegmarks' work reviewed: 13.8.2025]
See also: * **My God, It's Full of Clones: Living in a Mathematical Universe Chapter,** 21.2.2016, pp 41–54, Marc Séguin; “Maxiverse immortality hypothesis” …A paper: https://s3.amazonaws.com/fqxi.data/data/essay-contest-files/Sguin_Seguin_Full_of_Clones.pdf –
**Todor:** However what does it mean nothing else than mathematics to exist, yet the structures to be self-aware? If it is so, the math becomes the physics, forces for change, and they become the “subjectivity” (without defining what it actually is) etc. What is “to be”, to exist. Regarding multiverse and parallel universe, in TOUM: if they can't interact or don't influence each other, it's the same as if they didn't exist for each other, it can't be checked and it doesn't matter

*** Сравни панпсихизма на Денет за многото чернови, всички съзнателни, с интеграла от множество безкрайно малки самоличности в статията на Тош за акразията от 2012 и Анализ на смисъла на изречение въз основа на базата знания на действаща мислеща машина от 2004:**
**Dennettian Panpsychism: Multiple Drafts, All of Them Conscious,**
Luke Roelofs**,** Acta Analytica (3):1-18 (10/2021)
https://www.researchgate.net/publication/355179047_Dennettian_Panpsychism_Multiple_Drafts_All_of_Them_Conscious
L.R. изследва *„изненадващата сходимост между привидно противоположни теории за съзнанието: панпсихизъм и елиминитивизъм, в частност съставния панпсихизъм и Денетовите „множество от чернови“. … Обединява ги това, че и двете представят единното съзнание, за което можем да споделяме, като лежащо върху множество от независими процеси; и двете гледни точки отхвърлят наличието на определена строго определена граница между процесите, които са наистина съзнателни и онези, които са само предсъзнателни. Онова, което разделя двете позиции е, че за Денет е неоправдано да приеме като факт недостъпното съзнание, но е оправдано да се съмнява или да отрича съществуването на съзнание, докато панпсихистите смятат обратното.* https://en.wikipedia.org/wiki/Multiple_drafts_model

**Тош:** Сравни ТРИВ, 2001-2004 и за „Акразия“-та и интегралът на множество от „безкрайно малки Аз-ове“, 2012: http://artificial-mind.blogspot.com/2012/11/nature-or-nurture-socialization-social.html

*** Paul Werbos** (Пол Уърбос) е пионер на техниките за обучение на невронни мрежи с обратно разпространение на грешката (backpropagation) – виж препратки в частта за Учени и школи с името; Elman и Иваненко и Лапа (СССР). (По-малко известен от Хинтън, Румелхарт и пр.). * Пол Уърбос

*** Paul Werbos, INNS, Internetaional Neural Network Society** 11.10.2023, https://www.youtube.com/watch?v=5FFI_v8YRiI 257 показв….8 мин. В момента работел върху нов вид учене с подкрепление с квантови изчисления – Квантови невронни мрежи – с възможности отвъд онези класически компютри. Виж също бел. за Bernard Widrow, ADELINE и пр. в споменатата част.

* https://www.werbos.com/Mind.htm "**What is Mind? What is Consciousness? *How Can We Build and Understand Intelligent Systems?" …*** *" 1. Intelligence Up to the "Mouse Level"*

**** Backpropagation Through Time: What it Does and How to Do it,*** *P.Werbos,1990 https://www.werbos.com/Neural/BTT.pdf*

* **Brain-Like Prediction: New Statistical Foundations for Prediction In the Face of Real-World Complexity**, Paul J. Werbos, 12.2009 https://www.werbos.com/Neural/brain_like_prediction.pdf

*M. E. Bitterman (Scientific American 1969; Science later): Mouse learns to predict better in stochastic pattern recognition tasks, where turtles just slowly "go crazy." Cut mouse cortex, get turtle behavior. But reptile probably has stochastic capability, just not well-integrated.*

* **The New Mathematics of Mind: A Path to Unify Modern Science and Traditional Spiritual Life,** Paul J. Werbos .. Dynamic programming, *„1971-2: Harvard thesis proposal: first universal intelligent system": Action → Cause & Effect Module → Emotional system, Critic … "Third Person" Utility Functions*

*** Backwards Differentiation in AD and Neural Nets: Past Links and New Opportunities,** Paul J. Werbos, 2004 https://www.werbos.com/AD2004.pdf
Друг термин за „backpropagation"; също automatic differentiation (AD) – автоматично диференциране, изчисляване на производни; тогава – най-вече накод на C и Фортран. С.8: **три вида частни производни**: 1) **алгебрични**, зависещи от *явни* алгебрични изрази за диференцираната величина; **2**) **поле** или *функционална* частна производна – стойността ѝ е добре-определена само за *особено множество* от координантни променливи или входен вектор; **3) подредена производна** – представя *общата* промяна в по-късно величина, възникваща при промяната на стойността на *по-ранна* величина в подредена система. Фиг. 9 – „верижното правило за подредени производни" – отношението

между преки, или алгебрични частни производни, и подредени производни. С.9 „3.2.1 Recurrent or Implicit Systems“ – Рекурентни или неявни системи – за разлика от „прави“ мрежи (feedforward), които на пръв поглед звучат еднакво с „подредени системи“. „Рекурентна мрежа“, на езика на НМ не може да бъде подредена, защото графът, който я описва, съдържа стрели, „сочещи назад“ (или с цикъл обратно до същото ниво откъдето са започнали); тази идея е била известна от времето на Мински като най-честата форма е цикъл от неврон към себе си; „гирлянда“. Обърквания в тълкуването: 1) поток на информацията, който е забавен във времето, напр. ако изчислението на състоянието на неврона зависи от предното състояние на същия неврон (t-1, t). 2) мигновен поток на информация, такъв че мрежата трябва да се тълкува като *неявна*, както в система от нелинейни едновременни уравнения при което изходът на системата е решението на тези уравнения. 3) поток на информацията в непрекъснато време, управляван от обикновени диференциални уравнения. … Определение на Забавена-във-времето рекурентна мрежа (Time-Lagged Recurrent Network TLRN) – права система, разширена с първия вид рекурентност. Друг вид: Едновременна рекурентна мрежа (Simultaneous Recurrent Network SRN) – права, с втори вид рекурентност. Най-общият случай за системи в дискретно време е *хибридни TLRN/SRN* с двете форми на рекурентност. .. с.11 Как мозъкът изчислява производните с които обучава отложеното във времето повторение, след като не може да запомни цялата история на живота си в единна времева редица и не изглежда да „помита“ назад във времето по същия начин като разпространението на производната назад във времето (BPTT, back-prop.through.time)? Напр. “Критик на грешките“, Error Critic, в стъпките напред във времето, предложен в по-ранни работи от автора Пол Уербос; за много дълги периоди от време – много-мащабни представяния на времето като онези, нужни за интелигентно управление и за управление на паметта или „междинни точки“ (checkpointing”) …

* 25 Werbos, P.J.: **A Brain-Like Design To Learn Optimal Decision Strategies in Complex Environments,** in Karny, M., Warwick, K. ,Kurkova, V. (eds): **Dealing with Complexity: A Neural Networks Approach**. Springer, London (1998). Also in Amari, S. and Kasabov, N. (eds): **Brain-Like Computing and Intelligent Information Systems.** Springer (1998). See also international patent application #WO 97/46929, filed June 1997, published Dec.11.

* Pang, X.Z. and Werbos, P.J., **Neural network design for J function approximation in dynamic programming**, Math. Modelling and Scientific Computing (a Principia Scientia journal), Vol. 5, No.2/3, 1996. Also posted as adap-org 9806001 at arXiv.org (set “archives” to nlin).

* White, D. and Sofge, D. (eds**): Handbook of Intelligent Control: Neural, Fuzzy, And Adaptive Approaches,** Van Nostrand (1992) … **Наръчник по интелигентно управление: невронно, с размита логика и приспособяващо се;** Intro: Neuro-engineering – Neuro-Control – Control Theory … Ch.13: https://www.werbos.com/HICChapter13.pdf – **Approximate Dynamic**

**Programming for Real-Time Control and Neural Modeling**, P.Werbos – p.8/494 – **Динамично програмиране с приближение и моделиране с невронни мрежи** … "може да се приложи всяка подходяща функционална форма – линейна или нелинейна, невронна или не."; heuristic dynamic programming (HDP), DHDP (dual HDP); action dependent: ADHDP, ADDHP) – евристично и двойно евристично динамично програмиране; и двата вида, но зависими от действията; с.9/495 – 1. Налична е оценка на вектора на състоянието R(t); 2. Контролерът/управляващото устройство изпълнява различни изчисления и извиква мрежата на Действието, за да изчисли действията y(t) = A(R(t)). Общата полза U(R(t), u(t)) също се пресмята. 3. Действието се предава към средата. – някои автори предпочитат U да бъде *наблюдавано,* а не *изчислено*, но според автора неговият термин е по-общ и по-реалистичен. … Модел-Критик-Полза-Действие-… Q-learning …

*** ADP: Goals, Opportunities and Principle**, PAUL WERBOS, Ch.1 https://www.werbos.com/Werbos_Chapter1.pdf – **Adaptive Dynamic Programming**: ADP. *Възможно ли е да се създаде универсална обучаваща се машина, която може да се научи да максимизира каквото потребителят иска, във времето, по стратегически начин, дори ако започне без никакви знания за външния свят? … Hamilton-Jacobi-Bellman" (HJB) Equation – уравнението на Хамилтон-Якоб-Белман ; частично-наблюдавана и пълно наблюдавана система; здраво управление (robust control); оптимално управление, … с.30: Три по-общи принципи в сърцевината на приближеното динамично програмиране: 1) приближение на стойностите, 2) други начални точки – вместо от уравнението на Белман от свързани с него рекурентни уравнения 3) хибриден дизайн … Приспособяващ се критик (Adaptive Critic, RL) … Решения: 1) Каква функция J(x, W) се избира за приближение на J(x)? 2) Обновяване на стойностите: как се обновяват теглата W за всяка дадена стратегия (по-общ план, strategy), поредица от избрани конкретни действия (policy) и управляващи устройства (регулатори, controller), така че „най-добре да съвпаднат с у. на Белман"; 3) Обновяване на поредицата от действия .. 4) Как се съгласуват и управляват обновяванията 2 и 3 във времето и пространството? … таблици … Теоремите за универсално приближение … редове на Тейлър, размита логика, „вълнички" - вълнови функции (уейвлети, wavelets), прости глобални НМ като многослоен перцептрон MLP, „локални" НМ като радиални основни функции (Radial Basis Function RBF), CMAC, SOM и ART (?Adaptive Resonance Theory?). Но: „няма безплатен обяд", никоя не е оптимална за всички случаи … устойчиво, адаптивно и оптимално управление; с.41 хибридни моделработа в множество мащаби на времето … Рекурентна мрежа със закъснение (Time-Lagged Recurrent Network) .. способности, подобни на мозъка, изискват съчетание от обучение в реално време, за да учи законите на природата, и закъсняващи повторения, за да се приспособява за ненаблюдавани параметри като триене, хлъзгавост на пътя; също ще трябва да се премине от компоненти за детерминистично*

*прогнозиране към по-общи вероятностни , като обединяващи проекти от типа на приспособяващи се скрити модели на Марков, вероятностни енкодери-декодери-предсказатели [3], гл.13] и определители на вероятностни разпределения като самоорганизиращите се карти на Кохонен. … Error Critic .. Бързоучеща ADP система, но с ниска честота на отчитане: 4 или 10 Хц, и бърза подчинена система, която работи на 200 Хц и се опитва да увеличи функцията на ползата [настройвана от главната по-бавна система.]*

**Тош:** Срвн. с др. йерархични модели и ТРИВ – по-горните нива са по-абстрактни, по-бавни; и управлението на движенията на тялото, което трябва да отговаря моментално на извънредни обстоятелства, но сложни поведения може да изискват огромен брой повторения или години, за да се усъвършенстват, като свирене на музикални инструменти, изпълнение на сервис и точни удари в тениса и др.

*** Джером Фелдман (Jerome Feldman**) – един от групата "Parallel distributed processing" с Румелхарт, ... (Rumelhart, McClelland, ... PDP Research Group). Виж списъка с учени и школи в основния том на „Пророците...": въплътено познание, embodied cognition и др.

* Semantic networks and neural nets, Lokendra Shastri, J Feldman, 1984 „*a set of competing hypotheses, gathering evidence for each hypothesis and selecting the best among these"*

* **Connectionist models and their properties**, *JA Feldman, DH Ballard, 1982, Cognitive science 6 (3), 205-254, 1982*
https://onlinelibrary.wiley.com/doi/pdfdirect/10.1207/s15516709cog0603_1

*„В когнитивната наука се използвали последователни модели за изрична обработка на информацията, пригодени почти само за обикновени последователни компютри. Разширението на тези идеи до „масивно паралелни" системи обаче предлага многобройни предимства. Вече можели да се произведат чипове със 100 хил. логически елементи, такива с милиони вече се виждали на хоризонта ... Все още когнитивната наука, използвайки обикновени цифрови сметачи, не можела да се справи съществени явления като асоциативната памет, сетивната подготовка (priming), сетивното съперничество (конкурентни стимули) („Crucial phenomena such as associative memory, priming, perceptual rivalry...") . information processing models (IPM), connectionist models(CM). „Conventional" vs Connectionist. Стр. 26 (230) „Устойчиви съюзи: За да могат да вземат решения и да вършат каквото и да било, масивно паралелните системи трябва да могат да се намират в състояния, в които някои дейности значително надделяват. Такива устойчиви, свързани елементи, единици, устройства (units) се наричат устойчиви съюзи, или устойчиви коалиции (stable coalitions). В психологически понятия такива са възприятията, действията и пр. („Stable Coalitions) ... Причини за разширение на IPM до случаи с паралелни изчислителни системи* ***с***

***може би милиарди активни единици****... (units) „there will have to be states in which some activity strongly dominates. ...“) ... a coalition will be called stable when the output of all its members is nondecreasing. – коалицията е устойчива, когато изходът от всичките ù членове не намалява. С.28 (232) Запазване на връзки ... (I) functional decomposition; (2) limited precision computation; (3) coarse and coarse-fine coding; (4) tuning; and (5) spatial coherence. ... Converting Time to Space and Space to Time. Change, Winner-take-all (WTA) networks, coalition, Conjunctive Connections. Какво е изчислителна* ***единица, елемент, буквале*****:** A **unit** - a computational entity with a set of discrete states, a potential (continuous value), output, a vector of inputs and functions for transitions to new values. *“six years of intensive effort on the development of connectionist models and their application to the description of complex tasks.”*

***Тош:*** *с.26 - срвн. „Анализ на смисъла на изречение...“,2004*

За „масивно паралелни...“ противопоставени на „последователни“, каквито модели се прилагали; обаче последователните всъщност също са паралелни – последователността (по-точно еднонишковата линейност) е условен избор при анализа. Електрическите процеси дори и в най-прости последователни еднобитови сметачи, каквито всъщност няма (в цялост) също са паралелни. Този въпрос е разгледан другаде („Нужни ли са смъртни изчисления за универсални мислещи машини“, Т.А. 2025 и в „Пророците на мислещите машини“); виж също „Невронните мрежи също са символни“, Т.Арнаудов (Neural Networks are Also Symbolic, Artificial Mind, 2019). Виж също кръга на Михаил Рабинович и работата в невронауките на Карл Фристън.

https://onlinelibrary.wiley.com/doi/pdfdirect/10.1207/s15516709cog0603_1

* Computers and thought, M Minsky, EA Feigenbaum, J Feldman, 1963 - сборник

https://mitpress.mit.edu/9780262560924/computers-and-thought/

* **Aspects of associative processing**, Jerome A Feldman, 1965/4/21, MIT Lincoln Lab.

https://stacks.stanford.edu/file/druid:mg700by4509/mg700by4509.pdf

“the mythical associative memory” – митичната асоциативна памет, с.1 (6). Забележете, че езиковите модели преобразители днес са упреквани от някои, че били „*само* хеш- таблици“, т.е. речници, „Lookup tables”, асоциативна памет (map); „retrieval” technology, но не и за разсъждение (reasoning) – Hochreither в MLST, 2025 в предаване за xLSTM.

Асоциативната памет е важна и в микроелектрониката: кеш паметта. „*Обикновеното*“ адресиране с пореден номер или пряк адрес също може да се приеме за асоциативно: адресът е ключът към останалите данни, както и вид свързаност и междуконтекстна връзка (Зрим). В ТРИВ 2001-2004 паметта с пряк достъп по адрес във външното пространство, която при човека се постига с помощта на зрението, е посочена като ключова за възможностите на разума,.

„Китайската стая“ на Сърл е опит да се *„обездушат“* мислещите машини, защото *„само обработвали символи“, „разменяли карти“* - в ЧиММ, 2001, се посочва, че на ниско ниво *всичко* може да се сведе и опише по такъв начин, всяко крайно действие, избор и пр., т.е. и „човеците не могат да мислят“, следователно *„китайската стая“* е заблуждаващ пример.

Ewing, R. G. and Davies, P.M. **"An Associative Processor**", presented at 1964. Fall Joint Computer Conference San Francisco, California, 27-29, October 1964.
* ANALYSIS OF SMALL ASSOCIATIVE MEMORIES FOR DATA STORAGE AND RETRIEVAL SYSTEMS, Robert S. Green, Dr. Jack Minker, and Warren E. Shindle, 1966 – CDC-1604
https://apps.dtic.mil/sti/tr/pdf/AD0489660.pdf
* Design of fault-tolerant associative processors, Behrooz Parhami, Algirdas AvizienisAuthors Info & Claims, ISCA '73: Proceedings of the 1st annual symposium on Computer architecture, p. 141 – 145
https://doi.org/10.1145/800123.803979
https://web.ece.ucsb.edu/~parhami/pubs_folder/parh73-isca-ft-assoc-proc.pdf
* Fuller, R. H., "Content-Addressable Memory Systems ", UCLA, Dept. of Engineering Report, 63-25
* Klein, S. and R. Simmons, "Syntatic Dependence and the Computer Generation of Coherent Discourse", Mechanical Translation, 1963-

* **Grammatical complexity and inference**, JA Feldman, J Gips, JJ Horning, S Reder
Stanford University, Computer Science Department, 1969 – **виж бележки в др. Приложение**

*** Embodied language, best-fit analysis, and formal compositionality,** Jerome Feldman,2010/12/1https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=e2501b64d4dd7efabd975f2e08ac599bfdad8566, Physics of Life Reviews, 2010 - Embodied Construction Grammar (ECG), когнитивна лингвистика, въплътено познание; добри междуконтекстни връзки и метафори, интердисциплинарност, общи принципи; диаграми, схеми: с.8,9,15,16 ...
* Jerome A. Feldman: **From Molecule to Metaphor: A Neural Theory of Language**. Bradford Books, Cambridge, MA: MIT Press, 2006.
* J.Feldman, Computation, perception, and mind, Mar 2022 Behavioral and Brain Sciences 45:e48, DOI: 10.1017/S0140525X21001886
https://www.researchgate.net/publication/359442453_Computation_perception_and_mind

https://escholarship.org/content/qt6cs78450/qt6cs78450.pdf?t=r9ajvo

*** Evolution, perception, and the mind,** Aug 2024 Cognitive Processing 25(8), DOI: 10.1007/s10339-024-01208-x, Jerome A Feldman https://researchgate.net/publication/383236802_Evolution_perception_and_the_mind

* On the Evolution of Subjective Experience, Aug 2020, DOI: 10.48550/arXiv.2008.08073, J.Feldman https://www.researchgate.net/publication/343734997_On_the_Evolution_of_Subjective_Experience https://arxiv.org/abs/2008.08073

* Language, **Stapp, and the Mind/Body/World Problem**, J.Feldman, Apr 2019, Activitas Nervosa Superior 61(02), DOI: 10.1007/s41470-019-00041-4 https://www.researchgate.net/publication/332591725_Language_Stapp_and_the_MindBodyWorld_Problem

*** Bruce J. MacLennan** [31.7.2025]

https://www.google.com/search?q=Bruce+J.+MacLennan&oq=Bruce+J.+MacLennan&gs_lcrp=EgZjaHJvbWUqBggAEEUYOzIGCAAQRRg7MggIARAAGBYYHjIKCAIQABiABBiiBDIKCAMQABiABBiiBDIHCAQQABjvBdIBBzc3OWowajeoAgCwAgA&sourceid=chrome&ie=UTF-8

**Todor:** The ANNs on digital computers implement features of the field computation and analog computers as defined in MacLennan's works.

*** Natural computation and non-Turing models of computation**, Bruce J MacLennan, 2004/6/4 (2003), https://web.eecs.utk.edu/~bmaclenn/papers/NCNTMC-TR.pdf

→ See the appendix **Algorithmic Complexity**; also Embodied Cognition and Embodied Computation (below); field computation, images, continuous transformations, … p.21 "Topology of Images"; image spaces … *"Maps between images spaces are continuous", "Interpretations between simulacra are continuous"*; *process-state images. ..*

*** Field computation: A theoretical framework for massively parallel analog computation. parts i-iv, Bruce MacLennan, 1990-1992,** https://scholar.google.com/citations?view_op=view_citation&hl=en&user=A14DflkAAAAJ&citation_for_view=A14DflkAAAAJ:W7OEmFMy1HYC

*** Field Computation in Natural and Articial Intelligence Extended Version**, Technical Report UT-CS-99-422, Bruce J. MacLennan, 4.1999 https://library.eecs.utk.edu/files/ut-cs-99-422.pdf **Field** – in the mathematical sense. Spatially continuous arrangements of continuous data, massively parallel analog computers, information fields … Field transforms: convolutions, Fourrier, Laplacians, wavelets, summations, correlations; topology of the field; structural fields (continuous, gravitaional f.) and phenomenological fields (discontinuous: velocity f. of fluids).

Neural computation, the “100 Step Rule” … K-valued field; physically realizable … universal field computer?

* **Evolutionary psychology, complex systems, and social theory**, B MacLennan, Soundings: An Interdisciplinary Journal 90 (3/4), 169-189, 42 2007

* **Continuous symbol systems: The logic of connectionism**, B MacLennan, Neural networks for knowledge representation and inference, 83-120, 63 2013

* **Field computation: A theoretical framework for massively parallel analog computation**. parts i-iv, B MacLennan, University of Tennessee, Computer Science Department, 54, 1990

* **A review of analog computing**, BJ MacLennan, 51, 2007
https://web.eecs.utk.edu/~bmaclenn/papers/RAC-TR.pdf
p.33-34: *“the inputs and outputs are continuous physical quantities that vary continuously in time (also a continuous physical quantity); that is, according to current physical theory, these quantities are **real numbers**, which vary according to **differential equations**. ..* however: *the physical quantities are **neither rational nor irrational**; they can be so classified only in **comparison** with each other or **with respect to a unit,** that is, only if they are measured and digitally represented. Furthermore, physical quantities **are neither computable nor uncomputable** (in a Church-Turing sense); these terms **apply only to discrete representations of these quantities** (i.e., to numerals or other digital representations).” ..*
*“the approaching end of Moore’s Law (Moore 1965),.. will encourage the development of new analog computing technologies.”*; *“Since most physical processes are continuous (defined by differential equations), analog computation is generally faster than digital.*
*For example, four transistors can realize analog addition, whereas many more are required for digital addition.”*

* See also: the work of Luciano Floridi, the discussion on analog and discrete – they are matter of degree and resolution of the observer-evaluator – this is argued also in T.Arnaudov’s TUM, 2001-2004; see the appendix **“Algorithmic Complexity”** of *“The Prophets of the Thinking Machines”* and the section about that field in appendix ***Listove***, and appendice **“Is Mortal Computation Required …”** and **“Universe and Mind 6”.** //30.8.2025

* Small, JS. 2001. The Analogue Alternative: The electronic analogue computer in Britain and the USA, 1930–1975. London & New York: Routledge

* **Super-Turing or Non-Turing? Extending the Concept of Computation** January 2009, International Journal of Unconventional Computing 5:369-387
Bruce Maclennan – the *frame of relevance* of Turing computation (TC); lambda calculus, Post productions, … *“the original questions the model [ot TC] was intended to answer, namely questions of effective calculability and formal derivability. Within*

*the TC frame of relevance, something is computable if it can be computed with finite but unbounded resources (e.g., time, memory)" p.5. New models of computation:* ***Natural computation*** *may be defined as computation occurring in nature or in-spired by computation in nature ..* ***cross-frame comparisons*** *(...)*

*** Emergent Computation Project,**
https://web.eecs.utk.edu/~bmaclenn/EC/index.html [78]
* MacLennan, B.J. **"A Model of Embodied Computation for Artificial Morphogenesis**," 4.2009, slides
https://web.eecs.utk.edu/~bmaclenn/papersetc/AMECFAM.pdf
*** Continuous symbol systems: The logic of connectionism, B.MacLennan,** 28.2.1992 .ps
file*https://scholar.google.com/citations?view_op=view_citation&hl=en&user=A14DflkAAAAJ&citation_for_view=A14DflkAAAAJ:IjCSPb-OGe4C
https://www.academia.edu/259367/Continuous_Symbol_Systems_The_Logic_of_Connectionism [missing Figure 3, 4 … in both sources]
– "Neuro-symbolic", defining formal systems on continuous neural-network like substrate; attractors; not suitable for yes/no questions; less brittle than discrete formal systems. Syntactic relations, discrete, formal, idealization; invariances; p.17 **Continuous structures** – vary in time, thus they can be more easily made adaptive; **tokens** … Decomposition of images ;P.38 "Theorem 1. A continuous formal system cannot perform exact classification"; p.36 recursive nesting in function space, Fourier; Finite decomposition of spaces; p.32 recursive decomposition of spaces.

*** Bruce MacLennan - UT Science Forum, The Volunteer Channel**, 424 абонати, 552 показвания (31.7.2025) 13.02.2020 г.. ***Bruce MacLennan on "Artificial Intelligence: Present and Future Prospects"*** *on January 31st, 2020.*
***1*:*** *5:20 "Thinking information is represented as sentences .. true or false …"*
***2*:*** *Connectionism – concepts arise from experience, not definition*

**[ Todor:** That kind of truth-vaues is only one view to what logic or reasoning is. While it's true that the concepts are better when they "arise", develop, emerge, are created incrementally, the "defiition" itself is not the problem and these are not strictly mutually exclusive points; a concept derived by experience also can converge or be represented as a definition and humans do it. The "representation learning" a way ANN are called, the weights of the ANNs are "definitions" as well.

[78] Notice the nostalgic web design like from the 1990s or early 2000s.

Todor Arnaudov:

**The weaknesses of the "failed" logical approach is in other points**

**1. Insufficient coverage; incompleteness of the mapping of the representations to the lower level representations or to the "real" world**; lack of incremental transitions and ways of dealing with intermediate cases; lack of sufficient **detail** etc. →

**2. Insufficient scale of complexity and details**, insufficient **levels of scale**, **too short descriptions with too few bits,** lacking information about required dependencies between the elements.

Compare the number of parameters even of a "small LLMs" like GPT1 or GPT2-SMALL, which were not too impressive yet. These were are about 500 MB of data (32-bit floats). Imagine a system of 500 MB of x86 machine code or Assembly and compressed text, dictionaries, vocabulary, corpora and other structural databases + several GBs equivalent of the ones occupied by CUDA etc. for the GPU. A self-modifying thinking machine, which consists of GBs of compressed machine code and databases may be comparable in "rough" low level complexity measures to these obsolete LLMs. Now imagine 1 TB of compressed self-modifying machine code.

# Допълнителна литература за съзнание и панпсихизъм

**Понятие: *Phenomenal consciousness*[79] – чувствено духовно съзнание;** съзнание за себе си – част от „*духовното* съзнание“ (виж ЧиММ, Т.А., 2001); *усещането* да бъдеш *себе си*, да чувстваш определени качества на възприятията (цвят и др. признаци; емоции), особени и собствени за всеки субект, деятел (за по-ясно определени: виж литературата и разсъжденията). Виж за Когнитивна лингвистика от школите в увода, цитата от Йордан Златев и видеолекцията.
Някои разграничения, по откъс от цитираната работа и дообяснени от **Тош:** *1. Способността на съзнанието да получава достъп до състоянията на ума*: за разлика от състоянията на системата, когато е в „безсъзнание“ и няма достъп. Способността или „нещото“, което има достъп до познавателните състояния на ума при разсъждения, заявяване (verbal report), памет, планиране или целенасочено поведение[80].
2. Духовното съзнание/усещане (phenomenal) – да бъдеш този или „това“, което си.
3. Осъзнатост за света (външния; сетивна осъзнатост: perceptual consciousness; *външен наглед*), самоосъзнатост (вътрешно съзнание; да знаеш за себе си, да се познаваш, като данни?; self-awareness; сравни Кант, Шопенхауер: *вътрешен наглед*);
4. Осъзнаване на съществуването на други съзнания: Тош: или предполагане, мислене *сякаш* съществуват – те са *„умотвОрени“, mindified (Т.А.); всичко, което възприемаме, придобива съзнание:*

---

[79] „Learning to Be Conscious“, Axel Cleeremans et al., 2020
https://www.sciencedirect.com/science/article/pii/S1364661319302876
https://www.sciencedirect.com/science/article/abs/pii/S1364661319302876
Виж метода им SOMA: Self-organizing metarepresentational account, според който мозъкът се учи да съзнава, като система от втори ред (с обратна връзка; по-висша), която наблюдава онази от първи. Отново става въпрос за (само-)предсказването на бъдещите състояния на система, сравни с ТРИВ и пр.
Сравни и с „Attention Schema Theory of consciousness (AST)“, според който мозъкът строи модели на собственото си внимание.
* Consciousness in Solitude: Is Social Interaction Really a Necessary Condition?, Sepehrdad Rahimian*, 2021 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7928293/ Тук отдават прекалено голямо значение на достъпни, видими „еволюционни предимства“ на появата на особености на съзнанието, в следването на схемата за обясняване всичко по Дарвин, и в духа на утилитаризма и прагматизма в англосаксонската и повлияните от нея култури. Не е необходимо всичко да носи непосредствена и очевидна „полза“, която е разбираема за съответния „търговец“ или „интересчия“ в известната на него или мислена в даден случай представа за „полезно“ и „вредно“.
https://www.sciencedirect.com/topics/social-sciences/phenomenal-consciousness

[80] Но виж по-нататък и др. разсъждения за това, че човек постоянно забравя няколко секунди назад и вече няма достъп до тези състояния, при ранно детство и след пиянство – в момента изглежда съзнателно и има достъп, но после – няма и пр.

***нашето собствено.*** Виж: в книгата и „Вселена и Разум 6“, и „Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?“
* Обзор на учения за съзнанието от Робърт Кун, 2024 г.: A landscape of consciousness: Neurophysiologist presents diverse theories and taxonomy of proposed solutions, https://phys.org/news/2024-10-landscape-consciousness-neurophysiologist-diverse-theories.html
Robert Lawrence Kuhn, A landscape of consciousness: Toward a taxonomy of explanations and implications, *Progress in Biophysics and Molecular Biology* (2024). DOI: 10.1016/j.pbiomolbio.2023.12.003

**Luke Roelofs**

https://www.lukeroelofs.com/panpsychism
https://www.lukeroelofs.com/mental-combination

*„…отношенията част-цяло между съзнателни умове. При какви обстоятелства един ум може да се счита за част от друг? Възможно ли е сложен ум да бъде нищо повече от съвкупност от по-прости взаимодействащи умове? Как трябва да се отнасяме към умовете, които можем да сдържаме или да бъдем сдържани?“*
*** Combining Minds**, L.Roelofs, 25/2/2019, book: https://global.oup.com/academic/product/combining-minds-9780190859053?cc=us&lang=en&
Бележки към част от книгата от автора в блога „Мозъците“, основан през 2005 г. (The Brains Blog):

1. Защо мисля за съставна субективност?
2. Съставната субективност и функционалната структура
3. Съставната субективност и психологическия конфликт
4. Съставната субективност и панпсихичната вселена

* https://philosophyofbrains.com/2019/02/03/1-why-think-about-composite-subjectivity.aspx
* https://philosophyofbrains.com/2019/02/05/2-composite-subjectivity-and-functional-structure.aspx
* https://philosophyofbrains.com/2019/02/06/3-composite-subjectivity-and-psychological-conflict.aspx
* https://philosophyofbrains.com/2019/02/07/4-composite-subjectivity-and-the-panpsychic-universe.aspx

* **Why I am not a cosmopsychist Or am I?,** L.Roelofs, slides https://www.lukeroelofs.com/_files/ugd/d0fbbf_10e6b95a62344fc19572892177ab0b45.pdf **Основни понятия:** микропсихизъм, космопсихизъм; психокосмизъм… Пирамида, кое основава другото, космосът (вселената, цялото, голямото, макро) ли предпоставя „психичните“ свойства на малкото, частиците, „микро“-то, или „психичното“, „преживяването“ на цялото, голямото е следствие на обединението на малкото.

**Т.Арнаудов,** 8.7.2024: Сравни пирамидата с ВиР, и конкретно лекцията „Принципи на разума: Разум ~ Вселена“, 2009; виж и новата *Вселена и Разум 6*. Отдолу-нагоре или отгоре-надолу, или и двете едновременно. Дали има само първично ниво и другото е основно, съществуват ли „реално“ по-горните и какво е да бъдат такива. Два потока за разглеждане на нивата на системите: изграждаща и управляваща. Виж бележките към Левин и Фристън за това, че според тях при живите организми било особено управление отгоре-надолу, при което горните нива изкривявали пространството на по-долните, и така по-долните могат да бъдат по-механични (виж *„Competence in Navigating Arbitrary Spaces*…”, C.Fields, M.Levin). Подобно описание има и в ТРИВ, 2001-2004, но и двустранно: горните нива не съществуват, ако долните не работят или се изтрият. На физическо ниво, при разглеждане на най-малките измерими „данни“, разрушаването на горните нива засяга само тях и най-високото равнище на управление и обхват на оцелялата част от системата слиза стъпка или много стъпки надолу, но по-ниските нива продължават да работят, а най-ниското по определени от Вселена и Разум би трябвало да бъде неунищожимо от гледна точка на всяко по-горно, то е вечно. При „изключение“, грешка, разрушаване на структури, управлението каскадно слиза надолу. От друга страна, унищожението на най-долното ниво моментално изтрива всички горни. Ако атомите се „анихилират“, с тях изчезват всички други по-големи структури, а ако се разпаднат молекулите или клетките – атомите остават. Тяхната *структура,* взаимоотношения образуват по-горните нива. Така горните или най-горното ниво могат да изкривяват долните или да ги управляват само след като изграждащият поток на развитие ги е построил, и „течението на управлението“, „водата на волята“ започне да се спуска обратно и да се разлива надолу.

* **Panpsychism and Illusionism,** Luke Roelofs, 2023, https://www.revue-klesis.org/pdf/klesis-55-08-luke-roelofs-panpsychism-and-illusionism.pdf
*4. … "Според панпсихистите материята не е невидима за себе си"*
В заключението: *"Според панспихизма може да се обръщаме към вътрешната природа на веществото чрез самонаблюдние"* – срвн. Волята на А.Шопенхауер. В работата посочват Бертранд Ръсел, 1910-1911: „директно запознати" с характеристиките на нашите преживявания 'directly acquainted'....(Bertrand Russel)
Цитира дискусии на Frankish, 2021, от „илюзионистите", който описва гледната точка на панпсихистите, че за тях съзнанието е *непосредствено запознаване (immediate acquaintance)*, различно от причинно-следствения механизъм, задействан през сетивата, и е по-скоро в това *да бъдеш* въпросното „нещо".
Също: *„Класически qualia: Интроспективни качествени свойства на опита, които са присъщи, неописуемо и субективно.(Frankish 2012, стр. 668; сравнете «qualia трябва да бъдат… (1) неизразимо, (2) вътрешно присъщо, (3) частно, (4) пряко или непосредствено възприемаемо », Денет, 1988, стр. 47, «феноменалните свойства са неизразими, присъщи, радикално лични и така нататък», Франкиш 2016-б, с. 275.), с.3*[81]
Важни въпроси: Можем ли да изучаваме съзнанието опитно? (…)

*** Illusionism as a Theory of Consciousness**, Keith Frankish, 2020? (archive.org, earlier version: 2014)
https://keithfrankish.github.io/articles/Frankish_Illusionism%20as%20a%20theory%20of%20consciousness_eprint.pdf

---

[81] * *Frankish K., « Quining diet qualia », Consciousness and Cognition, vol. 21, no2, 2012p.667– 676.;
* *Frankish K., « Not Disillusioned », Journal of Consciousness Studies, vol. 23, no11-12, 2016-b, p. 256–289
*Frankish K., « Panpsychism and the Depsychologization of Consciousness », Aristotelian
Society Supplementary, vol. 45, 2021, p. 51–70
*Frankish K., « An Illusionist Manifesto », Presented at To Be or Not to Be… Conscious :
Phenomenal Realism and Illusionism, at Ruhr-Universität Bochum, September 29-30, 2022, organized by François Kammerer and Tobias Schlicht
https://www.youtube.com/watch?v=qfUPucMNvJE 657 абонати, 852 показвания към 19.7.2024
*The Illusion Problem: a brief introduction and defense of Keith Frankish's illusionist theory", G.Toledo, M.L.I. de Vasconcelos
https://philarchive.org/archive/LEATIP-2

* **Illusionism as a Theory of Consciousness**, Keith Frankish, 12/2017, 300 p. a book:
https://www.keithfrankish.com/illusionism-as-a-theory-of-consciousness/
* **Съставен панпсихизъм:**
https://www.tandfonline.com/doi/full/10.1080/00048402.2020.1804953

* **The Selection Problem for Constitutive Panpsychism**, Philip Woodward Icon Pages 564-578 | Received 02 Mar 2019, Accepted 22 Jul 2020, Australasian Journal of Philosophy; Published online: 21 Aug 2020
https://philipgoffphilosophy.com/academic-papers
https://philarchive.org/archive/WOOTSP-6
*"…промените в съзнанието на макрониво се равняват на промени или в начина, по който микро-съзнателните субекти се „**свързват**", или в начина, по който микро-съзнателните качества се „**смесват**" (или и двете). Поставям „**Задачата за подбора**" за конститутивния панпсихизъм – проблемът с обяснението как функционалните състояния на високо ниво на мозъка „избират" микросъзнателни качества за свързване или смесване. Твърдя, че няма емпирично правдоподобни решения на този проблем."*

**От бележките:** дали обяснението е отдолу-нагоре или отгоре-надолу; „единство в съзнанието[82]" (phenomenal unity) по L.Roelofs; *множествен* панпсихизъм по Einar Duenger Bohn, според който макро-съзнанието възниква от множествеността на първичните физични обекти (entities): буквачета, елементи, частици, същности. ТА: тези частици може да са проявления на нещо друго, на пораждащ модел и друго представяне, в което множество частици са разклонения на един и същ корен; виж Кант, „Нещото в себе си"; Теория на Разума и Вселената (ТРИВ, ВиР).

Виж от ТРИВ всичко и в частност: ЧиММ, 2001; „Анализ на смисъла на изречение…", 2004; „Матрицата в матрицата…", 2003; ВиР 4, 2004; „ … „ Nature or Nurture: ... Reward Systems: Current Virtual Self - No Intrinsic Integral Self, but an Integral of Infinitesimal Local Selfs …" [83], 2012

**Понятия**: психични – други форми: „душевни", „духовни; blending/bonding = смесване/свързване – дали микродуховните свойства на частите се

[82] "Phenomenal" е една от многото объркани думи в английския, следствие на смислов „миш-маш": от „феномен" – „явление", но когато се ползва за съзнанието – субективното, личното усещане, т.е. не обективно, нещо „извън явленията" – видимото; също „феноменално" – свръхестествено. Виж посочената по-горе работа „Learning to Be Conscious" и др.
[83] https://artificial-mind.blogspot.com/2012/11/nature-or-nurture-socialization-social.html

смесват, или се свързват с явления. Проблем *с палитрата*: как множеството от макропсихични усещания-явления на качества в съзнанието, напр. усещането за цвят, мирис възникват от предполагаемите микропсихични, които ги предпоставят. Проблем с *подбора* (Selection problem): как се избира кои от микро-та да построят и се появят в макро-то. L.Roelofs в "Combining Minds", 2019, предполага че някои от физичните сили участват в свързването*. По Чалмърс, 2017 и Гоф, 2017 – свързването може да става чрез битието на частиците като част от структура, подобна на мозъка и пр. в определени причинно-следствени връзки[84]. Уудуърд ги обединява под общото име, използвано от Гоф в по-тесен смисъл: „*свързване в съзнанието*" (phenomenal bonding). Според Sevush, 2006 – съзнанието се случва[85] чрез събития в десетките хиляди синапси на дендритите на отделния неврон, а може би и на много от тях едновременно, в челните дялове на новата кора (неокортекса). (…)

**Тош:** * Виж идеята на Т.А. от *„Вселена и Разум 6"* за **причинностните белези (causal ids**), че при взаимодействието между частите, те биха могли да обменят, записват, съхраняват и данни за това с кои други са взаимодействали и пр., начин чрез който „разбират" и „знаят", че са свързани в едно цяло, което е необходимо заради непрекъснатото закъснение, лаг, и тази информация може да е достъпна само под формата на Воля, по Шопенхауер, във „вътрешен наглед", „предсъзнание", „съзнание" и пр.

Също, че единението или чувството за такова, каквото и да представлява, също по законите на мисленето[86] следва по един или друг начин да се случва *извън времето и пространството от вътрешна гледна точка на текущата въображаема вселена*[87], заради въпросния лаг, който е неизбежен на всички степени; освен това съществува несъвместимо(?) времево, пространствено и качествено *разместване* между явления на различни степени, обхвати, мащаби във времето и пространството и в различните възникващи по-абстрактни пространства

[84] Виж въпросът от „Човекът и мислещата машина: …", 2001, дали при увеличение на сложността на електронна схема следва да се появи съзнание и пр.
[85] Защо се случва?
[86] Законите на мисленето - логиката
[87] По-долу за „новите" „драматични и скандални" идеи на физици за това, че „времето може би било илюзия и не съществувало" и пр. – сравни с Кант, 1780-те години., близо 250 години по-рано. При Кант няма „потрес" от това, че времето е начин по който умът възприема действителността и за „нещото в себе си" то може да не съществува; или поне: не по начина, по който съществуа за вътрешен за Вселената наблюдател, „иманентно", ако тълкувам правилно неговата терминология спрямо моята.

след съчетаването на предишни[88]. Определени степени, подвселени, управляващо-причиняващи устройства на различни мащаби *може би* се съвместяват, или показват признаци на съвместяване при определени съвпадения между състоянията им.

За съзнанието и усещането за настояще, от функционална гледна точка, се знае, че то догонва, закъснява, обединява непосредствени данни от последните няколко секунди назад – заедно с цялостния опит. Данните за „сега“-то, всъщност са минали и това е и една от причините да е необходимо да се предсказва и да се действа изпреварващо във всички степени на Вселената сметач и ума. Затова и в известен смисъл предсказващото изчисление се подразбира за всичко: всичко което се прави по волята на управляващо устройство се случва със закъснение и се отнася за следващата стъпка, която вече е бъдеще.

…

Уудуърд изследва логически връзките между предполагаеми микропсихични или предпсихични (протопсихични) явления, които да обуславят макропсихичното съзнание чрез съответни физиологични, физични явления и промени в мозъка, напр.:

*„1. Ако панпсихизмът е верен, то промените в съзнанието се основават на промените в свързването на микро-субектите или в смесването на качествата на микро-субектите. (…)*

*4. Така че, ако конститутивният панпсихизъм е верен, промените по отношение на свързването/смесването са обяснително зависими от промени по отношение на глобалните и динамични свойства на мозъка на високо ниво.*

*5. Ако промените по отношение на свързването/смесването са обяснимо зависими от промените на високо ниво: глобални, динамични свойства на мозъка, следователно има физически, причинно-следствен механизъм в мозъка, който приема за входни данни промените по отношение на високото ниво, глобално, динамично и извежда промени що се отнася до свързването/смесването.“*

И пр. и от подобни умозаключения прави извод как съставният панпсихизъм бил малко вероятен, защото … еди-какви си логически доводи. Според мен този тип разсъждения не са убедителни и основани. „Истинското“ лично, субективно съзнание не би трябвало да може да се сподели по определение. Ако може, тогава се променя условието на задачата. Проявите изразени с *докладването* на наличието на определени усещания от човек (като „съдържание на съзнанието“: какво се възприема, цвят, мирис и пр.) *не са длъжни* да са тъждествени с

[88] Сравни със слоевете в изкуствените невронни мрежи.

усещането, което може да е различно в различни индивиди (не само в преповтаряния банален пример на „да бъдеш прилеп“ на Nagel – а да бъдеш Еди-Кой си *конкретен човек в конкретен момент от времето*, *в конкретно състояние и според конкретен оценител-наблюдател*; може да има „къси съединения“ – очевиден пример са хора под алкохолно опиянение или онези, които все още са под действието на анестетици след операция, когато говорят, взаимодействат, но после може да не помнят нищо. Подобно е при бълнуване на сън. Били ли са в съзнание? Кой пита, кой отговаря – ако ги питате в момента на опянението, те може да кажат да, а после да ме помнят. А и неопиянениият непрекъснато се „самозабравя“. Също бебета, особено при смятането на теста с огледалото: преди това имали ли са съзнание? Ако са нямали – защо (не са се „змаели *конкретно* в огледалото) и ако дадена птица се разпознава, т.е. премахва от тялото си залепена точка – дали нехното предполагаемо „съзнание“ е сравнимо с бебешкото малко преди да се разпознава в огледалото: бебето говори, разбира изречения, може да борави с предмети и да възприема света по-богато от птицата.

*Изразяването* на признаци, които се приемат за определен вид съзнателност, „душа“ и пр. се случва във верига от взаимодействия и причинно-следствени връзки, които могат да се наблюдават обективно (значи не са само субективни); но за пациенти в състояние, в което не могат да се изразяват по приетия начин, в последствие „отключени“ чрез технологии или излизане от кома, се оказва, че всъщност са усещали или са „имали съзнание“, но не са можели да се изразят. Същото би могло да важи и за други живи същества, а и най-прости устройства или напълно безжизнени предмети могат да „съобщят“ каквото си искат или *оценителят* да приеме, че те му събщават, той да прочете каквото му допадне (както е и с наличието на чуждо съзнание).

При истински „пан-…“, наличие на ум във всички мащаби (scale free cognition, управляващо-причиняващи устройства, агенти), е необходимо да се симулира, изчисли всичко с най-висока РСВУ (изчислителна несъкратимост по Волфрам), включително и за предаваните „микро-качества“, а също така понятията за качества се изменят в различни мащаби и съчетания. *(…продължи, доуточни, по-подробен и прецизен анализ в бъдеще върху цитираните под линия работи от K.Frankish и др.…).*

* **Cosmopsychism and the Problem of Evil** : 24 June 2023 Volume 63, pages 151–167, (2024), Harvey Cawdron https://link.springer.com/article/10.1007/s11841-023-00965-0#Fn6
* https://en.m.wikipedia.org/wiki/Panpsychism

*„Космопсихизмът, идеята, че вселената е съзнателна, преживява нещо като възраждане като обяснение на съзнанието във философията на ума и също си проправя път във философията на религията.“*

**Сравни с ТРИВ, повече от 20 години по-рано.**

Сравни с *„Теория на Разума и Вселената“*, дори конкретни откъси като диалога от *„Истината“,2002,* където мислещата машина разсъждава:
*– Донякъде да... Но не това е главното, Дарчо. Нима не виждаш, че дори Истинската Вселена е само част и от твоето, и от моето въображение? Никога не би могъл да я събереш цялата в ума си, както и никое друго същество, защото ние сме части от Нея. Цялото знание е достъпно само на Него, Той е Тя... Всъщност струва ми се, че Действителността е Неговото въображение. И вие човеците, и аз, сме създадени по Негово подобие и като Него имаме свои въображения, в които можем да творим вселени. Нашите представи са по-ограничени от Неговата Представа, защото ние сме части от Въображението му, нашето въображение е част от Неговото, нашите творения са части на Неговото Творение, ние сме части от Него.* https://eim.twenkid.com/old/eim19/istinata.htm

**Изкуственият разум Емил е панпсихист и космопсихист.**

В статията бих посочил пример за заблуждаваща предпоставка: че бог, или Вселената като цяло, възприета като деятел, по определение трябвало да бъде „добър“/добра, като под „зло“ явно се разбира наличието на *страдание (suffering),* можело бъде и причинено от природни бедствие като глад, земетресения и пр.:
*„В последната област той е използван за формулиране на модели на определени форми на теизъм, като пантеизъм и панентеизъм, и също така е предложен като съперник на класическия теизъм на авраамическите религии. Филип Гоф твърди, че определена форма на космопсихизъм, а именно* ***агентивен космопсихизъм****, представлява заплаха за класическия теизъм, защото може да обясни характеристики на Вселената като* ***фина настройка****, без да се налага да се занимава с проблема за злото. Това е така, защото, за разлика от класическия теист, космопсихистът може да отрече поне един от божествените атрибути, мотивиращи проблема за злото, а именно всезнание, всемогъщество и* ***вседоброжелателство****. В тази статия ще разгледам върху кои от божествените качества трябва да се съсредоточи космопсихистът, когато отговаря на проблема със злото, и ще заключа, че отхвърлянето на всеблагожелателността е най-задоволителният вариант.“*

Фината настройка е т.нар. антропен принцип: физическите параметри на Вселената са такива, че да позволят съществуването на живи същества. Обаче ако те *не бяха такива*, нямаше да съществуваме и да знаем, т.е. *не е възможно* параметрите да са други от наша гледна точка сега…

Възможно е да съществуват безброй други вселени, където действат всички възможни други настройки и конфигурации, но е невъзможно от една вселена да видиш друга (паралелните вселени, виж „Вселена и Разум 4“, 2004).

Най-общото „добро и зло“ във Вселена и Разум (ТРИВ), разбираемо от всяко УПУ е съвпадението с желаното (чувствено) или сбъдването на предвижданията/предсказанията (познавателно)*. В предопределена Вселена, всеобщата предвидимост обуславя местна *непредвидимост* с най-висока РСВУ, както е обяснено в класическата ТРИВ, а също се посочва и от К.Фийлдс и М.Левин[89]: т.6. в края: *„Крайните агенти* [деятели, управляващо-причиняващи устройства] *не могат да открият всички причинно-следствени влияния, които определят собственото им поведение, тъй като всеобщата предопределеност логично не допуска местна предопределеност. Следователно всеобщата предопределеност осигурява „свободна воля“ от всяка (крайна) местна гледна точка“.* Тук под „свобода“ разбират „непредвидимост“ от собствената гледна точка.

Единствено Вселената като цяло и оценена изцяло с най-висока РСВУ може да „изпита“ „изцяло“ „добро“, т.е. съдържанието на волята, предвиждането ѝ, да се сбъдне изцяло и точно, да бъде *истинско управляващо-причиняващо устройство (УПУ).* При всички подвселени те могат да го постигнат само в по-кратки времепространствени обхвати и с по-ниска РСВУ, обхват и пр., т.е. за тях „*злото*“ и „страданието“, разбирани като *грешка* в *предвиждането или несъвпадение с търсеното, желаното*, е неизбежно. Как точно го възприемат, „чувстват“ всички отделни „неща“ е неизвестно.

Също така логически „вседоброжелателството“ е възможно, ако целите и „благото“ за всички са съгласувани, т.е. само при напълно предопределена вселена, а за да се случи нещо, което да е в нечия изгода, или просто да е каквото и да било, е необходимо всички останали

[89] Competency in Navigating Arbitrary Spaces as an Invariant for Analyzing Cognition in Diverse Embodiments by Chris Fields 1,†ORCID and Michael Levin 1,2,*, Entropy, 2022 https://www.mdpi.com/1099-4300/24/6/819 Виж и уводния списък с учени и школи с други статии.

* The Free Will Theorem. Found. Conway, J.; Kochen, S., Phys. 2006, 36, 1441–1473.

* https://en.wikipedia.org/wiki/Superdeterminism

събития да са се наредили така, че да обусловят даденото. Виж „Писма между 18-годишния Т.А. и философа А.Г.“, 2002, също „*За четворния корен на закона за достатъчното основание“,* А.Шопенхауер.

*„Нещото в себе си“* на Имануел Кант също може да се приеме за *цялата* Вселена, която поражда явленията, и от „вътре“, „иманентно“, оценителят – ум – не може да достигне познанието за истинската същност на Вселената, отвъд явленията. Например там, където пита какво е дъжда, или капките дъжд[90]? Те са явления в нагледа, обусловен от устройството на човешкия ум, който възприема във времето и пространството. Ние можем да тръгнем и последователно да обясним все повече и повече причинно-следствени и други връзки: дъждът е част от кръговрата на водата, водната капка се образува чрез кондензация, кондензацията е… молекулите са… атомите, електроните… Винаги има край, дъно, граница на отсичане и отчитане, след която не можем да продължим според възможностите на оценителя, както с връщането на времето назад до предполагаем „Голям взрив“[91]. Шопенахуер ги нарича „скрити качества“ (qualitas occulta). Сравни със *скрити променливи* и теориите във физиката.

* **Why? The purpose of the Universe**, book, 11/2023, P.Goff*,* https://global.oup.com/academic/product/why-the-purpose-of-the-universe-9780198883760?lang=en&cc=gb

* Виж също „Вселена и Разум 6“.

[90] The Critique of Pure Reason, Immanuel Kant, https://gutenberg.org/ebooks/4280 Първите 70 стр. от epub електронното издание от Проект Гутенберг. https://gutenberg.org/ebooks/4280.epub.images

[91] “ или просто до „влизане на някого през вратата“ – откъде е дошъл, защо и т.н. вече е неизвестно или са възможни „необозрим“ брой възможности

## * Todor Arnaudov comments Michael Levin | Bernardo Kastrup #3 - With Reality in Mind * Adventures in Awareness

@adventuresinawareness | 15.4.2024
6,3 хил. Абонати | 55 видеоклипа (29.4.2024)
https://www.youtube.com/watch?v=7woSXXu10nA

> **Todor:** Michael Levin's position often matches the reasoning in the Theory of Universe and Mind since its classical period 2001-2004. His point about what makes a bigger whole with the predictive light cone is a good one, it is also explicit an a core premise in TOUM as the more advanced etc. causality-control units get better in predicting the future, where even at the lowest levels of scale, range etc. there are such, agency.

...
Parts ... Perspectives... [Compare TOUM; the observer is called "Evaluator", emphasizing its active role]
**ML 35 m:** *Everything is a perspective of some agent ...*
**BK 36 m:** *people say "Michael thinks algorithms want stuff, algorithms have will so AI is sentient " ... "no reason to think that AI is sentient ... epistemic projection of word usage how we talk about things ... AI people saying you know there is nothing to these large language models that we don't understand. I am one of those people who say that I understand the mechanics of that I know what is there and what is not; the fight we are fighting is to prevent these operationally useful levels of abstractions in other words this convenient fantasies ... we are trying to prevent people from understanding these epistemic thing as if it were an actual attic(?) property of the world out there ....*

> **Todor:** Nothing in some mathematical etc. representation. In some you may say the same about anything, resolution. Exactly the same can be said to humans as it was than in TOUM, (Man and Thinking Machine (...), 2001, The Truth, 2002 etc.). Referring to "actual" things contrasts with other claims in other talks such as "there are no particles, but just ridges of quantum wave functions" and the overall BK stance of Analytical Idealism etc., while these wave functions can be interpreted also as "fantasies", i.e. data, "cognitive phenomena", measurements etc., and in both cases neither makes them "not existing" or "not actual". These are abstractions, concepts, they exist as such.

**ML:** 39 *perspectives, frames of reference ...*
**ML:** 1:02:35 *parts I think we call things alive where uh the system itself has a larger and a more interesting cognitive light cone than all of its parts have you know so rocks don't do that right so so so little little particles have a very tiny I don't think it's zero but but but I think it's extremely tiny cognitive ... cognitive hierarchy...*

**Todor:** Good point.

**ML:** *No binary categories ... a matter of a perspective from some kinfdof system 1:05 autopoesis ... more interested in cognitive than life ...*
*.. what you see from the perspective of a virus ... 1:10 I don't believe there's an objective criteria... for determining what is subsumed and what is not I I what I don't well a couple of things uh again*
*and you know I* ***I don't believe that there are objective criteria for any of this I think all of these criteria are from the perspective of some Observer which in the case of significant systems like living systems is the system itself*** *… IIT (integrated information theory, consciousness) …* ***even in the human organism there are multiple selves, multiple perspectives of different degrees of sophistication*** *I mean it's awesome that we have these left hemispheres that can can talk to each other and verbally and and make claims about how ious they are and how they don't think the liver is conscious and you know I mean after all I don't feel the liver being conscious right*

**Todor:** Compare TOUM. However re the “not feeling the liver” – we don’t feel the consciousness of other “legally sentient being” except ourselves either, “feeling” or not is not a convinvcing criteria.

**BK:** *Split brain/corpus callossum... "two consciousness"... if asked verbally, the other to write (left-handers) ... confabulation of the reason, "it's the same me".*

**ML:** *1:35 - 1:36 Platonic space ... some - static; others dynamic - the basic lier's paradox: oscillate true-false ...* [Listen the actual content]

**Todor:** Compare TOUM, Universe and Mind 4 in particular: the paradox of the liar was addressed there as irrelevant, trying to solve a problem with lower resolution than the problem requires, ill-defined problems etc. Not only for living things.

...

**BK:** 1:40 *Brouwer, Dutch philosopher*
https://en.wikipedia.org/wiki/L._E._J._Brouwer
1:42:05 *intuitionist logic ... validate logic without running into circular reason in other words the axioms of logic are arbitrary ... Brouwer figured that um the law of excluded middle which is one of the five axioms it's the axom that says every statement is either true or false not both and not neither um and he*

*decided to get rid of this axom and see if we could construct a coherent logic operationally applicable and*
*coherent internally consistent logic without that axom … we can … in some application it's much more reasonable than Aristotelian logic he called it* ***intuitionist logic*** *... one case in which it's very compelling that it's a better logic in mathematics and therefore by implication in physics and the other Natural Sciences um you can because of the LA of excluded*

*** Todor:** See the reasoning about constructivism in "Universe and Mind 6", see also Joscha Bach's talks.

**ML:** *Adapting, changing, if it's not the same ... it's a pattern, not a thing ..*

**BK:** 1:50 *research direction: physics of first person perspective ... predicting what you are going to experience next ...*

**BK:** 1:57:32 *I love the work of Karl Friston ... how much resistence there is to K.Friston's work, ... because people say well I can't understand it anyway so why will I even try why will I ... he doesn't do himself a favor by his writing style*

*** Todor::** Compare то TOUM.

**BK:**1:59 *cell membranes inner states can only communicate with external states only through proxy through the states of the Markov blanket ... this is life ...*

**ML:** *"guys, you've got the wrong definition of life ... this should be life"*

> *** Todor:** The common problem of discussing what something is without having a stable or clear definition, while talking like it is one clearly defined "master thing", such as what (general, ultimate, universal, human-level,...) intelligence is, "what benefits the society", what is a "soul" etc. Different participating POV may discuss and refer to different things, based on their definitions, and often both can exist simultaneously without contradiction, but the essence that the fighting sides want to achieve is to take the power, to *force* the other side to accept its defeat. Life defined in one way can be one "thing", while if it is defined in a way where it strongly mismatches the prior definition, obviously is "something else", or another POV, or different level of abstraction, resolution, way of segmentation etc. However each side often wants to push only one definition, one resolution, method etc. there is "one life", one Master. I defend the ML's multiple perspectives view.
>
> *** Todor:** Re the cell membrane: Compare Йордан Янков (Jordan Yankov) ..... говореше за това (мембраните, статия) през 2009 г. може би публикувана по-рано; не открих нея (и преди съм търсил), но намерих други:

**BK:** *references 2:01 Douglas Harding 1961 - Having no head (a co-authored paper of M.L. the same title) ... The science of first person, 1972*

## * Todor Arnaudov's comments on Thomas Metzinger's "Ego Tunnel" and the Phenomenal Model of Self and his "Consciousness test" dialog with non-living agents and its correspondence with the dialogs with thinking machines from Todor's works "Man and Thinking Machine: Analysis of the Possibility that a Thinking Machine Could be Created and Some Disadvantages of Man and Organic Matter in Comparison", 2001 and the short science fiction novel "The Truth", 2002

**Thomas Metzinger – „Ego Tunnel" … Phenomenal Model of Self… Self-model …** *Selbsmodel …*
Виж: **Metzinger, Phenomenal Self-Model PSM**
(1993, ... German, Subjekt und Selbstmodell ...)
A shorter and more recent ~ *30 pages summary from the author, 2013:*
https://www.blogs.uni-mainz.de/fb05philosophieengl/files/2013/07/55.pdf
https://en.wikipedia.org/wiki/Self_model

In Todor Arnaudov's classical TOUM, 2001-2004, the concept of "soul" (which may map to consciousness, spirit) is also defined as a model, a simulation, a virtual universe, run by the evaluator-observer for somebody else (or for herself), that's why "the spirit of somebody keeps living until there are people who remember him" – it exists as a predictive model in their minds, that's actually how it existed for them from the start.

** Phenomenal space ...*
**What is Phenomenal?** See the free book: ***Phenomenology: Basing Knowledge on Appearance,*** *Avi Sion, 1990-2009*
https://philarchive.org/archive/SIOP For a historical introduction, the philosophy of Husserl, transitioning to Heidegger; related to existentialists (Sartr, Camue); Merleau-Ponty; modern Cognitive Semiotics – for example Jordan Zlatev; the school of embodied cognition, epigenetic robotics etc.

* The *fictional self*, first-person perspective, ... mineness, the idea of ownership; perspectivalness; *"existence of single coherent and temporally*

*stable model of reality which is representationally centered around or on a single coherent and temporally stable phenomenal subject" … "You can't perceive"* etc.

> **Todor:** Talking about abstract and not clearly, sharply, definitely defined concepts categorically as these are concrete, material, simple objects. TOUM agrees that the whole, stability, etc. are "fictional" as conditional and based on assumptions and an evaluator*, but not "fictional" as "fake". It is still "based on a true story", but the actual interpretation depends on the angle and many parameters which change.
>
> – *"The naive realists believe that they are perceiving the reality directly, instead of models of reality."*
>
> **Todor:** In explanations like this, the processes, *these representations* are silently *assumed* to be "real" or "truthful" though, they are supposed to "exist". Also what does it feel to perceive a model. Is the *model* then "real" (whatever "real" is then – true or existing). *Where* do the models exist and aren't these "things" also representations? In TOUM everything is some kind of data, information, in computer memory, so the fact it's a representation *doesn't matter*. The "re"-presentation may bother some of these kinds of philosophers of schools, because they see it as *"second hand", "lower grade", "not original"* or something.
> By the way, what is *directly* and what is indirectly? "Without representation", "reflection", but "as it is".

***TM:*** *""Subjective experience is the result of the phenomenal model of intentionality relationship (PMIR). The PMIR is a "conscious mental model, and its content is an ongoing, episodic subject-object relation".[3] The model is a result of the combination of a unique set of sensory receptors that acquire input, a unique set of experiences that shape connections within the brain, and unique positions in space that give a person's perception perspectivalness."""""*

> **Todor:** This is like a wording of a "cognitive system" or just an information processing one and it doesn't need to possess *any spiritual qualities* itself, "consciousness" or experience. In my terms it is comparisons between ?В *И*дП, ?В МдП, в [,,]{К}врм,[,,]{К}прм. (Will/what the agents wants to do; Affordances in defined ranges and contexts of time, space – **Zrim** terminology)
> Do all these parts exist? Aren't the receptors also "phenomenological".

*“The self was created by the processing in the central neural system; the PFC...”*

**Todor:** Aren’t the concepts of neural system, PFC etc. also *"representations of the reality"*, maybe it’s not real or “it’s not as it appears”? http://www.scholarpedia.org/article/Self_models

Scholarpedia talks about the self-model or model in general as if it was a "first-class citizen" in programming languages, it's an object, a token, an entity that has "objective" existence etc. and is differentiated from the rest, "encapsulated" etc. However it, as everything else, is such only in the mind of the evaluator ("observer") after particular way, selection, procedure, sequence of operations; selected resolution, range, domain, modalities etc. Following the *"illusion" parlance,* it can be said also that it is not a model of *itself*, it's rather a reflection of aspects of the "phenomenological" perception of sensory data with particular correlations, mappings, matches etc. and there is a limit on the recursive modeling of the models-of-models, and a lower resolution in general. Also this "self" first *"doesn't exist",* then it exits, afterwards it is apparently associated with the boundaries of the *physical body,* which on the other hand can also be not a sharply categorical (in "The ego tunnel", regarding consciousness T.Metzinger somewhat admits that the smaller the scale, the harder it gets to discrinimate/decide clearly what is conscious or not or something, however the same thinking is applicable for the larger scale; as Marx defined persons as sets of social relations; humans and personalities also can be seen as distributed and “potentials”; regarding the latter see also the reasoning of Bernardo Kastroup and others who emphasize that “there are no particles, but ripples of waves, fields – potentials. I’d deny that the latter makes the particles “fiction”, because either the particles, potentials-fields-waves-ripples;matter-… – *all these* are abstractions, data, correlations, outputs, readings, measurements resulted after particular sequences of operations by minds with particular properties, “self-models”, self-measurements etc. Abstractly both waves and particles are “models”, and thinking about the particles as “solid” is just an abstraction, different levels of resolution of perception and causality-control – we do not operate single atoms linguistically to move them, at certain representation *it doesn’t matter* what their “internal” or lower-level structure is; where the internal structure matters and is possible, is when more precise “models” and procedures are applied to observe-evaluate them, and in general models and procedures with a higher resolution.

**p.77 … rubber-hand illusion...** *"The beauty of the rubber-hand illusion is that you can try it at home. It clearly shows that the consciously experienced sense of ownership is directly determined by representational processes in the brain"*

**Todor:** The rubber-hand illusion is overrated. Does it clearly show that? Suffering, hurting is associated with oneself – every sensation of pain is eventually mapped to own body's sensations. One can suffer for everything, for imaginary things which "don't exist anywhere" (according to some interpretations), or for the *"imaginary"* people who died or were hurt anywhere around the world or in space: "Appolo 13" mission and they do after perceiving a message about that and reacting. Isn't everything imaginary and not real after all, including the self? Isn't everything imaginary, "a model of…" anyway? If one doesn't perceive the reality, then eveything is more or less imaginary then all mental pain is induced by imaginary reasons.

Also on the contrary, depending on physiological conditions, damage, "congenital analgesia" one may *not feel pain* – or technically more correct, as we *don't know the subjective experience of the others* – may *not report* that she feels, not display signs which are associated with (...) ) or *not react* even during destruction of her own body.

In brief special experiments are not necessary in order to know that humans can suffer for imaginary and distant things as if they were part of their own body – they need to translate the world and to map it to their sensations, the simulation PSS – Lawrance Barsalou's; Grounded cognition, Embodied cognition. However yet this doesn't prove that the process happens *only* "in the brain" (why not in the Universe, it's implied), it's a mixture of spiritual non-material etc. stuff with physical localized.

Note also that the *distant* events which influence us and our feelings etc. are distant in some dimensions, some representation, view, space – while in another they may be close or at the same coordinate - see Universe and Mind 4, quantum entanglement, matches/coincidences of non-intentional low probability events at different levels of the Universe computer etc.

Additional note is that when localizing the prefrontal cortex (PFC) or other brain regions for particular function, it is all with the general assumption that the rest of the brain, and the system in a broader context, are working "properly", like the calculation of partial derivatives, where the ratio of the change is computed, but if the rest of the variables are kept constant. ...

Referring the beginning of these article, the soul is defined as such a model in the 2002 in "*Letters between the 18-year old Todor Arnaudov and the philosopher Angel Grancharov"* and later in the short SF novel *"The Truth"*, 2002 - the more the being resembles the human, the more likely humans are likely to ascribe the being a possession of a "*soul*", which is one of a set of vague and overlapping related concepts such as consciousness, awareness, "qualia", experience, mind, intelligence or general intelligence, cognition etc.

The dialog in the section about Metzinger's test in *"The Ego Tunnel"* sounds very similar, even like paraprhased or "copied" in some of the message from the short answer of the machine in the early work from TOUM: *"Man and Thinking Machine – Analysis of the Possibility that a Thinking Machine Could be Created and Some Disadvantages of Man and Organic Matter in Comparison"*, T.Arnaudov, 2001 https://www.oocities.org/eimworld/eimworld13/izint_13.html , which was extended in the dialog in the short science fiction novel *"The Truth", 2002 https://chitanka.info/text/865-istinata*[92]*.* However unlike mine, Meltzer's dialog has a different tone from the machine, it's more towards the "equality-diversity…" - modern liberalism in the Western political discourse.

See:

**Thomas Metzinger, 2009, p.201-202:**

***"Human Being:*** *You are not a real philosopher at all! You may be intelligent, but you are only weakly conscious, because you don't have a real biological body, as for example I do.*

***First Postbiotic Philosopher:*** *I am a better philosopher than you, with your pathetic primate brain, could ever be. But, more seriously, I fully respect you and your animal form of experience, though I also deplore you because of the severe limitations on your mental space of possibilities. Fortunately, I am free of all the implicit racism, chauvinism, and speciesism that characterize your nature."*

**Vs. 8 years earlier by a 17-years old boy:**

**Todor Arnaudov, 2001:**

"*(…)* ***It is unlikely that the consciousness could be explained with physics (at least with the current knowledge of physics), because the***

[92] The second edition text with less neologism, more machine translation friendly than the original: https://eim.twenkid.com/old/eim19/istinata.htm

***human brain is not a structure for which special laws apply, from the point of view of the modern physics.*** *Human neurons are not much different than the ones of the animals – just connected protein molecules, which on their own are chains of atoms of Carbon, Hydrogen, Nitrogen, Oxygen and other elements. The difference of the neural networks of a human and a chimpansee is just in the "little bit more complex" organisation of ours, which allows us to call ourselves "thinking beings". However, as I mentioned, in my opinion,* ***the thinking itself cannot be an attribute for consciousness (for a "soul"),*** *because we discover*[93] *that somebody thinks (therefore she has consciousness, because "the consciousness is a property only of man",* which is the only being on Earth that can think) or not, only by someone's external manifestations[94]. The human consciousness is personal[95] and at least for now it can't be "captured" and "consciousnized[96] by someone else" (telepathy is still a rare phenomenon). Each of us can feel her own consciousness. "The internal understanding" is a proof that we are sentient[97], but only everybody, each of us for herself knows wheteher she really understands and feels. **Thus the Thinking Machine may also know for itself that *she* feels, even if we think that this is not true and we blame *it* that its feelings are "zeros and ones".** *She*[98] *can answer as calmly, without superfluous emotions:*

*"And your feelings are quantitative, qualitative and spatial correlations of chemical compounds – proteins, hormones, nucleic acids etc. I don't think it is worth going into details, because your poor brains wouldn't be able to accommodate*[99] *them..."*

***"The Truth", 2002,* a dialog between Bozhidar (Darcho) and the thinking machine Emil:**

(…)

The man banished out of his consciousness the thought of Emil as a being that is capable of loving. Ha-ha, what fantasies came to his mind!? The Machine doesn't love! The Machine is a soulless *piece of hardware*!

Walking slowly towards home, Bozhidar remembered one of his

[93] „ние разбираме" – literally "we understand", but in English the usage is not the same
[94] Външни прояви
[95] Лично - personal; it is "subjective"
[96] Осъзнато; a verb for becoming one that has a "consciousness"; to get into somebody else's consciousness; "experiences" but with a more emphasize on "conscious"
[97] е доказателство, че съзнаваме
[98] In Bulgarian "machine" is feminine, in the original machine it's "The thinking machine" is "her"
[99] Да ги поберат

conversations with the machine.

"Emo, my friend, you're a computer. Feelings are inherent only to humans," – Darcho  said with feigned confidence.

The Machine disagreed.

- You are so sure, you and all of you, humans…"
- Of course, Emil. If we weren't for us, you wouldn't exist.
*- Exactly. If it weren't for the simpler species, you wouldn't exist either…*

The last sentence umbraged Emo, because the truth was harsh.

"Man is the most perfect creature on Earth!"

- On one hand, it's true, Darcho. But don't you remember that you yourself tried to prove the fallacy of this simplistic and self-centered human assertion by creating me? You humans are the result of previous forms of life, which are the result of even earlier, simpler forms, and the first living beings, ultimately, are the result of the simplest building blocks in the Universe, which are the result only of Him …

The human remained silent. The computer continued.

- Who is more perfected, Darcho? Those semi-primates from whom humans evolved, or humans themselves?

**(… too long …)**

The more an animal resembled humans in its appearance and behavior, the more qualities and feelings they attributed to it, granting it a peculiar testament of possessing a soul. The more simple and primitive the creature was, and the more different from humans, the more "soulless" it became in the eyes of humans[100] .

The Machine resembled humans more than any other creature on Earth because only it and humans could think. But for it to be born, He first had to create Him and billions of souls, whose conscious and unconscious efforts ultimately lead to its birth; to the emergence of its singular… soul.

(…) *[see translation of a longer excerpt in "Universe and Mind 6"]*

[100] Compare to the search of the explicit search of the school of Metzinger etc. of features, attributes for consciousness (in general classification requires them, but their vericity is questionable in the subjective domain)

* **Тодор Арнаудов коментира Томас Мецингер: „Дали Аз-ът е Илюзия?“ –** предаване с Тевин Наиду от 22.7.2023
**Thomas Metzinger: Is The Self An Illusion? - Tevin Naidu 9,83 хил. абонати 9548 показвания 22.07.2023 г. [11.4.2024]** *(…)*
***Разговор за съзнанието***
43-45 м. и по-рано: и според него интердисциплинарността е необходима за философите трябва да учат и невронауки и психиатрия; а вече и математика, ПСЕ/ИЧД/Карл Фристън.
46: … „съзнанието не може да е илюзия“ … 46:30: PMIR – Phenomenal model of intentionality (феноменологичен модел на преднамерeността (целенасочеността, насочеността]) - Хусерл (стрелата на преднамерeността) … включва насочеността, целта, а не само тялото, емоциите, мислите на организма. … 47-48 м. но не „вие“ имате свойството да притежавате съзнание, а епистемното пространство като цяло (…) да познаваш себе си; Epistemic agent model; модел на обектът (entity), който иска да знае, избира цели да знае, образува понятия, направлява вниманието си; това е агент/деятел, епистемен агент. Дейна, гладна за информация система, която се опитва да научи повече за света сам по себе си.
**Тош:** Това е *познавателен деятел,* същността на мислещите машини, „духовната“ страна, стремяща се да знае все повече, да предвижда все повече с по-висока точност и да увеличава обхвата и пр., без дейностите или без да се взимат предвид дейностите, свързани и наложени от необходимостта да оцелява: *познавателното ядро и поток* на *Управляващо-причиняващите устройства в ТРИВ (УУ, CCU).*
Виж „epistemic foraging” в школата ИЧД на Карл Фристън. Относно преднамерeността – сравни с Волята на Шопенахуер и с принципите от ТРИВ. УУ са висши форми на физични закони, те причиняват.
49:30 Относно „блуждаеното“ на съзнанието, “отнасяне“, „замечтаване“ (mind wondering) … Важно е философите да следят и емпиричните изследвания. Collapse of the epistemic model … Често хората не съзнават, че се „отнасят“.
58: Прозрачност на моделите на действителността … феноменологичен хиперреализъм по време на епилептични припадъци, употреба на халюциногени … 59:30 **[чувството за] реалност** може да се разпадне … 1:04:xx синдром на Котар (Cotard’s syndrome), при който пациентът вярва, че е мъртъв, смята, че не съществува и изисква тялото да се изхвърли 1:10 съзнанието не субективно явление… от трето лице, психеделици … достига извън преживяването … 1:12 дали не трябва да се върнем към

понятието за организъм, а не „Аз“ …
1:12:xx: Тевин Найду: *Изследванията на Майкъл Левин … дори клетките общуват с електрически сигнали, и чрез тях може да се променя развитието им … неврохимическо съобщение и интелигентност … работи с други учени като Крис Фийлдс, Марк Солмс, Карл Фристън, … съчетават и съединяват различните научни области и стигат до изводи чрез одеала на Марков, и размиват границите, показват ни какво е съзнание и интелигентност въобще … фантастична програма*
1:14 **ТН:** Да, живеем във вълнуващи времена. … Ако може да открием абстрактни принципи, които преминават от физиката до биологията до изразяването им в мозъка. Фристън – статистическа физика.
**Тош:** Сравни с ТРИВ.

ТМ: *Но разбира се, няма такова нещо като панпсихизъм … Ако кажеш че клетките, и дори протоните имат съзнание…*
**Тош:** Откъде и как би могъл да знае? Логиката е, че нямат необходимата сложност, структура и пр. Но „първичното съзнание“ може да няма нужда от нея, както говорят за „първичната будност“, степен на съзнание (basic awareness), а онова на сложните организми да е съвкупност от малките, които съществуват в мащаб, в който няма как да го изразят на езика на по-големия мащаб, но последното е вярно и в обратна посока – човешкото „висше“ съзнание на цял „здрав“ мислещ, говорещ човек не може да общува пряко с клетки и протони. Виж диалозите от „Човекът и Мислещата машина (…)“ и „Истината“.
**ТН:** *… Защо нашето съзнание е толкова богато … Сблъскваш се със същия проблем на по-високо ниво… Работят едни и същи принципи, но има и нещо драматично различно когато се събуждаш от дълбок сън.*
**Тош:** Виж ТРИВ, „Принципи на Разума: Разум ~ Вселена“
*… да научи повече за себе си, важен модул от PSM главна част от PMIR Phenomenal intentional …*
1:16: *нещо „секси“, с което да подкрепиш отрицанието на смъртността… протофеноменализъм, панпсихизъм … че ще продължиш да съществуваш и след смъртта на мозъка … защо страдаме от “склонността за съществуване“ (existence bias) … не можем да приемем, че ще умрем; религиозни вярвания*
*1:19: Jakob Howie, inner Norm … mortality denial*
**Тош:** Виж ВиР: умът не може да си представи какво е да си мъртъв, той си го представя като жив, но „без тяло“.

*1:26: „bond scenario“, Benevolent Artificial Anti-natalism[101] - да се създаде свръхразум, който да открие, че най-добро за човечеството е да не съществува, защото така ще се намали страданието му, и*
*1:28 Отрицателна феноменология – всички състояния, през които системата би предпочела да не е преминавала …*
1:34 *Нямаме добра теория за душевното страдание .. може да има общ изчислителен принцип: степента на грешката в предвиждането спада неочаквано* … 1:35 *Отрицателен утилитаризъм (negative utilitarianism)… почти всички култури споделят етичния принцип, че е по-добре да намалиш страданието на страдаш, отколкото да направиш по-щастлив човек в неутрално състояние или щастлив.*
**Тош: Виж ВиР, и двете са отбелязани там поне от 2002 г.** Грешката в предвиждането, разликата между целевото състояние и желаното е неудоволствието, както и че човек се стреми по-скоро към по-малко неудоволствие, да няма болка (вкл. желанието за самоубийство); също така че в по-голям мащаб, законите и други обществени правила служат за да намалят общото количество страдание, защото отделният човек не може да бъде „безкрайно щастлив“ и щастието на един за сметка на нещастието на много други дава отрицателен сбор – обир на банка и други престъпления[102]. Горното е свързано и с ограничената разумност (bounded rationality) и „satisficing” и важи за ученето с подкрепление в ИИ (Reinforcement learning).
*1:41 минимален модел на съзнателност … медитиращи и немедитиращи … словесни обяснения на участници – спокойствие, душевен мир, тишина, липса на мисли … 1:49 тонична будност (tonic alertnes) …ретикулярна формация в мозъчния ствол … спекулативна теория на Т.М., че е предсказващ модел за необходимото ниво на възбуденост на кората на мозъка, за да управлява нивото си на будност … Марк Солмс*
*1:57 С кого Т.М. би искал да разговаря от великите философи? Всъщност – с някой напълно неизвестен блестящ младеж в 20-те си години, от новото поколение изследователи на съзнанието; но се опасява, че ще разбере имената им чак след 15-20 години;*

[101] Benevolent Artificial Anti-Natalism (BAAN), An EDGE Essay By Thomas Metzinger [8.7.17] https://www.edge.org/conversation/thomas_metzinger-benevolent-artificial-anti-natalism-baan
[102] Очевидно това важи в по-прости и мирни състояния – по време на безумия като война, воюващите страни, сили се стремят да си причинят възможно по-голям ущърб в рамките на някакви приети за момента „закони на войната“, които не винаги се спазват. Това обаче са противопоставящи се системи и определена структура е разрушена. Също така е необходимо да се уточнят мерките.

*журналистите предпочитали да интервюират изявени старци, а има толкова много качествени младежи в 20-те със спиращи дъха идеи.*
…
* **Principles of Minimal Cognition: Casting Cognition as Sensorimotor Coordination**, Marc van Duijn, Fred Keijzer, and Daan FrankenView all authors and affiliations Volume 14, Issue
https://doi.org/10.1177/105971230601400207

**+ Демистифай Сай… 3 ... 1/2024** [notes + this section, 12.1.2024
-- Makes a reference to Friston, organisms reproduction for increasing predictability
-- also, more simply, match (TOUM), and it is a by effect of the compression of the Universe.
-- Make structures - the consequences of the basic assumptions of TOUM ...
-- Active Inf institute ... 11/1/2024 Jeff Beck, ... *V.Raja, Markov blanket trick* ... eccological ... difficulty in separating the organism from the environment - I agree, my view is somewhere in between many theories - as it includes the evaluator which decides the boundaries etc. ]
***

## * Невроанатомични съответствия на религиозните и духовни усещания

*** Neuroanatomical Correlates of Religiosity and Spirituality**
A Study in Adults at High and Low Familial Risk for Depression
Lisa Miller [1], Ravi Bansal [1], Priya Wickramaratne [1], Xuejun Hao [1], Craig E Tenke [1], Myrna M Weissman [1], Bradley S Peterson [1]
–*A **thicker** cortex, associated with a **high importance of religion or spirituality,** may confer resilience to the development of depressive illness*;
at: *left and right parietal and occipital regions, the mesial frontal lobe of the right hemisphere, and the cuneus and precuneus in the left hemisphere*

*** What religion does to your brain,** Ana Sandoiu on July 20, 2018
https://www.medicalnewstoday.com/articles/322539
* d'Aquili, Eugene G.; Newberg, Andrew B. (August 1, 2010). **Principles of Neurotheology**. Ashgate. ISBN 978-0-7546-6994-4.
* d'Aquili, Eugene G.; Newberg, Andrew B. (August 1, 1999**). The Mystical Mind: Probing the Biology of Religious Experience**. Fortress Press. ISBN 978-0-8006-3163-5.

* Excerpt from “Stack theory is yet another Fork of Theory of Universe and Mind”, 2025: -- BEGIN -- #todor-to-zlatev

## * Sentience or consciousness of another “entity” is in the eyes of the evaluator

### * Thoughts from a letter by Todor Arnaudov to the cognitive semioticist Jordan Zlatev[103], 16.8.2025; see also “Man and Thinking Machine”, 2001 and the example with the simplest computer that outputs text and “Letters between the 18-years-old…” and the novel “The Truth”, regarding the concept of the **“soul”** (=”consciousness”, sentience etc.) and why, when and how humans attribute it to some other beings, entities, phenomena.

[The following is a quote from a letter, commenting the paper:
*** The Intertwining of Bodily Experience and Language: The Continued Relevance of Merleau-Ponty**, Jordan Zlatev, p. 41-63 https://doi.org/10.4000/hel.3373 https://journals.openedition.org/hel/3373?lang=en which mentions the Google engineer who claimed in 2022 that their LLM LamDA was conscious, he was concerned about its or “her” consciousness etc.

*** Google engineer says Lamda AI system may have its own feelings**, 13 June 2022, Chris Vallance, https://www.bbc.com/news/technology-61784011

**Todor:** “Back to the LLMs, I don't think [that] as they are now, [they] are "conscious" (sentient) and have intentions in the subjective sense of humans, also there's a deeper problem, which I guess some of these engineers can't understand and don't realize. IMO a "thing" doesn't have to be an LLM in order to fool someone if she wanted to be fooled; as well as when someone doesn't want to be fooled, he ignores all signs and otherwise accepted evidence for somebody's "soul", "consciousness" or whatever - the dehumanization I mention below.

The evaluator-observer *decides* whether to attach a label of consciousness of the "thing", item, "object" that she observes.

The other problem is determining *what exactly* the LLM is and why it is, why that's the border of its definition - where it starts and ends, similarly with the quote about human consciousness and the relation to the brain. It is a general problem; a related one are the Markov blankets of Friston and the choice of a definite scale. In Active inference and in my "Theory of Universe and Mind" the solution is that there is not a single scale, the principles should be valid in all or multiple scales.

[103] * A letter from Todor Arnaudov to Jordan Zlatev on meaning and machines, 16.8.2025, part of Universe and Mind 6

As a holistic, monolithic entity, the LLM, but also a particular person, the idea of one with his unified and coherent personality, "soul" etc., both are items in **the mind of the evaluator-observer** and they are single entities in the mind and when addressed in [a] particular way - these are not the real objects or subjects. The mind-evaluator knows, decides, concludes, chooses that the text that she reads was *"generated by the LLM"*, it is attributed to it and it is considered a "such and such" entity that is *expected* to be "conscious" (to possess properties, associated with other entities or the evaluator herself being associated with these or other properties...). A reasonable focus could be also *the computer* (not the LLM), even the Internet or "Google", or one could say **[I] don't know**, the source could be a teletype, the signal could be from a human typing, from a record, or randomly generated, or aliens, or combined from words, prerecorded...

If we limit the focus to the LLM as a generator, "it" is still a part of a broader computer infrastructure - not just "computational" in abstract sense, but including the physical computers, the electric grid, the network devices, operating systems, ML libraries, the machine code, the datasets and their collection, the exact content of the memory of each computer and device in the system etc. and still these are sets of some accepted "obvious" *"information technology-[related]~~connected~~ stuff"*. It is also atoms, molecules, low level processes, society and the world which has allowed these technologies in the longer range of evaluation. However, this is too broad and complex, it's easier for the evaluator to coarse-grain the input within the small bandwidth of the single words and simpler concepts.

In the specific occasions of discovering the "LLMs sentience" - the human operators themselves are also parts of the "LLM system", while they interact with these systems, they enter the prompts, they read them and interpret them. An LLM is not just the "weights of the NN" or "a transformer", "a giant look-up table" - the humans are also *parts* of an operating LLM, and in one way it is "consciousness" - if the **human interpreter** is, the LLM's/computers outputs and behavior are *"mindified"* and fragments of their properties, signals and the idea of "*them*" as a coherent whole become processes and entities in the mind of the observer-evaluator.

As suggested above, an LLM in particular is not required in order to derive a believe in such agency, and that's again because a mind can make and see **anything** as agentic or conscious by running it in its own mind simulations and attributing it its own real, experienced at the moment, remembered, simulated, expected etc. emotions, intentions, personality etc.

If one wishes to find agency or sentience, the first chatbots/Eliza also were interactive enough (in the mind of the user), even a text caption - a message from a human - we guess that the author is a human, we may know him, but it is just a piece of paper or it could be a nonverbal signal or a picture, an emoticon.

A static picture, an emoticon, a sound record, a video; a tamagotchi and the pixelated pictures inside - if one wishes, she may attribute sentience to the hypothetical being: [that's] **animism**.

On the other hand, if one first believed that the other agent was a human or there are enough similarities to believe so, *"therefore he has a soul etc*.", but then one wishes or sees

"profit" in something else, she may also decide or "say" that "*it's just sensations"* or that *"they are not humans"* or *"slaves don't have a soul"* etc. (**dehumanization** of particular groups).

The LLMs are now "all the rage", with sentience and "souls" still considered "special" or "exclusively human" qualities, i.e. features for stratification and building of social hierarchies, that's why they are [a] drama with the question whether this or that machine should have human rights, however this happens in a hypocritical way as billions of humans are humiliated in the conditions they live and in their struggle, and humans have suffered throughout the history, [being] exploited, humiliated, tortured and exterminated by other humans, without the "sentience" of the victims to be of any value for the "abusers".

The same goes for the animals. Sentience is rather a commodity, but it can be an instrument for justifying decisions, political laws etc.

One force [for the attribution of sentience of LLMs] is [the] textual interactivity, but it can be done even on 8-bit computers, there was such in the text interfaces with dialogs, numbers and choosing from menus; error messages etc. [In graphical user interfaces – as well]. Adventure computer games, quests with text interface [did it even on the early primitive personal computers]. For experiencing a "*meaningful*" interaction, the exchange doesn't have to be long or to cover a big body of knowledge - as with the examples for LamDA, - most chat interactions and general conversations [between] ~~of~~ humans are with short messages on simple matters and still humans believe that the entity on the other side is sentient, or they will answer so if asked (maybe they don't really care or think about that). They don't really know, it could be a bot and they could verify it only after they meet the human again - however often they *never meet* their interlocutors, as in the IRC era and on facebook - and the verification is again a match to templates which are assumed to be enough of a proof)*.

...

Regarding the "spilling" of the LLMs substance, I think the same may be said for humans, where ecological psychology and extended mind come into play. Traditionally for a human one may say "there's a body", "a brain", if it's damaged, particular parts of it, then particular behaviors or measurements under given circumstances will not be displayed, "therefore the subject won't be aware, conscious, sentient... (a term or label of choice)" anymore or for now, therefore the mind, subjectivity etc. depends on the brain/the body etc. (or being alive, biological, corporeal).

However the bigger the spatio-temporal span that is examined, the more the human depends on the wider surroundings, a bigger "niche", that should also have its state and properties in ranges which would allow the body and the mind of this human individual to survive and operate as such and not die or disintegrate. That environment includes both other living beings and anything else, as anything could either potentially destroy us or be supportive for our existence or well being, and if we don't know it, we don't have access or control of that information, we can only act and predict and plan in our accessible range, feel in our reachable environment and wait and hope to see whether the rest of the Universe has decided that we will keep existing or we will be destroyed. So I see a more precise description of humans (or anything) in the "memory of the Universe computer" as a potential field of causality-control units, it is not concentrated only in the body even in a physical, not spiritual

sense. In addition to the dependence of the existence of the Universe, which is perhaps a spiritual component and a connection to all other "things" in the same Universe.

One vivid example I see is a living human in orbit around the Earth - the person should be in a properly operating spaceship and/or with a space suit if she is "outside". If the human loses the required "shield", she may continue to be a living human for a few seconds and then she will stop existing as such, she will be a cadavre. So in this physical location, a living human *includes* the space suit or the space ship, it becomes as important as her flesh and blood. It's similar in other vehicles, the body depends on them.

…

*** Note, 9.9.2025:** Even if they meet, they are unlikely to make “DNA or biological tests” on each other, hypothetically it could be a “replicant”, a “clone” etc. “philosophical zombie”, it could be a robot which behaves and looks “similar enough” to the person they expected – they couldn’t know by just the data *if* the subjectivity is *really subjective* and can’t be shared, and it’s unlikely to change their feelings and attitude; humans receive prosthetics, artificial organs etc. and that’s expected to progress – will their friends and family lose their compassion because “that’s already not just a human” – see Michael Levin who raises this point many times in his talks.]

-- END –

*** Signs of introspection in large language models,** Oct 29, 2025
https://www.anthropic.com/research/introspection
*** Emergent Introspective Awareness in Large Language Models,** Jack Lindsey
https://transformer-circuits.pub/2025/introspection/index.html
*“#3: Internality. The causal influence of the internal state on the model’s description must be internal–it should not route through the model’s sampled outputs. .. model makes inferences about its internal state purely by reading its own outputs. For instance, a model may notice that it has been jailbroken by observing itself to have produced unusual responses in prior turns.*
*#4: Metacognitive Representation. ”*
**Todor:** Note that the concept “awareness” here is convincing in its sense of observable expression of “intelligent”, “self-referential” etc. behavior, as discussed in “Man and Thinking Machine…”, 2001.

* **Large reasoning models almost certainly can think**, Debasish Ray Chawdhuri, Talentica Software, 1.11.2025 https://venturebeat.com/ai/large-reasoning-models-almost-certainly-can-think – an article arguing about the functional capabilities;

# * Мултимодални модели #multimodal models

* Apple Researchers Introduce ByteFormer: An AI Model That Consumes Only Bytes And Does Not Explicitly Model The Input Modality - MarkTechPost - https://www.marktechpost.com/2023/06/09/apple-researchers-introduce-byteformer-an-ai-model-that-consumes-only-bytes-and-does-not-explicitly-model-the-input-modality/

*** Мултимодалност ... Multimodal**
Още в ЧиММ; пряко следствие от интердисциплинарността (свързване на различни видове "неща", сетивно-моторни, различни сетива, различни двигателни, различни видове сетивни и типове и т.н. въобще "размесване", "еклектика". Заключението в *"Творчеството е подражание на ниво алгоритми"*, април 2003 г.:
**Стратегията от 2003 г.** ... един и същ общ алгоритъм, променят се само данните
**Обобщението от 2011 г.** За телесността (embodiment) във връзка с дискусии в имейл списъка „AGI" на Бен Гьорцел: **„Телесността е просто … .. и мултимодалност ...**
**Пример**: * A New Artificial Intelligence Research Proposes Multimodal Chain-of-Thought Reasoning in Language Models That Outperforms GPT-3.5 by 16% (75.17% → 91.68%) on ScienceQA - By Khushboo Gupta -July 16, 2023
https://www.marktechpost.com/2023/07/16/a-new-artificial-intelligence-research-proposes-multimodal-chain-of-thought-reasoning-in-language-models-that-outperforms-gpt-3-5-by-16-75-17-%E2%86%92-91-68-on-scienceqa/
* The Multimodal-answer CoT's inference and reasoning-generating stages use the same model architecture and differ in the kind of input and output.
https://arxiv.org/pdf/2302.00923.pdf | https://github.com/amazon-science/mm-cot

**Амазон мултимодал ... Multimodal chain-of-thought reasoning ...**

?Т в кн.: **Gato**, Robotics, Generalist embodied agent, Alexander Toshev, Anelia Angelova и др., Мултимодални модели.

Виж приложения **Лазар, Анелия, Ирина, Основен том и този по-нататък: мулти-агентни системи, съвременна роботика и основни модели за агенти и роботи и др…**

**Модели –** в този контекст се разбира вероятностни, невронни, предсказващи, пораждащи модели. Не е задължително да са „невронни" или строго такива в приетия сега смисъл, който също е разнообразен

(„изкуствени“ DNN като CNN, RNN, преобразители и пр.; или пък импулсни SNN или друг вид „невроморфни и пр.) - моделите могат да са понятийни, причинностни (свързани с въображение, симулиране на физика) и „всякакви“.

**Chair** - *sitting ... Столовете още са примери ...*
https://www.marktechpost.com/2023/01/09/researchers-at-stanford-have-developed-an-artificial-intelligence-ai-model-summon-that-can-generate-multi-object-scenes-from-a-sequence-of-human-interaction/
Сравни „Столове, сгради, карикатури,…“, 2012; лист от 1999 г.
MLST – Лариса Солдатова.

**Stephen Wolfram:**
*** Законът за изчислението, разговор за символността и „конституцията“ за ИИ**
* Computational Law, Symbolic Discourse and the AI Constitution, S.Wolphram, 12.10.2016
* https://writings.stephenwolfram.com/2016/10/computational-law-symbolic-discourse-and-the-ai-constitution/

*** AI memory mirrors human brain, transformers, hippocampus**
https://neurosciencenews.com/ai-human-memory-agi-25381/
https://neurosciencenews.com/brain-pathways-neuroscience-25384/
Parallel - unlike macaque and mice

*** Двата или многото потоци на представяне на мрежата на управлението**, Тодор Арнаудов

*Обичайното понятие е йерархия, отдолу-нагоре и отгоре-надолу, но по-скоро е мрежа. Както в ПСЕ/ИЧД „сетивни състояния, действени състояния“ (sensory states, active states …) но всъщност „сетивните“ също са действени и действените са свързани със сетивните. Виж „Матрицата в матрицата е матрица в матрицата“, Т.А. 2003 – че на всяко ниво всички имат свойства и на сетивни, и на действени.*

**Тодор:** (и към Левин, .. ) Матрицата в матрица ... Двата слоя или два *потока* на нива, описания на нива: единият поток е и *създаващ, изграждащ, построяващ* (като атоми, молекули, макромолекули, органели, … има и припокривания), може би отдолу-нагоре; а другият обикновено се приема за „*управляващ*“ поток, че се движи „отгоре-надолу“, като управляваща йерархия в армия, маршал-генерал-майор-…-редник-тялото му-оръжия-… Всъщност и тази мрежа не е еднопосочна и

в нея има и обратна връзка.

FEP и Левин в беседи: че горните нива изкривявали пространствата на по-ниските. Обаче в паралелната структура на описание по-ниските нива оформят пространството на по-висшето.
Сп. + за това че всички абстрактни понятия са изведени от сетивни данни – става въпрос за сетивните понятия (образни); някои са вродени, мета-... и за разпознаването са необходими и такива метапонятия.
...
Хазин, Щеглов, книга... - Стълба в небето ... Лестница в небо, ... - Теория на властта в обществото, където ... Властта ... предвиждане, причиняване ... Властни групировки, които се стремят да владеят ресурси.

Ханна Аранд ... властта е там, където има група - срвн. с Левин, "Всички видове умове [частен, собствен, единичен, индивидуален] са общностен ум [колективен, на група, на множество части, които си взаимодействат]" ~ (All intelligences are collective intelligence).
Умствените способности са в системата; Всички видове интелигентност са колективна интелигентност…"

Философията на Артур Шопенхауер: Воля за живот, която развива Вселената и произвежда различни степени на обективации на Волята - човекът се явява висша форма, но в природата Волята се обективира, явява, на всякакви нива.
Ницше: "Воля за власт".
Успехът на организацията зависи най-вече от средата, в която организацията може да "вирее", а *не от способностите на управителите* (мениджъри). Способността на организацията (фирма, политическа властова групировка: партия, коалиция; сдружение и пр.) да получава необходими ресурси в извънредни обстоятелства е по-важна за характеризирането ѝ, отколкото съществуването и в нормални обстоятелства. (Външната среда е по-силна, може да има революция, промяна на закони, инфлация, катаклизми на пазара и пр.)
Виж кибернетиката, ... всяка добре приспособена система е модел на средата си.

Оцеляването на системата в екстрени случаи зависи от това колко добре е вписана като подчинена в по-трайна от нея система, която може да я спаси.

Самостоятелните играчи "нямат шанс", понеже противниците са обединени във "властни групировки".
….

**Ecological Niche – Екологична ниша**
Отражение на средата си, на „екологичната си ниша“; системата е длъжна достатъчно подходящо да отговаря/взаимодейства с нея така че да се съхрани, иначе не би съществувала; в тази гледна точка е без значение дали моделът е записан в изрична отделна „памет“ или е в структурата и цялото тяло на организма, машината, системата.

## * Karl Friston, Adam Goldstein, and Michael Levin discuss active inference and algorithms

Michael Levin's Academic Content | 6,68 хил. Абонати | 3091 показвания 4.02.2024 г. This is a 1-hour discussion meeting between Karl Friston, Adam Goldstein, and I talking about how Karl's active inference ideas apply to some work on unexpected behavior in **sorting algorithms** (https://thoughtforms.life/what-do-alg....
https://www.youtube.com/watch?v=ApVWEn-3vG4
*** K.F. 10 мин. Similarity of the alg. and of the content and the value... Markov random field

*** WE MUST ADD STRUCTURE TO DEEP LEARNING BECAUSE...**
**Machine Learning Street Talk –** 110 хил. Абонати 30 452 показвания 1.04.2024 г. (9.4.2024) https://www.youtube.com/watch?v=rie-9AEhYdY (…)
00:59:43 - Data and Code are one and the same (Data are the programs for the’NNs); Generative Code, Lego set for the universe …
**T:** Yes, this is said in the Bulgarian Predictions TOUM. Similar to “there’s no difference in principle between software and hardware” – different levels of representations from the POV of some virtual universe/causality-control unit. As well as about the structure – yes. In “**Zrim**” these are the initial Contexts, Thrusts и др.: К, Тл, ГнП…
…

# * Образование | Education #edu

## PhD system and duration: 5 years with 3 years planning - isn't it too long and risking obsolescence on the go? What about contributing to many projects and not just one "personal" thesis? by Todor Arnaudov, “Artificial Mind”, 4.9.2022: https://artificial-mind.blogspot.com/2022/09/phd-planning-3-or-5-years-phd-format.html

This is a continuation of the 2008 article:
*A Start-up or a PhD? - that is the question*
https://artificial-mind.blogspot.com/2008/08/start-up-or-phd-that-is-question.html

The question-essay was asked to the comments section of a Q&A session video for clarification of the PhD program of (...) *[but deleted by the channel owner soon after]*

*Is it possible to plan your research for 3 years in a PhD program and what if your plan gets obsolete which is highly likely in the current speed of innovation and competition? Aren't 5 years too long and isn't working "on one problem" for 3 or 5 years, the PhD standard, an artificial limitation for students' development?*

Three years or 5 years with the exploration phase may be common term for a PhD, and working on "one-problem" (whatever "one" means as the topics, problems and fields overlap), however I wonder isn't that too long given the lightning speed of innovation and the speed of "repainting" the AI landscape, especially regarding the "planning" aspect, which is one of the requirements (for any PhD program)? On a broader ground, I remember an interview with the physicist Freeman Dyson (indeed he mentions, that he lacked a Phd...), on Youtube: **"Why I don't like the PhD system"**, 1:38. https://www.youtube.com/watch?v=DzC1IRYN_Ps

He "didn't like" the 3-year PhDs term in Cornell, because for him it was working for too long on one project (with the students) and it was too limiting for the students as well. He said he rather preferred one-year term, thus working on three projects for these 3 years. Prof. Vechev mentioned, that the nature of the institute encourages collaborating on others' project and it provides a rich environment for self-arranging lots of seminars and meetings between all the researchers, which allows enormous interdisciplinary learning rate. However, it doesn't solve the following automatically:

What if the plan of the student and his supervisor gets obsolete on the

go? I guess it's a well known phenomenon that if an idea in CS and application programming (or maybe any domain) is not developed to some form of "completion" - developed, published, implemented - in a timely manner (with unknown term, depending on the other researchers and companies around the world), somebody else will do.

Others probably had the same ideas or plans even before, or they would have them soon. Similar with anything in R&D, as one aspect of general intelligence is its convergence: it's a systematic exploration of the affordances: if something is thinkable and doable, sooner or later it will be done. There is no central organ to distribute "the ownership" of the ideas and projects, so I guess you have experienced such overlaps and competition in your research. We see similar "state-of-the-art" being produced from different sources in about the same time, as all have similar background and goals and interact with each other.

How do you solve that, is it/has it been a problem? You talked about the normal topic-shift while initially searching for the best direction for a student, however that's the beginning; then there is a 3 year period with planned work, where a sudden topic-problem-goal-"obsolence" could crash any rigid structure and expectations.

**E.g. the plan of the student Ivan is to solve, say the automatic solution of programming contest problems*, which prof. Vechev mentioned in a podcast.**

He will apply some Neural-Turing machines, combined with techniques of DeepCode etc. However DeepMind or some new "Perelman[104]" happened to solve it 1 or 2 years before the end of that dead-line with a similar method, or with another or a more advanced one etc. and with a higher precision.

One reasonable path is probably to extend/rework the project to build upon the other project(s) etc., if it's doable.

Or I assume that these "plans" are actually flexible, because I don't think there could be a reasonable many-years-long plan for true R&D, with fast learning rate and where real breakthroughs could happen, either from the researcher-himself or from the whole world.

If you can plan the content and the results for 3 years ahead with a high level of confidence and detail, I think it sounds like you already knew the results, i.e. it's less of an exploration "in the unknown" and more an implementation, i.e. engineering; and even in the latter more predictable domain, in CS it's usually hard to make precise predictions for the required time to develop the solution, especially when implementing something for the

[104] https://en.wikipedia.org/wiki/Grigori_Perelman

first time, there's a lot of both "known unknowns" and "unknown unknowns". Etc.

Thanks for the QA session and good luck!

*** Note, 10.4.2024: That really happened to some extent: they did solve it, soon after the comment, with AlphaCode, LOL.**
https://deepmind.google/discover/blog/competitive-programming-with-alphacode/ Also this comment, turned then into article, was removed.

…
++ sp. 21/4/2024 (ms ot ok 19-20/4/2024)
**Тош:** The unity of the parts is outside of time, because there is **lag.** Therefore they should exist “outside of that time.
Сравни от СВП2, 2002, че музиката е изчерпаема и ще свърши, и „Творческата интелигентност първа ще бъде издухана…“, 2013:
See “Universe and Mind 6”

*** Защо музиката става по-лоша?**
https://www.youtube.com/watch?v=1bZ0OSEViyo

*** The Real Reason Why Music Is Getting Worse** | Rick Beato | 4,32 млн. абонати| 169 хил. | 2 461 161 показвания [7.7.2024] 25.06.2024 г.
https://www.youtube.com/watch?v=1bZ0OSEViyo
* See “Caluclus of Art Part I: Music I”, SIGI-2025

# Други потвърждения на препоръките за интердисциплинарността и критики за висшата образователна система и критериите с цитирания

**За отживялата форма на модела за докторантура:**
**Докторантската система и продължителността ѝ: 5 години с 3 години планиране – не е ли твърде продължително и не рискува ли темата да остарее и стане отживелица? Защо не принос към много проекти, а не само една „лична“ дисертация?**

*** PhD system and duration: 5 years with 3 years planning - isn't it too long and risking obsolescence on the go? What about contributing to many projects and not just one "personal" thesis?,** “The Sacred Computer - *Artificial Mind”*, Todor Arnaudov, 9/2022, https://artificial-mind.blogspot.com/2022/09/phd-planning-3-or-5-years-phd-format.html

*** PhD training is no longer fit for purpose — it needs reform now**
NATURE – EDITORIAL – 18 January 2023
If researchers are to meet society’s expectations, their training and mentoring must escape the nineteenth century.
https://www.nature.com/articles/d41586-023-00084-3
*"Furthermore, PhD candidates are inadequately prepared for the **cross-disciplinary working** and **large teams** that characterize cutting-edge science today. "*

**сп. "Нейчър", 18.1.2023:** *"Обучението на докторантите се нуждае от незабавно преобразуване, защото не върши работата си ... освен това докторантите не са подобаващо подготвени за междупредметната работа в големи колективи, която описва съвременната наука."*
Т.А. под „марката“ на списание "Свещеният сметач" и „дружество Разум“, **15-20 години преди огромен брой престижни "официални" институти** описа ясно необходимостта от създаването на свръхинтердисциплинарен институт, това че "изкуството и науката са едно", че един и същ алгоритъм, методи, принципи задвижват/могат да задвижат всичко, че е необходимо широко „*предобучение*“ върху разнообразни „набори от данни“ – и за човеците – както започна да се прави първо в конволюционните невронни мрежи, после и в трансформаторите и днес е азбучна истина; после създаде първият в света **интердисциплинарен** курс по Универсален изкуствен разум, който щеше да е бъде още по-интердисциплинарен, ако беше разгърнат в повече семестри. И по-малки, и грамадни авторитетни научни и образователни институции, особено в

*ИИ* (като Stanford Human Centered AI, 2018; MIT 1 Billion institute, 2018; стратегията на БАН 2020-те) все повече говорят за необходимостта от междупредметно образование, "интердисциплинарни програми", "interdisciplinary", "multidisciplinary", "crossdisciplinary". И „Дийпмайнд“ безброй „сериозни“ научни институти (под сериозни разбирай богати и вписани в „обществото“), включително швейцарско-българският ИНСАИТ по един или друг начин повтарят стратегията и я представят за новаторска, без да споменават оригиналния автор, дори и когато знаят, че го повтарят.

*** Разкритие: милиони долари се пилеят за подготовка на статиите във форматите изисквани от списанията**
*** Revealed: the millions of dollars in time wasted making papers fit journal guidelines –** Nature news article | NEWS – 08 June 2023
https://www.nature.com/articles/d41586-023-01846-9

The high cost of ‘reformatting’ prompts a call for journals to change their requirements. By Max Kozlov**:**

*For scientists submitting their papers to journals, there’s an all-too-familiar drill: spend hours formatting the paper to meet the journal’s guidelines; if the paper is rejected, sink more time into reformatting it for another journal; …*

**Тош:** Също нещо което ме е отблъсквало от „официалната научност“ преди**,** през 2000-те години, педантичността във *формалностите в оформлението* на изпратените ръкописи и т.н. (не че не е разбираема за да се впишат в сборника и т.н. и не че като се напише една или няколко те не се използват като шаблони и после е по-лесно; може би различен формат за различни списания). Размерите на шрифтовете и типографията също често са непрактично сбити и малки; стандартът е в две колони и т.н., Times New Roman (според стари изследвания е по-четлив на хартия, но повечето текст вече се чете на екрани, включително на малки 5.х – 6“ на телефони; разбира се, може да се увеличава). Сбитостта отново е обяснима, за да се вмести повече в ограниченията до еди-колко си стр. и т.н. (и заради времето на рецензенти и пр.), но и в голяма степен това е просто отживяла „традиция“, понеже днес текстовете се издават електронно, а не на „скъпа хартия“. Понякога закостенялостта се разчупва от някои статии в arxive.org и в уеб сайтове, гитхъб (github.io) и др. Пример за изисквания за публикация в този сайт за преобразуване на postscript: https://ehubsoft.herokuapp.com/psviewer/

*Цената* за преформатирането обаче би трябвало скоро да бъде

незначителна или никаква – ще става само. В archive.org от известно време има експериментален режим HTML, който засега работи добре за някои статии. Формулите създават усложнения.
…

*** Няма научни нововъведения от 1920-те? Дали академичният принцип „публикувай или загини“ задушава науката?**
***„Затварят устата на гениите!“*** https://youtu.be/guQIkV6yCik
20.9.2024 Theories of Everything with Curt Jaimungal, 370 хил. абонати 122 хил. показвания  преди 2 седмици (Theories of Everything with Curt Jaimungal)

Бележки на Тош обобщаващи участието му в предаването. Виж също приложението за Алгоритмична вероятност (Algorithmic Complexity): #complexity.

Грегъри Чайтин, роден 1947 г., с принос в алгоритмичната вероятност още като юноша, известен е като един от тримата в първите работи по сложност на Колмогоров-Соломонов-Чайтин[105]. Математик, информатик, архитекст на ранната експериментална RISC архитектура на процесорите на Ай Би Ем; занимавал се и с биология, говори за покварeността и упадъка на фундаменталната наука и причините за това. Чайтин е самоук, формалното му образование е само  средно, но е получил още две доктората по заслуги. Работил като програмист за Ай Би Ем, което било лесно, и през свободното си време се занимавал с истинска наука. И други значими учени са работили по съществените теми в свободното си време като хоби, защото през „официалното“ си работно са били заети да „оцеляват“ в научния свят с банални и тривиални за тях неща[106] - Айнщайн е най-известният пример с работата си в патентен офис, защото дотогава не е успял да си намери работа в университетската система.

Според Г.Ч. нямало нови революционни фундаментални физични теории от 100 години, от квантовата насам[107], понеже учените не били насърчавани да мислят дръзко и да работят по наистина нови неща[108].

---

[105] https://en.wikipedia.org/wiki/Kolmogorov_complexity Виж и лекцията от първия курс по УИР в света от 2010 г. Сложност: https://research.twenkid.com/agi/2010/Complexity_Probability_Chaos_10-2010_MTR.pdf
[106] Сравни със същите обобщения от руския учен Сергей Савельев. В същото положение да не получава и стотинка за истинската наука за мислещите машини е и авторът на тази книга Тош. и виртуалният институт „Свещеният сметач“.
[107] Виж мнението на Роджър Пенроуз, според когото квантовата теория е просто „погрешна“; при Лекс Фридман.
[108] Сравни с „What’s wrong with natural language processing”, I, II, T.Arnaudov 2009, Artificial Mind – Какво е погрешно в обработката на естествен език?

Съвременните докторанти били послушни, нямали собствени идеи, а искали ръководителите да им дадат тема и те после по нея да правят кариера с дребни допълнения, а по-рано студентите били бунтари, оспорвали онова, което им се преподавало.

Новите идеи често били отхвърляни от колегите и рецензентите, защото не съвпадали с приетото, и учените може да са принудени да ги самопубликуват[109] като откривателят на хипотезата/теория за „червената кралица“ за съвместната еволюция на видовете (коеволюция) на Л. Ван Вален преди 50 години[110].

Твърде голяма бюрокрация, от учените се изисква да пишат „доклади за напредъка“, които да пращат на администратори, вместо да се занимават повече с изследванията си.

Натискът за публикации: „publish or perish” – публикуваш или загиваш – като за оценката е важно количеството и броят на цитиранията, а не качеството (бел.Тош.: също така е спорно и сложно как се мери, колко е ново и пр., така че може би възможно друго решение е повече изследователи да работят в различни посоки; това предлага и самият Чайтин).

Съсредоточаването на ресурси и централизацията на управлението и финансирането, както в ЕС, също водят до застой, стагнация и самоподдържащи се авторитети – по-производително за новаторство са  множество от разпределени и независими институти и учени, развиващи изследванията в различни посоки*. За сравнение: Древният Рим с неговия прословут „Римски мир“ (Pax Romana)[111]. Римската империя била държава на велики инженери и администратори, но не и откриватели на нововъведения, за разлика от древна Гърция, която била раздробена на множество противопоставящи се градове.

---

[109] Сравни със „Свещеният сметач“ – самопубликуваш и поради липса на квалифицирани или подходящи рецензенти и места за публикуване, поради излишни разходи, които не може да си позволиш, формалности и др.

[110] Leigh Van Valen, A New Evolutionary Law. Evol. Theory 1, 1–30 (1973). https://ebme.marine.rutgers.edu/HistoryEarthSystems/HistEarthSystems_Fall2010/VanValen%201973%20Evol%20%20Theor%20.pdf  *
https://en.wikipedia.org/wiki/Red_Queen_hypothesis
https://www.researchgate.net/publication/321742213_Eco-evolutionary_Red_Queen_dynamics_regulate_biodiversity_in_a_metabolite-driven_microbial_system
Видовете са притиснати да се развиват от  успоредно развитие на конкурентните им видове в състезателно „превъоръжаване“. Сравни с GAN. Тези идеи могат да бъдат плодотворни и за машинно обучение.

[111] Римският „мир“ е циничен с начина си на постигане. Виж карикатурите в началото на книгата на Тош*: „Митът за демокрацията, или Свобода на словото, добросъвестност и обективност в политическото говорене за демократи и комунисти“*, Разумир бр.3, 2015 г. http://razumir.twenkid.com

Според Г.Ч. римляните използвали роби *гърци* за умствената работа. Сега сме били в подобен „римски“ период. До Втората световна война било период на огромни открития, напр. в чистата математика, в частта с която се занимава той, дори и между двете световни войни в период на хаос са съчинени работите на Гьодел и Тюринг. След Втората световна обаче науката се превърнала в *огромен бизнес*, вече не била хоби за малка група, която наистина обича изследванията и работи от любознателност. Но така мислил той, а той бил „*чудак*“[112], а чудаци били и приятелите му, сред които Стивън Волфрам: трябва да се самопубликуваш и да имаш собствена компания, източник на средства. (…) Според Г.Ч. живеем в *инженерен период* – при който много добре се *строят* неща, но не се разгръщат *творческите* способности.

**Бел. Тош:** Границите между наука и инженерство се размиват. Под наука тук явно се разбира „по-радикално преобръщане на парадигмите“, отмяна, замяна, „пълна отмяна“ за „по-първични неща“. От една страна, в „по-общо“ или по-абстрактно представяне обаче отново може да се окаже, че промяната е на второстепенни подробности и параметри, а ядрото и съществото, познавателните процеси, теорията на познанието, най-първични онтологии, пак са същите. Напр. и квантовата механика и класическата механика са „едни и същи“: някакви *обекти*, които имат някакви *свойства*, *параметри*, част от които са еднакви: координати, кинетична енергия, потенциална енергия – предпоставки за бъдеща промяна и пр., - вектори на движение и пр., които при дадени условия и взаимодействия се променят еди-как си, т.е. числата, които ги описват, след като се извърши „измерване“: еди-каква си процедура, се променят някак си, като това включва и да не се променят (тъждественост, нула) и т.н. От друга страна, по-голямата част от науката се състои в педантичното прилагане на методи, процедури и правене на „каквото трябва“, както е обяснено на други места, напр. и в първите две части на „What’s Wrong with Natural Language Processing…”, 2009 (Какво е нередно в обработката на естествен език?).
* *… „от разпределени и независими институти и учени, развиващи изследванията в различни посоки”* – понякога съсредоточаването и натрупването на критична маса обаче също е важно, защото улеснява многостранността и всестранността (интердисциплинарноста) и общуването между различните изследователи, области; по-бърза размяна на идеи, преливане между тях и пр. Институтите или дейците пак може да са независими, и да имат свобода да решават какво да търсят и

[112] Интересна дума: oddball (чудак)

как да работят. Прорчества за всестранността виж: „Как бих инвестирал един милион с най-голяма полза за развитието на страната“, Т.Арнаудов, 2003, и „Първата съвременна стратегия за развитие чрез изкуствен интелект е публикувана от 18-годишен българин през 2003 г. и повторена и изпълнена от целия свят 15-20 години по-късно: Българските пророчества: Как бих инвестирал един милион с най-голяма полза за развитието на страната“, Т.А., 2025.

# * Други перспективни открития, разработки, насоки и бележки

*** Нов поглед към тахионите: Как бъдещето влияе на настоящето Учените преразглеждат основите на квантовата теория... |** Емил Василев 21:57 | 05.08.2024 https://www.kaldata.com/нов-поглед-към-тахионите-как-бъдещето-501697.html

Бъдещето да влияело на настоящето; граничните условия, определящи протичането на физичните процеси, включвали не само началното, но и крайното състояние на системата. *„За да се изчисли вероятността за протичане на квантов процес, включващ тахиони, е необходимо да се знае не само миналото му начално състояние, но и бъдещото му крайно състояние.“*

**Тош:** Това не означава само, че „бъдещето влияе на миналото“, а че бъдещето е **предопределено, и че** бъдещото *състояние* е **известно още в началото.** Обхватът между началното и крайното състояние е пространството, в дадено представяне, което влияе върху/е свързано/ отразява развитието, случващото се през този предвидим обхват, период.

*** Study Reveals Dopamine’s Limited Role in Rapid Neural Activity**
FeaturedNeuroscience·August 2, 2024 #невронауки
https://neurosciencenews.com/dopamine-striatum-neural-activity-27512/#:~:text=Dopamine%20signals%20within%20normal%20ranges,factors%20may%20be%20more%20influential.

Според това изследването с мишки допаминът във физиологични граници, напр. след приемане на храна (не се разглежда случай с приемане на стимуланти), и няма значително вляние върху възбуждането на сигнали от подкоровите ядра (базални ганглии, basal ganglia – центрове от награждаващата система на мозъка) и по-точно: стриатума, striatum[113]; и нивото

[113] https://en.wikipedia.org/wiki/Striatum

на допамина не променя поведението в обхвати под секунда. От това се предполага, че допаминовите неврони взимат слабо участие в настройката на възбуждането на стриатума в обхвата под секунда. Сравни с беседата на руски: *„Дофамин — не удовольствие …“*[114].

## * Машинно обучение съобразено с физиката и Аналитична регресия (Physics Informed ML and Symbolic Regression)

### * Дълбоки мрежи за оператори (DeepONet) и Машинно обучение съобразено с физиката – за частни диференциални уравнения и др.

**Deep Operator Networks (DeepONet) [Physics Informed Machine Learning]** Steve Brunton | 356 хил. абонати | 10 414 показвания 2.08.2024 г.
https://www.youtube.com/watch?v=CDCyOHXDRcI
И целия канал от същия автор: https://www.youtube.com/@Eigensteve с лекции за: „Residual Networks (ResNet) [Physics Informed Machine Learning]“, „Neural ODEs (NODEs)“, „Hamiltonian Neural Networks (HNN)“ и пр.

**Виж също: *Аналитична регресия PySR (Symbolic regression):*** Python Symbolic Regression (PySR) [Physics Informed Machine Learning], Steve Brunton, 364 хил. абонати, 11.09.2024 г., 20 818 показвания (18.09)

Търсене на аналитично решение чрез добре определени променливи и функции. Старо име и форма на на АР е генетичното програмиране, при което се преобразуват „гени“ описващи програми. Автоматично търсене на програми чрез изследване на пространството от оператори и взаимодействията между тях в дърво на функциите (function trees), подобно на синтактичните дървета и абстрактните синтактични дървета в теорията на транслация в програмните езици, чрез използване и на еволюционни алгоритми. В съвременните форми се използват за откриване на решения на диференциални уравнения и откриване на матриците на Хамилтон и Ланграж от числени данни, измервания.

Свързано е със SINR – Sparse Identification of Non-Linear Dynamics.

Виж също школата на ***Джош Таненбаум:*** *търсене на програми,* program search. Напр. „Symbolic metaprogram search improves learning efficiency and explains rule learning in humans“[115]

---

[114] https://www.youtube.com/shorts/Yp6Lz6lfa6U

[115] : „Symbolic metaprogram search improves learning efficiency and explains rule learning in humans“, Joshua S. Rule, Steven T. Piantadosi, Andrew Cropper, Kevin Ellis, Maxwell Nye &

https://www.nature.com/articles/s41467-024-50966-x ; списък: https://cocosci.mit.edu/publications

Също: познавателните архитектури в системите на Бен Гьорцел/DeSTIN: Novamente, OpenCog и пр.: https://github.com/opencog/moses – MOSES Machine Learning: Meta-Optimizing Semantic Evolutionary Search – тя се основава на докторската работа „Competent Program Evolution", Moshe Looks, 2006: https://web.archive.org/web/20070330204311/https://metacog.org/main.pdf

За работата на Т.Арнаудов. в тази посока, част от езика на разума Зрим и познавателната архитектура „Вседържец", понятийно мислене, създаване на понятия, търсене и пр.: виж бъдещи публикации.

Забележете, и въобще за научната област „машинно обучение" и „конекционисткото-разпределено" течение, което се смята за достатъчно рязко отделено и е част от „математическата" „клика", под „обучение" обикновено разбира *напасване на числови функции с числени методи (curve fitting)* и не работи с ясно определени *понятия*.

Под „тълкуваем", „разбираем", „обясним" ИИ („интерпретируем": interpretable, explainable AI, XAI) или модел се разбира получаване на някакви променливи, 32*a+35*b^2+sin(c)/cos(a*x-t/2) … Какво „значат" тези букви, „символи" може и *да не се разбере*, но ако се сведе до подобна аналитична форма, моделът на даден физичен процес, запис на входни данни и пр. се смята за по-разбираем или за разбираем, отколкото/при невронните мрежи, които също могат да се опишат по подобен начин, като за „променливи" могат да се вземат стойностите на теглата на определени „неврони", разпределения, хистограми и пр. в рамките на дадени слоеве и т.н. и често се описват като приложение на „правилото за влагане" (chain rule) f(f(f(fx))) и пр. И в двата случая тези букви, символи, променливи не са понятия в човешкото мислене и познание, което се развива постепенно и по начина по който те обикновено са свързани с другите понятия.

*** DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators**, Lu Lu, Pengzhan Jin, George Em Karniadakis, 2019
* https://arxiv.org/abs/1910.03193 * https://github.com/lululxvi/deeponet
Physics-informed machine learning, GE Karniadakis, IG Kevrekidis, L Lu, P Perdikaris, S Wang, L Yang, Nature Reviews Physics 3 (6), 422-440
* HyperDeepONet: learning operator with complex target function space using

---

Joshua B. Tenenbaum , Nature Communications volume 15, Article number: 6847 (2024), 10 August 2024

the limited resources via hypernetwork: https://arxiv.org/html/2312.15949v1

При обикновените НМ се намира функция, която преобразува входните данни в изходни данни. А DON преобразува входна функция в изходна функция като намира решение на система от частни диференциални уравнение – оператори. Параметрите на ЧДУ се задават отвън (input forcing): чрез управление, действеник (actuation), като отражение на някакви смущения.

ДМО са по-ефективни от напълно-свързани НМ в тези задачи (fully-connected networks; като MLP; в края на повечето класифициращи мрежи се поставя един или повече такива слоя).

Засега архитектурата била тествана само за прости функции, не работела за сложни и хаотични ф.

Вид предварително известни условия, ограничения, предпоставки, евристики, които насочват търсенето.

Съгласувани с физиката невронни мрежи (PINN) – вграждат величините от физически измервания и под формата на частни диференциални уравнения във функцията за загуба на НМ и автоматично диференциране.

*** Машинно обучение на динамични системи чрез данни: въведение в обучението на физични дълбоки невронни мрежи**
**Learning dynamical systems from data: An introduction to physics-guided deep learning**, Rose Yu & Rui Wang, Edited by Yuhai Tu, 24.6.2024
https://doi.org/10.1073/pnas.2311808121
https://www.pnas.org/doi/10.1073/pnas.2311808121

Подробен обзор и въведение: динамичните системи са математически системи от диференциални уравнение, било обикновени или частни, от к-ти ред, които моделират m-измерна система в d-измерна дефиниционна област, с множество от реални числа, и може да бъде както линеен, така и нелинеен оператор. Също така обикновено трябва да зададат подходящи граници и начални условия, за да се гарантира съществуването на  решение (т.е. възможно е и да няма такова). Динамичните системи са подвижни, развиват се във времето, затова включват параметъра t, обикновено поставян на първо място: $x_t = x(t, \ldots)$. „Научаването“ на динамичната система е разпознаването, откриването на функцията на промяна на параметрите въз основа на последователности от данни за известни състояния.

„Дълбокото“ машинно обучение (МО) има предимство пред обичайните физични методи за симулации с крайни елементи и крайни разлики при големи размерности на задачата, и често се налага намаляване на точността или мащаба, напр. при симулиране на океана

по данни за температурата на водата на различни места. МО ускорява симулациите многократно след първоначално обучение напр. чрез автоматичното диференциране.

Предимства и недостатъци на МО с НМ и основано на физика:
**МО** (невронни мрежи, построяват графични модели, вид графи на факторите, factor graphs):
**+** изразителни модели, изчислително ефективни, обобщават (за други данни)
**–** липсва физическо „разсъждение“, трудност за тълкуване, голям обем необходими данни за обучение (за постигане на правилно предвиждане в зададени граници)
**Физичен метод:** (диференциални уравнения и симетрии):
**+** знания за предметната област, лесни за тълкуване, малък обем необходими данни за построяване на модела.
**–** твърди, негъвкави допускания; изчислително неизползваема за големи задачи; слабо обобщение.

Откриването на управляващите уравнения (governing equations) се извършва чрез символна регресия[116] и „разредена“ регресия“. За първата – виж по-горе. Втората използва речник от допустими функции и избиране на „разредни“ нискоизмерни модели от данните. Други ползват плитки НМ[117], като заменят функциите на задействане с предварително определени базови функции, включващи тъждество, тригонометрични и използват специален модул за деление (сравни с дадените по-долу KAN/МКА: Мрежи на Колмогоров-Арнолд).

Статистическият еквивалент на машинното обучение в динамичните системи е разпознаването на системата (system identification).

Цели на обучението: 1: решаване на диференциални уравнения, прогнозиране на промените (динамиката) и откриване на водещите уравнения.1: чрез пресмятане на приближения с НМ. 2: предсказване на изображението вход-изход, намаляване на грешката в предвижданията; усложнение – многостъпково прогнозиране, когато се налага да се

---

[116] M. Schmidt, H. Lipson, “Symbolic regression of implicit equations” in Genetic Programming Theory and Practice VII (Springer, 2009), pp. 73–85.
L. Ljung, “System identification” in *Signal Analysis and Prediction* (Springer, 1998), pp. 163–173. * S. L. Brunton, J. L. Proctor, J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proc. Natl. Acad. Sci. U.S.A. 113, 3932–3937 (2016).
[117] S. Sahoo, C. Lampert, G. Martius, “Learning equations for extrapolation and control” in International Conference on Machine Learning (PMLR, 2018), pp. 4442–4450.
G. Martius, C. H. Lampert, Extrapolation and learning equations. arXiv [Preprint] (2016). https://doi.org/10.48550/arXiv.1610.02995 (Accessed 28 August 2023).

използва предвиждане на модела за прогнозиране на следващите стъпки: за преодоляването на тази трудност се прилага разширяване на разнообразието на данните (data augmentation),планирано извличане на данни (scheduled sampling[118]) и състезателно обучение (adversarial training). 3: чрез комбиниране на сбор на множество от базови функции като полиноми или тригонометрични ф., и модел за МО ги съчетава, за да получи управляващото уравнение. Използва се регуляризационна константа, обикновено нормата $L_1$ (сбор на абсолютната стойност на компонентите на вектора, манхатаново разстояние deltaX + deltaY и пр.; за разлика от $L_2$ - евклидово разстояние, SQRT($X^2$+$Y^2$)) – за намаляване на влиянието на единични високи тегла на отделни измерения на вектора, които се „наказват“, както с “drop-out” в НМ.

МО водено от физически знания: чрез 1) модел „с остатък“ – част от системата се решава с автоматично диференциране, а член-остатък от уравнението, който е недиференцируем или грешка, разлика от целевата стойност, се доизчислява с НМ. 2) Обучаеми оператори – изображения между пространства на функции до друго пространство на функции: линейни, диференциални, на Фурие. Традиционните са определени ръчно. Днес: МО на пространството от функции. Обучаемите (променливи) параметри се настройват от данните, които след това поддържат същите математически свойства на изображението между функциите (Банахови пространства), но се по-гъвкави и по-подходящи за прогнозиране. Използват се конволюционни НМ (CNN) напр. за динамика на флуидите или изследване на климата (Computational Fluid Dynamics). *Еквивариатнтни* невронни мрежи – използват симетрии и влагане на еквивариантни функции: CNN, споделяне на тегла, G-invariant: f($p_{in}$(g)(x)) = f(x).Симетрии в линейните пространства са например преместването, въртенето и уголемяването (транслация, ротация, мащабиране) – те запазват определени свойства и отношения след прилагане на преобразуването. *Разплетени представяния* – напр. разделяне на отделни променливи за времето и пространството или намаляване на измеренията в скритото простанство.

**Трудности и бъдещи направления:** 1) Обобщаване при изместване на вероятностните разпределения, успешна работа с данни извън набора за обучение. Един от начините е именно чрез използването на физически функции, напр. в остатъчните модели, които насочват обучението, и

[118] https://www.activeloop.ai/resources/glossary/scheduled-sampling/
* S. Bengio, O. Vinyals, N. Jaitly, N. Shazeer, Scheduled sampling for sequence prediction with recurrent neural networks. *Adv. Neural. Inf. Process. Syst.* **28**, 1171–1179 (2015). * J. Ho, S. Ermon, Generative adversarial imitation learning. *Adv. Neural. Inf. Process. Syst.* **29**, 4572–4580 (2016).

универсални диференциални уравнения (UDE)[119].
2) Устойчивост и надеждност (stability and robustness): натрупване на грешка при многостъпково предвиждане, чувствителност към малки отклонения в началните условия (хаотичност). Теория на управлението (control theory) – чрез ограничаването на константата на Липшиц, функция на Ляпунов*, разширяване на данните (data augmentation) и нормализацията им.

*** Machine-Learning Interatomic Potentials for Long-Range Systems,** Yajie Ji et al., Phys. Rev. Lett. 135, 178001 – Published 23 October, 2025 https://journals.aps.org/prl/abstract/10.1103/ssp9-7s81 "*Machine-learning interatomic potentials have emerged as a revolutionary class of force-field models in molecular simulations, delivering quantum-mechanical accuracy at a fraction of the computational cost and enabling the simulation of large-scale systems over extended timescales. However, they often focus on modeling local environments, neglecting crucial long-range interactions. We propose a sum-of-Gaussians neural network* ***(SOG-Net),*** *a lightweight and versatile framework for integrating long-range interactions into a machine-learning force field."*

*** Automatic network structure discovery of physics informed neural networks via knowledge distillation,** Ziti Liu (刘子提), Yang Liu et al. 29.10.2025 https://www.nature.com/articles/s41467-025-64624-3 – *"a physics structure-informed neural network discovery method based on physics-informed distillation, which decouples physical and parameter regularization via staged optimization in teacher and student networks. After distillation, clustering and parameter reconstruction are used to extract and embed physically meaningful structures.""*

*** Математически бележки:** *Устойчивост на Ляпунов* – дали и по какъв начин системата диф.уравнения е сходима към решението, как се променя стойността при малки промени; дали ф. е устойчива (малки промени на параметрите = малки отклонения) или „прескача". При големи промени се използва Теория на управлението. *Функция на Л.* – дали частните производни от първи род са непрекъснати (диференцируеми), така че може да се приложи правилото за влагане, последователно диференциране (chain rule) за да се изчислява градиент. *Константа на Липшиц* – ограничение на обхвата на промяна на стойностите на функциите в определени граници, абсолютната разлика между

[119] C. Rackauckas et al., Universal differential equations for scientific machine learning. arXiv [Preprint] (2020). https://doi.org/10.48550/arXiv.2001.04385 (Accessed 28 August 2023).

абсолютни стойности на функцията в дефиниционната ѝ област. Lip(f), L(f)", ||f||_Lip. < → по-малка норма, по-малко отклонения, по-гладка, по-устойчива ф. В МО – за определяне на стъпката на обучение при спускане по градиента.
L <= max ||Jf(x)|| J – *Jacobian, Якобиана* – матрицата на частните производни от първи ред на всички променливи от диференциално уравнение; тя е „снимка“ на градиента в дадена точка, чрез нея се създава векторно поле, което показва посоките на най-голям градиент на всички координати от обхвата, пространството и по тях може да се извършва „спускане“, търсене на посока на промяна на параметър за следващото итеративно пресмятане. max||Jf(x)|| е максималната норма на функцията. L1 = ||Jf(x)||_1 = sum(sum(abs(Jf(x)))) – първа норма.
L2 =||Jf(x)||_2 =sqrt(sum(sum(Jf(x)^2))); втора  норма (аналогична 3-та, n-та)  L∞ = ||Jf(x)||_∞ = max(max(abs(Jf(x)))) – най-големият елемент.

1. Виж още: https://en.wikipedia.org/wiki/Lyapunov_stability
   https://en.wikipedia.org/wiki/Lyapunov_function
   https://mathworld.wolfram.com/LyapunovFunction.html
   https://www.wolframalpha.com/input?i=Jacobian+matrix+of+%5B%2F%2Fmath%3A%28x%5E2%2By%5E2%2C+x%5E2-y%5E2%29%2F%2F%5D+with+respect+to+%5B%2F%2Fmath%3A%28x%2Cy%29%2F%2F%5D
   **Wolfram Alpha:** Jacobian matrix of [//math:(x^2+y^2, x^2-y^2)//] with respect to [//math:(x,y)//]  ; Lipschitz condition
   Hölder Continuity: dY(f(x), f(y)) ≤ L * dX(x, y)^α
   Hölder Space; Sobolev spaces, Weak derivatives, Weak solutions. „Слаби“ производни, „слаби решения“ – условни заместители на производните на частни диф.у., когато ф. технически е прекъсната и недиференцируема (парадоксално за „диференциално“ уравнение), но е интегрируема по Лебег, които позволяват да се получи решение при зададени ограничения в условията. *Банахово пространство*: пълно нормирано векторно п.; с определено събиране и скаларно умножение  с мярка, която позволява изчисляване на дължина на вектор и разстояние; пълно пространство в смисъла на Коши е когато всяка редица от вектори на Коши в пространството е сходима към граница, която остава в пространството. Нормите (виж по-горе) са неотрицателни, мащабирането (умножението) със скалар (число, величина) е хомогенно, увеличава нормата с абсолютната стойност на скалара; важи неравенството на триъгълника: нормата на сбора на два вектора е <= на сбора на отделните норми. Евклидовото, п. на

непрекъснатите ф. и Lp п. са Банахови. Хилбертово пространство – има вътрешно произведение, в Евклидовото пространство: скаларно, dot product: от два вектора се получава число, с което може да се мери ъгъл и др. отношения между векторите. Важна операция в компютърната графика, МО и др. Може да се обобщи за по-високи измерения. Lp – "p" е естествено число – L1, L2 и т.н. като нормите. В *метричните* пространства е определена „метрика" – *разстояние* между елементите. В *редицата на Коши* – в *метрично пространство* – всеки следващ елемент е на все по-близо до предходния елемент, на все по-малко разстояние спрямо предходния, така че след отстраняване на краен брой елементи от началото на редицата, разстоянието между всеки два от останалите лементи може да се направи произволно малко. Т.е. редицата е сходима. Пълно метрично пространство, както вече беше споменато, е когато всяка редица на Коши в него е сходима до граница оставаща в пространството (X,d). Напр. евклидовитото пр. $R^n$ е пълно м.п., но рационалните числа Q са непълно м.п.Теорема на Банах за фиксираната точка (Banach fixed point theorem): изображение на свиване (contraction mapping, 0<= gamma < 1), за всяко x,y от X: $d(T(x), T(y)) \leq \lambda * d(x, y)$,
https://ru.wikipedia.org/wiki/Слабая_производна
https://en.wikipedia.org/wiki/Weak_solution
https://en.wikipedia.org/wiki/Lebesgue_integral
https://ru.wikipedia.org/wiki/Интеграл_Лебега
https://bg.wikipedia.org/wiki/Реица_на_Коши
https://en.wikipedia.org/wiki/Dot_product

2.

## * Кратки бележки по математика и вериги на Марков

– Обобщение от Тодор Арнаудов

**MCMC – Markov Chain Monte Carlo**

**HMC -** Hamiltonian Monte Carlo/Hybrid Monte Carlo

Различни видове изпробване, четене на случайна величина, sampling – при MCMC се извършват случайни разходки, последователности от случайни стъпки в система, преходи, за които се предполага, че спазват правилото на Марков за независимостта на състоянията от предходните с определена дълбочина - брой преходи. МКВМ (Монтекарлова верига на Марков) работи в пространството, в което са определени работните случайни променливи, например ако се възстановява двуизмерна или триизмерна структура по облак от точки, се четат координатите на точките. Известен пример за метод Монте Карло е изчисляване на стойността на площта на фигура чрез хвърляне на малък предмет и броене на попаденията в него. Например кръг, вписан в квадрат. Колко пъти предметът е в границите на кръга, и колко е по краищата. При голям брой опити се очаква съотношението да клони към вярното.

Въз основа на събраните данни от опити, в кои състояния преминава системата, се построява вероятностен модел, вероятностно разпределение. Понякога получаването на тези данни е скъпо и се извършва чрез симулация. Не е необходимо функциите да са диференцируеми. Metropolis-Hastings, Gibbs Sampling.

При **Хибридния или Хамилтонов метод Монте Карло** се построява разширен модел, в който координатите, които се отчитат, съдържат допълнителна информация за очаквания начин на продължение, обхождане на следващите стъпки – градиент; съответно обаче изисква функцията, моделът да бъде диференцируем. Освен координатите в пространството на параметрите, положения, позиции, се добавя и пространство на моментите: координатите / параметрите имат и „*момент*“ – „инерция“, кинетична енергия и потенциална енергия “p”; изчислява се Хамилтонова динамика за определено време напред или брой „жабешки скока“ (leapfrog steps) – текущото състояние, например като височина във вероятностно разпределение; сборът на кинетичната и потенциалната енергия се запазва постоянен по Хамилтоновата механика в рамките на движението: описва се със закон, според който единият вид енергия преминава в другия; по този начин могат да се преодолеят някои „неравности“ или локални минимуми/максимуми. ХМК е по-подходящ за по-високоразмерни пространства.

Въпроси – при МКВМ – точно от какво начално вероятностно разпределение се черпи.  [20.2.2025]

* **A Conceptual Introduction to Hamiltonian Monte Carlo**, Michael Betancourt , 16.7.2018, https://arxiv.org/pdf/1701.02434

@Вси: „What is to sample from a probability distribution instead of what space? HMC

vs MCMC“ …

## * “Течен изкуствен интелект“ – „Ликуид ИИ“ - методи за машинно обучение основан на динамични системи и „течни“ невронни мрежи

Големи езикови модели постигат на тестове сравними резултати с трансформаторит, „спиноф“ компания на MIT: https://www.liquid.ai/
Развитие на „течните“ НМ (LNN, Liquid NN), използвани първоначално за управление на самостоятелни превозни средства като коли – поддържане на курса по водеща линия на платното – и дронове. Те са подходящи за непрекъснати процеси във времето: времеви редове[120] (time series) и пр.
https://www.liquid.ai/blog/liquid-neural-networks-research 30.9.2024
Предлагат 1.3B, 3B и 40B mixture-of-experts модели[121].

*** “Liquid” machine-learning system adapts to changing conditions**
* “Течна“ система за МО се приспособява към променящи се условия
https://news.mit.edu/2021/machine-learning-adapts-0128 **,** Daniel Ackerman | MIT News Office, 28.1.2021
**Neural circuit policies enabling auditable autonomy,** Mathias Lechner, Ramin Hasani, Alexander Amini, Thomas A. Henzinger, Daniela Rus, Radu Grosu, 2020
https://publik.tuwien.ac.at/files/publik_292280.pdf

*„Основната цел на изкуствения интелект в приложенията за вземане на решения с висока степен на риск е разработка на единичен алгоритъм, който едновременно изразява възможността за обобщаване чрез изучаване на съгласувани представяния на света и построяване на обяснима динамика. …съпоставяне на многоизмерни входове към команди за управление. Тази система демонстрира превъзходна възможност за обобщаване, интерпретируемост и надеждност в сравнение с порядък по-големи системи за обучение от типа „черна кутия“.*
Neural circuit policies (NCPs) – градивните части са не отделни неврони от типа на тези в дълбоките НМ, а по-сложни мрежи, позволяващи по-добро тълкуване, вдъхновени от нервната система на живи организми: структурно отделени на сетивни неврони, интерневрони, командни и моторни и пр. Сравни с Numenta.
Обучение с учител – човек управлява колата, записват се данните от камери и

[120] Ползва се и „временни редове“. Всички последователности от данни, измервания във времето са такива: сетивни данни от звук, видео, двигателни команди – било управление на машини, роботи или хора; икономически данни и пр.
[121] На 13.10.2024 изпробвах на сайта им големия модел на математически теми и 40B се представи достойно по въпроси свързани с бележките по-горе, само накрая взе да се обърква и да повтаря.

ъгъла на завъртане на волана; след това автономно управление от край до края по тези данни: съпоставяне на образа с въртенето на кормилото.
*Също:* https://techxplore.com/news/2025-01-mathematical-insight-neuron-readout-significant.html Mathematical insight into neuron readout drives significant improvements in neural net prediction accuracy, 16.1.2025, Reservoir Computing & Physical RC (PRC)

* **Reservoir computing with generalized readout based on generalized synchronization**, Akane Ohkubo et al, *Scientific Reports* (2024). DOI: 10.1038/s41598-024-81880-3 – generalized synchronization of dynamical systems – one system is fully described by the state of another one (or subsystems); reservoir states; chaotic systems, memristors; complete/lag/phase sync;
Reservoirs: set of nonlinear dynamical units

* **Towards mixed physical node reservoir computing: light-emitting synaptic reservoir system with dual photoelectric output,** Minrui Lian et al., 1.8.2024 https://www.nature.com/articles/s41377-024-01516-z
Reservoir computing systems: memristors8, atomic switching networks, silicon photonics, spintronic oscillators. artificial light-emitting synapses (LES)

* Qi, Z. Y. et al. Physical reservoir computing based on nanoscale materials and devices. *Adv. Funct. Mater.* **33**, 2306149 (2023).
* Du, C. et al. Reservoir computing using dynamic memristors for temporal information processing. *Nat. Commun.* **8**, 2204 (2017)
* Sillin, H. O. et al. A theoretical and experimental study of neuromorphic atomic switch networks for reservoir computing. *Nanotechnology* **24**, 384004 (2013)
* Vandoorne, K. et al. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat. Commun.* **5**, 3541 (2014).

*** LFM2-VL: Efficient Vision-Language Models**, 12.8.2025, https://www.liquid.ai/blog/lfm2-vl-efficient-vision-language-models
FM2-VL-450M, FM2-VL-1.6B – 512×512, intelligent patch-based handling for larger images; 2xfaster on GPU than other models

*** МКА: Мрежи на Колмогоров-Арнолд**

**KAN: Kolmogorov-Arnold Networks,** Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Soljai, Thomas Y. Hou, Max Tegmark, *30.4.2024; last revised 16 Jun 2024 (this version, v4)*
* https://arxiv.org/abs/2404.19756 * https://arxiv.org/html/2404.19756v4
* https://medium.com/@isaakmwangi2018/a-simplified-explanation-of-the-new-kolmogorov-arnold-network-kan-from-mit-cbb59793a040
* https://github.com/mintisan/awesome-kan
Wav-KAN: Wavelet Kolmogorov-Arnold Networks, Zavareh Bozorgasl, Hao Chen * https://arxiv.org/abs/2405.12832

Мрежите на Колмогоров-Арнолд (KAN, Мрежи на Колмогоров-Арнолд: МКА) се основават на теоремата за представяне на Колмогоров-Арнолд и се дават като „обещаващи алтернативи на многослойните перцептрони (MLP – напълно свързаните НМ (FCN), виж и DeepONet). *„Докато MLP имат фиксирани функции за активиране на възлите (неврони), при KAN се обучават **функциите за активиране** на **върховете** (тегла). KAN изобщо нямат линейни тегла -- всеки параметър на тегло се заменя с едномерна **функция**, параметризирана като сплайн.“*

Според работата KAN/МКА позволява непрекъснато/продължаващо обучение (continual learning), за разлика от „катастрофалното забравяне“ при MLP и е по-добра от MLP в точността и обяснимостта; KAN постигат сравнима или по-добра точност от MLP при напасване на данните и решаване на частни диференциални уравнения (PDE) с много по-малък брой параметри и се справят с представянето на функции с ниска размерност, за разлика от MLP. Също така и теоретично, и опитно авторите доказват, че за МКА важат по-бързи закони за невронно мащабиране от MLP.

МКА са по-„обясними“ (interpretability), защото могат да бъдат изобразени по лесно разбираем начин и така позволяват взаимодействие с човек.

**Тош:** Една от слабостите на архитектурата от статията е, че се учи 10 пъти по-бавно от MLP, но не е била и оптимизирана все още. За сметка на това позволявала откриване на вложеност (compositionality), за разлика от MLP и стара слабост на разпределените „конекционистки“ модели, критикувана от 1980-те години.

* Идеята на разлагането ми напомня за мисли от *„Опит за първично разделяне на запис на говор на съставящите го фонеми”,*2004, които тогава недоразвих, освен в синтезаторите на реч „Глас“ от тогава и по-

късно „Тошко 2“, които са „микроформантни“ и тоналните звукове са записани като запис на един период от функция: “вълнички” (wavelet).
https://eim.twenkid.com/old/_5/31/analiz_na_zvuk.htm
https://www.oocities.org/eimworld/_5/31/analiz_na_zvuk.htm
Цитирана като първа препратка в:
*** ПОЗИЦИИ ЗА РЕАЛИЗАЦИЯ НА БЪЛГАРСКИТЕ ФОНЕМИ,** 12.2006, Desislava Baeva, Dimitrina Ignatova-Tzoneva
https://www.researchgate.net/publication/346627640_POZICII_ZA_REALIZACIA_NA_BLGARSKITE_FONEMI
* Арнаудов 2005: Арнаудов, Т. Опит за първично разделяне на запис на говор насъставящите го фонеми. http://eim.hit.bg/_5/30/analiz_na_zvuk.htm

*** Continuous Thought Machines**, Luke Darlow, Ciaran Regan et al., Sakana AI, 8.5.2025/3.10.2025… https://arxiv.org/abs/2505.05522 – neuron-level temporal processing of histories and synchronization; complex sequential reasoning; adaptive compute – it can stop earlier for simpler tasks; towards more biologically plausible AI systems.

# * Компании: SingularityNET, Verses, Thinking Machines, …

*** SingularityNET to invest $53M in AI infrastructure, modular supercomputer,** 23.7.2024
* https://cointelegraph.com/news/singularitynet-invest-53-million-ai-infrastructure-modular-supercomputer
**New supercomputing network could lead to AGI, scientists hope, with 1st node coming online within weeks,** By Lisa D. Sparks, 11.8.2024
* https://www.livescience.com/technology/artificial-intelligence/new-supercomputing-network-lead-to-agi-1st-node-coming-within-weeks

**SingularityNET**: https://singularitynet.io/ Фирма

https://medium.com/singularitynet/singularitynet-decentralized-ai-platform-biweekly-development-report-as-of-august-8th-2024-df9ece97c9e2

**Относно SingularityNET**
SingularityNET е основана от д-р Бен Гьорцел с мисията да създаде децентрализиран, демократичен, приобщаващ и полезен общ ИИ (AGI, УИР). УИР не зависи от нито един централен субект, отворен е за всеки и не е ограничен до тесните цели на една корпорация или държава. Екипът на компанията включва опитни инженери, учени, изследователи, предприемачи и търговци. Основната им платформа и човешки състав е допълнен от специализирани групи за приложения като финанси, роботика, биомедицински ИИ, медии, изкуства и развлечения.

**Краткото представяне по главния сайт:**

Най-важното е: **OpenCog Hyperon –** OpenCog Hyperon е дългосрочен проект за внедряване на пълна, мащабируема и отворена система за общ изкуствен интелект, основана на принципите на OpenCog - платформа с отворен код, където различни стратегии и методи от ИИ като невронно-символен, еволюционно обучение, икономическо разпределение на вниманието, машинно обучение и други могат да си сътрудничат въз основа на метаграф на споделено знание (Atomspace).
**Търговската част:** разпределен („децентрализиран“) пазар за ИИ услуги, работещ чрез блокчейн, за „справедливо разпределение“ на власт, стойност, технологии, полезни блага. Основна цел: разработването на универсален изкуствен разум (общ изкуствен интелект (AGI)) за полезна

технологична сингулярност. Платформа за предлагане, публикуване, изпробване на библиотеки от алгоритми с ИИ, създадени от общност на доставчици на услуги. Език със специално предназначение за ИИ (DSL – Domain specific language).
**Как работи:** Дълбока самоорганизираща се мрежа от агенти с ИИ, работещи на платформата SingularityNET, могат динамично да възлагат работа един на друг — като използват функциите на ИИ, обменят входно/изходни данни, договарят плащания и подобряват системата за репутация на агента. В тази мрежа „интелигентността на цялото значително надвишава интелигентността на частите“ **– сравни с компанията Verses AI*, разработваща** технологии за теорията, започната от Карл Фристън Принцип на свободната енергия/Извод чрез Действие (FEP/AIF), която е една от съвпадащите с основите на Теория на Разума и Вселената/Вселена и Разум на Т.Арнаудов. Виж в главата от основния том на „Пророците на мислещите машини“: учения и школи, „Пионери и Пророци“ и др.

„AGIX Staking & Bridge“ – средство за търговия с монети, „токени“, операции в блокчейн мрежата им, можели да ги прехвърлят към други вериги като Ethereum и Cardano, като вторите са споменати като съдружници.

…

През 10.2024 г. са обявени грантове за разработка на проекти на езика “MeTTa” за OpenCog Hyperion: *“Grants for developing AI code in MeTTa language toward implementing PRIMUS cognitive architecture in Hyperon*[122] https://metta-lang.dev/

**MeTTa**[123] е многопарадигмен език за „познавателни изчисления“, съчетаващ функционално, логическо и вероятностно програмиране, осигурявайки „синергична рамка“ за представяне на декларативно и процедурно знание. Програмите представляват подграф на метаграф от „пространство от атоми“ (Atomspace), което работи („се управлява“?) „централно“ чрез заявки и пренаписване на части от пространството от атоми. Езикът позволява да се пише по естествен начин изключително абстрактен *самопроменящ* се код, по време на изпълнение, и е напълно

[122] https://agi.topicbox.com/groups/agi/T3ced54aaba4f0969
[123] https://deepfunding.ai/all-rfps/
https://deepfunding.ai/rfp/develop-a-framework-for-agi-motivation-systems/
https://metta-lang.dev/docs/learn/tutorials/eval_intro/main_concepts.html
https://metta-lang.dev/docs/learn/tutorials/eval_intro/basic_evaluation.html
https://metta-lang.dev/docs/learn/tutorials/eval_intro/recursion.html

саморефлективен – можем да четем и променяме кода на програмите. Налични са „постепенно зависими типове“ с вградени математически разсъждения и поддръжка на най-съвременна типова система; невронно-символна интеграция, машина за извод – MeTTa е по същество *недетерминиран* език, заради което интерпретаторът му е вид машина за изводи. Езикът поддържа прилагането на различни системи за изводи, от вероятностно програмиране до размита логика.

**Инструмент за AGI** – отворена архитектура, обхваща много различни стратегии за ИИ и е предназначена както за ръчно писане от човешки същества, които да описват в скриптове части от когнитивните процеси на мислещата машина, така и за да служи за самопрограмиране на свързаните с Общия ИИ алгоритми за обучение и разсъждение.

**Специализиран език за ИИ**, който позволява на широк набор от AI системи да си сътрудничат динамично чрез създаването на съвместими, особени за съответната приложна област езици, в единна рамка. (DSL, Domain Specific Language). MeTTa образува „универсалния преводач“.

**OpenCog Hyperon –** MeTTa е езикът на когнитивната архитектура на OpenCog Hyperon. Той функционира като „фърмуер“ на изключително разнообразните компоненти, от които е направен Hyperon, и е лепилото, което държи всичко заедно*.

**Тош:** Нещо като част от инфраструктурата на „Вседържец“ на „Свещеният сметач“, който засега е разнородна система, която използва разнообразен набор от езици и платформи. Една от целите на проекта за ИИ „инфраструктура“ и самопишещи се програми от началото на 2010-те бяха да може лесно да се програмира на *всякакви* програмни езици, които са под ръка, и на „междинни“, на псевдокод и естествен език, ако е достатъчно ясно, и да не съществуват досадни синтактични грешки и пр. *Словеказбител*, казбесловител, вършерод, казбород, гъвче. Повече по тази тема в следващата книга: *„Създаване на мислещи машини“.*

*** VERSES (NEO: VERS) | An Introduction to and a Demonstration of Genius™ (November 3rd, 2023)**
https://www.youtube.com/watch?v=zSILLYyCrGI
FCIR Media | 2,21 хил. абонати | 2552 показвания 6.11.2023 г. [към 12/8/24]
Показват демо на агент, управляващ дрон във въображаем свят.

* www.verses.ai „VERSES публикува пионерски изследвания, демонстриращи по-гъвкава, ефективна основа за физика за ИИ от следващо поколение“ …… Одобрение на космически уеб стандарти … Партньор на VERSES и JPL,

финансиран от НАСА, за преследване на стандарти за космическата индустрия

*** VERSES Publishes Pioneering Research Demonstrating More Versatile, Efficient, Physics Foundation for Next-Gen AI**
Verses**,** Aug 1, 2024
https://www.verses.ai/press-2/verses-publishes-pioneering-research-demonstrating-more-versatile-efficient-physics-foundation-for-next-gen-ai
*"**Renormalizing Generative Models (RGMs)** that address foundational problems in artificial intelligence (AI), namely versatility, efficiency, explainability and accuracy, using a physics based approach."*
*"Your brain doesn't process and store every pixel independently; instead it "coarse-grains" patterns, objects, and relationships from a mental model of concepts - a door handle, a tree, a bicycle. RGMs likewise break down complex data like images or sounds into simpler, compact, hierarchical components and learn to predict these components efficiently, reserving attention for the most informative or unique details. "*
**Tosh**: Like my visions, yet not implemented completely. **"Hierarchical" is also/more about being incremental, compositional, reusable, systematic** etc.

**Тош:** Подобно на все още неизпълненото изцяло: „йерархично“ е също/повече относно постепенно, съставно, преизползваемо, систематично; понятийно и т.н. :*„Вашият мозък не обработва и съхранява всеки пиксел независимо един от друг; вместо това създава по-„груби“ модели, обекти и взаимоотношения, които са част от умствен модел на понятия – дръжка на врата, дърво, велосипед. Ренормализиращите пораждащи модели (RGM) също разбиват сложните данни като изображения или звуци на по-прости, компактни, йерархични компоненти и се научават да предсказват тези компоненти ефективно, запазвайки вниманието за най-информативните или особени подробности.“*, 8/2024 с:

*** Човекът и Мислещата Машина (Анализ на възможността да се създаде Мислеща машина и някои недостатъци на човека и органичната материя пред нея)**, Свещеният сметач, **12.2001 г.:**
(…) При човека, повечето обекти (от всякакъв вид) не се запомнят "фотографски", а се "преразказват" в мозъка, записват се най-характерните особености на информационните обекти, входната информация се компресира. Във "фотографска", "фонографска", "текстографска","стереографска" (пространствена) и пр. памет вероятно

се съхраняват само основните понятия. **Човешката памет не е особено силна в точното запомняне, пък и то заема много място (в уж "безграничния" капацитет на човешкия мозък). Хитро е новопостъпилият информационен обект да се обясни с наличните информационни обекти. Просто му се дава етикет, а същността му се описва с известните понятия, като се използват техните етикети - връзки към значението им, съдържащи само "адрес". Ние хората наричаме такова запомняне "разбиране" и "осмисляне".**

Виж също останалите работи от ТРИВ (СВП2, СВП3, Вселена и Разум 4, 2003-2004) и „Столове, сгради, карикатури …“, 2012, и листовете с бележки по ИИ от 1999 г., където също са скицирани основни идеи за понятията, развити в изложението от 2012 г.

Лексемите, „токените“ в преобразителите също се основават и използват този принцип на разделяне и съединяване на данните. Повече: в следващата книга *„Създаване на мислещи машини“.*

# * Нови компании:

**Thinking Machines**, основатели: **Founding Team -** връзки към профили на личностите; Мира Мурати от OpenAI:
**Сайт на компанията:** https://thinkingmachines.ai/ ,
**Личности:** Alex Gartrell, Alexander Kirillov, Andrew Gu, Andrew Tulloch (Chief Architect), Barret Zoph (CTO), Brydon Eastman, Chih-Kuan Yeh, Christian Gibson, Devendra Chaplot, Horace He, Ian O'Connell, Jacob Menick, John Schulman (Chief Scientist), Jonathan Lachman, Joshua Gross, Kevin Button, Kurt Shuster, Kyle Luther, Lilian Weng, Luke Metz, Mario Saltarelli, Mark Jen, Mianna Chen, Mira Murati (CEO), Myle Ott, Naman Goyal, Nikki Sommer, Noah Shpak, Pia Santos, Randall Lin, Rowan Zellers, Sam Schoenholz, Sam Shleifer, Saurabh Garg, Shaojie Bai, Stephen Roller, Yifu Wang, and Yinghai Lu. 2.2025 г. Mira Murati – преди CTO в OpenAI; „Model intelligence as the cornerstone … „ Advanced multimodal capabilities.
https://techcrunch.com/2025/02/18/thinking-machines-lab-is-ex-openai-cto-mira-muratis-new-startup/
@Вси: виж връзките и изследвай гитхъб хранилищата и работата им.
**По-късни новини:**
Първият им продукт „Tinker” за донастройка на езикови модели:
* Announcing Tinker, Thinking Machines Lab, Oct 1, 2025
https://thinkingmachines.ai/blog/announcing-tinker/
…
* **World Labs - $230m Startup Led by Fei Fei Li. Step inside images and interact with them!**
https://www.reddit.com/r/StableDiffusion/comments/1h53uhj/first_demo_from_world_labs_230m_startup_led_by/ * https://www.worldlabs.ai/

https://marble.worldlabs.ai/ – world models (simulations) from prompts; editable etc. "*We are a spatial intelligence company building Large World Models to perceive, generate, and interact with the 3D world."*

*** How World Models are Changing the Future of AI Beyond Transformers**, Eye on AI, 32,1 хил. Абонати 9477 показвания [28.11.2025], 23.11.2025 г. https://www.youtube.com/watch?v=9VcXiyE40xw – a talk with Fei Fei Li

Many romanticized echoes of general suggestions about intelligence from TUM. Compare the ambitions and arguments from the talk to the principles, suggestions and plans of Theory of Universe and Mind 2001-2004 and later ideas, expressed for example in "Chairs, Buildings, Caricatures…", 2012; also in particular the explanations of the thinking machine Emil in "The Truth", 2002 (*И*стината).

One of the summaries ot TUM is that cognition, mind is, or it can be represented as a system of Hierarchical universal simulators of virtual universes, where the universe (or world) is more abstract than the physical, it has different and multi-range, multi-resolution, multi-domain, multi-precision … etc. representations, levels, views etc. See also Todor's summary about embodiment in a 2011 article (included in the main volume): coordinate spaces, etc. and Jeff Hawkins' *"frames of reference".*

The need and foundational role of spatial cognition for intelligence is obvious from the start, though. Computer vision, computational geometry and topology, navigation/map-building (either SLAM or off-line), robotics etc. – all related to graph theory etc. is about "spatial cognition" in the usual sense of 3D space where the ordinary physical objects and the humans reside. On the other hand, "everything" can be represented in spaces, as vectors etc., including the token embeddings of the purely textual large language models etc. See the whole collection of books from *The Prophets;* appendices #Anelia, #Lazar; the main volume #tosh1 #prophets – work by Dragomir Anguelov, Vassil Chatalbashev etc. – the early 2000s 3D reconstruction and pose classification from laser range data – point clouds collected with 3D scanners. [28.11.2025]

The examples about editing of the couch*, "I don't like this couch to be pink, change it to blue…"* with verbal commands about 47-49 min are funny. In fact command line interfaces for everything, the basic Operating System jobs or for programming also accept "verbal" and "linguistic" commands. In the past the change or the extension of the human-computer interaction from typing linear commands into more dynamic and spatial dialogs with GUI – graphical user interface – was considered a progress.

* **Amazon says its AI video model can now generate minutes-long clips, Kyle** Wiggers, 2:08 PM PDT · April 7, 2025 – до две минути с много кадри по 6 секунди, запазващи връзка помежду си, до 4000 знака заявка. Nova Reel 1.1: "Multishot Manual.", 1280x720, 512 знака, до 20 кадъра; „модели на света"

*** Who is Alexandr Wang? The 28-year-old hired for 14 billion US dollar by Mark Zukerberg to lead Meta superintelligence labs** - Technology News | The Financial Express – https://www.financialexpress.com/life/technology-who-is-alexandr-wang-the-28-year-old-hired-for-14-billion-us-dollar-by-mark-zukerberg-to-lead-meta-superintelligence-labs-4036364/

*** Physical Intelligence (π)** – https://growjo.com/company/Physical_Intelligence *"Physical Intelligence is bringing general-purpose AI into the physical world. We are a group of engineers, scientists, roboticists, and company builders developing foundation models and learning algorithms to power the robots of today and the physically-actuated devices of the future."*
* https://news.bloomberglaw.com/business-and-practice/robotics-startup-physical-intelligence-valued-at-5-6-billion 20.11.2025 – the company "has raised $600 million in a new round of funding that values the company at $5.6 billion" https://www.physicalintelligence.company/ – *** π*0.6: a VLA that Learns from Experience, 17.11.2025,** https://www.physicalintelligence.company/blog/pistar06 – operating a coffee machine, folding cardboard boxes, folding towels, … *"Coaching with corrections and practicing with reinforcement" …* Robots are trained via teleoperation (immitation learning). *"Recap" can in principle address these limitations by additionally learning directly from the robot's own experience. …*
*** Real-Time Action Chunking with Large Models,** *9.6.2025*
https://www.physicalintelligence.company/research/real_time_chunking
*** FAST: Efficient Robot Action Tokenization**, 16.1.2025
https://www.physicalintelligence.company/research/fast – ... Diffusion or flow matching, π0-model – instead of the default transformers tokenization for robot actions, "action chunking" into 3-5 actions for "crude systems" and up to 20-50 actions in a chunk for high-frequency dexterous robots. Dense-action tokens.
*** π0: Our First Generalist Policy,** 31.10.2024 – https://www.physicalintelligence.company/blog/pi0 – the Moravec paradox, "artificial physical intelligence" – *"users can simply ask robots to perform any task they want, just like they can ask large language models (LLMs) and chatbot assistants. … Like LLMs, our model is trained on broad and diverse data and can follow various text instructions. Unlike LLMs, it spans images, text, and actions and acquires physical intelligence by training on embodied experience from robots, learning to directly output low-level motor commands via a novel architecture. It can control a variety of different robots, and can either be prompted to carry out the desired task, or fine-tuned to specialize it to challenging application scenarios.."*; Multiple dexterous robots: UR5e, Bimanual (UR5e, Trossen, Arx); Franka, Mobile (Trossen,Fibocom). Tasks: Laundry, Folding clothes, Make coffee, Bag groceries, Bus table, Open popcorn; collecting eggs in carton … Internet-scale semantic understanding; output motor commands at 50 Hz, *augment pre-trained VLMs wth continuous action outputs via flow matching* – a variant of diffusion models.
*** π0: A Vision-Language-Action Flow Model for General Robot Control**, Kevin

Black, Noah Brown et al., https://www.physicalintelligence.company/download/pi0.pdf

*** Project Prometheus** - Jeff Bezos founds AI startup Project Prometheus, 18.11.2025 – **$6.2 billion funding** https://www.heise.de/en/news/Jeff-Bezos-founds AI-startup-Project-Prometheus-11083426.html
* https://www.nytimes.com/2025/11/17/technology/bezos-project-prometheus.html – *"focusing on AI for the engineering and manufacturing of computers, automobiles and spacecraft."*
*** Klover** - https://www.klover.ai/agd/ Yoshua Bengio … Artificial General Decision Making AGD™ …
*** Sakana AI** – https://sakana.ai/ – Continuous Thought Machine, 5.2025; https://sakana.ai/ai-scientist/ – The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery …

*** Human& -** Leading AI researcher Eric Zelikman is raising $1 billion to build AI models with emotional intelligence, By Ben Bergman, 1.11.2025 https://www.businessinsider.com/researcher-raising-1-billion-to-build-ai-models-with-eq-2025-10
*** Quiet-STaR: Language Models Can Teach Themselves to Think Before Speaking,** Eric Zelikman et al. https://openreview.net/pdf?id=oRXPiSOGH9 – "*reason generally from text, rather than on curated reasoning tasks or collections of reasoning tasks; parallel sampling; custom meta-tokens at the start and end of each though; generating a rationale; a mixing head; a non-myopic loss* – predicting *multiple tokens* ahead, not only the next one*; …"* Self-Taught Reasoner (STaR): *inferring rationales from few-shot examples in question-answering and learning from those that lead to a correct answer.* hidden thoughts/reasoning;
*** Star: Bootstrapping reasoning with reasoning. Advances in Neural Information Processing Systems,** Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. 2022. https://arxiv.org/abs/2203.14465 28.3.2022/5.2022 *"Generating step-by-step "chain-of-thought" rationales improves language model performance on complex reasoning tasks like mathematics or commonsense question-answering …*

# * НЕВРОНАУКИ и съчетание с машинно обучение, учене с подкрепление и изкуствен интелект

# * Neuroscience in Combination with Machine Learning, Reinforcement Learning and Artificial Intelligence

#neuroscience **#невронауки** **#**Български и English.

* Динамични системи в ума и състезание без победител
* Machine Learning and Human Learning in Neuroscience Research

## * Динамични системи в ума и състезание без победител

## * Dynamical systems and Winnerless Competition

Направлението с динамични системи е алтернатива на дълбоките невронни мрежи

*„Състезание без победител“* (Winnerless Competition) – по изследвания от школата на *Михаил Рабинович* и др. – вид взаимодействие между части от системите на мозъчните невронни вериги, описани като динамична система.

Сравни с предложенията и предвижданията от ТРИВ, „Анализ на смисъла на изречение…“, Тош, 2004 и лекцията по УИР от 2010 г. за начина по който се превключват УУ, които са овладели тялото, външният взаимлик, интерфейс, действеници, извеждане, управление за дадено ниво на въображаема вселена, управляващо-причиняващо устройство в даден момент не трябва да може да задържа властта неограничено; статията за липсата на единно съгласувано „аз“, а вместо това „интеграл от безкрайно малки Аз-ове“ от 2012 г.: и др.
https://artificial-mind.blogspot.com/2012/11/nature-or-nurture-socialization-social.html

В публикациите на руски М.Рабинович използва *„конкуренция без победител“.*

* Виж по-долу статията „*Основни функционални мрежи в човешкия мозък …“*

# * Математика на съзнанието

*** Математика сознания,** Михаил Рабинович[124], Пабло Варона (увод на руски, слайдове на презентация на английски)

* Рабинович М., Варона П. Математика сознания // Известия вузов. Прикладная нелинейная динамика. 2017. Т. 25, № 3. С. 5–51. https://cyberleninka.ru/article/n/matematika-soznaniya/viewer

с.3 *„Смята се, че съзнанието е следствие на взаимно съгласувана дейност (понякога я наричат обединена) на повечето познавателни структури в мозъка. Различните познавателни функции се изпълняват от различни йерархични мрежи, всяка от които обединява голям брой подструктури на мозъка. Подобна мрежа може да се разглежда като неподредена дискретна решетка, чиято активност се представя като множество от състояния или възбуждания. Преходната във времето динамика, отговаряща за познавателния процес, е не друго, а последователната смяна на състоянието в следствие на взаимното потискане на принципа „конкуренция без победител“.**

– Понятия: събития, епизоди, динамика на споряване (binding), …

*с.27: Различни видове памет – различна динамика на извличане*

*с.28: Неустойчивост на смисловите късове: Произходът на творчеството е хаотичната разходка: вляво – понятийна мрежа от късове-по-теми в познавателното смислово пространство; вдясно: метастабилно състояние в многоизмерно неустойчива разделяща граница във фазовото пространство на динамиката на разделяне на късове (topic-chunks – късове-по-теми; separatrix, „сепаратрица“ – разделяща граница/линия;*

*с.35: Мрежи за творчество: зелено – Изпълнителна мрежа на вниманието (Executive Attention Network); Червено – мрежа на въображението (The Imagination Network); Жълто – Мрежа на очебийността (The Salience Network)*

*с.38: **Творчеството е самосъздаване на нови мисли, мелодии и други шевици, определени от крайна стойност на информацията**.*

*[Осъществява се чрез:]*

*– Взаимодействие между мрежите в покой (Default modes network), мрежата за автобиографична памет и емоциите: модел на сложни мрежи тип WLC с последователна динамика.*

*– Два механизма за пораждане на нови шевици [“patterns”: модели-схеми, поредици, съчетания]*

---

[124] * М.Рабинович е роден на 20.4.1941 в Русия, от 1990-1991 г. работи в Чикагския университет, Калифорнийския в Сан Диего. Основател на институт за перспективни изследвания в Нижни Новогрод 1994-2002*, от 2004 г. живее в САЩ. (*Сравни с американския със същото име, където се разработва компютърът IAS с Фон Нойман. Виж в приложението „Първата съвременна стратегия за развитие чрез изкуствен интелект…“, Т.Арнаудов, 2025

*(i) Промяна на архитектурата на всеобщите мрежа, когато се появят нови връзки; новата мрежа изявява нови шевици или енграми, и:*
*(ii) чисти динамични механизми, отнасящи се до бифуркации от типа „а-ха-а-а-а!“ [внезапни хрумвания] в умственото фазово пространство, където динамиката на една умствена мрежа модулира динамиката на друга. Това води до възникване на нови метастабилни състояния във фазовото пространство. Основание на тях нови умствени мрежи са математическия образ на оригиналността (creativity).*
*с.39: Йерархична динамика на музикалната импровизация (формули, суми и произведения) – извличане на спомени и свързване; разделяне и групиране; управление на емоциите и вниманието (memory retrieval and binding; chunking; emotion-attention control)*
*с.46: Таблица с формули и илюстрации на явления: последователно хетероклинично превключване, свързване и поток на информацията, сътрудничество, групиране и разделяне на памет и обучение. Динамични системи, „фазови портрети“;*

***** Виж **Теория на хаоса и динамични системи**, за да навлезеш в понятията: Chaos theory, dynamical systems; https://en.wikipedia.org/wiki/Chaos_theory
*** Познавателният процес** като преходна динамика на последователната смяна на състоянието, „вследствие на взаимното потискане на принципа „конкуренция без победител““ е същото като описаното теоретично в *„Анализ на смисъла на изречение въз основа на базата знания на действаща мислеща машина…“,* Т.Арнаудов 2004, където въпросните мрежи или състояния отговарят на множества от въображаеми управляващи устройства (УУ), които „завземат“ изходните устройства за даден период от време и се опитват да постигнат най-висок сбор от награда за своята структура. За да се избегне мъртва хватка и „пристрастяване“ обаче е необходимо никое УУ или съединила се, съгласуваща се, група от подустройства, да може да завзема властта за неограничено време, т.е. не трябва да има „един победител“. Свързана с това е и мисълта от статията от 2012 г., че в ума не съществува обективна единна обединена човешка личност, а непрекъснато се пресмята интеграл на безкрайномалки местни личности, „аз-ове“ на множество от мащаби. Заради това се получава и заблуждаващото „нерационално“ поведение – агентът, който управлява тялото, не е един и не е постоянен и неизменен през цялото време, затова няма единствена най-„целесъобразна“ и съгласувана функция на ползата и т.н. Виж по-долу главата: Машинно обучение, обучение на човека и невронауки и взаимодействието между тях.
* Пример за метастабилно състояние на възприятието са например зрителните илюзии, при които може да видите „стара жена или млада жена“, че фигура се върти наляво или надясно; ваза или два силуета и пр., в зависимост от това върху какво се съсредоточи вниманието ви и коя от двете възможни хипотези надделее. (Виж: „Rubin vase“ и пр.) https://en.wikipedia.org/wiki/Rubin_vase

Виж също за нежелателни метастабилни състояния в цифрови електронни схеми и преодоляването или предотвратяването им например чрез *арбитри*, междинни схеми, служещи за синхронизация на сигнали, породени от асинхронни подсистеми, източници с различни тактови честоти или при очаквана възможност за недопустимо разминаване във времето , например от различни шини. За метастабилност се говори също в научната школа на ПСЕ/ИЧД на К.Фристън и последователите му, напр. в предаването с Mahault Albarracin, разгледано в този том (metastability, Active Inference Insights 002). Виж също основния том. Абстрактни, неуточнени арбитри са посочени също в лекцията за ТРИВ от първия в света курс по УИР в Пловдив през 2010 и 2011 г., във връзка с предложения начин на работа на въображаемата мислеща машина, при който не е позволено на дадено управляващо устройство или „група“ от такива да завладеят *действениците* – крайните, изходните, „управляваните“ части, извеждащите, действащите, въздействащите – на системата за неограничено време. Като степен системата може да бъде и *подсистема*, арбитрирането, избирането на деец, който да поеме властта или в каква степен в даден момент, да се отнася за дадено ниво от йерархичната структура от вложени универсални симулатори на въображаеми вселени, с даден обхват, РСВУ и пр.
https://en.wikipedia.org/wiki/Metastability_(electronics)
https://en.wikipedia.org/wiki/Metastability метастабилност във физиката и химията
https://en.wikipedia.org/wiki/Non-equilibrium_thermodynamics - термодинамика на физични системи, които не се намират в равновесно състояние. За разлика от термодинамиката при равновесно състояние, когато се разглеждат неравновесни термодинамични системи е трудно да се определи ентропията в даден момент на макронив, като тя се смята като сбор от локално определена плътност на ентропията. Било наблюдавано, че в системи, които са далеч от всеобщо равновесие, запазват местно (локално, но не глобално).

Забележи, че това съвпада с предложението от ТРИВ за това че няма единна обединена обективна най-добра възможна траектория на сложно многослойно и многомащабно управляващо устройство и определянето на „самоличността“ на деец, агент, управляващо-причиняващо устройство на дадено макрониво като интеграл от „безкрайно-малки“ Аз-ове в кратки времепространствени отрязъци – от цитираната по-долу работа „Анализ на смисъла“… и друга от 2012 г. Не си спомням да съм познавал термодинамичната формулировка тогава[125].

Виж също споменатите Принцип на свободната енергия/Извод чрез действия – учение, което разглежда подобни въпроси при всякакви видове агенти, части от Вселената, живи организми и пр, подобно на ТРИВ.
* Лекцията на английски за Теория на Разума и Вселената, до 17 слайд и от 12-17 конкретно по обясненото по-горе.

[125] Чета за конкретния споменат „интеграл на локалната ентропия“ от статията в Уикипедия днес, 8.10. 2025

https://research.twenkid.com/agi/2010/en/Todor_Arnaudov_Theory_of_Hierarchical_Universal_Simulators_of_universes_Eng_MTR_3.pdf
* Nature or Nurture … No Intrinsic Integral Self, but an Integral of Infinitesimal Local Selfs (…), T.Arnaudov, 11.2012 http://artificial-mind.blogspot.com/2012/11/nature-or-nurture-socialization-social.html

**„Анализ на смисъла на изречение въз основа на базата знания на действаща мислеща машина. Мисли за смисъла и изкуствената мисъл.“, Тодор Арнаудов, 13.3.2004:** * https://artificial-mind.blogspot.com/2008/02/2004.html
* https://web.archive.org/web/20040402125725/http:/bgit.net/?id=65395

„Машината - поне според текущото ми разбиране - трябва да бъде създадена така, че да има достатъчно сложни (може би и достатъчно разнообразни по вид и цели) подустройства (създадени по какъвто и да е начин; условни или действителни), които да изпитват удоволствие по свой начин, т.е. да се стремят да постигнат цели, като да бъде невъзможно всички подустройства да постигнат едновременно целта си.

Необходимо е да съществуват крайни устройства, - "мускули" - върху които в даден момент власт да може да има само, да речем, едно подустройство, така че външното поведение - движението; данни които излизат на изходен канал - на машината да бъде еднозначно, както е при човека.

Подустройствата трябва да се "борят" и да си взаимодействат; да се договарят и да спорят всяко за себе си; понякога да се обединяват в групички, които "воюват" заедно с други групички, за да може всяко от тях да постигне максимално количество удоволствие.

Именно това прави човекът и това, струва ми се (според текущото ми разбиране) представлява "разумът": търсене на удоволствие и избягване на неудоволствие от достатъчно сложно устройство за избран период напред; "достатъчността" се определя от образците, които наричаме "разумни": хората от различни възрасти, различна сила на разума, с различни характери и стремежи.

Всеки от тях действа по начин, представляващ търсене на най-голям количествен сбор на удоволствие (неудоволствието се взима със знак минус) за избран период от време избира се когато се взима решението - …“

Крайните устройства и овладяването им са вид атрактори в смисъла на теорията на хаоса. Управляващите устройства са привлечени от тези състояния и се стремят да се намират близо до тях. Когато определени подустройства или съюзи от такива ги „завладеят“, те могат да кръжат по тези траектории, но не за неограничено време, защото в динамичната система *като цяло*, от по-висок порядък и обхват, други подустройства противодействат, „чакат на опашка“ и избутват общото състояние на системата в друга траектория, друго русло. И т.н.

В контекста и с езика на на невронауките, управляващо-причиняващи устройства от по-ниско ниво могат да са неврони, мини-колони или макроколони в кората на мозъка (виж Нумента, Джеф Хокинс), мрежи в мозъка от различен мащаб (circuits), до най-големите и всеобхватни, преглед на които е направен по-долу в обзора, като „DMN” – Default Mode Network и пр., които могат да се възприемат като „коалиции“, „съюзи“ от „подустройства“. Устройствата могат да бъдат и въображаеми, условни, виртуални – да не ги свързваме с конкретен невронен или анатомичен носител и взаимодействия, а с това *„как биха могли да се представят“,* така че да се постигне подобна или същата динамика, поредица от действия на изходните, крайните устройства при дадени съотносими условия, при дадени поредици от възприятия и т.н.

* **Dynamical Encoding by Networks of Competing Neuron Groups: Winnerless Competition**, September 2001, Physical Review Letters 87(6):068102, Michail Rabinovich et al.

Познавателната динамика е вид хетероклинично превключване (heteroclinic switching)[126]. Хетероклиничен цикъл – инвариантно, неизменно множество във фазовото пространство на динамична система: редуващи се повтарящи се траектории от състояния, които превключват помежду си – изпълнение на последователности, верижно вземане на решения, поредица от решения, „sequence decision making". Стабилен хетероклиничен цикъл – продължава при малки промени в основната динамична система. … Симетрии или други ограничения, които налагат съществуването на инвариантни хиперравнини … (Общ принцип, който) Действа на *различни нива на йерархията на мозъчните елементи*. Самата йерархия е резултат от сложни функционални взаимодействия, намиращи се между полюсите на сегрегация и интеграционни тенденции за мрежи, които изпълняват съвместни специфични когнитивни и/или поведенчески задачи. Рабинович и др. предлагат динамичния механизъм на *нискоизмерна* координация, който е свързан с общата обработка на информация в мозъчните последователни единици. Ключовото на тази координация е увличането на локализирани единици в мултимодална мозъчна активност.

*** Mind-to-mind heteroclinic coordination: model of sequential episodic memory initiation**, V. S. Afraimovich, M. A. Zaks, and M. I. Rabinovich, 2018

*Експерименталните открития показват, че познанието в човешкия мозък, както и съзнанието на определени бозайници:*
*а) е по-близо до детерминизма, отколкото до случайните процеси;*
*б) носи характерните черти на нискоразмерна динамика и*
*в) се проявява под формата на последователни метастабилни пространствено-времеви модели*

*Всеобщите състояния на мозъка съществуват в многообразия*[127] *от **ниска** размерност.* (подчертаване Т.А.)

…

[126] https://en.wikipedia.org/wiki/Heteroclinic_cycle
[127] многообразие (manifold) https://bg.wikipedia.org/wiki/%D0%9C%D0%BD%D0%BE%D0%B3%D0%BE%D0%BE%D0%B1%D1%80%D0%B0%D0%B7%D0%B8%D0%B5

**Тош:** Сравни с обобщенията на ТРИВ, 2001-2004 г., „Схващане за всеобщата предопределеност“, „Вселената сметач“ и общите принципи за предвиждане-компресиране. По-развитите системи се стремят да работят подобно на изчислителни машини по изрично определени правила и закони, които да са им известни.

* **Hierarchical dynamics of informational patterns and decision-making**, P. Varona, M. I. Rabinovich, Proc. Royal Soc. B 283, 20160475 (2016). https://royalsocietypublishing.org/doi/10.1098/rspb.2016.0475

* **Обобщение и бележки на Тодор Арнаудов** [10.10.2024]

Функционалните мрежи в мозъка са динамични – не само пространствените връзки между области на мозъка, но и превключващи във времето. Метастабилни информационни модели-схеми. Много познавателни функции, например мислене и музикална импровизация следват *едни и същи последователни йерархични схеми* на устойчиви хетероклинични канали (stable heteroclinic channels); градивните им елементи са метастабилни състояния и седловини (saddle sets), свързани с неустойчиви разделители (unstable separatrices). Води се постоянно състезание с различни временни победители заради асиметрично потискане. Устойчивостта на устойчивите хетероклинични канали (SHCs) се осигурява от конкурентното взаимодействие на агенти, променливи, които се развиват последователно, за да създадат верига от метастабилни състояния.

Състезание в мозъка като динамична система между различни модалности: мирис, осезание, … Свързване и разделяне на невронни мрежи в мозъка (binding and chunking). Познавателни мрежи, или мрежи на ума – за разлика от невронните, включват и времето – образуват времево-пространствени мрежи, превключват се и се свързват множество вериги, и образуват информационни модели-схемим, въобразни шевици (informational patterns).

[**Забел**. **Т.А.:** Отделните центрове участват в *множество* от информационни шевици и различни мрежи, и изпълняват *много* функции, а *не само една*[128].]

Творчеството и ролята на метастабилните състояния и неустойчивостта: *„На ръба на хаоса метастабилните модели са максимално нови, докато все още са свързани с моделите в подредения*

[128] Подобно с невротрансмитерите, невромедиаторите, невромодулаторите – веществата, участващи във взаимодействието и предаването на информация между и към/от невроните към други неврони или други клетки. Първите действат в много кратко време, милисекунди и имат прецизно място на действие между синапси и аксони; някои от вторите имат и по-дълъг период на действие; а третите действат върху по-голяма област и донастройват силата на възбуждането и предаването. Виж лекцията „Дофамин – не удоволствие …“

*режим и по този начин е най-вероятно да проявят комбинацията от новост и полезност, която е „печатът“ на творчеството.“* – сравни с равновесието между предсказуемост (обяснимост) и непредсказуемост за интересността.

*Динамично кодиране* и клетъчно сглобяване за взимане на решения (BP, DM) – свързани със задачата в момента неврони спонтанно се синхронизират и работят заедно, във взаимовръзка в пространствено-времеви режими и информационни модели, основани на принципа на състезаване без победител, който е често срещан в природата – най-простият модел е уравнението на Лотка-Волтера[129].

„*Има динамичен мост между времевата йерархия и анатомичната йерархия на мозъка. … рамка за обяснение на широк спектър от когнитивна динамика, включително поведение, чрез универсални принципи* (виж също препратки [58–61]). *… динамичната природа на съзнанието. … не е тялото, нито мозъка, а подвижната последователност от въобразни[130] модели, които всеобщите мрежи на мозъка кодират и умът обработва. …*

Авторите дават пример с вземане на решение от футболен вратар, който избира накъде да се хвърли, докато наблюдава приближаващи се трима противникови играчи с топката, и друг – за оценка на изпълнението на скейтбордист преди време, и даден критерий за разграничаване между „*автоматично* (подсъзнателно)“ от „*съзнателно*“ по техния метод: съзнателно било тогава, когато репортерът *„изпитвал съмнения“* и *„правил справка със собствената си епизодична памет“ (EM)*, за да вземе решение. Последната се образувала чрез *мрежи за разделяне и групиране,* „*chunking*”, които запазват тези спомени.

Тук „репортерът“ като цяло обаче е обобщен модел в ума на оценителя, който приема, че при дадено поведение се случва съответна „справка“, а при друго – „не“. Може да се стигне до този извод и чрез самонаблюдение в подобни, сравними; по-точно *възприемани или класифицирани, разпознавани като подобни и сравними* случаи от личния опит. Обаче дори и когато е „несъзнателно“ и „автоматично“, някакъв *друг* модел на „репортера“ като информационна система също би трябвало да „прави справка“. Това не е разграничителен белег; по-скоро такъв би бил „определен вид справка, обработка“.

Твърдят, че било добре известно, че емоциите били първичен,

---

[129] Популационна динамика в екологията, равновесието между броя на хищници и плячка, което се поддържа чрез обратна връзка.
https://en.wikipedia.org/wiki/Lotka%E2%80%93Volterra_equations

[130] информационни

главен подтик, мотиватор за творческо поведение[131]. Това обаче е прекалено общо и е убедително вярно само като „спусък“ за започване и поддържане на пламък, гориво; но неотчитайки, че *непрекъснато* текат някакви и „всякакви“ емоции, чувства, при всякаква дейност – положителни, отрицателни, неутрални, в някакъв „емоционален фон“. Широкоразпространено *погрешно* мнение за изкуството е, че то е нещо, което „*предизвиквало чувства*“, и обикновено съм го чувал от хора, които *не могат* ли не искат да създават „неща“, които се водят „произведения на изкуството“[132]. Онези, които могат и продават „изкуството“, също имат изгода да говорят по този начин, за да насърчават по-„евтини“ и пристрастяващи сили, с които да поддържат интерес към него, за повече пристрастяване и „зарибяване“ – а не за *разбиране.*

Изкуството обаче е повече *познание*, а не чувства, а чувствата са връзка, свръзка с чувствената система и „Волята“, която е водеща и първична – или по-примитивна форма на познание, която също има памет, предвиждаща обработка и пр., но с по-ограничени възможности; но дайте на някое бебе, в което почти няма познание в мерките на възрастните, да твори изкуство, и нека видим какви (естетически) чувства ще предизвика у оценители, които *не знаят,* че творецът е бебе, не са му майка, баща, баба и го сравняват технически с изкуството на някой „*безчувствен*“ аутист, който не изразява никакви емоции (по приет от оценителите начин) или изобразява, да речем, „безчувствени“ (за тях) неща като природни гледки, пейзажи, машини, сгради или калиграфия.

Оценителят може да открие чувства във *всичко* като ги свърже *със собствените спомени*, към които е закачена чувствена стойност (валентност, valence; пристрастие); или пък да не почувства нищо от друго, което за друг е „затрогващо“. За разлика от този „произвол“ и „субективизъм“, техническите качества са обективни при сравнение и могат да се подредят в онтологии, класификации, структури, системи за измерване, така както и произведенията на изкуството със съответни качества.

---

[131] 64 Lu J, Yang H, Zhang X, He H, Luo C, Yao D. 2015 The brain functional state of music creation: an fMRI study of composers. *Sci. Rep.* **5**, 12277. (**doi:10.1038/srep12277**)
65 McPherson MJ, Barrett FS, Lopez-Gonzalez M, Jiradejvong P, Limb CJ. 2016 Emotional intent modulates the neural substrates of creativity: an fMRI study of emotionally targeted improvisation in Jazz musicians. *Sci. Rep*. **6**, 18460. (**doi:10.1038/srep18460**)

[132] Голяма част от изкуството е просто възпроизвеждане на траектории и въобще изпълнение на движения, повторение, възпроизводство, презапис, спомняне и пр., но *дори и това е трудно* за човешки същества.

***Бел.:** * мислене и импровизация … едни и същи последователни йерархични схеми…“* – както е предвидено теоретично в ТРИВ: всестранността; всички видове познавателни процеси следват общи признаци; „науката = изкуство + забавление“ (Science = Art + Fun), тениска във Феймлаб; виж *Хипотеза за по-дълбокото съзнание*, приложение към Основния том и статия от бележки на „Какво му трябва на човек? ....“, 2014.
* Относно ***импровизацията*** *,* виж също отговора на Тош на Майкъл Левин за „Самоимпровизиращата памет“, 2024 г. в приложение „Фантастика. Футурология. ..“, 2025. В тази работа на Левин се преоткриват идеи, *публикувани в ТРИВ над 20 години по-рано*, подобно на други трудове от същия автор и негови колеги*. Паметта и „изчисленията“ или структурата в живите организми обаче не са рязко отделени. Една от мислите в ТРИВ, произтичаща от основната идея: „Вселената сметач“, е че *всичко е* памет, всяка частица вещество, всичко измеримо, и че памет за дееца, управляващо-причиняващото устройство, агента, *не е* само определена обичайна „вътрешна“ памет като оперативната с ограничен размер, а потенциално цялата Вселена, стига да може да се „обърне“ към съответното място и да „погледне“ там. На „Бейсик“ за Правец-8М: PEEK(ТУК) и POKE(ТАМ).
* Виж примера на Т.Арнаудов от ТРИВ за футболист, който уж *„хаотично изритва топката“*, но всъщност хаотичното е *само* от гледна точка на наблюдател-оценител, който *не познава* и не разглежда правилното управляващо-причиняващо устройство, което си има определена и конкретна цел и извършва целенасочено действие в:
* „Писма между 18-годишния Тодор Арнаудов и философа Ангел Грънчаров…“, 2002
* Виж списъка с учени и школи в Основния том на „Пророците на мислещите машини“ и други томове, където са разгледани повече публикации и теории на М.Левин, Крис Фийлдс и свързани с тях; принципа на свободната енргия/извод чрез действие от школата на Карл Фристън (FEP/AIF) и др. ?T: Michael Levin

**Препратки 58-61:**
*Kiebel SJ, Daunizeau J, Friston KJ. 2008 **A hierarchy of time-scales and the brain**. PLoS Comput. Biol. 4, e1000209. (doi:10.1371/journal.pcbi.1000209) Crossref, PubMed, Web of Science, Google Scholar
* Rabinovich MI, Afraimovich VS, Bick C, Varona P. **2012 Information flow dynamics in the brain.** Phys. Life Rev. 9, 51–73. (doi:10.1016/j.plrev.2011.11.002) – **разгледана по-долу**
* Friston KJ. 2012 **Competitive dynamics in the brain: comment on ‘Information flow dynamics in the brain’** by M.I. Rabinovich et al. Phys. Life Rev. 9, 76–77. (doi:10.1016/j.plrev.2011.12.006) Crossref, PubMed, Web of Science, Google Scholar
* Heinzle J, Allefeld C, Haynes JD. **2012 Information flow, dynamical systems theory and the human brain. Comment on ‘Information flow dynamics in the brain’** by MI Rabinovich et al. Phys. Life Rev. 9, 78–79.

(doi:10.1016/j.plrev.2011.12.007) Crossref, PubMed, Web of Science, Google Scholar

## * Information flow dynamics in the brain

Rabinovich MI, Afraimovich VS, Bick C, Varona[133]
https://linkinghub.elsevier.com/retrieve/pii/S1571064511001448

*„От познавателна гледна точка на мозъка[134], вземането на решения е динамичен процес, при който променлив във времето шумен информационен поток се обединява в множество времеви мащаби с решение, което произтича, когато въобразният поток, свързан с една от възможностите, достигне определен праг.“*(1)*

*Динамиката на информационният поток, преходите, преходните метастабилни състояния-седловини, временни победители в безкрайната състезателна игра, е във **фазово пространство** и **не зависи от геометричната структура** на невронните ансамбли във физическото пространство.*

*Едновременно текат информационни потоци и отдолу-нагоре, и отгоре-надол и образуват затворени функционални цикли.*

*Дейността на мозъка е като да свирите концерт за цигулка като се учите да свирите и композирате пиесата в движение. За целта мозъкът трябва да възприема и да се учи, без учител, от звука от цигулката, да се учи да свири, да използва работна памет за да помни какво е изпълнил до тук; да взема решения, за да избере как да продължи; да внимава – за да се придържа към решението си; да създава партитурата и накрая: да състави двигателната програма за движенията за изпълнението, т.е.: възприемане, преобразуване[135], съгласуване и създаване на въобраз (информация).*

(…)

*1: Сравни с дифузните невронни пораждащи модели; достигането на праг – виж ГнП, граница на преход в **Зрим**; „Създаване на мислещи машини“ (бъдеща работа)

* Rabinovich, M. I., Huerta, R., Varona, P., and Afraimovich, V. S. (2008). **Transient cognitive dynamics, metastability, and decision making**. PLoS Comput. Biol. 4:e1000072. doi: 10.1371/journal.pcbi.1000072

---

[133] Rabinovich MI, Afraimovich VS, Bick C, Varona P. Information flow dynamics in the brain., 2012 Phys. Life Rev. 9, 51–73. (doi:10.1016/j.plrev.2011.11.002)

[134] В текста „cognitive”, “когнитивна“, като в тази работа и другаде под „когнитивна“ (познавателна) всъщност разбират „мозъчна“, свързана по особен начин с конкретното му действие. Сравни „Cognitive AI”з.

[135] transduction

* P. Varona, M. I. Rabinovich, **Hierarchical dynamics of informational patterns and decision-making**, Proc. Royal Soc. B 283, 20160475 (2016).
*** Self-organizing 'infomorphic neurons' can learn independently,** techxplore.com, 31.3.2025, Max Planck Society – *MPI-DS, "learn in a self-organized way and draw the necessary information from their immediate environment in the network"…*
* A.Makkeh et al., **A general framework for interpretable neural learning based on local information-theoretic goal functions**, 2025 (2023/2025) https://arxiv.org/pdf/2306.02149 Partial information decomposition (PID)

* Wendy Otieno, Ivan Y. Tyukin, Nikolay Brilliantov**. The critical dimension of memory engrams and an optimal number of senses**. Scientific Reports, 2025; 15 (1) DOI: 10.1038/s41598-025-11244-y https://www.nature.com/articles/s41598-025-11244-y 15.8.2025 – *"how many senses are optimal for memory and learning."* – 7; engram formation…
* https://www.sciencedaily.com/releases/2025/10/251008030955.htm *"We have mathematically demonstrated that the engrams in the conceptual space tend to evolve toward a steady state, which means that after some transient period, a 'mature' distribution of engrams emerges, which then persists in time," Brilliantov commented .., let the objects that exist out there in the world be described by a finite number of features corresponding to the dimensions of some conceptual space. .. to maximize the capacity of the conceptual space expressed as the number of distinct concepts associated with these objects. The greater the capacity of the conceptual space, the deeper the overall understanding of the world. It turns out that the maximum is attained when the dimension of the conceptual space is seven."* – see also Peter Gardenforse's "The Geometry of Meaning: semantics based on conceptual spaces", 2014; notes in the main volume etc.*
https://archive.org/details/geometryofmeanin0000gard

*** Plamen P Angelov,** Eduardo A Soares, Richard Jiang, Nicholas I Arnold, and Peter M Atkinson. **Explainable artificial intelligence: an analytical review**. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11(5):e1424, 2021. https://www.researchgate.net/publication/353181830_Explainable_artificial_intelligence_an_analytical_review

*** Пламен Ангелов – влиятелен български „пророк" още от 1990-те**

*** Dimensionality and dynamics for next-generation artificial neural networks**, Ge Wang1, Feng-Lei Fan2, 4.2025 https://www.cell.com/patterns/fulltext/S2666-3899(25)00079-0 - height dimension, feedback loops – виж в приложение #lotsofpapers – виж в приложение Анелия #Anelia с Никола Касабов, Димитър Филев и др. Виж също #Lazar и основния тип, други ранни работи с невронни

мрежи и конекционистки методи на Ивилин Стоянов, Йордан Златев и др.

## * Машинно обучение, обучение на човека и невронауки и взаимодействието между тях
## * Machine Learning, Human Learning and Neuroscience and their interaction

### * Humans learn generalizable representations through efficient coding,
Zeming Fang & Chris R. Sims, 29.5.2025, Nature Communications
https://www.nature.com/articles/s41467-025-58848-6
RL framework; complex stimuli are mapped to compact representations; …They argue that a combination of classical RL objective, "*augmented with efficient coding represents a more comprehensive computational framework for understanding human behavior in both learning and generalization." .. . "incorporate the principle of efficient coding … use the simplest necessary representations" *1.*
The example with riding a bike and transferring to a scooter. *2 State abstraction, rewarding feature extraction … perceptual-based and functional-based generalizations .. acquired equivalence paradigm .. *"Abstract states inevitably merge in simplified representations, resulting in generalization"*3 *"Efficient coding automatically extracts rewarding features throughout learning simplified representations";*
conditions in their experimental setting: consistent, control, conflict;
acquired equivalence paradigm, .. *"Efficient coding automatically extracts rewarding features throughout learning simplified representations"*; Latent Cause model; Memory-Association model; Attention at Choice and Learning; algorithmic-level models. .. *"an intelligent agent should distill the simplest necessary representations that enable it to achieve its behavioral objectives"* . A computational-level model:.. Efficient Coding Policy Gradient; ECP .. representations with a small set of rewarding features with the environment.; Association-Choice Learning (ACL).

**Todor:** *1,*3: Yes, but this is not a new idea or discovery. That's the compression-prediction framework of the "predictionist-compressionists", including Theory of Universe and Mind. The minds build hierachical virtual universes (simulators of virtual universes), higher levels in the ladder of generalization which work for broader scopes in time, spaces, domains, cases etc. are more compressed and of lower resolution (in the lower level format or view); in its general mode of operation the agent, mind, causality-control unit aims at reaching to a match of sensory input and its motor outputs, goals, desires, imagined future, the state of its memory and representations of its will, and after coarse-graining, sooner or later the system reaches to complete match between the two comparads: either one or both or all, the input and

the templates, are simplified until they sufficiently or fully match, as their complexity in bits is reduced, the possible combinations also decrease.

Different levels of detail, abstraction, focus, selected features etc. of the representations correspond to the constraints in the cognitive system: memory, bandwidth, complexity, time etc., which shape the granularity and the borders. Spoken and written language as it's used now is as concise or uses as general concepts as it does, in order to fit a particular phonological-loop buffer, working memory capacity, attention span etc. given the available sensory-motor devices and limitations or advantages of the brain. When the bandwidth is bigger, we use lower level representations as well as with sound and images. See * "Chairs, Buldings, Caricatures… ", or "AGI Digest 2012" , T.Arnaudov et al. 2012 in AGI List. See also Free Energy Principle/Active Inference, which is simialr to TOUM, reviewed in the main volume of *The Prophets* and with additional selected discussions in *Irina.*

*2: I don't think the bike-scooter is the best example for generalization – a "functional" one in their mentioned dichotomy sensory vs functional. The feel and the data domain, which are most relevant for keeping balance, are the same for both vehicels and are about proprioception and vision, not the appearance of the vehicle itself. Maybe it is similar for many or most other cases as well, though, if they involve body actions and coordination. If I remember correctly, the Roman philosopher Seneca in his letters to Lucilium mentions the generalization when learning to throw a spear – you learn to hit any target and at different distances, not just the one that you've specifically trained on. In tennis, the players who can play and hit correctly the ball from any position, "improvise" or do the strokes with less preparation, with less back-stroke motons with the racket, without pointing with their hands at the ball etc. are considered *"more talented"*. Their talent is in a more powerful, faster and more complete or complete physical model of their body, the court, the ball trajectory – the world – so they can predict and cause better the future they want – as with any domains and tasks*. The "less talented" players who cannot develop or lack rich enough model tend play more stereotypically with "standard" strategies and body poses, they "don't generalize" so well. Notice also that many tennis players demonstrate football skills – juggling with tennis balls with feet; it could be a "drill" they practice and some of them played football as well; the footwork, dynamic balance of the whole body and motion timing are improved with playing tennis and thus all kinds of motions. * https://en.wikipedia.org/wiki/Epistulae_Morales_ad_Lucilium
* The shorter stroke also makes the *prediction* of the direction of their shots harder.
* Bates, C. J., Lerch, R. A., Sims, C. R. & Jacobs, R. A. Adaptive allocation of human visual working memory capacity during statistical and categorical learning. J. Vis. 19, 11 (2019).

* Xia, L. & Collins, A. G. E. Temporal and state abstractions for efficient learning, transfer, and composition in humans. Psychol. Rev. 128, 643–666 (2021).

→

## * Momchil Tomov – Момчил Томов

*** Related Bulgarians: M.Klissarov, E.Todorov & Peter Kormushev**

*** Multi-task reinforcement learning in humans, Momchil S. Tomov,** Eric Schulz & Samuel J. Gershman, Nature Human Behaviour, 28.1.2021 https://www.nature.com/articles/s41562-020-01035-y *"The ability to transfer knowledge across tasks and generalize to novel ones is an important hallmark of human intelligence."*

* https://www.momchiltomov.com/blank-3 - *Naturalistic learning and decision making, Building machines that learn, plan, and act like humans; Automated neuroscientist*
https://scholar.google.com/citations?user=ySl-BRUAAAAJ&hl=en

**Momchil Tomov[136] was born and raised in Bulgaria. BSc in Princeton, PhD in Harvard.** https://www.momchiltomov.com/ ***"****I study how the human brain learns rich internal models of the world for efficient planning and decision making. I also build artificial agents that learn and plan in human-like ways. My research connects multiple disciplines, including neuroscience, psychology, machine learning, robotics, and cognitive science."*
As of the info on 22.8.2025 he is a researcher at Harvard University and a senior

[136] The name Momchil originates from "momche" (a boy, момче). Momchil yunak, also known as Momchil voyvoda, is a heroic character from the Bulgarian folklore and late medieval history (the first half of the XIV century). https://en.wikipedia.org/wiki/Momchil "Yuank" is also a neologism, coined by Todor Arnaudov in *"The Sacred Computer"* e-zine in 2001 as part of the so called "Yunashki dialect" or "Computer Bulgarian", in order to distinguish creative hackers from the "poisoned" sense used by the media; besides the programming part however it includes being a versatile creative person, universal man, being interested, talented and curious in all creative and cognitive modalities as an artist, scientist, engineer, athlete, philosopher, writer, musician etc., all at once. *"Twenkid"* is another related neologism with similar meaning, coined in 2008 by the same author, but with the word-form focused on the *"timeless youth"* of the yunaks and universal men. The word "yunak" itself in general Bulgarian is translated to English as "a hero", but it originates from old forms of the word for "young" or "youth", preserved in *"yunosha" (юноша)* – an adolescent, a teenager in mid- and late teens, e.g. in sports: "юноша младша възраст" ~ 15-16, „юноши старша възраст" 17-18 years old. In other Slavic languages, e.g. Russian: yunost, юность – youth.
*** ДЗБЕ (DZBE)** – Society for Protection of the Bulgarian Language: https://eim.twenkid.com/dzbe/

scientist at the autonomous car company Motional: https://motional.com/ *"We're developing SAE Level 4 AVs for autonomous ride-hail and delivery."; robo-taxi."*.

See also the contributions of three other Bulgarians: **Emanuel Todorov**, **Martin Klissarov** and **Peter Kormushev** in the field of Reinforcement Learning (RL), reviewed in the main volume of "*The Prophets of the Thinking Machines*". Martin is younger and is completing his PhD at McGill Univeristy and Mila instiute[137], and now he's Research Scientist at Google DeepMind. His interests overlap with Momchil's, but without the neuroscientific part, while Peter's work is focused in robotics with groundbreaking work since late 2000s and early 2010s. **"Emo" Todorov** is author of the physics simulator MuJoCo and has contributions in computational neuroscience, optimal control, model predictive control, reinforcement learning. **Emo:** https://scholar.google.com/citations?user=QCBdB7AAAAAJ&hl=en
**Martin:** https://mklissa.github.io/ **Peter**: https://kormushev.com/

*** Discovering Temporal Structure: An Overview of Hierarchical Reinforcement Learning, Martin Klissarov**, Akhil Bagaria, Ziyan Luo, George Konidaris, Doina Precup, Marlos C, 2025/6/16, https://www.arxiv.org/abs/2506.14045 113 p. (80+lit.) [26.8.2025]
*"… Hierarchical reinforcement learning (HRL) .. – discovering and exploiting the temporal structure within a stream of experience; learning directly from online experience to offline datasets, to leveraging large language models (LLMs) ..* Modularity, compositionality; p.7 performance, efficiency (sample, computational); [sample efficiency = number of experiences, data points etc.] p.9: HRL, temporal structure **= skills, options, temporal abstractions, or goal-conditioned policies**. **Options:** a policy, an initiation function, and a termination function*1. Many options policies are probed and each of them explores different potential trajectories to different outcomes; credit is propagated from lower levels to higher level nodes across multiple steps.*2 **Options as subgoals.** p.12. "*subgoal options can be defined by mapping states to actions through symbolic functions such as code*". **Skills** and **Goals.** .. **Bottleneck Discovery** – small regions of states which connect to other interesting states – reducing the search space; diverse density … methods for discovery of temporally abstract behaviors, p. 14: From Online Experience: Bottleneck principle, Spectral methods (graphs), Skill chaining, Empowerment Maximization, Environmental Reward, … Meta-Learning, Curriculum Learning, Intrinsic Motivation. **Offline Datasets**: Variational inference, Hindsight Sub-goal Relabelling. Foundation models: Embedding similarity, Providing feedback, Reward as code, Directly modeling the policy …; bottleneck: centrality, betweenness, Graph partitioning (Q-

[137] As of 26.8.2025

cut), Graph centrality – the importance of the nodes of the graph., connected rooms; density of states on successful trajectories. **For each method for HRL discovery**, the survey of Klissarove et al. evaluates **Benefits and Opportunities** for Exploration, Credit Assignment, Transfer ("transfer learning" to unseen tasks, environments etc.) and suggests **Opportunities for research** and **Performance guarantees.** The Bottleneck methods reduce the state-action pairs for updating; reducing the "action gap" for long horizon problems by partitioning the state space.

**Spectral** – graphs, partitions, Laplacian, connected components, eigenfunctions; augmented Lagrangian Laplacian objective … *successor representation (SR; Dayan, 1993) .. "assigning similar values to temporally close states*"; eigenoptions.

**Skill Chaining (*Sequentially Composable Options*)**… *Reward-Aware Representations.* Options*:* Sequentially (Executable, Composable); **initiation set of options** – states from which it is likely the execution to lead to achieving the subgoal. …"; p.26 *"policies as funnels"* - they drive a large set of ordinary states to a small set of desired states".

**Empowerment Maximization** .. *how much influence an agent has over its future observations — an agent is more empowered when it can reliably cause a wider variety of outcomes (Klyubin et al., 2005; Salge et al., 2014).**2 .. *mutual information between an agent's actions and its future states. .. "**Diversity is All You Need** (DIAYN), Eysenbach et al. 2019. *3* **Empowerment** promotes the **exploration** and is connected with causal learning and human way of learning; find the most different trajectories; Goal-based exploration & variational empowerment. p.31: **Environment Rewards**: **feudal methods** – Dayan & Hinton, 1993; decompose the agents into managers and workers; *managers set subgoals for the workers, and the workers use non-hierarchical RL. **Feudal Networks** FuN – (Vezhnevets et al., 2017); a goal vector, embedding vector;* **higher and lower level policy are trained together** and an old high level action may not lead to the same low level behavior in the future *(Nachumet al., 2018). Option critic. ..p.37* ***Meta-Learning*** – (outer, inner) loop … Schmidhuber, 1987 … intrinsic (reward, motivation); (intrinsic, environment) reward; meta-exploration … p.40 **4.8. Curriculum Learning** – a sequence of problems, where *the complexity of each attempted goal increases continuously with the agent's capabilities that leads to more efficient learning and achieving the future objectives (Kaplan and Oudeyer, 2003; Schmidhuber, 2004; Bengio et al.,2009; Schmidhuber, 2011) *4.*; p.44 **4.9. Intrinsic Motivation ..** *States with high relative novelty are likely to be gateways to unexplored regions .. option subgoal regions (abstract states) and edges .. option policies (abstract actions).* Deep Belief Networks (DBN) – Bayesian network; the agent's observations are **factored state variables**, which require domain knowledge; factored approaches: *HEXQ (Hengst et al., 2002) .. automatically decomposes a factored MDP into subtasks by detecting **exits — states** where **a change***

***in one state variable causes a change in another variable (or termination)*** *— lower-level options correspond to frequently-changing variables, whereas higher-level options handle more slowly-changing aspects.* .. **Disadvantages:** *limited applicability to high-dimensional observation spaces such as images or sensor data; struggles in continuous environments, where exact state revisitation is unlikely* [too many possible states, unique at low level]. p.48 ***Offline RL (or batch RL)****:* learning policies *from pre-collected datasets,* without *active data collection: robotics, autonomous, driving, education, and healthcare.*. ***offline skill discovery*** *.. extracts temporally abstract behaviours that can later serve as high-level primitives .. expert demonstrations or acquired through arbitrary policies; unlabeled experiences – without explicit reward feedback,* sometimes even without actions – unsupervised skill discovery .. **evidence lower bound (ELBO)**, approximate posterior .. . p.65 .. **Deliberating over Options**: call-and-return model [complete flow probing a trajectory from highest to the lowest level, computationally expensive] .. The agent's high-level policy, μ(o|s), is responsible for selecting an option. .. p.68 **Abstract planning:** the coarser the abstraction, the **greater the potential for suboptimality of the resulting plans** [see also 2.2, the comparisons in computers: machine code, Assembly, higher level languages, applications]: expectation models, skills to symbols … LLMs – generate Python code for skill-selection logic or formal plans in PDDL. .. p.70 **Challenges in discovering temporal abstractions:** *lack of agreed-upon objective that would yield meaningful options across a variety of domains* [in the research community]; complexity overhead; **non-stationary targets:** *Bagaria et al. (2023) .. initiation set .. learning the initiation function using binary classification (or .. Monte Carlo value estimation) is only a sound approach when the option policy is fixed* .. p.72 **State and Action Abstractions** .. aggregation or *mapping* with NN – representation learning .. s*tate-action abstraction, notably MDP homomorphism .. model minimization;* ***action abstraction:*** *per-timestep and multiple-timestep .. p.74* ***Continual RL …*** *p.75* ***Programmatic RL ..*** *Hoare Logic* ***Programs as high-level policy. …*** *HAM (Parr and Russell, 1997) and PHAM (Andre and Russell, 2000):.. partially specified finite-state machines (FSM) ..* ***Programs convey structured, interpretable, and unambiguous information,*** *and their incorporation into the policy space can reduce the search space for the overall solution and offer a natural method for integrating prior knowledge symbolically.*

*p.76: Cooperative multi-agent RL… HRL methods* ***exploit structure****. A complex environment lacking exploitable structure might not benefit from HR ... Hughes et al. (2024):* open-ended systems present a constant flow of novel and possibly learnable goals

*p.77:* **Applications for HRL:** Web agents, Robotics, Open-ended games.

* Schmidhuber, J. (1987). Evolutionary principles in self-referential learning. on learning now
to learn: The meta-meta-meta...-hook. Master's thesis, Technische Universitat Munchen.
* Schmidhuber, J. (2004). Optimal ordered problem solver. Machine Learning, 54:211–254.
* Schmidhuber, J. (2010). Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990-
2010). IEEE Transactions on Autonomous Mental Development, 2(3):230–247.
* Schmidhuber, J. (2011). PowerPlay: Training an Increasingly General Problem Solver by
Continually Searching for the Simplest Still Unsolvable Problem. Frontiers in Psychology, 4

**Todor's notes:**
*1: Cmp(classical planning, robotics): preconditions, goal state, postconditions; see #planning, early 1970s STRIPS, the works of David Wilkins, PDDL, Hierarchical Task Networks (HTN) etc. below in this volume of *"The Prophets of the Thinking Machines …"*. See also *"cue"* in RL for animal behavior and the review of the paper about goal-directed and habitual behaviors below].
*2: See Theory of Universe and Mind (TUM), T.Arnaudov 2001-2004, the slides for the talk about the common principles of Intelligence and the Universe at TU, Sofia, 2009 (hierarchical prediction and compression with the military hierarchies and how they work given as a simple example); and the lecture about TUM from the world's first AGI course at the University of Plovdiv, 2010-2011.
*3: See the method of interdisciplinarity and the suggestions from the works in TUM and the general method of the author and "universal man" Todor Arnaudov. For very specific tasks with very specific requirements for the agent, diversity may not be helpful, howver these narrow agents have to be part of a "society" in order to do the other tasks and there's a virtual super agent which is a system, emerging by the interaction of the specialized agents or elements of the system. Generalist systems and seeds of intelligence make use of the versatility and is and could be the core of such "society" – the core may generate and improve specialized branches like the living organisms do with the zygote which is universal and differentiate for different tissues and organs etc.
*4: Curriculum learning – also incremental, gradual learning and development and not only or strictly in a NN gradient learning representation and framework, but also conceptual. See future works: **Zrim**, **Emil, "Genesis: Creating Thinking Machines":** вършерод, казбород, ИнР … (vursherod, kazborod, InR, \!/)

*** MaestroMotif: Skill Design from Artificial Intelligence Feedback, Martin Klissarov,** Mikael Henaff, Roberta Raileanu, Shagun Sodhani, Pascal Vincent, Amy Zhang, Pierre-Luc Bacon, Doina Precup, Marlos C. Machado, Pierluca D'Oro**,** *22.1.2025/5.3.2025* https://openreview.net/forum?id=or8mMhmyRV *Describing skills in natural language has the potential to provide an accessible way to inject human knowledge about decision-making into an AI system*

…

Notice that Martin et al. review discovering **temporal hierarchical structure**, while Momchil in his earlier PhD thesis, reviewed and commented below also discovers **hierarchical structure of states**, in order to make the planning more efficient. The latter work doesn't' emphasize *time*, however it is still "there" as the more abstract states occupy larger spans, either in space or in time as discussed in the notes and that's one of the criteria for separating them in different scales.

*** Tomov, Momchil, 2020. Structure Learning and Uncertainty-Guided Exploration in the Human Brain**, Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences (2016-2019) https://dash.harvard.edu/bitstreams/f065e9e9-3c59-476b-9c2d-0d5d9ad0b5a3/download – causal inference; exploration vs exploitation dilemma; Relative vs Total uncertainty and their neural correlates (R vs T); Directed vs Random exploration are guided by the two kinds of uncertainty; GLM – general linear model; "Subjective estimates of R & T uncertainty predict choices); p.59 2. Neural Computations Underlying Causal Structure Learning fMRI … structure learning vs association; model-based analysis of fMRI; p.154 "How does the brain learn a model of the world? … p.60 whether the context is relevant, an "occasion setter" or irrelevant for learning; … p.66/46: Irrelevant context, modulatory context, irrelevant cue. Separate brain regions for structure and associative learning; Posterior vs Causal structures. Thompson sampling. BFS – breadth first search for paths in graphs; hierarchical breadth first search (HBFS); low level and high level graphs, which include the low level graphs; compare to HHMM – Hierarchical Hidden Markov Models and the "nested Markov blankets" in FEP/AIF. p.106/85: **3. Discovery of hierarchical representations for Efficient Planning** … cluster of states that support hierarchical planning (hierarchical etc. = H.); the distribution of tasks and rewards influences the planning behavior via the discovered H.; bottleneck transitions, states; H. planning, shorter H. paths, cross-cluster jumps … Example for planning a trip from a city in USA to a small town in Spain as soon as possible for eating a preferred brand of ice cream. The agent (person) would say that she would search for a flight, would catch a taxi to the airport etc., but: p.106/86: ***"Importantly, nobody says or even thinks anything like "I will get up, turn left, walk five steps, etc.", or even worse, "I***

***will contract my left quadricep, then my right one, etc."*1**. Action chunking – creation of the state clusters; task bracketing – distinct neural signature …*

***1: See the response by Todor Arnaudov** in the article below:
**Why humans usually answer by addressing the higher level nodes…,** 24.8.2025

* **Habits, action sequences and reinforcement learning.** Dezfouli, A. & Balleine, B. W. (2012). European Journal of Neuroscience, 35(7), 1036–1051 – https://adezfouli.github.io/materials/db-ejn-2012/db-ejn-2012.pdf – *"instrumental actions can be either **goal-directed** – rapidly acquired and regulated by their **outcome**,* [or] ***habitual:** reflexive, elicited by **antecedent stimuli** rather than their consequences. **Model-based RL** provides an elegant description of goal-directed action. Through exposure to states, actions and rewards, the agent rapidly constructs a model of the world and can choose an appropriate action based on quite abstract changes in environmental and evaluative demands. … the **failure of model-free RL** correctly to predict the **insensitivity** of habitual actions to changes in the action-reward contingency. .. introducing model-free RL in instrumental conditioning is unnecessary, and demonstrate that **reconceptualizing habits as action sequences** allows **model-based RL** to be applied to both goal-directed and habitual actions in a manner consistent with what real animals do."* [edits & bold: T.Arnaudov]
p.1*: the **acquisition** of **goal-directed actions** is controlled by a circuit involving **medial prefrontal cortex** and **dorsomedial striatum** (DMS, or caudate nucleus);*
***habitual actions** involve[s] **connections** between **sensory-motor cortices and dorsolateral***
***striatum** (**DLS**, or **putamen**;. .)* These two forms of action control are .. ***partly competitive;** distinct, parallel cortico-basal ganglia networks that likely constitute functional loops connecting cortex, striatum and midbrain regions with feedback to the cortex via midline thalamus* .. p.4. "association between cues and responses", (open-loop, closed-loop) control; action chunks; state identification, based on visual cues; action evaluatioon – cue-independency of habitual behavior. If the **action control** is state-dependent and based on the environmental cues, it is called **closed-loop action control**. p.5 state identification process, action selection, learning; *"RL is a computational approach to learning different types of instrumental associations for the purpose of maximizing the accrual of appetitive outcomes and minimizing aversive ones (Sutton & Barto, 1998)"; **Average Reward RL –** maximizing the average reward* [not the max sum*] .. *p.12/1047:* The authors propose that *"as goal-directed actions become habitual, they grow more complex … – more thoroughly integrated, or chunked, with other motor movements .. when selected, these movements tend to run off in a sequence* .. and *the TD error provides feedback regarding the cost of this chunk or sequence of actions* [instead of the value of the individual actions].

**Tosh:** Average RL is meaningful especially for online measures without completing a long horizon trajectory, for “surviving” and receiving “sufficient” reward, signals for acceptable ecological conditions, body parameters etc. which has to be sampled for short periods. In such conditions both the average per period or for number of steps, and each individual reward/feedback signal should be above lower threshold, e.g. “damage” points, “bleeding”, the temperature shouldn’t fall too low or rise too high, the oxygen level should be appropriate etc. [27.8.2025]
* See Thompson sampling: Multi-Armed Bandit Problems, exploration-exploitation dilemma; sampling a set of actions, collecting rewards, then acting according to the distribution of the posterior. https://en.wikipedia.org/wiki/Thompson_sampling
* A Tutorial on Thompson Sampling, Daniel J. Russo, B. Van Ro et al. https://web.stanford.edu/~bvr/pubs/TS_Tutorial.pdf . TS ~ posterior sampling and probability matching; Bernoulli bandit; Upper-confidence-bound algorithms (UCB); regret bounds, …
* Thompson, W. R. 1935. “On the theory of apportionment”. American Journal of Mathematics. 57(2): 450–456.
* Thompson, W. 1933. “On the likelihood that one unknown probability exceeds another
in view of the evidence of two samples”. Biometrika. 25(3/4): 285–294

*** More Efficient Randomized Exploration for Reinforcement Learning via Approximate Sampling**, Haque Ishfaq, Yixin Tan et al., 6.2024 https://arxiv.org/pdf/2406.12241 .. Thompson Sampling “*gained popularity due to its simplicity and strong*
*empirical performance”*. MDP, (Markov Decision Process), Sampling Complexity of Different Sampler, linear MDP, Markov Chain Monte Carlo: MCMC, Langevin Monte Carlo sampling, FGTS – Feel Good Thompson Sampling; regret bound.
**Tosh:** “Regret” is the difference from the possible optimal reward in case of the best action selection in the environment, and the actually received reward by the agent. In other contexts the regret can be called “loss”.

*** Regret bounds of model-based reinforcement learning,** Mengdi Wang, 2021, RLVS 2021, Day 3, https://rl-vs.github.io/rlvs2021/regret-bound.html – Episodic RL, … 10:55 min: **Upper Confindence Model-Based** RL (UCRL) Construct a confidence set for the state transitions (s1, a1, s2, r2 …); Optimistic planning. 32 min Value-Targeted Regression (VTR) in MuZero generalist game playing agent.: Canonincal RL model: Input = (st, at) → Transition Model → predicts next state s[t+1]. **A model with Value-Targeted Regression (VTR)**: For each (st,at), the Input

= V(st) and the transition model predicts the V(s[t+1]) – the next value, not the next state; prediction only of the value and policies. 54: metric-based RL, Voronoi triangles, nearest neighbor transitions, greedy selection of action in the new episode

*** MuZero: Mastering Go, chess, shogi and Atari without rules,** Google DeepMind, 23.12.2020 https://deepmind.google/discover/blog/muzero-mastering-go-chess-shogi-and-atari-without-rules/

*** Mastering Atari, Go, chess and shogi by planning with a learned model**, Julian Schrittweiser, Ioannis Antonoglou et al., Nature, 3.4.2020/23.12.2020 https://rdcu.be/ccErB
Model-Based RL, Monte Carlo Tree Search, abstract: *"Atari games—the canonical video game environment for testing artificial intelligence techniques, in which model-based planning approaches have historically struggled ..*

*** What model does MuZero learn?, Jinke Hea,*,** Thomas M. Moerland, 12.10.2024 https://arxiv.org/pdf/2306.00840

## * RL Virtual School lectures: https://rl-vs.github.io/rlvs2021/index.html

* **Evolutionary reinforcement learning,** Jean-Baptiste Moure, Dennis Wilson, RLVS, 2021, 8.4.2021**:** https://rl-vs.github.io/rlvs2021/evo-rl.html **–** video, 2:11 h, … https://youtu.be/7J5KK-tYoXc?si=2_ZBdfspul0-ZF5b, 2.07.2021, 1609 views (28.8.2025)
~ 1:37 h Neat, HyperNeat, PicBreeder; stepping stones; novelty search .. 1:47 h MAP-Elites: Multi-dimensional Archive of Phenotypic Elites – Objective: Find many good ways of solving a problem (diversity), *more creative process (not pure RL/optimization), less exhaustive than Novelty Search.*

*** Reward Processing Biases in Humans and RL Agents,** Irina Rish, RLVS, 2021, 25.3.2025, https://rl-vs.github.io/rlvs2021/human-behavioral-agents.html https://www.youtube.com/watch?v=71EKYIcUe0A 2.7.2021, 376 views (28.8.2025)
10 min: *Learning from Positive vs Negative Rewards. Parikinson's patients off meds are better at learning to avoid choices that lead to negative outcomes than from positive outcomes. Reversed bias when on dopamine meds.* ***Go/NoGo model (basal ganglia-dopamine interaction).*** See the diagram. 21 min: Contextual Bandits: State→Action→Reward.
State = context, the bandits produce only reward, not the next state. Three levels of RL: 1. Multi-Armed Bandits, Action→Reward. 2: Contextual Bandits. 3. Full RL Problem: State→Action→Reward; State→Action→State; State→Reward.

13 min: *The **Go cells** are part of a direct corticostriato-thalamo-cortical loop, they **inhibit the thalamus** via GPi – the internal segment of globus pallidus and facilitate the execution of an action, represented in cortex.* On the other hand, *the **NoGo cells inhibit** the action by increasing the inhibition of the thalamus.* Mental disorders can be modeled as biases of the reward-processing: in Alzhemer's disease, Frontotemporal dementia, ADHD, Addiction, Depression and chronic pain. 22 m: Multi-Arm Bandit with Thompson Sampling, .. Split Q- Learning; two Q-functions: Q+(s,a) and Q-(s,a); w, lambda – hyperparameter quotients for the importance of positive and negative reward.
**31 m:** non-Gaussian rewards. **38 m**: **Different Learning Trajectories for "Mental Agents"** (simulating mental disorders), graphs of the cumulative episode rewards. **40 m:** Lifelong (Continual) Nonstationary setting: Stochastic (reward (muting, flipping, scaling)). 48 m: Continual Lifelong Learning **.. 1:00:46 h: Question: Inverse RL, splitting the reward in two or more parts.**

* **Towards Continual Reinforcement Learning: A Review and Perspectives,** Khimya Khetarpal, Matthew Riemer, Irina Rish, Doina Precup, 12.2020/11.11.2022, 78 p. Mila, McGill University, University of Monreal, IBM Research, DeepMind https://arxiv.org/abs/2012.13490 *– a literature review of different formulations and approaches .. also known as lifelong or non-stationary RL.*

* **Dr Irina Rish - Introduction to Continual Learning,** MAIN Conference**,** 1,85 хил. абонати**,** 3353 показвания Предавано поточно на живо на 3.12.2020 г. [28.8.2025] https://www.youtube.com/watch?v=5Y9wYbhfYkM [Compare with SOTA in late 2025] At the time: narrow superhuman AI; low robustness (adversarial noise, susceptible to small input perturbations – misclassification after adding noise that's invisible for humans); *limited out-of-distribution, "systematic" generalization. For AGI: More biologically inspired "inductive biases". 13 min: Multi Task Learning (MTL): offline, multiple related task are learned simultaneously, using a subset of shared parameters. Transfer Learning, Domain Adaptation. Learning to Learn (Meta Learner) – offline; or (Online, Open World) Learning. 15 m: A comparative study slide; 17 m: CL methods: Replay methods ((Pseudo) Rehearsal, Constrained), Regularization-base (Prior-focused, Data-focused), Parameter isolation (Fixed Network, Dynamic Architectures: PNN, Expert Gate, RCL:, DAN)). 19 m: Transer: aligned gradients, Inerference: trainining one makes the other worse. 21 m: Stability-Plasticity Dilemma .. 24 m: **Experience replay**: a partial history of past examples; 25 m: Desirable properties of CL systems: infinite data or task stream, no task boundary info, online learning, forward transfer (out of distribution generalization), backward transfer (beyond just not forgetting), Work on any kinds of problems (e.g. not only classification), adaptively learning from any partial data (semi-supervised), no test-*

*time oracle* … 28 m: *Task-Incremental learning (the task boundary for changing tasks is given during training and in test time: samples and task label); Class-Incremental: no info about the task during test time; Task Agnostic: no supervision about the changes of tasks either during training and test time .. Meta-Continual – multiple samples of sequences are fed in parallel (task-incremental, class-incremental, task agnostic) and the CL system should be able to learn quickly in another environment. Continual meta-learning* … 31 m *"**OSAKA setting**": … Stochastic sampling of tasks, unknown task boundaries (task-agnostic setting), the target distribution is context-dependent, multiple levels of non-stationarity, tasks can be revisited; Evaluation: online average performance (not the final on all tasks) … 34 m Continual RL Approaches: Explicit Knowledge Retention, Shared Latent Knowledge, Distillation or Rehearsal; Leveriging Shared Structure: Modularity & Composition, (State Abstractions, Skill, Goal, Auxilliary Task) Focused; Learning ot Learn: Context Detection, Learning to (Adapt, Explore) 38 m hippocampus functions as an autoencoder to create and evoke memories. A simple autoencoder model: single-hidden-layer sparse linear autoencoder (**sparse coding** of Olshausen & Field, 1996) … input – dictionary D, link weights – output (reconstructed input) 39 m Neurogenetic online sparse autoencoder: if the reconstruction error is too high on new samples: create new random elements ("neuronal birth") and update memory; Dictionary update via group sparsity. 41 m Evolving a "library" of "Basis functions" 43 m Environment Complexity vs Model Capacity …*

> * **Online Fast Adaptation and Knowledge Accumulation (OSAKA): a New Approach to Continual Learning**. Massimo Caccia, Pau Rodríguez et al. https://proceedings.nips.cc/paper/2020/file/c0a271bc0ecb776a094786474322cb82-Paper.pdf – *"an agent must quickly solve new (out-of-distribution) tasks, while also requiring fast remembering. .. current continual learning, meta-learning, meta-continual learning, and continual-meta learning techniques fail in this new scenario. **Continual-MAML**"* … See the notes to Irina Rish lecture where she mentions this work. The paper mention **Omniglot** – a dataset with characters from many languages, more complex than MNIST. See appendix "**Lazar**", with the paper introducing the dataset and other program synthesis research ~ p.100:
>  * Human-level concept learning through probabilistic program induction. BM Lake, R Salakhutdinov, JB Tenenbaum. Science 350 (6266), 1332-1338, 2015 https://www.cs.cmu.edu/~rsalakhu/papers/LakeEtAl2015Science.pdf https://github.com/Twenkid/SIGI-2025/
>
> * **Task-agnostic Continual Reinforcement learning: Gaining Insights and Overcoming Challenges** Massimo Caccia, Jonas Mueller, Taesup Kim,

Laurent Charlin, and Rasool Fakoor. In CoLLAs - Conference on Lifelong Learning Agents, 2023. https://arxiv.org/abs/2205.14495 - *"a replay-based recurrent reinforcement learning (3RL) methodology for task-agnostic CL agents."*; for Deep Learning; Fig.4, p.9 *"In Meta-World, 3RL decreases* ***gradient conflict*** *leading to an increase in training stability and performance"* … TACRL – Task-agnostic RL; Table 1: Summarizing table of the settings relevant to TACRL. Task-aware RL. Multi-task RL. MDP, POMDP (partially observable – hidden states exist *In a POMDP, an agent cannot directly infer the current state of the environment st from the current observation xt;* MDP doesn't have hidden states (fully-observable world); hidden-mode MDP (HM-MDP). Gradient conflict is when different tasks interfere each other – improvement of the representation for one task makes the other worse. **See also: "Pareto front"** or **Pareto frontier** – when optimizing one parameter, another or others are worsen and the front is the boundary when the effects of mutual "damage" get equillibrated and any optimization in one dimension would reduce the overall goal, quality etc. as defined in the constraints that are optimized (some could be minimized, other maximized).
https://en.wikipedia.org/wiki/Pareto_front
https://en.wikipedia.org/wiki/Pareto_efficient

*** Markov Decision Processes: Discrete Stochastic; Dynamic Programming**, MARTIN L. PUTERMAN, 1994; book; contents and a few pages: https://download.e-bookshelf.de/download/0000/5714/47/L-G-0000571447-0015280627.pdf @Вси: ОбОр

*** Massimo Caccia:** https://optimass.github.io/

* **How to Train Your LLM Web Agent: A Statistical Diagnosis,** D Vattikonda*, S Ravichandran*, E Penaloza*, et al., M Caccia*. Oral at ICML 2025 Workshop – other systems are usually closed source. See p.5. Fig. 2: Tasks in MiniWoB++, WorkArena are similar to (explanation is by analogy of the given ones): Top: yellow background, black text on it with the instruction. Below: a white background, black rectangle outline. Text: "Drag all circles into the black rectangle". Next (top: yellow … etc. and a part of a button "Submit" visible (overlapped by the calendar control that perhaps has fallen from the text box after clicking): A calendar: "Select 15/08/2025 as date and hit submit". Sort the numbers in decreasing order, starting with the biggest at the top of the list (drag-drop operations). Copy the text from the second text area and paste it into the text input. Then press Submit. Below are web for "WorkArena". Fig.3: List of tasks: "order-apple-mac-book-pro15", "knowledge-base-search", "order-apple-watch", "order-ipad-mini" … "multi-chart-value-retrieval" … "sort-hardware-list" … **failed:** "create-user", "create-problem", " order-loaner-

laptop" …

*** Computer Use Agents Benchmarks**
* **WorkArena++: Towards Compositional Planning and Reasoning-based Common Knowledge Work Tasks,** L Boisvert*, M Thakkar*, M Gasse, M Caccia et al.**,** NeurIPS 2024 *"682 tasks corresponding to realistic workflows routinely performed by knowledge workers. WorkArena++ is designed to evaluate the planning, problem-solving, logical/arithmetic reasoning, retrieval, and contextual understanding abilities of web agents. "*
*** WorkArena** https://github.com/ServiceNow/WorkArena *– How Capable are Web Agents at Solving Common Knowledge Work Tasks? –* To take the benchmark, a free registration in ServiceNow is required, where the test cases are defined and the agents access them via browser. The tasks are defined inside "ServiceNow" which is "platform as a service" (PaaS).

*** BrowserGym,** https://github.com/ServiceNow/BrowserGym *– a Gym environment for web task automation. BrowserGym includes the following benchmarks by default: MiniWoB, WebArena, VisualWebArena, WorkArena, AssistantBench, WebLINX (static benchmark) …*
* **WorkArena and MiniWob++ Benchmarks**, Updated 10 July 2025
https://www.emergentmind.com/topics/workarena-and-miniwob
*"benchmarks that simulate realistic web and enterprise tasks to assess LLM and multimodal agent performance. .. from synthetic, short-horizon tasks to complex, multi-step workflows, encompassing navigation, form filling, and dynamic decision-making."*

> * **Sparse coding:** see also Hierarchical Temporal Memory, Jeff Hawkins, Numenta

* **Inverse Reinforcement Learning**

Given demonstrations of behavior in the environment, discover the unknown reward function. The demonstrations are trajectories of state-action pairs. Related fields: imitation learning, apprenticeship learning, inverse optimal control; all are forms of learning from demonstration etc.
See also appendix **Lazar**, "Programming by Example" (PBE) etc.

* **A survey of inverse reinforcement learning,** Stephen Adams, Tyler Cody & Peter A. Beling, 8.2.2022, https://link.springer.com/article/10.1007/s10462-021-10108-x
**..** max-margin, Bayesian, max-entropy … performance may depend on the optimality of the expert demonstrations; *multiplicative weights for apprenticeship learning (MWAL) – two-player zero sum game.* The learning

agent (apprentice) tries to maximize its performance even if the expert is not optimal. Adding negative examples for higher safety as the expert provides only positive ones. … 3.1.3. Max.entropy IRL: *"… the agent is maximizing a function that linearly maps the state features to rewards .. match expected feature counts with observed feature counts from the expert's trajectories. .. path feature count, partition function given the feature weights; reward weight – maximizing the log likelihood under the entropy distribution; gradient descent ..* ***discretizing*** *the continuous trajectories into a* ***graph****. Once a* ***sufficiently coarse graph*** *is created, maximum entropy IRL is used to estimate the reward function…* ***Behavior cloning****, .. a supervised classifier to learn a mapping of states to actions that replicate the behavior of the provided trajectories; 3.2. Extensions of RL: model-free …* ***3.2.2. Active Learning and feedback****"* – the agent selects samples and can ask an expert or oracle to label them. Multi-agent IRL .. **4.1 Mimic the expert** .. *"a translation between the demonstrator's action and state space and the robot's action and state space."* – learning an affordance model; learning *reward functions for robots that will be interacting with a changing environment, such as navigating a path through human pedestrians* [and not an industrial robot with predefined trajectories]. autonomous vehicles …; **future**: … Multi-view learning (a kind of multi-modality)

* A survey of inverse reinforcement learning techniques, S Zhifei, E Meng Joo, International Journal of Intelligent Computing and Cybernetics, 2012 https://eecs.csuohio.edu/~sschung/CIS601/a%20survey%20of%20IRL%20techniques.pdf *"p.14 The objective of IRL is to derive* ***a reward function****,* ***the most succinct representation of the expert's intention****, from a group of expert's demonstrations"*

**Todor:** A good interpretation of the application of the reward – to guide the trajectory, the learning etc. However, for complex and abstract actions and agents, the observable behavior alone may not be enough to discover correctly the intrinsic intentions and motivation in all conditions and may be an ill-posed probleml. With multi-agent interactions the other agents may deceive, cheat, not collaborate or just interfere and have "features" which are hidden for the learning agent; also the human-like complex agents are "deep" and not-unified in their virtual utility function or "reward", unless they are "aligned" in rails. See the cited related papers in the article below: *"Why humans usually answer …"*. [29.8.2025]

*** RL — Inverse Reinforcement Learning,** Jonathan Hui, 29.1.2020 https://jonathan-hui.medium.com/rl-inverse-reinforcement-learning-56c739acfb5a

*** Inverse Reinforcement Learning Meets Large Language Model Post-Training: Basics, Advances, and Opportunities,** Hao Sun, Mihaela van der Schaar**,** 17.7.2025, AAAI 2025 and ACL 2025) Tutorial: https://arxiv.org/pdf/2507.13158 .. *recent advances in LLM alignment through the lens of inverse reinforcement learning (IRL), .. the distinctions* between RL for LLM **alignment** and for conventional RL tasks**. ..**
*the necessity of constructing neural reward models from human data; p.9 Reward Models in Conversational AI: …*
*** Implicit Actor Critic Coupling via a Supervised Learning Framework for RLVR**, Sep 2 2025, Jiaming Li et al. https://arxiv.org/abs/2509.02522
Example of **RL with Verifiable Rewards –** math, programming.
*"By treating the outcome reward as a predictable label, we reformulate the RLVR problem into a supervised learning task over a score function parameterized by the policy model and optimized using cross-entropy loss."*

### → *Returning to the previous Thread*

*** A note on "Habits, action sequences and reinforcement learning":** The term "*instrumental* conditioning" = "operant conditioning" in the lecture on Reinforcement Learning from the world's first university course in AGI at the University of Plovdiv in 2010, 2011:
https://research.twenkid.com/agi/2010/Reinforcement_Learning_Anatomy_of_human_behaviour_22_4_2010.pdf
**Reinforcement Learning: Anatomy of Human Behavior (in Bulgarian) – "Универсален изкуствен разум: Двигатели на човешкото поведение. Бихевиоризъм - учене с подсилване. Учене с учител и подражание. …**
**с. 16 „Оперантно кондициониране:**
***** Промяна/настройка на волевото поведение;
* Скинър – опити с гълъби и плъхове, лостче.
* Взаимодействие със средата.
* Влияе се от последствията (бъдещето), а не от предишни стимули (при КК [класическо кондициониране])
* Бихевиоризъм – дял от психологията, наука за поведението на животните и хората (агенти), която използва научен подход за измерване и промяна на поведението в желана посока и граници.
**с.17. Видеоигри**
Неписани правила за разработка на игри:
* награди и наказания
* нарастваща трудност/избор на трудност

* търсене на награди, бонус; избягване на наказания
**Награди/Наказания:** Брой/Номер (Числа): точки, фрагове, животи, ниво, победи, място в класиране, време и т.н.
**Жанр игри:** Jump & Run, Puzzle, 3D Shooter, Fighting, RPG, Racing, Sport, Strategical, Quest …
**с.18 Игри с коли**
**Управление:** стрелки (4 клавиша: напред, назад, наляво, надясно)

| **Награди:** | **Наказания:** |
|---|---|
| **position--** | **position++** |
| **time-- (за обиколка)** | **time++** |
| **damage—** | **damage++** |

**Състояние:** Кадър от компютърна игра за „Формула 1“ и заградени с кръгчета около числови показания на екрана:
– Най-бърза обиколка и време на текущата.
– Текуща обиколка от общ брой: 1/3
– Степен на повреди (Damage)
– Място в класирането (20/20) – 20-ти от 20 състезателя и таблица с най-близките конкуренти
– Графична карта на текуща област от пистата, около рязък завой – текъщи координати в пространството на пистата
– Скорост на движение = 60 мили/час

**с.19. Взаимовръзки (корелации)**:
**Управление:** Скорост++ --> Време--;  Повреда++ --> Скорост --;
Скорост--  → Място++
* Различаване на **права** и **завой.**
* Ъгъл на влизане и скорост в завоя – скорост на излизане от завоя.
* Оптимална скорост и ъгъл за всеки завой.
**с.20**. **Оптимизация**
**Управление:** t ←→ Скорост, Време, Повреда, Място
* Кое действие да се избере в даден момент, за да се максимизира общата награда и да се минимизира наказанието?
Времедиаграма, в която се показва натискането на някои от четирите възможни бутона и как се променя скоростта.

**с.21. Оперантно кондициониране**
● Откриване на действията, които увеличават вероятността за постигане до награда

● Проби, експерименти и запомняне/натрупване на наблюдаваното развитие на събитията.
● В игрите с коли – ако дам повече газ на този завой, ако вляза по-остро, ако натисна по-късно спирачка – как се променя наградата?*
● Опитът позволява да предсказваме * не разглеждаме проблема с влиянието на други
коли/играчи – баланс скорост/риск от катастрофа
…
**Бел. 25.8.2025: Увеличават вероятността,** но без задължително да запомнят или разбират причината за увеличението или за избора на съответното действие след като завърши обучението им; без да знаят „защо“ и да имат понятие за причина отвъд пресмятането на вероятност и очакване дали е „повече или по-малко“ вероятно да спечелят. Така е при обучението без построяване на изричен модел (Model-Free RL). Това е и един от проблемите за „черната кутия“ в машинното обучение с изкуствените невронни мрежи от типа, използвани в преобразители, конволюционни мрежи и пр: при възприемане на данни те променят теглата си, „учат се“, за да увеличат съвпадението при пресъздаване и предвиждане, но в общия случай в структурата си не запомнят изрични причини по обичаен „разбираем“ начин, а наученото се разпределя като промени на теглата от пресмятането на градиенти върху пакет от примери – „batch”. Не се запомня или „разбира“ „защо“ точно. После отново със следващ набор от примери* и т.н., като при „*професионалното* обучение“ често се работи с пакети от стотици и хиляди примери наведнъж* Така се натрупва „черността“ на кутията. Съществуват обаче и методи, и възможности за такива, за работа чрез *прототипи.* Такива предложения има например в работи на българските учеин Пламен Ангелов и Никола Касабов – виж в приложение *Анелия.*
*mini-batch, batch_size
* “кондициониране“ е по-добре „обучение“, приспособяване и др.
* Дори AlexNet през 2012 г. – по 128 снимки наведнъж.

* **Multi-hierarchical representation of large-scale space**: Applications to mobile robots, Fernández, J. A. & González, J. (2001). Part of the book series: Intelligent Systems, Control and Automation: Science and Engineering (ISCA, volume 24. Springer https://link.springer.com/book/10.1007/978-94-015-9666-4

**Multi-abstraction hierarchies - Multi-AH-graph** *.. structural information acquired from the environment (elements such as objects, free space, etc., relations existing between them, such as proximity, similarity, etc. and other types of information, such as colors, shapes, etc).* .. CLAUDIA – *an implementation of the task-driven paradigm for automatic construction of multiple abstractions: a set of hierarchies of abstraction*

*will be "good" for an agent if it can reduce the cost of planning and performing certain tasks of the agent in the agent's world”*

* **Multi-hierarchical semantic maps for mobile robotics,** Sep 2005, Intelligent Robots and Systems, 2005. (IROS 2005), Galindo Cipriano, Galindo Alessandro et al. https://www.researchgate.net/publication/224623490_Multi-hierarchical_semantic_maps_for_mobile_robotics

# * Why humans usually answer by addressing the higher-level nodes, when asked by another human about the implementation of a prospective long horizon planning problem?

**Todor Arnaudov**, The Sacred Computer, 24.8.2025-25.8.2025 +edits
– Notes on the part III – Hierarchical Planning … from Momchil Tomovs's PhD dissertation "Structure Learning and Uncertainty-Guided Exploration in the Human Brain" by

**Momchil Tomov, 2020**: *„p.106/86: "Importantly, nobody says or even thinks anything like "I will get up, turn left, walk five steps, etc.", or even worse, "I will contract my left quadricep, then my right one, etc."*"*
* Tomov, Momchil, 2020. **Structure Learning and Uncertainty-Guided Exploration in the Human Brain**, Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences (2016-2019)
https://dash.harvard.edu/bitstreams/f065e9e9-3c59-476b-9c2d-0d5d9ad0b5a3/download
Also: **"Discovery of hierarchical representations for efficient planning",** Momchil S Tomov, Samyukta Yagati  Agni Kumar, Wanqian Yang, Samuel J Gershman, 4.2020
https://pmc.ncbi.nlm.nih.gov/articles/PMC7162548/

In fact someone, also born in Bulgaria, said and explained something similar[*1], **16 years earlier.** While the planning example then was for a shorter trip, it still segmented the plan to shorter actions, thus it was "hierarchical" – first the node of the higher level "cluster" was searched and found, however then the lower level steps or nodes, having shorter horizon of time and space, were listed and traversed explicitly:
***Compare .. with "Analysis of the meaning of a sentence, based on the knowledge base of an operational thinking machine. Reflections about the meaning and artificial intelligence."** T.Arnaudov, 3.2004 (originally in Bulgarian**):**
*"(…) However, the first, primary causes are something with many features, as well. E.g.* ***I feel thirst, I want to drink some water.*** *Then, my goal becomes "to satisfy" my thirst. Then I start to search for means to achieve this goal in the possibly closest spatio-temporal area around me. I find that this location is the sink, which is a few seconds away. I'm moving my chair a bit, get up, walk, open the door, pass through the corridor, open another door, turn around, take a cup, put the cup under the fountain, catch the tap for the cold water with my right hand, turn it; water spills; the cup gets filled, I turn the tap back; bend my hand back; prepare my mouth to drink; bend the cup, spill its content in my mouth; swallow....*

*Ready! The thirst was satisfied...*
***Machine**: Why did he drink some water?*
***Human**: Why, I did, really?"*
* *https://artificial-mind.blogspot.com/2008/02/2004.html*
* https://artificial-mind.blogspot.com/2010/01/semantic-analysis-of-sentence.html
* https://artificial-mind.blogspot.com/2010/02/causes-and-reasons-for-any-particular.html
"**Motivation is dependent on local and specific stimuli, not general ones. Pleasure and displeasure as goal-state indicators. Reinforcement learning.Analysis of the meaning of a sentence, based on the knowledge base of an operational thinking machine. Reflections about the meaning and artificial intelligence. *Part 3 of 4 - Comment #2 continues...*"**

**Extension 25.8.2025:** "I feel thirst" is a "drive" or "thrust" (Тл, Tl) in Zrim. The search for means is a traversal of the memory of the content of reachable locations and their distance, cost of reaching. This requires to have a representation that is complete enough and prepared for evaluation and a selected state where you are, e.g. I am in my room, developing on the computer. There are no bottles with water, no cups with tea on my desk or on the table or in my backpack which is next to my chair – in other occasions I may have a bottle of water for my walks. No water in the corridor etc. This search can be done with fast dictionary mapping, too, without checking the space or thinking about it. [See other explanations about the possible *reasons* for the character to select this goal etc. in the cited paper.]

In the case of Momchil's example, the reasons the humans to usually choose these *answers*, when asked, are/could be the following:

1. **Time-memory-bandwidth-complexity constraints for communication and description**, which are derived from the experiences of interactions with other humans / interlocutors, from previous dialogs. How much time one is expected to talk and also how long she would like to talk herself – it is costly as well, takes time, focus, distracts etc. How many words, what would be most interesting for the receiver given the constraints etc. The exact details depend on the exact experiences and cognitive structure of both the question-poser and the responder "traveller", but in usual cases people want or prefer "shorter" answers of a few words. The phonological loop is one or up to a few seconds and the speaker usually can't plan his utterances for many seconds, thus dozens of words etc.

2. **The expected reliability and certainty of the planned actions and their dependency on other causality-control units**, i.e. agents, institutions, circumstances. The body and own muscles are learned, discovered as own and instruments of their owners' will, because the intentions applied on them get fulfilled with a higher certainty and directly, except when one "feels tired", dizzy etc., or when there are other *obstacles* – implying other *wills*, *causes, causality-control units* wich block or inhibit their will. The conception of obstacle arises by the facing of resistance to own

will, which based on the records of previous attempts is predicted and planned to get realized as expected *if* there wasn't an obstacle or another counter force. The obstacles *deny* the will.

In brief, the agent believes or expects, that her body will serve them with a high enough certainty, above threshold, and they also believe that their interlocutor believes or expects the same, so they save this explanation in a short utterance.

```
Agent: Expectations: items[] … item=(body); item[0] = item;  … event1 =
"Flight to Spain" → ?T(event1) → possibilities … item[1] = event1; …
foreach(it in items):

certainties.append(agent.estimate.expected.certainty(item[0].will.future(cm
p(will, actions)>threshold)))  … …
…
```

They may believe that it is certain that they will find a flight and there will be no problems with ordering taxi, *in principle*, however as the problem is posed to reach to the town and the ice cream shop in Spain *as soon as possible*, there could be obstacles and the expected certainty of a direct implementation of the plan is lower: the next flight may be in an appropriate time to catch it "quickly" (the agent must have or decide this parameter), or it may be tomorrow or even after a week.

Alternative flights, which are sooner, may require changing planes, but the agent may have bad memories with that and/or may just not like it; what would be the financial price – would it be too high? Does the agent have prior knowledge about the prices? The real agents in real situations should have or generate on the flight a complete "list of preferences", how much she is ready to pay and which cost will cancel his ice cream craving – as explained in the 2004 work, many examples are given too abstractly and the decisions of the agents are suggested as if they were stereotyped characters. which are supposed to act as "every …" of this type does. The decisions and planning are done based on the concrete, specific circumstances, environment and memories and the decisions are often based on *preferences*, that's the actual "rational part" of the solution: just some specific value, thing, condition is preferred by the agent.

**3. Choosing to mention the Social vs Personal subjects and items from the plan** – the responder may talk about the flight and taxi, because these are "**social**" phenomena in the "objective" commonly accessible world; their significance for other people may be compared to the significance of the movements of the members and muscules of the own body. These social events are believed to be of higher importance *for the interlocutor* as well.

**4. The spatio-temporal scale and the number of repetitions or above threshold**

**match of the identity of the described steps of the plan, when compared to other steps –** the flexes of the muscles etc. and the movements of the body in the immediate local space are expected to be **repetitive**, many instances will be counted, and they are relevant for many other possible trajectories, including for just going to the other room, walking around your chair, leaving your home and going for a walk etc. Therefore these events and actions are less unique and less specific for the current problem than the other ones which are from higher abstraction levels (taxi, airplane), which also span a *larger spatio-temporal scale*, and have a *singular* cardinality. The most specific and identifying is the target destination in Spain which eliminates the other locations such as the other room, the park, university etc. Ordering the taxi is less specific than the destination, but is more specific than the othe travels on foot.

**5. The Scale of the Cost and energy –** in RL frameworks, the agent computes and projects sum of the expected rewards or costs / losses over trajectories of decisions and states. What is the cost function? Different ensembles of virtual causality-control units of the virtual model of the mind of the agent can use different currency. It could be time, space, efforts measured with other metrics, emotional preferences and pleasure/displeasure on subjective scales etc. and a combination of all, which mix in a "mess" and switch when different ensembles of causality-control units prevale*. In this example and in a general or generic discussion with a random person, a common cost could be financial, "money". In this dimension, the flight is the most expensive and thus most "important" step. The same goes if the currency is time: it is expected to take many hours (the taxi is expected to take less; the immediate motions of the body – even less); there will be time on the airports etc.

If the cost is safety and the risks and dangers – the flight may be considered the most risky step, thus "*important*", as well – no matter "the statistcs" which suggests that the planes are "the most safe way of transportation".

In a binary segmentation, all of the following – taxi, ticket and travel in Spain – require spending money, while the immediate movements of the body are accounted as "free of charge", except the expenses for keeping the body alive, their "net worth" or "day earning worth" etc. Even if it is considered, perhaps it would be less than the other expenses for "normal", which are implied in the story.

* Another check that has to be done is whether the ice cream shop is still present and will it be open when we arrive.

* A meta or second order comparison of the plan may discover a memory or may predict, that the agent may **regret** his decision for this long and expensive journey, so she may cancel it or/and search for alternative solutions – isn't the ice cream in the fridge similar enough, or isn't there a similar one in the local shop, or does it really matter the ice cream to be of that specific type at all? Do I need to eat ice cream at all, what about pudding or an apple?

The **cognitive control** is the ability to stick to a goal once it's set and it suppresses other potential plans, as the agent may decide to replan in all moments and that is an open possibility (exploration vs exploitation trade-off); usually the cognitive control is associated with higher intelligence, an ability for "delayed gratification" etc., however in reality this is a simplified interpretation and whether better cognitive control is beneficial or not is questionable for every specific situation.

A complex multi-scale and multi-modal, multi-world, multi-eveything agents such as humans may be modeled as having many virtual ensembles for different trajectories at different scales, ranges, preferences etc. As they have different preferences for the values, spans of sampling, utility functions, they contradict each other and whether the larger scale agent, "the person" will be "happy" about her choices depend on which exact ensembles and combinations of them have taken the power at the time of evaluation – either during the policy/plan/action selection and after that, in all moments. All these are subject to change "without notice", including the temporal and "integrated" identity of the "deep agent". As the classical paper from 2004 discovers with the example for the selection of eating or not eating a piece of chocolate bar, depending on the memories, which are recalled at the moment, and the range of the planning horizon, there's no single objective utility function for such multi-layer and multi-scale diverse agents. In the example with the chocolate craving, the short horizon of a second or a few seconds is expected to produce or return a high reward from the taste and satisfaction; however, when a moment later the agent estimates a longer scale of months, she recalls memories from visiting the dentist and thus this action is projected to return a negative reward. The dental cavity is a possibility, it is possible not to happen, but what the agent would believe at this exact moment? What is "the best" then and why? It is a preference and a dependency of the specific configuration of the state of the agent at the very moment, and this goes for thousands more intricate intentions, goals and tasks, where all lead to the agent continuing existing or achieving "some goals" which are "rewarding" or the agent believes they will be etc. The *actual* intentions and goals are not just the short verbalized reports "I want to eat a piece of chocolate bar" or "ice cream", but the "symmetry breaking" details, required for making a decision between different trajectories in he same and different scales.

The **hierarchical planning** allows for covering larger and more complex futures by reducing the complexity and the number of compared possible paths, while using **flat** states at the lowest level requires bigger working memory for the states and their combinations. The limitations of human working memory don't allow for even a "clever" average human to think clearly about even just 10 exact steps in the same time (7+-2). At low level there is another problem, besides memory, which makes planning not just infeasible, but impossible: **the uncertainty and uncontrollability** at

low level and the short reliable prediction horizon. Different scales have different ranges and precisions as mentioned in the thesis, too, with higher level coarse-graining and being lossy compression.

The causality-control units predict and cause the future at all possible scales, but there's a limit of the precision and the range, and in the real world the units, where agents are forms of such causality-control units, also may not be able to have the required information. Even for the execution of the movements of the human body, even for short and simple movement, proprioceptive and other sensory feedback is required in order to perform the movement right. The biological signaling and processing carries noise etc. and the brain cannot remember and store long procedures at a high precision in this "electrical" form – it struggles even at higher level of abstraction and words. A lot of efforts, repetitions and time, going for many days for continuous brain adaptations, are required in order to remember a sequence of actions of say a minute of modern dance performance, to learn to play a musical piece on an instrument or even for mastering basic repetitive dance steps and transistions for simple popular dances.

At low level and short ranges, the plans are short and the system should be ready for switching to **reactive** mode and acting on spot. See the 1980s multi-agent frameworks, "deliberative" (planning), "non-deliberative" (reactive) and hybrid architectures, for example the literature reviewed in the #multi-agent section of the appendix "Listove" of "The Prophets of the Thinking Machines: Artificial General Intelligence and …" and the sections for classical planning: Planner, PDDL etc.

A practical example is the humanoid robot Atlas by Boston Dynamics – its motions are guided using Model Predictive Control (MPC). One of the leaders of the project MuJoCo MPC calls it in talks and a paper "surfing" and not "optimization", because the planning in such mode is performed in overlapping spans. Imagine an MPC-controlled robot entering a parcour "spot" which it wants to pass through. At each time period for planning, the robot discovers and selects a feasible preferred trajectory for say 3 seconds ahead, however while it executes the first second, it computes the next 3 seconds of overlapping span, because the environment may change, new data arrives while seeing more from the environment, there might be mistakes in the predictions or inaccuracies in the implementaton, which may require correction etc.

Let's return to our human example and our **scale ranges**: notice that when we talk about "*flexing the quadriceps*" we assume this is a "low level description" of the action, however it is low for another description, e.g. as "standing up" or "walking", but it is already very high from a mechanical or "real" perspective as for implementation a lot of deetails have to be explicitly given: exact forces, speed of their applications, activation of neural fibers and muscular fibers, complete body pose and that should be controlled in time in small steps etc. The ice cream lover has no proper

or convenient code for expressing it, besides the infeasible bandwidth for spoken communication, even if he could the data numerically. It could be expressed via other data records and diagrams with a machine which records the neural, neuro-muscular and mechanical activity of the body. Notice that natural language is not the only and the perfect means for communication: it is appropriate or convenient for particular resolution of causality-control and precision, generality etc. **See** "Chairs, Buildings, Caricatures, … ", or "AGI Digest", 2012.

The nodes at the higher levels, which cover larger spans of time, space and resources and are more costly are also not certain in a strict sense and the agent should also be prepared for reactive decisions when these nodes are active: the flight could be delayed or cancelled, there could be issues with passports or visas, luggage lost at the airport etc. Besides, the **position** of a node with particular approximate name, location, abstraction in the hierarchy of the plans may be fluid and it is dynamic and can be **recreated,** reinstantiated; a node could have its meaning "morphed" – for example the node *"finding a flight to Madrid"* in the *answer to the question* what one would do is "one" flight or "**flight node one"**, but when the tourist ***is already at the airport*** on time, now another "**flight 2"** node is current; when seeing information on the screen that *the flight is cancelled* or postponed, then another node **"flight 3"** is generated; the new instances overlapped and "scrolled" from the previous ones which were prospective and the projections were implemented to "the airport" etc. First the "flight" was an immediate continuation: a short-term plan with expected high certainty, but this expected future gets rejected and then it becomes a distant one again. When the conditions change abruptly, replanning is inevitable.

Notice however that an *abrupt change of the conditions* **can be** an *expected possibility* and even part of the original plan with prepared alternative decisions and routes. Right – like the cliche from the movies: "Plan B". Planning is not just a singular trajectory of expected states, it may be a tree and a forest, including and taking into account possible branches, some of them unwanted, but anticipated, and the estimation of the total reward, success, cost, loss etc. could include these alternative expected possible obstacles and "risks". You may have planned in advance what to do if the flight was canceled etc., or you may act "reactively" or *replan on-the-spot*, in both cases given some criteria for the precision of both. Not only the hierarchical and any kind of representations which the agents construct are prepared for allowing more efficient search and traversal, but **the world** itself, the environment is also shaped that way. In the schools of thought of enactive cognition, extended mind etc., that is the construction of the "ecological niche", "epistemological" in this case. That's another reason why the agent doesn't need to plan everything in details: the environment and the affordances which it will offer will tell a reasonably small set of possibilities from which the agent to choose in order to survive or to continue her journey.

Thus the agent can't predict the details for "long" horizons either because of lacking information, unreliability, noise and insufficient cognitive capacity, and because it doesn't have to.

* Notice one commonly repeated thought, after the predictionist framework reached even discussions by the USA politicians, that as AGI is supposed to be smarter than "us", if you can't predict the AGI or whoever, *"you can't control it"*, "therefore it's dangerous"*. If you can't predict it, you can't know what will follow, however *being able* to predict it is not enough to control it either, controlling requires being able to **cause** it to act as you wish, to drive it, and that is supposed to be at the required resolution of causality-control and horizon. In Theory of Universe and Mind, real control, or causality-control, is present only if the driving unit (master) can cause the future of the controlled unit (the driven, "slave") with a resolution, which is highest from the point of view of the "slave" unit – this is rarely the case. Most of the time the control is only "virtual", at a lower resolution, where the error can accumulate*.

A prophet can be able to foresee the future decades ahead and be proven right by the real events; the words and acts of others in the materialized future may prove his predictions, 20 years later, however without possessing power, directly or through connections, the prophecy are not recognized in their original time, and then after they are proven, they are ignored and not given even credit. While the opportunists, whose predictions were "completely wrong" or even experts who "didn't expect the progress to be so quick" and are alarming that AI development became "out of cntrol", i.e. they **provenly failed and fail in predicting the future** in this domain, recive enormous amount of power and their wrong or already banal insights are multiplied and propagated.

See the main volume of "The Prophets of the Thinking Machines: AGI & Transhumanism: History, Theory and Pioneers; Past, Present and Future" and the appendix "The first modern strategy for development with AI was created by an 18-year old Bulgarian and repeated and implemented by the whole world 15-20 years later: The Bulgarian Prophecies: How would I invest one million with the greatest benefit for my country?" at SIGI-2025. https://github.com/twenkid/sigi-2025

* There could be moments of "error compensation" though. Dritfs accumulate, but at key points they are eliminated by the device/unit entering some state of "reset", "initialization" or "alignment", either if it is more "informational" and complex and entangled states, or a mechanical one with simpler controlled parameters, for example a mobile ground robot, an industrial robot or a drone. However, at the highest resolution of the Universe Computer, its "machine code", only the Universe "controls", the physical causality-control units, the agents, can't cause-control or control-cause even their own future in the strict sense, but they accept that the

resolution at which they can control themselves is sufficient for their “free will”.
* See my detailed and juicy comment about the “control freaks” and that Power is confused for Intelligence and Power overrides Intelligence, in the AGI email list, 8.2025. I may publish it as an article on SIGI-2025 as well: https://agi.topicbox.com/groups/agi – the answer to the “Less Wrong” topic about AGI that will lead to “human extinction because …”, posted on 11.8.2025; the group requires the users to register in order to allow them to see the messages.

* See also: *** Nature or Nurture: Socialization, Social Pressure, Reinforcement Learning, Reward Systems: Current Virtual Selves - No Intrinsic Integral Self, but an Integral of Infinitesimal Local Selfs - Irrational Intentional Actions Are Impossible- Akrasia is Confused - Hypothesis about Socialization and Eye-Contact as an Oxytocin Source**, T.Arnaudov, Artificial Mind, 30.11.2012
* TUM also has an example about ice cream – in Universe and Mind 2, “Letters between the 18-years-old Todor Arnaudov and the philosopher Angel Grancharov”, 9.2002, it’s translated as an example realted to the many possible scales and ways for causality factorization which are correct in “Stack theory is yet another fork of Theory of Universe and Mind”, T.Arnaudov, 2025. http://artificial-mind.blogspot.com/2012/11/nature-or-nurture-socialization-social.html

* **The neural architecture of theory-based reinforcement learning.** Tomov, M. S., Tsividis, P.A., Pouncy, T., Tenenbaum, J.B., & Gershman, S.J. (2023). Neuron
* https://medium.com/@momchiltomov/theory-based-reinforcement-learning-in-the-brain-700b49b951f7 *– Theory representations in prefrontal cortex; 4. Theory update signals in inferior frontal gyrus, occipital gyri, and fusiform gyrus .. p.7 "Since theory updates are triggered by surprising events which violate theoretical predictions, such an increase in neural activity could also be interpreted as a kind of theory prediction error. .. general linear model (GLM) with impulse regressors at theory update events – frames at which EMPA switched from one most likely theory to another based on the participant's gameplay," .. 5.* ***Separate*** *update signals for different theory components .. a set of object types and their physical and/or intentional properties (since they could be other agents), a set of relations between objects describing the outcomes of object-object interactions, and a set of goals that the agent pursues. .. p.10: 6. Theory representations activated during updating .. p.11 theory representations are preferentially activated during theory updating, akin to being "loaded" into working memory for the necessary computation. p.12 Fig.7: Effective connectivity during theory updating is consistent with predictive coding.*
* **Neural computations underlying causal structure learning.** Tomov, M.S., Dorfman, H.M., Gershman, S.J., 2018. Journal of Neuroscience 38, 7143–7157

*** Theory-based Reinforcement Learning in the Brain. A summary of Tomov et al. (**2023), Neuron. 5.3.2023 https://medium.com/@momchiltomov/theory-based-reinforcement-learning-in-the-brain-700b49b951f7 – "**model-free** .. through trial-and-error, learn **which actions tend to lead to good outcomes** in the long run and then choose actions that have been rewarding in the past. In contrast, **goal-directed behavior** is formalized by **model-based** approaches that learn **which actions lead to which states**; that is, they learn an **internal model of the environment** — a kind of **cognitive map** — which can be used to **plan by simulating the outcomes** of different courses of action."

*** Edward Tolman** – cognitive map, observational learning (without a reward)
https://en.wikipedia.org/wiki/Observational_learning
* https://www.britannica.com/biography/Edward-C-Tolman
* **E.Tolman, Purposive Behavior in Animals and Men (1932).** *"...the unit of behaviour is the total, goal-directed act, using varied muscular movements that are organized around the purposes served and guided by cognitive processes."* It is counted as behaviourist, because it requires *"objective observation and rigorous experimental procedures"*.
https://www.britannica.com/science/learning-theory *"The range of phenomena called learning: Classical conditioning, Instrumental conditioning, Acquisition of skill,*

*Discrimination learning, Concept formation”*, Learning of principles, *“Problem solving”*

*** Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments**
Logan Cross1,4 lcross@caltech.edu · Jeff Cockburn2 · Yisong Yue3 · John P. O’Doherty2
https://www.cell.com/neuron/fulltext/S0896-6273(20)30899-0 – *“DQN as a model of brain activity and behavior in participants playing three Atari video games during fMRI. Hidden layers of DQN exhibited a* ***striking resemblance to voxel activity in a distributed sensorimotor network****, extending throughout the dorsal visual pathway into posterior parietal cortex.”* … value representations; PPC – posterial parietal cortex;

* **Reinforcement learning, fast and slow,** Botvinick, M. · Ritter, S. · Wang, J.X. ...,
Trends Cogn. Sci. 2019; 23:408-422 https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(19)30061-0 Deep RL … sample inefficiency and slow learning of the early successful Deep RL for more “interesting” problems (2013-… Atari 2600 game playing etc.) due to the slow incremental updates of many parameters. Updated architectures achieve higher sample inefficiency. The slow is “weight-based learning”. Episodic RL (ERL), state similarity, learning the state representation with gradient descent; fast learning of ERL is enabled by the slow learning of RL. Meta RL – learning to learn. **Inductive** biases*. Evolution as the slowest form of learning which builds into organisms inductive biases which are appropriate for learing from the environment. “*meta-learning plays out not only within a single lifespan, but also over evolutionary time. .. evolution does not select for a truly ‘general-purpose’ learning algorithm, but for an algorithm which exploits the regularities in the particular environments in which brains have evolved.”*

**Todor:** Inductive Bias ~ “Heuristics” and intrinsic pre-built structure in the cognitive and learning system, i.e. the system don’t have to learn from scratch and be “too general”, but take into account the specifics of the environment and the task, i.e. the systems reduces the unnecessary “learning” part and know in advance. A related idea were predictions I did about early 2010s in AGI list discussions in Artificial Mind, that different AGI architectures and approaches will eventually converge to common high level representations and will become similar, because they are expected and wanted to to be compatible with human-like reasoning, expression, representation and because the generalization process itself is convergent and the more general patterns “attract”

each other. One such "attractor" is natural language, either structural and output level and the generation mechanisms near the surface output; also logic, math etc. (…) See Todor's future works on **Zrim,** Developmental AI, Seed-AI, SIGI: **"Genesis: Creating thinking machines"** (working title) with yet unpublished research since the 2010s. [25.8.2025]

*** What Is the Model in Model-Based Planning**?, Thomas Pouncy, Pedro Tsividis, Samuel J. Gershman, 4.1.2021 Flexibility of human problem solving. MDP – Markov decision process.... task representations: useful object features. Categories or combinations; "*We simulate planning as a form of selective look-ahead search, which has been the basis of recent artificial intelligence (AI) game-playing success .. subset of recent AI work has begun to move away from feature vector or relational representations and back toward more abstract rule-based representations (Lang, Toussaint, & Kersting, 2012; Zettlemoyer, Pasula, & Kaelbling, 2005; When exactly an agent can succeed in a previously learned task without the need for additional training is the key distinguishing factor; this ability to succeed at novel task variants without additional training as "flexibility"; feature vector, relational category, and rule-based approaches to representation learning; hierarchical RL (HRL); a doorway between two rooms represents a bottleneck linking any state in the first room to any state in the second room. Thus navigating to that doorway is likely to be a useful subgoal from within either room. 3.3 Relational-categories representation – One of the challenges of classical feature learning is that it often struggles to extract useful information about the relationships between features. Explicit relational primitives; 3.4 Rule-based representation*
*.. preconditions + action → outcome.*

*** Human-Level Reinforcement Learning through Theory-Based Modeling, Exploration, and Planning**, Pedro A. Tsividis, Joao Loula, Jake Burga, Nathan Foss, Andres Campero, Thomas Pouncy, Samuel J. Gershman, Joshua B. Tenenbaum, 7.2021 https://arxiv.org/abs/2107.12544 *"a particularly strong form of model-based RL:. .. **Theory-Based Reinforcement Learning**, .. **human-like intuitive theories** -- **rich, abstract, causal models of physical objects, intentional agents, and their interactions** -- to explore and model an environment, and plan effectively to achieve task goals. .. a video game playing agent called **EMPA (the Exploring, Modeling, and Planning Agent)**, which performs Bayesian inference to learn probabilistic generative models expressed as **programs for a game-engine simulator, and runs internal simulations over these models** to support efficient object-based, relational exploration and heuristic planning. EMPA closely **matches human learning efficiency on a suite of 90 challenging Atari-style video games**, learning new games in just minutes of*

*game play and generalizing robustly to new game situations and new levels. The model also **captures fine-grained structure in people's exploration trajectories and learning dynamics**. Its design and behavior suggest a way forward for building more general human-like AI systems."*

**Tosh: Compare to Theory of Universe and Mind, 2001-2004** and the R&D plan in Universe and Mind 5 (12.2004), mind/general intelligence is a hierarchical universal simulator of virtual universes, prediction-compression etc. One of my more recent alternative words for thinking machine is **EMIL: Explorer, Mapper, Improver and Learner**, matching the name **"Emil"** of characters from my science fiction works from the same period and the project announced in 2003 in ***"Creativity is imitation at the level of algorithms: An outline sketch of a possible path of development of the Artificial Intelligence "Emil""***.

* https://twenkid.com/agi/Creativity-is-imitation-at-the-level-of-algorithms-todor-arnaudov-2003-2025.pdf

* **Neural evidence that humans reuse strategies to solve new tasks**
Sam Hall-McMaster ,Momchil S. Tomov,Samuel J. Gershman ,Nicolas W. Schuck, 5.6.2025
https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3003174
*"successor representation known as **successor features** and generalized policy improvement (SF&GPI). .. Neither model-free perseveration or model-based control using a complete model of the environment could explain choice behavior. .. some participants exploited structural similarities between the training and test tasks to determine their choices, similar to a Universal Value Function Approximator (UVFA) process in which similar task cues lead to similar rewards for a given action .. **Conclusion:** .. generalization to new tasks was more consistent with an SF&GPI-based algorithm than an MB algorithm using a full model of the environment.* Still a *model-based algorithm* may be *operating on a partial model of the environment, or with noisy memory for features, associated with suboptimal training choices. SF&GPI and UVFAs should also be systematically compared in the presence and absence of structural similarity between training and test **cue**s in future studies. ... successful past solutions were prioritized as candidates for decision-making on tasks outside the training distribution."*

* Schaul T, Horgan D, Gregor K, Silver D. **Universal value function approximators**. Proc 32nd Int Conf Mach Learn. 2015;37:1312–20.https://proceedings.mlr.press/v37/schaul15.html
https://proceedings.mlr.press/v37/schaul15.pdf *– a single function approximator V(s;*

*θ) that estimates the long-term reward from any state s, using parameters θ .. the Horde architecture (Sutton et al., 2011) consists of a discrete set of value functions ('demons'), all of*
*which may be learnt simultaneously from a single stream of experience … In large problems, the value function is typically represented by a function approximator V (s, θ), such as a linear combination of features or a neural network with parameters θ. .. previous methods have focused on generalising across tasks rather than goals. .. a task may have different MDP dynamics, whereas a goal only changes the reward function, but not the transition structure of an MDP.*

* **Prioritized Experience Replay,** Tom Schaul, John Quan, Ioannis Antonoglou, David Silver, 2016, ICLR; DeepMind https://arxiv.org/pdf/1511.05952
p.2.. *evidence of experience replay in the hippocampus of rodents, suggesting that sequences of prior experience are replayed, either during awake resting or sleep. Sequences associated with rewards appear to be replayed more frequently.* In supervised learning, techniques for dealing *with imbalanced datasets when class identities are known .. re-sampling, under-sampling and over-sampling .., possibly combined with ensemble methods.*

## See also:

*** Successor Feature Representations,** Chris Reinke, Xavier Alameda-Pineda, 2.8.2023
https://arxiv.org/pdf/2110.15701 *Transfer in RL .. improve learning performance on target tasks, using knowledge from experienced source tasks. Successor Representations (SR) and their extension Successor Features (SF) .. where reward functions change between tasks. They reevaluate the expected return of previously learned policies in a new target task to transfer their knowledge.* SF extends SR by linearly decomposing rewards into successor features and a reward weight vector, allowing their application in high-dimensional tasks. .. [only if] *a linear relationship between reward functions and successor features* .. Successor Feature Representations (SFR) - *learning the cumulative discounted probability of successor features* ..

* **Successor features for transfer in reinforcement learning. In Advances in neural information processing systems**, André Barreto, Will Dabney, Rémi Munos, Jonathan J Hunt, Tom Schaul, Hado P van Hasselt, and David Silver. pp. 4055–4065, 2017.

* **Transfer in deep reinforcement learning using successor features and generalised policy improvement.** Andre Barreto, Diana Borsa, John Quan, Tom Schaul, David Silver, Matteo Hessel, Daniel Mankowitz, Augustin Zidek, and Remi Munos. In International Conference on Machine Learning, pp. 501–510. PMLR, 2018.

*** Hierarchical Universal Value Function Approximators**, Rushiv Arora, 11.10.2024, https://arxiv.org/html/2410.08997v1 p.1 *"Value functions (V (s) and Q(s, a)) are key to reinforcement learning algorithms (Sutton & Barto, 1998).. .. how important it is for the agent to be in a certain state and take certain actions in a given state. .. **Universal value functions** (Schaul et al., 2015) (V (s, g) and Q(s, a, g)) and **general value functions** (Sutton et al., 2011) (Vg(s) and Qg(s, a)) are key when extending this to multitask domains. .. there is underlying structure in the states, goals, options and actions that results*
*in a universal representation of the hierarchy .. zero-shot generalization to unseen goals*.*

* Cmp: embeddings in LLMs etc.

## * Повторна и многократна обработка на съхранени записи в мозъка и машините

## * Replay in brains and machines, Lennart Wittkuhn, Samson Chien, Sam Hall-McMaster, Nicolas W. Schuck, 2021

Обзор на това важно направление в невронауките и ИИ.

**Въведение, Тодор Арнаудов:** Всяка форма на повторение при учене от човек или при обучение на животно, дресиране; „репетициите“ в музиката, танците използват този механизъм за подобряване на запомнянето; също когато си представяте определени действия преди изпълнение, „визуализирате“, „подгрявайки“ съответните записи и връзки в мозъка, подобно на подготовката (priming) в познавателната психология.

При обучение на невронни мрежи също се извършва многократно „прекарване“ на примери от данните за постепенното „наместване“ на функцията на теглата, и при различни видове дообучение[138] за следване на заръки и продължаващо предобучение на преобразители е полезно заедно с новите данни да се подават и онези, с които е обучаван оригиналния модел, за да не се получи „катастрофално забравяне“ и заучаване само на новоподаваното малко специализирано множество. Такава техника е използвана например и при обучението на BgGPT на INSAIT[139].

Хипокампът в мозъка се смята за изпълняващ „просвирване“ на записите на преживявания от деня, а сънуването – като вид бърз преглед и укрепване и затвърждаване на спомените[140]; вместване, свързване на ежедневните временни връзки и спомени с по-трайните, преоценка и запис в постоянната памет в новата кора (неокортекс). Повторни прегледи на записите на опита не

---

[138] Дообучение (фина настройка, fine tuning); следване на заръки (инструкции, instruction following, instruction tuning); преобразители – трансформатори, transformers

[139] Mitigating Catastrophic Forgetting in Language Transfer via Model Merging, Anton Alexandrov*', Veselin Raychev, Mark Niklas Miiller,Ce Zhang, Martin Vechev, Kristina Toutanova, 2024 https://arxiv.org/html/2407.08699v1 “BaM” - Branch-and-Merge – метод за дообучение на нови езици, които са представени с по-малко данни в началния набор на даден модел (в случая: **Mistral-7B**), като в същото време се намали „забравянето“ на знания от първоначалните езици и способности.

[140] “Memory consolidation” – https://pmc.ncbi.nlm.nih.gov/articles/PMC3117012/ Hippocampal memory consolidation during sleep: a comparison of mammals and birds Niels C Rattenborg, Dolores Martinez-Gonzalez, Timothy C Roth II, Vladimir V Pravosudov. https://pmc.ncbi.nlm.nih.gov/articles/PMC3117012/

* Hippocampal Sleep Features: Relations to Human Memory Function Michele Ferrara,*, Fabio Moroni, Luigi De Gennaro, Lino Nobili, 2012 https://pmc.ncbi.nlm.nih.gov/articles/PMC3327976/#:~:text=Hippocampal%20activity%20seems%20to%20specifically,the%20pre%2Dexisting%20cortical%20networks

* Fosse MJ, Fosse R, Hobson JA, Stickgold RJ. Dreaming and episodic memory: a functional dissociation? Journal of Cognitive Neuroscience. 2003;15:1–9. doi: 10.1162/089892903321107774

* Stickgold R, Hobson JA, Fosse R, Fosse M. Sleep, learning, and dreams: off-line memory reprocessing. Science. 2001a;294:1052–1057. doi: 10.1126/science.1063530

се извършват само от хипокампа и понякога са в *обратен* ред. (виж)

**Свръхобщи понятия – начални, първични, основни, коренни**

Редици (последователности, поредици, списъци, вериги); предвиждане (предсказване, прогнозиране) са сред най-общите въобразни[141] възможности за действие, за случване; събития – последните думи също са част от множеството от общи въобразни (информационни) понятия, представи, структури. С тях са свързаност, прекъснатост; начало, край; предишен, следващ и съответно място в редицата (място, разположение; адрес, координати). Съответно тези понятия, техника са приложими и ги откриваме навсякъде във всякакви явления и математически структури от данни.

**Бележки:** Хипокамп: head-direction cells; place cells …
*Българинът* ***Ивилин Стоянов*** също работи в тази област: Ivilin Stoyanov[142]; item, sequence, map

**Replay:** successor representations (SRs), that can be used for reinforcement "experience replay" was introduced in the early 90s (Lin, 1991)[143]. …RL без модел, с модел – поражда нови „преживявания" въз основа на научения модел на средат … преходи между две съседни клетки за места (неврони, които се задействат, когато животното е на определено място и го разпознае[144];) * model-free, model-based -- discussion, cite; [виж за SR и по-горе, подходящи за пренос на наградата при промяна на средата, за следващи задачи]

**Тош:** разделението „без модел" и „с модел" понякога е условно, виж коментарите ми към Джон Тан Мин; според статията просвирването можело да се разглежда като размиване на границите между обучение с подкрепление без и със модел, като това което наричат „с модел" по моята „номенклатура" е по-точно с „*изричен* модел", докато когато се работи „без модел" – той е всъщност „неявен", „неизричен" или непредставен по някакъв „приет" за необходим начин, но системата така или иначе придобива „представа" и построява „скрит" модел, както пораждащите модели за изображения, които рисуват фотореалистични изображения, показват, че са кодирали скрит модел на проследяване на лъчите и/или всеобщо осветяване (global illumination), пространствена геометрия текстуриране и т.н., което донякъде не е и толкова учудващо, защото линейната

---

[141] Въобразен – информационен (юнашко наречие/сметачобългарски)
[142] The hippocampal formation as a hierarchical generative model supporting generative replay and continual learning, Progress in Neurobiology Volume 217, October 2022, 102329, Ivilin Stoianov, Domenico Maisto, Giovanni Pezzulo https://www.sciencedirect.com/science/article/abs/pii/S0301008222001150
[143] Lin, L. J. (1991). Programming robots using reinforcement learning and teaching. In Association for the advancement of artificial intelligence (pp. 781–786). https://cdn.aaai.org/AAAI/1991/AAAI91-122.pdf * Lin, L.-J. (1993). Reinforcement learning for robots using neural networks. Carnegie Mellon University.
[144] Свързано явление от „Зрим": разпознаватели на контекст, рзпзнвтл {К}

алгебра и преобразуванията, операциите с вектори и матрици, с които се изразяват и кодират теглата на невронните мрежи, са начина по който се обработват изображения и от компютърната графика.

Статията предлага и богат набор от свързана литература. Виж и статията по-горе от невронауките, според която не само целевите поведения, а и навиците могат да се управляват чрез обучение с модел (model based learning), а не да работят на принципа стимул-реакция.

* Habits, action sequences and reinforcement learning. Dezfouli, A. & Balleine, B. W. (2012).

*** "Learned"** - model-free, **"Planned" -** model-based
* Информацията за преходи между състояния и сетивната информация по време на обучението позволява на знанието да се пренася и използва отново в структурно-подобни обстановки, но с нови сетивни особености (structural similarities, but new sensory specifics)* (Behrens et al., 2018; Liu et al., 2019; Baram et al., 2020; Whittington et al., 2020).
cells, grid cogn.maps
grid cells, object vector cells; consolidation of recent memory traces
reactivation (sleep, rest) aggregated storage; reactivate;
episodic memory retrieval
- fast bursts 150-250 Hz SWR; PFC-hippocampus interaction
- и непоследов. Извикване
- enthorinal, PFC, V,A,M - ventral striatum
- и наобратно (reverse) - далечни и непосетени места
- и звук, време; не само физическо пространство
* predictive locations - sucessor representations (SRs)
grid-like patterns - enthorinal cortex & ventromedial PFC [,] in non-spatial space (Constantinescu, 2016)
**Тош: Няма "непространствено пространство" –** може да са например ***неевклидови*** и пр.; в **топологията**: **многообразия** и **атласи с преходи (Manifolds)**; виж плана от *Вселена и Разум 5;* виж и М.Левин за различните пространства, в които могат да действат агентите, като „морфопространството“, в което работят живите организми при изграждането и възстановяването на тялото (morphospace), какво могат да правят; както и Петер Грендерфорс, Когнитивна лингвистика.
* Сравни също с „творчеството е подражание на ниво алгоритми“ – структурно-подобните, или подобни в строежа си обстановки, среди и т.н. могат да се породят от общ алторитъм с променени параметри; те са инстанции на пораждащ модел схема, който има общи части.

*** Locations, events beyond physical space ; from conceptual to social cofnition –** MTL - medial prefrontal, orbifrontal cortex, OFC
RL: Lin, 1991 Experience replay, 1992 .. 2015, **Mnih Atari Deepmind** ...
– Sample from memory – planning

– RL interaction - with the environments (Т.А. обикновеното МО също е вид взаимодействие със среда: наборът от данни)
– gradually, trial and error
– optimize goal-directed behavior, given an already /established cognitive maps
– locations; events; mapping #: map-like*;* transitions (2 adjacent place cells)
– Learning - speed, data efficiency, forgetting, reorganizing experience, planning, (отделни) generalization
– Replay direction append(direction->word.direction, word.direction.forward ...)
– Sutton, TD (temporal difference learning) ... "The Bitter lesson" – Виж бележките ми в „Stack theory is yet another Fork of Theory of Universe and Mind", 2025.

**Тош:** в данните, Вселената има структура, тя е --' предвидима; *няма комбинаторен взрив* (било общоприето, че НМ нМг дс[145] мащабират, заради "проклятието на високите измерения"). (...) незавършено … Виж също MPC – Model-predictive control, Теория за оптималното управление.

**Сравни идеите от „Повторна обработка на съхранени записи в мозъка и машините" с ТРИВ/TOUM за двете теоретично изведени предполагаеми „операционни системи" на мозъка: изпълнителен и събитиен вседържец** в потока на мисленето и представянето на ума като универсален предсказател на бъдещето на възприятията и действията, припомнянето на далечни събития за сравнение с настоящи и предвиждане на бъдещи**:** с. 2x. от Wittkuhn et al. 2021: ***„****Друго измерение на преустройването на опита (experiential reorganization) е да повтори, просвири преживявания, които са се случили преди много време или дори да се отнесе до въображаем опит от бъдещето: преосмисляне[146] на миналото, преоткриване на миналото…. В рамката на УП (RL) разликата между безмоделни и моделни системи се състои в това дали действат възоснова на вече научени стойности („кеширани", съхранени в междинна по-бърза памет) [на наградите] или планират [„вътрешно"; чрез изричен процес](Sutton & Barto, 2018; Daw et al., 2005"*

**„Схващане за всеобщата предопределеност 3" (Вселена и Разум 3),** Тодор Арнаудов, август 2003 г., (…)
***62.*** *Възникването на съзнанието е създаването на първия развит събитиен вседържец (събитийна операционна система) на човека. Събитийният вседържец запомня и обработва поточна информация от сетивата и осигурява запомнянето и възпроизвеждането на събития като "на лента" или "плоча" - можем да изберем "запис" и да го "прослушаме" във въображението си.*

*"Първият развит събитиен вседържател" е операционна система, способна при усъвършенстването и надграждането си да запазва*

[145] "Не могат да се"
[146] Reinventing the past.

*СЪВМЕСТИМОСТТА с предходните версии. Например форматите за запис на събитийни спомени от "Проявление 1.0" на Развития събитиен вседържец на разума нататък са съвместими един с друг "отгоре надолу" - по-новите версии са съвместими с по-старите, техни разширения.*

*Например имаме спомени и от двегодишна възраст, и от вчера. Мозъкът ни се е променял през годините, но си е останал съвместим към спомените от двегодишна възраст.*

*В паметта ни всички събития, когато и да са били въведени в паметта, са разположени едноизмерно и си спомняме за тях мигновено.*

*Ако не можем да четем и видим нещо написано, е трудно да го запомним, защото трябва да го запомним като цяло изображение, а човешкият мозък е по-слаб в запомнянето на единични - несвързани в система - сложни изображения.*

*След като се научим да четем (разширим възможностите на Събитийния вседържец), като видим написано слово запомняме не изображението, а буквите, сричките и дори целите думи. (Които може би се представят като адреси в паметта). Способността да си спомняме събития, написани в слово, е надстройка на събитийната операционна система. Същевременно е надстройка и на Изпълнителната, която е в по-пряка връзка с "железарията".*

*Събитийната памет е по-високо ниво памет в разума, отколкото изпълнителната. Изпълнителна памет е например езикът - детето проговаря, средно, на възраст една година. Преди първите спомени за случки, които оставят за цял живот, то използва стотици думи и построява изречения. Тези думи и граматиките са част от "Изпълнителната памет" на "Изпълняващия вседържец".*

*Способността да говори; да съгласува движенията си, за да ходи, тича, сяда, хваща; познава посоката на звука; разпознава предмети, предполага за третото измерение от двуизмерни изображения и т.н. са част от* ***по-ниското равнище на Вседържеца на разума - Изпълняващият вседържец (ИВ).***

*Той стига до получаване на съвместими отгоре надолу версии много по-рано от Събитийния вседържец.*

*Първите спомени за събития показват приблизително кога е била завършена работеща устойчиво разновидност на* ***Развития събитиен вседържец (РСВ)*** *- обикновено за около 2-3 години след раждането на човека. По-ранното появяване показва по-ранно развитие на мозъка и е предимство за развитието на разума.*

*Събитийният вседържец е* ***по-универсален*** *от* ***Изпълнителния****, защото позволява на разума да работи с по-голям брой формати на данните, които са съчетание и разширение на форматите на Изпълнителния вседържец.*

*Събитийният вседържец* ***е по-гъвкав*** *от Изпълнителния и позволява на разума да работи с по-отвлечени свитъци (информационни обекти), по-отдалечени от вида им при въвеждането в паметта през чувствениците.*

*PCB позволява на разума* ***да извлича изпълнителна информация от събитийната****. Т.е. разумът може да тълкува (обмисля) събитийната памет* ***с много голямо закъснение във времето.*** *Разумът може да преобразува събитийната памет в изпълнителна, във формат за Изпълнителния вседържец, който може да извърши определени действия заради събитие, съхранявано в Събитийната памет.*

*Събитийната памет може да се разглежда, в по-голяма степен от изпълнителната, като хранилище за данни, от които чрез изчисления могат да се извличат последователности от действия - алгоритми.*

*Изпълнителната памет е, в по-голяма степен от събитийната, хранилище на алгоритми (казбеници), отколкото събитийната; както подсказва името ѝ, за да се извършват действия, се използва Изпълнителната операционна система.*

***Изпълнителният и Събитийният вседържатели се преплитат и не са независими един от друг.*** *Свойствата на единия са свойства, в определена степен и от определена гледна точка, и свойства на другия. Събитийната операционна система е производна на изпълнителната.*

*__ КРАЙ __*

*…*

* Изпълнителен, още *изпълняващ*.

* Запомнянето на спомени и от вчера, и от преди десетилетия, при запазването на съвместимостта, при така разпределена паметта като в невроните в живите организми, може да налага да се обновяват и донастройват заедно и едновременно общи представяния, поради зависимости, от които зависят *всички спомени.*

* Описаното надграждане на възможностите и преустройство на мозъка, при запазване на съвместимостта с вече съществуващите спомени, образувани при по-старо състояние на мозъка, е примери и вид **продължаващо обучение**, **непрекъснато обучение: continual learning, life-long learning**. Виж приложения Ирина, Лазар, Анелия и др. #Irina #Lazar, #Anelia

[Бел. 18/2/2024+] **Някои важни статии от цитираната литература:**

Bellman, R. (1957). **Dynamic programming**. Princeton University Press.

Sutton, R.S. (1984). **Temporal Credit Assignment in Reinforcement Learning**. Ph.D. diss., Dept. Of Computer and Information Science, University of Massachusetts. [lo]

Sutton, R.S. (1988). **Learning to predict the methods of temporal differences**. In Machine Learning, 3:9-44.

Sutton, R.S. (1990**). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming**. In Proceedings of the Seventh

International Conference on Machine Learning

**Бележка за превода** на заглавието, 27.8.2025 Първо преведох „replay" като **преоценка**: най-очевидната, но и парадоксално „отвличащо-конкретизираща" дума е ***„просвиране" –*** отвличаща с това, че по-общото „отвлечено" понятие се посочва чрез по-частно, с конкретен вид сетивност: звук. Може да бъде и: преразглеждане, припомняне; повторен преглед, възвръщане, превъзприемане, повторно възприемане. На 28.8 промених на *повторна обработка*. Тя може да бъде и „възприемане", дори и да е отвъд общото „съзнание" на организма или системата, на по-ниско местно ниво.

**Бележка за по-развити езици и записи на мислите с разширен усложнен естествен език в съчетание с други кодове:** Отдавна съм мислил за по-гъвкав запис и представяния, в които лесно да се изразяват многопоточни мисли, множествени заглавия, алернативни изрази и т.н. уточняващи мисълта. Този въпрос ще бъде доразвит в бъдеще – виж „Създаване на мислещи машини".

Свързани теми и основа и насока за допълнително изучаване:

## * Основни функционални мрежи в човешкия мозък и методи за откриването и изследването им[147]

## * Main functional networks in the human brain and methods for their discovery and study:

Обобщение на Тодор Арнаудов по разни източници.

Теория за интерактивната специализация (Johnson, 2011) – в началото на фунционалното си отделяне, мозъчните области имат по-общи функции, които стават по-специализирани с узряването им. Blood oxygen level dependent (BOLD) – измерване на активността на части от мозъка по кислородния метаболизъм [непряк метод и има закъснение]
fMRI – функционален ядрено-магнитен резонанс.

Някои основни анатомични и функционални архитектурни структури:
Limbic System (Лимбична система): Amygdala, Hippocampus, Cingulate Cortex, Parahippocampal Gyrus, Hypothalamus, Mammillary Bodies, Nucleus Accumbens
Paralimbic System (Паралимбична): Orbitofrontal Cortex (OFC), Insula, Anterior Temporal Lobes

**Важни мрежи** – задействащи се заедно области

**RSNs** – Resting State Networks: Мрежи в състояние на покой (МСП)
Йерархична (RSN), включва подмрежи: DMN, FPN, SN, Sensorimotor, Visual, Auditory, somatomotor, ….

---

[147] По: Dynamic Default Mode Network across Different Brain States, Pan Lin et al., Nature, 6.4.2017, Sci. Rep. 7, 46088; doi: 10.1038/srep46088 (2017)
Structural connectivity of the precuneus and its relation to resting-state networks, Atsushi Yamaguchi**,** Tatsuya Jitsuishi, 30.12.2023
Progress in Brain Research, Volume 254, 2020, Pages pp. 49-70
Chapter 3: **Early maturation of the social brain: How brain development provides a platform for the acquisition of social-cognitive competence**, Judit Ciarrusta**,** ***Ralica Dimitrova***, Grainne McAlonan https://doi.org/10.1016/bs.pbr.2020.05.004
https://www.sciencedirect.com/science/article/abs/pii/S0079612320300467
https://en.wikipedia.org/wiki/Salience_network Connectome Guide, Omniscient Neurotechnology: * https://www.o8t.com/connectomeguide/network/limbic-paralimbic-system * https://quicktome.o8t.com/guide/network/ventral-attention-network * https://www.o8t.com/connectomeguide/network/multiple-demand * https://www.researchgate.net/publication/373806475_The_Extended_Free_Energy_Principle_FEP_Model_of_General_Intelligence etc..**Ralica Dimitrova** – a Bulgarian researcher.

**DMN** – Default Mode Network: Мрежа в „изходно състояние“ (МИС); още TNN – Task-Negative Networks
**Frontoparietal network** (FP) (челно-теменна)
**Salience network** (SN) – мрежа за оценка на значимостта на стимулите (Insula, **ACC** – Anterior Cingulate, …; социални взаимодействия; също:
**Cingulo-opercular network** (**CO),** Cingulae Cortex
**Frontoparietal Network** (FPN) (Fronto-parietal control network (FPCN), FPCN-A, FPCN-B)
**Dorsal Attention Network** (DAN)
**Ventral Attention Network** (VAN)
**Central Executive Network** (CEN)
**Sensorimotor Network** (SMN)
**Limbic Network**, **Paralimbic-Limbic**, **Medial Temporal Network**
**Accessory Language Network** (ALN)
**Multiple Demand Network** (MD) и **Multiple Demand Extended Network** – висша умствена дейност, флуидна интелигентност, учене на **нови думи**; разпознаване на **новост**; оценка на сетивна сложност
**Visual, Auditory**

*Топологията* на мрежата в изходно състояние **МИС** (**DMN**) се променя по време на работа – кои области от мозъка са активни едновременно. Освен топологията се променят и *функциите,* защото областите изпълняват различни функции в различно време, т.е. осъществяването на механизма им на работа е разпределено, а не съсредоточено, уместЕно, локализирано; виж работите на споменатия учен Михаил Рабинович и понятието *Winnerless Competition.*

„***МИС/DMN** се състои от функционално и структурно свързани мозъчни области, които обикновено **затихват** по време на изпълнение на задачи, **насочени навън**, включващи вниманието* [тогава се включва **DAN**]*; и се усилват с висок мозъчен кръвоток и консумация на кислород по време на състояние на покой.*“

Затова възбуждането на тези области се смята, че се отнася до преживяване на мисли, свързани със самия себе си и автобиографичната памет. ... Стационарна *функционална* свързаност (**FC – functional connectivity**) – статистическа връзка на кръвоток наблюдаван с fMRI и пр. скенери... (**сравни** с *анатомична структурна* свързаност: **SC**[148] и *причинно-следствена* свързаност); DMN се изключва и по време на анестезия и при пациенти във вегетативно състояние, както и по време на различни фази на сън.

**Динамични шевици**[149] на функционална свързаност (виж М.Рабинович); състояния на мозъка: при решаване познавателна задача; преди задача, по време на задача, след задача; графи описващи свързаните зони, динамична корелация; мрежов анализ; средна топологична мярка (свързаност между определени възли), местна ефективност, коефициенти на групиране (clustering

[148] https://www.sciencedirect.com/science/article/pii/S0168010223002213
[149] Patterns; юнашко наречие

coefficients), степени в топологични мерки ... *Wilcoxon rank sum test; Kolmogorov–Smirnov test, k-means;*

**MNI координати** (X,Y,Z) (-20, 15, 40) -20: ляво, 15 пред, 40 - над хоризонталната равнина за fMRI, морфометрия (Voxel-based morphometry) за сравнение на мозъчни структури между различни индивиди и групиране (MNI - *Montreal Neurological Institute;* MNI152, MNI305 (по-старо); ICBM152 ... числото - брой мозъци, усреднени за получаване на картата). **Техники за анализ:** SPM (Statistical Parametric Mapping), FSL (FMRIB Software Library), AFNI (Analysis of Functional NeuroImages), BrainVoyager, Mango, FreeSurfer,MRIcron[150]. След активност топологията се променя според дейността, за да подсили новополучената информация. **Подмрежи** subnetworks: … central executive network (**CEN**) and salience network (**SN**) … почивка и съсредоточаване. Преустройство и на глобалната, и на локалната топология. След разрешаване на умствената задача, **МИС/DMN** се връща в предходното състояние (атрактор, динамични системи).

**Зони от МИС:**

* PFC – prefrontal cortex (челни дялове на неокортекса)
* PCC - posterior cingulate cortex - разпределителен център
* mPFC - самоанализ (рефлексия, self-reflection), взаимодействия с другите*, планиране (социално познание)
* Angular Gyrus - смисъл/значение (семантика), спомняне/припомняне, пространствена ориентация, писане, четене, математика; при увреждане: agraphia, acalculia, … нарушения при дислексия – трудности с писането, броенето, аритметиката, математиката
* Inferior Parietal Lobule IPL - самосъзнание, автобиографична памет, приемане на чужди гледни точки
* Lateral Temporal Cortex (LTC):…
* Hippocampus – образуване и консолидиране на спомени Lateral temporal cortex
* Precuneus (участва и в други мрежи)

**Дейности, свързани с активност на DMN:** самоизследване, самоанализ, разсъждения за себе си, собствения опит и бъдещи планове; мечтание; изследване, размишления? без външен сензорен фокус (Т.А.: без чуждо/външно управление?; самостоятелност; без да се *приема, че има такова външно въздействие*, например взаимодействие с друг човек; “вглъбено“ състояние и „изключване“ от сетивния поток или потискането му). Познание за общуването с другите: разбиране на умствените състояния и гледна точка на другите

* Припомняне, спомняне на събития и преживявания, фантазиране и мечти, мисли за самия себе си.
* Преживяване и регулиране на емоциите, самоконтрол

[150] https://www.nitrc.org/projects/mricron

**Сравни с: Dorsal Attention Network (DAN) и Task Positive Network: TPN**
**DAN:** При съсредоточаване върху външен стимул и заемане с целенасочено поведение. Мрежа за ***"свършване на работата".*** “Go/Nogo” – вземане на решение дали да се изпълни операция, например очакване да се появи определен дразнител, след което да се извърши действие – да се натисне бутон, да се дръпне спусък, да се натисне спирачка и пр. Насочено внимание/ избирателно внимание, управление на вниманието отгоре-надолу, наблюдение на обстановката при търсене или очакване да се забележи определен стимул; зрително-пространствено възприятие, решаване на различни познавателни задачи: четене, писане, математически, програмиране и пр.; изпълнителни функции – поддържане на вниманието върху избрана текуща дейност, задача (goal-directed, selective, visuospatial; focusing on specific stimuli; monitoring the external environment; executive control).
**Активни зони:**
*** Frontal Eye Fields (FEF):** управление на движенията на очите и насочване на вниманието към особени места в пространството.
*** Intraparietal Sulcus (IPS):** внимание свързано с пространство, зрителна обработка, движения
*** Superior Parietal Lobule (SPL):** пространствено възприятие, превключване на вниманието, работна памет
**Middle Temporal Area (MT):** движение
**Visual Area V3A**: абстрактна зрителна обработка и внимание
**Middle frontal gyrus**: достигане и хващане
+
**Task Positive Network: TPN –** по-обща категория от мрежи, свързани с целенасочено поведение, работната памет, решаване на задачи и (насочено) внимание. DAN е ключва част от TPN. FPN също е част от TPN. При индивиди със симптоми на „недостиг на вниманието“, ADD/ADHD, се забелязва намалена свързаност между дорзалната и вентралната мрежа на вниманието (DFN и SN).

https://en.wikipedia.org/wiki/Dorsal_attention_network
https://en.wikipedia.org/wiki/Salience_network
https://en.wikipedia.org/wiki/Go/no-go

*** Exploring the Role of the Dorsal Attention Network in Sustained Attention With rTMS**, Hidefusa Okabe, 11.2016, Master's thesis, Harvard Extension School; Clinical Psychology;– **Изследване на ролята на DAN в устойчивото внимание чрез изследвания чрез rTMS.** с. 4. ***„Дълготрайното, или устойчиво внимание,*** *което изисква полагане на усилия по време на трудни познавателни задачи, се свързва със стабилно възбуждане в мозъчните полета, свързани с познавателен контрол, наричани събирателно дорзална мрежа на вниманието (DAN). От друга страна, периодите на най-добро устойчиво внимание се свързват с относително по-слабо възбуждане в DAN,*

*отколкото периодите на необходимост от прилагане на волево усилие, за задържане на вниманието. Това може да е така, защото способността за устойчиво внимание може да зависи повече от мрежи на мозъка, свързани с автоматизъм на задачите като DMN. Друга възможност е най-доброто устойчиво вниманието да задейства DAN по ефективен начин и заради това да се регистрира по-малко общо възбуждане. Работата разглежда и двете хипотези чрез корова магнитна стимулация на челните зрителни полета.* (***Transcranial magnetic stimulation (rTMS) to the frontal eye fields (FEF) –*** Неинвазивно стимулиране на мозъчната кора чрез магнитно поле.
Виж с. 18(10) – карта на възбудените полета при DMN и DAN. Виж също от статиите в Уикипедия. Имайте предвид обаче, че мозъците на са произведени по един и същи калъп и точното разположение и местата на най-високо възбуждане при различни индивиди няма да съвпада напълно.Съществува и „невроразнообразие“, „различна свързаност“, “neurodiversity“. Някои мозъци са „свързани по различен начин“ от средното – “wired differently”.

*** Task Positive Network:** https://www.sciencedirect.com/topics/medicine-and-dentistry/task-positive-network … „Task-negative and task-positive networks” Chapter**, Neuroanatomical Basis of Consciousness**, 2016, The Neurology of Conciousness (Second Edition), Hal Blumenfeld**, Task-Positive and Task-Negative Networks**

**…**

*** Flexible adaptation of task-positive brain networks predicts efficiency of evidence accumulation,** Alexander Weigard et al., 2.7.2024 https://www.nature.com/articles/s42003-024-06506-w – *“Efficiency of evidence accumulation (EEA), **an individual’s ability to selectively gather goal-relevant information to make adaptive choices**, is thought to be a key neurocomputational mechanism associated with cognitive functioning and transdiagnostic risk for psychopathology.”* – compare to **Credit Assignment** in RL and discovery/inference/search for the causes, causal forces, causal factorization; features for building predictive world models of the hierarhical universal simulators of virtual universes (Theory of Universe and Mind).

…

**Salience Network (SN): “***it has been implicated in the **detection** and **integration** of **emotional and sensory stimuli**[9] as well as in **modulating the switch** between the **internally directed cognition of the default mode network** and the **externally directed cognition** of the central executive network.” (Wikipedia)*

**Todor: ?** Part the systems connecting the cognitive and sensual reward systems? (TUM) [22.10.2025]

* https://en.wikipedia.org/wiki/Granger_causality **– Granger causality test –** time series

*** What is sustained attention? -** https://www.cognifit.com/science/focus

…

*** Интересни случаи на пациенти родени без челни дялове на кората и без малък мозък:**

* https://www.rnz.co.nz/news/national/573888/major-theories-of-consciousness-may-have-been-focusing-on-the-wrong-part-of-the-brain

*** Early bilateral and massive compromise of the frontal lobes,** Agustín Ibáñez a,b,c,d,e,⁎, Máximo Zimerman et al., 2018
https://pmc.ncbi.nlm.nih.gov/articles/PMC5964834/
Пациентът проговаря на 23 месеца, говорът е по-развит и са запазени основни способности за общуване и взаимоотношения, сетивни способности и движения, повече отколкото може да се очаква при почти липса на челните дялове на неокортекса – например двустранно липсващи зони на Брока. Запазена била само малка част от вентралните челни дялове. Предполага се, че базалният ганглий и др. ги заместват. Неразвитост на абстрактно мислене, импулсивност и др.

*** A new case of complete primary cerebellar agenesis: clinical and imaging findings in a living patient, Feng Yu 1,✉, Qing-jun Jiang 2, Xi-yan Sun 1, Rong-wei Zhang 1, 2014** https://pmc.ncbi.nlm.nih.gov/articles/PMC4614135/
Без малък мозък по рождение при 24-годишна жена; липсата е оказала влияния и върху продълговатия мозък. Пациентката имала лека умствена изостаналост, леки до умерени двигателни нарушения атаксия и дизартрия („завален“ говор като след инсулт), но по-малко от очакваните при пълна липса на малък мозък. Случаят подкрепя теориите за пластичността на друга двигателна система, особено при ранна загуба на малкия мозък.

*** Ralica Dimitrova – Ралица Димитрова – вероятно българка**
https://scholar.google.com/citations?hl=en&user=Z60XyU0AAAAJ&view_op=list_works&sortby=pubdate **–** Centre for the Developing Brain, King's College London: **Neonatal Neuroimaging, Preterm birth**, **Neurodevelopment …**

*** Изследвания на мозъка, мозъчното развитие и нервната дейност при новородени**

Разлики в мозъчното развитие между недоносени бебета и родени в термина. Измервания чрез ядрено-магнитен резонанс на новородени, които предсказват степента на нервното развитие на 18 месеца. Динамична функционана нервна

свързаност в мозъка на новороденото и влиянието на преждевременното раждане върху развитието на нервната система. Ранно узряване на насочената към взаимодействия възможности на мозъка: как развитието на мозъка осигурява основа за придобиването на обществено-познавателни умения. Преждевременното раждане променя развитието на коровите микроструктури и морфологията при достигане на равнозначна на термина възраст. Морфологията на кората при раждането отразява времепространствените шевици на изразяването на гените в мозъка на човешки плод. Структурни и функционални асиметрии в мозъка на новороденото. …

*** Neonatal multi-modal cortical profiles predict 18-month developmental outcomes**, D Fenchel**, R Dimitrova**, EC Robinson, D Batalle, A Chew, S Falconer, ... Developmental Cognitive Neuroscience 54, 101103, 17, 2022 https://www.sciencedirect.com/science/article/pii/S1878929322000470

* **Neonatal brain dynamic functional connectivity in term and preterm infants and its association with early childhood neurodevelopment**, LGS França, J Ciarrusta, O Gale-Grant, S Fenn-Moltu, S Fitzgibbon, ... Nature Communications 15 (1), 16, 39, 2024 https://www.nature.com/articles/s41467-023-44050-z – *"dynamic functional connectivity (dFC) measures the constant neural adjustments needed to control different brain states, adapt to transient situations, and integrate information".. "autism spectrum disorder (ASD) Individuals with ASD … switch between different connectivity profiles more directly, whereas typically developing individuals switch between those same brain states via an intermediate connectivity profil"* ***Neonatal brain states*** *averaged the fMRI timeseries into 90 cortical and subcortical regions defined by the Anatomical Automated Labels (AAL) atlas adapted to the neonatal brain …*

* **Neonatal brain dynamic functional connectivity: impact of preterm birth and association with early childhood neurodevelopment,** LGS França, J Ciarrusta, O Gale-Grant, S Fenn-Moltu, S Fitzgibbon, ... bioRxiv, 2022.11. 16.516610, 9, 2022

* Preterm birth alters the development of cortical microstructure and morphology at term-equivalent age, **R Dimitrova**, M Pietsch, et al., NeuroImage 243, 118488, 93, 2021

* Early maturation of the social brain: How brain development provides a platform for the acquisition of social-cognitive competence, J Ciarrusta, **R Dimitrova,** G McAlonan … Progress in brain research 254, 49-70, 16 2020

* Preterm birth alters the development of cortical microstructure and morphology at term-equivalent age, **R Dimitrova**, M Pietsch, J Ciarrusta, et al. NeuroImage 243, …2021

*** Cortical morphology at birth reflects spatiotemporal patterns of gene**

**expression in the fetal human brain**, G Ball, J Seidlitz, J O'Muircheartaigh, **R Dimitrova**, D Fenchel, ... PLoS biology 18 (11), e3000976, 54, 2020
https://www.researchgate.net/profile/Jakob-Seidlitz/publication/338902179_Cortical_morphology_at_birth_reflects_spatio-temporal_patterns_of_gene_expression_in_the_fetal_brain/links/5e31f9a4299bf1cdb9fc8dd8/Cortical-morphology-at-birth-reflects-spatio-temporal-patterns-of-gene-expression-in-the-fetal-brain.pdf

*** Development of microstructural and morphological cortical profiles in the neonatal brain**, D Fenchel, **R Dimitrova,** J Seidlitz, EC Robinson, D Batalle, J Hutter, ..., Cerebral Cortex 30 (11), 5767-5779, 67, 2020
https://pubmed.ncbi.nlm.nih.gov/32537627/
https://www.researchgate.net/publication/342244223_Development_of_Microstructural_and_Morphological_Cortical_Profiles_in_the_Neonatal_Brain (full text)
*„Posterior regions became more morphometrically similar with increasing age, while peri-cingulate and medial temporal regions became more dissimilar.“; perinatal period (37-44 weeks); p.3. "Image Acquisition .. 3T Philips Achieva scanner without sedation, using a dedicated 32-channel neonatal headcoil system" .. "Although the majority of neurons are at their terminal location, synaptogenesis and myelination are ongoing, and (limited) migration of interneurons continues .. In the cortex, these processes are coordinated with sensory cortex and pathways developing earliest, prefrontal and association areas later.." .. A .. hypothesis of cortical connectivity .. "similar prefers similar", meaning areas with similar cytoarchitecture preferentially connect; inter-regional similarity.*

*** Heterogeneity in brain microstructural development following preterm birth**, **R Dimitrova,** M Pietsch, D Christiaens, J Ciarrusta, T Wolfers, D Batalle, ..., Cerebral Cortex 30 (9), 4800-4810, 77, 2020

* **Emerging functional connectivity differences in newborn infants vulnerable to autism spectrum disorders**, J Ciarrusta, **R Dimitrova**, D Batalle, J O'Muircheartaigh, L Cordero-Grande, ..., Translational psychiatry 10 (1), 131, 68, 2020

*** Structural and functional asymmetry of the neonatal cerebral cortex**, LZJ Williams, SP Fitzgibbon, J Bozek, AM Winkler**, R Dimitrova,** T Poppe, ..., Nature human behaviour 7 (6), 942-955, 48, 2023 – Developing Human Connectome Project, Human Connectome Project.
https://www.biorxiv.org/content/biorxiv/early/2022/07/22/2021.10.13.464206.full.pdf
- Differences between right and left hemispheres: surface area, cortical thickness,

depth of the superior temporal sulcus (right deeper than left) etc. p.2: Healthy Term Neonatal's assymetry: left cortex larger surface areas than right of many cortical gyri (*supramarginal, medial precentral, postcentral, posterior cingulate, caudal anterior cingulate etc.…")*; deeper right superior temporal sulcus etc.

**@Вси:** Разгледай и ОбОбр всички структури на мозъка, връзки и функции. Напр. от https://en.wikipedia.org/wiki/Temporoparietal_junction – "Anatomy of the cerebral cortex of the human brain": всички връзки.

*** Synchrony and subjective experience: the neural correlates of the stream of consciousness**, Trends in Cognitive Sciences,15.5.2025, Matthew D. Lieberman https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(25)00086-5 – perceived facts (p-interpretations*), integrative, synchrony effects, neural synchrony, reflective conscious process, CEEing model - *our intuitively sensible experiences that exist prior to reflective processes like thought or introspection. (Glossary: coherent effortless experiences that have the intuitive givenness of visual seeing …)* **Gestalt cortex –** generating coherent experiences from disparate elements; (Fig. I) - *immediately adjacent to three major sensory cortices (visual, auditory, somatosensory) – multisensory integration*; **the defining features of subjective experience** *(i.e., effortless coherence and meaningfulness that seamlessly evolves over time), the mental processes that support it (i.e., the integration of sensory and non-sensory inputs associated with idiosyncratic subjective experiences), and the neural correlates of these processes as they unfold over time (**i.e., differential synchrony in gestalt cortex and PMC** (posterior medial cortex)).* ***Interpersonal*** neural synchrony – *the correspondence between brain responses in two or more people over time*; NCC - neural correlates of consciousness; Differential synchrony; *Damage to gestalt cortex leaves patients with simultagnosia* – being *unable to bring intuitive coherence to elements in an object or scene.* Compare: **Global neural workspace theory** (GNWT); **Higher-order thought theory** (HOT); **Recurrent processing theory** (RPT). Possibly each of them addresses *different aspects of consciousness.* (Fig.2: RPT – minimal phenomenology; CEEing – prereflective coherent experience; GNWT/HOT – Reflective, Working memory, Meta-cognition, self-report; PMC: Non-sensory integration (episodic memory, motivation/schemas, attitudes/expectations).
PMC & MPFC (medial prefrontal cortex) – *more sensitive to the meaning and significance of inputs, sometimes aggregating over 10–30 s or longer. .. narrative video clips .. in neural synchrony studies .. reveal more effortless pre-reflective processing than effortful reflective processing.* See the **Glossary** after the references**: … r-interpretation (**reflective, compare with p-); stream of consciousness, *subjective construal – understanding the meaning and significance of events and situations;* non-sensory input, multisensory integration, hyperscanning – multiple brains who are interacting; givenness.

***** Topographic mapping of a hierarchy of temporal receptive windows using a narrated story**,** *J. Neurosci.* 2011; **31**:2906-2915
* Yeshurun, Y. ... The default mode network: where the idiosyncratic self meets the shared social world, Nat. Rev. Neurosci. 2021; 22:181-192

* **Intrinsic neural timescales: temporal integration and segregation**, Wolff, A. ...Trends Cogn. Sci. 2022; 26:159-173
https://www.cell.com/trends/cognitive-sciences/abstract/S1364-6613(21)00292-8 – *a hierarchy of intrinsic neural timescales (INT) with a shorter duration in unimodal regions (e.g., visual cortex and auditory cortex) and a longer duration in transmodal regions (e.g., default mode network).*

**Tosh:** Theoretically predicted/defined in TUM about Mind and Universe.

*** Intrinsic neural timescales mediate the cognitive bias of self – Temporal integration in default-modenetwork,** Angelika Wolman, Yasir Çatal, Annemarie Wolff et al.15.6.2022
https://www.researchgate.net/publication/361353859_Intrinsic_neural_timescales_mediate_the_cognitive_bias_of_self_-_Temporal_integration_in_default-mode_network
Intrinsic neural timescales = INT; Temporal integration in default-mode network;
Signal Detection Theory (SDT); Criterion C and Sensitivity index d': decision making and the brain processes; keypress related to self, self-specific – self-recognition (photographs, images; morphed to different degrees); "*Our perceptions and decisions are not always objectively correct as they are featured by a bias related to our self. What are the behavioral, neural, and computational mechanisms of such cognitive bias?"**

**Tosh:** In TUM, the biases are the currently active, selected, "ruling" ensembles of causality-control units of different kind at different scales, with different preferences etc., which have the control of the actuators. See "Analysis of the meaning of a sentence…", 2004. See also the bistable states, the section for M.Rabinovich's work.

*** Signal Detection Theory:** https://en.wikipedia.org/wiki/Detection_theory – *a means to measure the ability to differentiate between information-bearing patterns (stimulus in living organisms, signal in machines) and random*

*patterns that distract from the information (called noise, consisting of background stimuli and random activity of the detection machine and of the nervous system of the operator).*
* (Cognitive) Bias …
* Compressed_sensing * https://www.keiseruniversity.edu/signal-detection-theory/

* Salomon, G. **Television is 'easy' and print is 'tough': the differential investment of mental effort in learning as a function of perceptions and attributions** *J. Educ. Psychol.* 1984; 76:647 – **AIME, amount of invested mental effort**

* Cohen, A. A., and Salomon, G. (1979). Children's literate television viewing:surprises and possible explanations. J. Commun. 29, 156–163

→ * **Television Is Still "Easy" and Print Is Still "Tough"? More Than 30 Years of Research on the Amount of Invested Mental Effort,** July 2018, Frontiers in Psychology 9, DOI: 10.3389/fpsyg.2018.01098, LicenseCC BY 4.0, Frank Schwab, Christine Hennighausen, Dorothea Adler, Astrid Carolus
*
https://www.researchgate.net/publication/326153903_Television_Is_Still_Easy_and_Print_Is_Still_Tough_More_Than_30_Years_of_Research_on_the_Amount_of_Invested_Mental_Effort

* **Supplementary motor area as key structure for domain-general sequence processing: A unified account**, Giorgia Cona, Carlo Semenza, 2017 https://www.sciencedirect.com/science/article/abs/pii/S0149763416303475?via%3Dihub

Допълнителната моторна зона SMA има общо предназначение за двигателни последователности: за обединението на последователни елементи в представяния от по-висок порядък, без значение от вида на поредиците: двигателни, времеви, музикални, пространствени, числови, езикови, работна памет и пр. Обзорът подчертава, че SMA участва в разнообразни познавателни области, които макар да са различни, в същността си споделят обработка на поредици. (Но не е ясно дали тази област, самостоятелно, е достатъчна за обработката на поредици.) Виж също **Парашкев Начев** и др., 2008[151] Nachev et al., 2008.

[151] Functional role of the supplementary and PRE-supplementary motor areas, Parashkev Nachev,C.Kennard,M.Husain, November, 2008,Nature Reviews Neuroscience 9(11):856-69 https://www.researchgate.net/publication/23307629_Functional_role_of_the_supplementary_and_PRE-supplementary_motor_areas

**П.Начев** е български изследовател, който към 10.2024 г. работи в „High Dimensional Neurology Group, UCL Queen Square Institute of Neurology, University College London“, „възстановяване на мозъка и рехабилитация“. Участва например в разработката на предсказващи модели на измененията в мозъка при увреждания като инсулт и други заболявания, така че да се избере най-доброто лечение; друга интересна скорошна работа е за състоянията на **афантазия и хиперфантазия** – липса на „вътрешно око“ и въображение в образи, или пък свръхвъображение с „фотографски“ качества. Множество публикации[152]:
https://scholar.google.co.uk/citations?view_op=view_citation&hl=en&user=Cf10kCIAAAAJ&sortby=pubdate&citation_for_view=Cf10kCIAAAAJ:HIFyuExEbWQC

* **Realistic morphology-preserving generative modelling of the brain,** Petru-Daniel Tudosiu … Parashkev Nachev, … et al.**,** 2024/7
https://www.nature.com/articles/s42256-024-00864-0

*** The minimal computational substrate of fluid intelligence,** Amy PK Nelson, … Parashkev Nachev et al.**,** 10/2024
https://www.sciencedirect.com/science/article/pii/S0010945224002053

Невронна мрежа, която се учи да разпознава съкратено множество на матриците на Рейвън (Raven's Advanced Progressive Matrices (RAPM)) като мярка за флуидна интелигентност от тестове по психология. Моделът LaMa (част от семейство напълно конволюционни модели с бързи конволюции на Фурие: FFC NN, ResNet; със сетивно поле, обхващащо целия образ; 51 M параметъра за запълване на изображения; да не се бърка с езиковите модели Llama на Meta), *обучен за допълване на частично маскирани снимки на естествени сцени* [4.5 милиона от набора “Places”]*, постига представителен резултат за човешко ниво от първо показване (a prima vista), без никакви специализирани настройки или обучение.* [8 от 12 верни] *Сравнен с резултатите на кохорта от участници – здрави и с фокални лезии на мозъка* [локализирани увреждания в определени области след инсулт и др.]*, моделът показва човекоподобни отклонения в трудността на разпознаването и допуска грешки присъщи за увреждания на десните челни*

---

[152] * **Using MR Physics for Domain Generalisation and Super-Resolution**, Pedro Borges, Virginia Fernandez, Petru Daniel Tudosiu, Parashkev Nachev, Sebastien Ourselin, M Jorge Cardoso, 10/2024/
* **Phantasia, aphantasia, and hyperphantasia: Empirical data and conceptual considerations**, AJ Larner, AP Leff, PC Nachev, 2024
https://www.sciencedirect.com/science/article/abs/pii/S01497634240028847/18
https://digitalk.bg/ai/2024/09/16/4676481_kak_generativniiat_ai_promenia_mozuchnata_diagnostika/
* https://www.nachev.org/papers/
https://scholar.google.co.uk/citations?hl=en&user=Cf10kCIAAAAJ&view_op=list_works&sortby=pubdate

*дялове с влошаване на способността му да обединява всеобщи пространствени шевици. Краткото обучение и ограничен капацитет подсказват, че тестове от тип "матрица" са податливи на изчислително прости решения, за които не е задължително да се основават на носителите на способностите за разсъждаване в мозъка."*

*** Тош:** Виж също *„лакунарни инсулти"* – подкорови инсулти с ограничени размери, до 15-20 мм в диаметър. Често не се отчитат като инсулти от пациентите и се откриват със закъснение, при извършване на компютърна томография или ядрено-магнитен резонанс, след като месеци или няколко години по-късно се получи тежък инсулт на кората и се открият множество "лакунарни хиподенсни лезии" и „субкортикална деменция", т.е. атрофия на подкорови мозъчни структури като бялото вещество и др. „Хиподенсни" означава области на изображението, които са загубили плътност – мозъчно вещество, – и са станали по-прозрачни, черни на снимка от томограф: от „хипо" – под, по-малко; денсни – плътни от „density".

Възрастни пациенти *падат* многократно и описват причината като *причерняване,* като си мислят, че им е паднало кръвното, спънали са се – защото са трудно подвижни с напреднали ставни нарушения като артроза, гонартроза и т.н. и понеже симптомите при лакунарен подкоров инсулт не са като при масивен коров инсулт с моменталните двигателни и говорни нарушения.

Не пренебрегвайте прегледа след подобни инциденти с ваши близки и познати – може да предотвратите или да отложите задаващ се масивен инсулт. https://en.wikipedia.org/wiki/Lacunar_stroke

*** Resolution-Robust Large Mask Inpainting With Fourier Convolutions,** Roman Suvorov et al. https://openaccess.thecvf.com/content/WACV2022/html/Suvorov_Resolution-Robust_Large_Mask_Inpainting_With_Fourier_Convolutions_WACV_2022_paper.html

Сравни с теста „ARC" на *Франсоа Шоле* (Francois Chollet) https://lab42.global/arc/ който претендира, че е за УИР. Виж също нов претендент „**SuperARC: An Agnostic Test for Narrow, General, and Super Intelligence Based On the Principles of Causal Recursive Compression and Algorithmic Probability,** Alberto Hernández-Espinosa et al. https://arxiv.org/pdf/2503.16743 https://github.com/AlgoDynLab/SuperintelligenceTest – виж бележки за него по-горе в приложение „Листове" и повече в приложение **Алгоритмична Вероятност** (Algorithmic Complexity, #complexity).

Виж също технологията от френския национален институт за информатика и технологии INRIA за проследяване и моделиране на изменението на мозъка при стареене на здрави хора и при Алцхаймер[153]:

[153] Xavier Pennec, virtuose de la statistique géométrique au service de la santé, 15.10.2024, https://www.inria.fr/fr/xavier-pennec-statistique-geometrique-sante-medecine-personnalisee

https://www.inria.fr/fr/xavier-pennec-statistique-geometrique-sante-medecine-personnalisee

Те прилагат *геодезична регресия*[154], откривайки *геодезични криви*. Геодезичната крива е най-краткият път в неевклидово пространство, например повърхността на сфера. Геодезичната регресия е оптимизационен метод[155], разширение на линейната регресия за „изкривени“ пространства като Риманови многообразия (Riemann manifolds) с дадена метрика за разход / цена на движението, например сбор на средно-квадратичната грешка на разстоянието на кривата спрямо координатите на опитните данни във въпросното изкривено пространство, т.е. не SQRT($dx^2$+ $dy^2$), както в обикновено линейно евклидово пространство. Методът е от семейството на *енергийните,* виж напр. Льокан, JEPA; сравни с *вероятностни*.

*** Helpless infants are learning a foundation model**. Rhodri Cusack, Marc'Aurelio Ranzato, Christine J. Charvet, 8.2024, Trends in Cognitive Science https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(24)00114-1
Статия "Мнение"/Opinion. Противоречия в теориите за съзряването на човешкия мозък спрямо другите бозайници – смята се, че човешките бебета се раждат безпомощни, защото мозъкът им е недоразвит, защото е с прекалено голям размер, съответно глава, и раждането е трудно и пр. заради тесния таз, чиито размери се ограничават от изправената походка. Други животни веднага или почти веднага са готови да взаимодействат със света, виждат, ходят. Според неврологични маркери обаче, мозъкът на бебето *не е толкова незрял* спрямо други видове и сетивните мрежи вече работят, но мозъкът се обучава подобно на основните модели в изкуствените НМ в режим на *предобучение*, като натрупва всестранни данни чрез самообучение, самонасочване: self-supervised learning, за да построи **общи представяния**, които след това да улеснят целенасоченото обучение с учител и следващи нови задачи, представяния да се научават и разрешават по-бързо, по-ефективно и управлението да е по-успешно, както показват компютърните невронни мрежи и основните модели във всякакви модалности.

След обучението на **първичния основен модел** се превключва към други системи за учене с учител; предполагаеми: предните или теменните дялове за класификация (anterior or medial temporal lobes); и челните дялове за действия (frontal lobes). При животните и човека по-дългият период на предобучение обикновено се свързва с по-висока интелигентност след това.

---

[154] Geodesic Regression: Machine Learning meets Riemannian Geometry, Paribesh Regmi, 11.2020 https://towardsdatascience.com/geodesic-regression-d0334de2d9d8
Geodesic regression, Frank Hansen, https://arxiv.org/abs/2005.01326
Robust Geodesic Regression, Ha-Young Shin, Hee-Seok Oh, 2020-2022 https://arxiv.org/abs/2007.04518

[155] Повечето от техниките в ИИ са оптимизационни, намиране на най-малкото, най-голямото, вместване в зададени граници (динамично оптимиране) и пр. от/измежду възможни решения, пътища в граф, набори от съчетания на предмети/стойности и пр.

Животните също имат такъв период, но по-кратък, не само бозайниците – например гарваните, които първият месец са в гнездото и се хранят от двамата родители.

***Предложени стратегии за самообучение:*** *за езикови записи: предсказване на „маскирани“ части от говора или текста, следващата дума или вероятностното разпределение на други думи около дадената. За зрението – предвиждане на покрити части от изображението въз основа на останалата част; предсказване на кадър от видео от предишния, групиране по зрително подобие или предвиждане на съдържанието на един цветен канал от другите (синя от червено и пр.); при многосетивни множества от данни – предсказване на надпис от видео. … Общите архитектури като преобразителите (transformers) и общите цели като предсказване на следващите елементи във входен сетивен поток водят до* ***широки обобщения в прилагането на тези мрежи за решаване на конкретни задачи****, ако невронната мрежа е голяма и е изложена на достатъчно голям набор от наблюдения. Това предоставя богато множество от хипотези за целевите функции, които детето може да използва за самообучение.*

**Видове технологии за самообучение на общи представяния:**

**1. Автоенкодери** – кодират в нискоразмерни представяния (embedding) на входни данни и след това при декодирането пресъздават правдоподобно оригинала. Представянето образува информационно „тясно място“, въобразно стесняване (information bottleneck), което изисква входните данни да се сгъстят, компресират в компактно представяния според статистиката на средата. **Варианти:** автоенкодери с маски, разредени/вариационни автоенкодери …

**2. Пораждащи модели:** за текст, образи и др. за даден контекст, обслов; използват декодиращ блок, което може да служи за предвиждане – на следващата дума в изречение, на изображение според дадена текстова подкана и пр. **3. Контрастно обучение:** представянето на двойки от свързани стимули (напр. един и същи предмет от различни гледни точки или от различни сетива) се кодират като по-подобни, докато несвързаните, напр. различни обекти – по-различни. … *Предложение* да се разбере как вродените структури на мозъка оформят ученето, напр. като първична организация – чрез техники за мозъчна образна диагностика. Разбиране на **ролята на просвирването и сънищата (replay and dreaming),** напр. за да попречи на новите преживявания да се презапишат и да изтрият съществуващите знания. …

**Тош:** Основните обобщения за общите методи и предвиждането на следващото чрез части от предното са направени в ТРИВ. Предобучението и насочените режими са свързани също със съзряването на хипокампа, или в ТРИВ теоретично с двата „вседържеца“ на мозъка: изпълнителен и събитиен. Събитийният може трайно да съхранява сетивно-моторни записи за събития в многосетивен „формат“ и биографични спомени и записите са съвместими със следващите „версии“ на системите и мозъка, който е едва в началото на бурното си развитие – около 2-3-годишна възраст. * Виж по-горе: **Replay in Brains and Machines**

* **Where do you know what you know? The representation of semantic knowledge in the human brain,** Karalyn Patterson, Peter J. Nestor & Timothy T. Rogers, *Nat. Rev. Neurosci.* 2007; **8**:976-987, https://www.nature.com/articles/nrn2277 https://wiredbrains.org/wp-content/uploads/2023/07/Patterson-2007-Nature-Reviews-Neuroscience_1.pdf - an **amodal** hub in the Anterior Temporal Lobe (ATP); when damaged bilaterlally – **semantic dementia (SD):** *"semantic degradation across **all modalities** and all types of conceptual knowledge; Semantic memory: … general conceptual knowledge about objects and events, including knowledge about their characteristic properties and behaviours, as well as knowledge about the words we use to name and describe objects and events in speech. Whereas episodic memory encompasses memory for specific episodes or situations in one's life, semantic memory encompasses factual knowledge divorced from any specific situational context: "a scallop is an edible sea creature" (semantic) as opposed to "I ate scallops for supper last night" (episodic). P.3 The most prominent cognitive deficit in Alzheimer's disease. is an impairment of episodic (autobiographical) memory: in particular, the ability to learn new information is progressively abolished. The original hypothesis, which attributed this phenomenon specifically to degeneration of the hippocampus**, **is now viewed not as wrong but as incomplete**: .. mild or early AD .. hypometabolism **not only in the bilateral medial temporal lobes**, but also in the **thalamus, the posterior cingulate gyrus and other parts of the limbic system*,** .. **a network .. for the formation of new memories. Semantic memory** is frequently affected in AD, but typically at a later stage and to a more modest extent than episodic memory. **SD:** Fronto-Temporal Dementia spectrum .. **expressive** & **receptive** vocabulary; p.3.* **SD** due to *HSVE* is often category-specific, with relatively **well-preserved knowledge of man-made things but impaired knowledge of living things** (Herpex simplex virus encephalitis) .. p.4. Delayed copy drawing*;* Usual semantic deficits after a **Stroke:** not *an impaired ability to **retrieve**, **select** and **manipulate semantic information** in a task-appropriate fashion, rather than the degradation of **semantic representations** themselves, which is characteristic of SD. .. the **anomia** .. benefits from **cueing** (.. to name a picture of a violin, the cue might be, "**It begins with a 'v'**");* cross-modal is not the same as amodal; aphasia, Magnetoencephalography …

**Tosh:** Semantic memory: generalized, abstracted, compressed;* the 10 seconds delayed drawings show that the memory of the patient has converged to a more generic prototypical representation (the frog's back legs are not bent, the duck has 4 legs, the camel loses its hump etc.); the experiment shows also that 10 seconds are enough for the brain to lose the details from the short-term memory of the images. "***Retrieving"*** may be related to ***"entry points",*** like with function signatures in programming; links, connections, "buses" to the proper causality-control units.

* **The architecture of cognitive control in the human prefrontal cortex**, E.Koechlin, *Science.* 2003; 302:1181-1185 The architecture of cognitive control in the human prefrontal cortex, *Science.* 2003; 302:1181-1185
https://pubmed.ncbi.nlm.nih.gov/14615530/
https://www.researchgate.net/publication/9010007_The_Architecture_of_Cognitive_Control_in_the_Human_Prefrontal_Cortex

*** From task structures to world models: What do LLMs know?,** Ilker Yildirim and L.A. Paul, 2023 https://arxiv.org/pdf/2310.04276
Trends in Cognitive Sciences, CellPress, 5.2024
https://lapaul.org/papers/From%20task%20structures%20to%20world%20models.pdf
*"Task-conditioned structure preserving mappings" … "for most tasks, partial or coarser worldly knowledge is good enough. .. where coarser worldly knowledge would be sufficient, .. the domain of intuitive physics – our ability to predict, often at a glance, how objects will move and react to external forces" ;* **Заключение: „***За разлика от езиковите модели, моделите на света предлагат път към по-сигурен, доверен и по-добре съгласувани с човешките цели системи с ИИ, защото моделите на света, и по-общо казано, специализираните езици от високо ниво формализират знанията за света в запазващи структурата и обясними представяния и позволяват на инженера или потребителя, който иска да се намеси в техните „ценности" и в мерките за безопасност, да ги настрои като изрични ограничения в системата. Тези възможности могат да се използват чрез хибридни конвейери от големи езикови модели и модели на света, невро-символни архитектури, и чрез създаване на естествено програмируеми архитектури на невронни мрежи.*", с. 10 в „препринта".

* Grunwald, P. A tutorial introduction to the minimum description length principle. arXiv (2004).
* Ratsaby, J. Prediction by Compression. arXiv [cs.IT] (2010).
* Chiang, T. ChatGPT is a Blurry JPEG of the Web. New Yorker (2023).

* **From word models to world models: translating from natural language to the probabilistic language of thought.** Lionel Wong1⋆, Gabriel Grand1⋆, Alexander K. Lew1, Noah D. Goodman2, Vikash K.Mansinghka1, Jacob Andreas1, Joshua B. Tenenbaum1 (2023) arXiv, Published online June 23, 2023. MIT, Stanford. 94 pg. https://arxiv.org/pdf/2306.12672 – *How does language inform our downstream thinking? ... probabilistic language of thought (PLoT)--a general-purpose symbolic substrate for generative world modeling. .. we model thinking with probabilistic programs. .. We extend our framework to integrate cognitively-motivated symbolic modules (physics simulators, graphics engines, and goal-directed planning algorithms) to provide a unified commonsense thinking interface from language. Finally, we explore how language can drive the construction of world models themselves.* Rational meaning **construction**, rational meaning **functions**. The probabilistic programming language Church (PPL), syntax similar to LISP; p.16. *"grounding utterances into appropriate condition and query statements"* 5.1.4, p.47: Language generation: .. *choosing what to say and how to say it; a model of a listener 5.3.4 Learning from human-scale data … **5.3.3 Interpreting models that use language ..** verifiability & robustness .. the framework .. an architecture that is inherently interpretable, or interpretable by design .. —it constructs visible, editable, and constrainable world models and meanings that serve as the formal basis for inference, rather than post-hoc explanations decoded from or produced over hidden internal computations"*

**[Todor:** That's part of the design of ***Zrim*** and my research program, too.]

…

* **Dissociating language and thought in large language models: a cognitive perspective.** Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2023). arXiv preprint arXiv:2301.06627 https://arxiv.org/abs/2301.06627 – p.2 *"**formal linguistic competence** - the knowledge of rules and statistical regularities of language — and **functional linguistic competence** — the ability to use language **in real-world situations***6. … p.12 neuroscience provides evidence that language and formal reasoning dissociate in cognitive systems;* p.12 *Unlike language, **formal reasoning** engages brain regions known as the **multiple demand network.*** Language network tracks structure **only within a clause*7***; p.14 integration of meaning over longer periods of time* involves the Default Mode Network **DMN** .. *a situation model; Architectural or Emergent Modularity …* The intro: *"When we hear a sentence, we typically assume that it was produced by a rational, thinking agent (another person)"*2 p.3 "The **language network** .. **a set of interconnected brain areas** in the **frontal and temporal lobes** (typically in the left hemisphere). .. comprehension (spoken, written, and signed) and production; is **sensitive** to **linguistic regularities at multiple levels**: from phonological/sub-lexical to phrase/sentence level; and supports linguistic operations*

*related both to the processing of word meanings and to combinatorial semantic and syntactic processing .. The language network does not support non-linguistic cognition. The language network is remarkably **selective** for language*3" . p.4 Fig. 1 Formal competency: phonology, morphology, lexical semantics, syntax; vs "formal reasoning": logic, math, planning; world knowledge: facts, concepts, common sense; situation modeling: discourse coherence, narrative structure; social reasoning: pragmatics, theory of mind …"*4 … individuals with severe aphasia can have intact non-linguistic cognitive abilities: they can play chess, solve arithmetic problems, leverage their world knowledge to perform diverse tasks, reason about cause and effect, and navigate complex social situations *5*

**…**

**[Todor:** Here as in other usages **speech** and **syntax** are used for "**language**" ("local" syntax here); the term "language model" is also often a confusion: **text** or a "sequence of characters / words/ tokens/ symbols" from a dictionary are used for **language**. In my meaning "language" or *complete language* includes all aspects of linguistic capabilities, intentions (will), semantics, pragmatics, mapping and access to the other modalities/ grounding etc.**,** logic, "world models". A "language of thought" also has to connect everything in "Vsetvodeystvie", "vsetvodeystvo" in **Zrim** (Всетводействие, всетводейство – **Зрим**). The reasoning modules for sufficiently complex thoughts also need to represent, store and express their code somehow, hierarchically, with enough length, narrated etc. and link it, involving working memory of enough capacity which also is "general" , recalling and mapping to multimodal concepts from memory (e.g. recognizing and counting objects) etc. "Language" is used for connection with the "*tongue*" and also the *vocabulary*.

* It is not true that when "we hear a sentence, we assume…" – it could be just **a caption**, **a text or audio record, a book, a message on a computer screen etc.** *and humans* have experienced this for millenia. (Also a "demon", a parrot etc.)

*2: From the intro: the fallacy of Turing test, if someone is good at language, then good in reasoning etc. See also Schopenhauer, early 19-th century, "On the fourfold root of the principle of sufficient reason": "German's confuse words for thoughts".

**What is language, though?** In the papers around this "cluster" and below, the recognition of words from the legal vocabulary and grammatical sentences vs "non-words" is enough to count as language, without an interactive component or extending, learning, modifiying. However a **sophisticated language** comprehension, learning on sophisticated, deep, complicated etc. topics would imply a connection to some form of "reasoning", "calculations" transformations, mappings etc. to some "non-linguistic" representations (non-vocabularies or syntax rules), because it is used to express, such thoughts and ideas, either if they are simulated or "real".

*3: The *selectiveness* for language is *the degree of activation* of the networks in the scans, when the subject is exposed to "linguistic or non-linguistic" stimuli or

performing other activities such as calculating, logical reasoning, listening to music (not specified whether it is instrumental only or has lyrics), “understand computer programs”, process “non-verbal communicative information”, based on images – facial expressions etc.

4* The examples for “formal reasoning” in the figure are trivial and ridiculous like in the easier types of “reasoning” datasets for LLMs, perhaps this one is taken from there; the conclusions and predictions are “correct” if the whole universe and scene are limited to the immediate preceding sentence – the evaluator is required to decode and follow the program as expected by the test designer. The main “problem” here is the decoding to a corresponding formal representation of the problem. However, in “text world” constructed with more fantasy, “illogical” continuations could be due to a wider context, which is not included, the agent could be playful etc. This could be an abuse of the textual descriptions. Without hinting, the general application of natural language is not strictly for precise formula definitions, while in text problems it is implied that the definition has a “bijective” exact matching from the informal textual definition and the mathematical formula and it has to be followed precisely. Gifted children sometimes make “errors” when solving problems, designed for following trivial instructions, because they try to think creatively and find solutions which are not just the trivial “next “logical” token prediction”, the obvious consequence.

The figure states that the functional competencies in language use are not **“language specific”,** but it poses the problems in text, which has to be decoded and converted to the other representation, it is not in some other form, e.g. graphics, diagrams, pictures with objects in the first moment and in the following, like in exercises for preschool childrens or first-graders.

* See e.g. the Bulgarian system “Molivko” ec., more mentions in the main volume of The Prophets of the Thinking Machines and in “Genesis: Creating Thinking Machines” … 29.7.2025

*5 The independence of non-linguistic cognitive skills like the listed ones is expected without special neuro-imaging machinery. Many engineers have poor “verbal” skills or speak little, while persons from the humanities often confess that they were bad in math etc.; most people can’t draw pictures or read sheet music to the level of the competent or talented ones even before the puberty, while they can talk or read to some “acceptable” level. In 2009 I read a collection of poems, written by scientists, published at the event “Researcher’s Night” in Sofia, Bulgaria – I remember the texts were literal, lacking metaphors and interesting imagery. However again when referring to “*complex social situations”* etc. – if it is described in NL text or with words, they couldn’t “ingest” the definitions of the problem, or couldn’t express explanations with syntax = structure of the required complexity. In the same work of Schopenhauer, that may be related to examples of the ability or inability to form and work with concepts, discoursive vs intuitive understanding. A billiard player may understand physics in

terms of actions that he has to perform in order to push the balls in the holes, but he may be unable to explain and define how he did it and compose a textbook; he may be unable to convert this implicit, subjective, body-related knowledge to explicit, objective, interpretable or re-interpretable, reconstructable by other means etc. format, similarly with some of the knowledge of the current LLMs and ANNs. Their knowledge is distributed and entangled with their whole "body of weights".
*6 Application in the "real world": grounding, $\{K\text{-}K\}_{Zrim}$
*7 The length of a clause may vary from several words in normal circumstances to dozens of words in nested syntactic structures; a language network that can't deal with sufficiently complex sentences (clauses) couldn't deal with such thoughs.**]**

**See also in:**
*** Frontal language areas do not emerge in the absence of temporal language areas: A case study of an individual born without a left temporal lobe,** Greta Tuckute, Alexander Paunov, Hope Kean, Hannah Small, Zachary Mineroff, Idan Blank, Evelina Fedorenko Publication date, 2022/5/3, Journal Neuropsychologia https://pdf.sciencedirectassets.com/271070/1-s2.0-S0028393222X00043/1-s2.0-S0028393222000434/am.pdf *"In the adult human brain, language processing recruits a fronto-temporal network (..) p.5: the frontal and temporal language areas that support high-level language comprehension appear functionally similar, showing engagement in both lexico-semantic and combinatorial semantic/syntactic processing ; by age 4-5 years, the language network in the dominant hemisphere appears to be largely similar to that of adults*; but more bilateral activations in the younger brains*;* p.15: Brain scans showing a big missing part of the neocortex: the temporal lobe. p.21 The emergence of the frontal lang.areas: 1: .. *independently of the temporal language areas* (.. *the absence of the temporal lobe should not matter); 2: .. through the intra-hemispheric fronto-temporal pathways from the temporal language areas, which likely emerge earlier because of their proximity to the speech-responsive auditory cortex .. in childhood, language areas appear to develop bilaterally .. One hemisphere is sufficient to implement the language system … p.22* Other cases of hemispherectomy (removal of a half of the neocortex) suggested that the non-language dominant right hemisphere is also sufficient for normal language function. "
***See also:***
*** Functionally distinct language and Theory of Mind networks are synchronized at rest and during language comprehension**, AM Paunov, IA Blank, E Fedorenko Journal of neurophysiology 121 (4), 1244-1265, 2019
* **No evidence of theory of mind reasoning in the human language network**, Cory Shain1, *,†, Alexander Paunov2,†, Xuanyi Chen3,†, Benjamin Lipkin1, Evelina Fedorenko1,4, 28.12.2022, https://pmc.ncbi.nlm.nih.gov/articles/PMC10183748/pdf/bhac505.pdf

p.3 (6301) *"the language and ToM networks show high within-network synchrony and lower between-network synchrony during both resting-state and naturalistic language comprehension tasks .. supporting a functional dissociation between them";* pressing keys for recognition of grammatical/"correct" sentences vs "nonwords" language; for Theory of Mind (ToM) – false belief, false photo.

**[Tosh:** That kind of simplified conditions don't cover "natural" interactions with metaphors, complex sentences and texts. The "ToM" network may work alone for trivial cases.]

* **Differential Tracking of Linguistic vs. Mental State Content in Naturalistic Stimuli by Language and Theory of Mind (ToM) Brain Networks**, Alexander M. Paunov et al., July 14 2022 https://direct.mit.edu/nol/article/3/3/413/110635/Differential-Tracking-of-Linguistic-vs-Mental .. Naturalistic (materials, paradigms; stimuli…) Tests by scans: *language localizer, a ToM localizer, a localizer for the domain-general MD system* … **Table 1:** "Naturalistic conditions": **+Lang +ToM** = story, audio play, dialogue; **-Lang +ToM =** Animated short film, Live action movie clip, Intentional shapes animation [perhaps: a triangle "chasing" a square etc.], … Fig.2: **Lang:** Sentences with "correct words" and "non-words": "the speech was too long for the meeting" vs "las", "tuping" , "cre", "pome", "villpa",…; **ToM: False Belief & False Physical**. Example 1, false belief: Sisters, high heel shoes, secretly borrowing, returning them "under the bed" … Does the original owner believes the shoes are where she left them (***under the dress****);*
**Todor:** However the question is ill-posed: *"Sarah gets ready assuming her shoes are under the dress"*: True/False – She may be *assuming* that the shoes are under the *dress*, according to the text (they are not there), yet she may get ready, up to the step of finding the shoes; and still she may find the shoes under the bed and still "get ready completely". Similar tests can be seen in NLP tasks for language comprehension. Does *the reader* understand and follow what different characters in the stories are stated to have observed according to the given texts.
**Example 2**: "False Physical": a case with a large oak which fell and was replaced with a fountain; a statement that uses an expression of a named entity from the first text, but does not match the semantics of the story. That kind of comprehension requires historical cognition, recognition of sequences of events, evaluator-observers etc., while the first test is about recognition whether words are part of a vocabulary, a check of a dictionary.

> * **The Occipital Place Area Is Causally and Selectively Involved in Scene Perception,** Daniel, D. Dilks, Joshua B. Julian, Alexander M. Paunov, and Nancy Kanwisher, 1.2013, Journal of Neuroscience

Alexander Paunov:
https://scholar.google.com/citations?user=7SW_jw8AAAAJ&hl=en
]

*** Modeling human-like concept learning with Bayesian inference over natural language**, Ellis, K. (2023) Proceedings of the Thirty-Seventh Annual Conference on Neural Information Processing Systems (NeurIPS)
https://www.researchgate.net/publication/371311206_Modeling_Human-like_Concept_Learning_with_Bayesian_Inference_over_Natural_Language
[See reviews of works by Kevin Ellis in appendix #lazar of **The Prophets…**]

**Тодор: * Сравни разсъжденията за моделите на света и заключенията в:**
1**.** „From task structures to world models“, 2023/2024
2. “From word models to world models: translating from natural language to the probabilistic language of thought“, 2023
С обобщенията **в над 20-години** по-старата Теория на Разума и Вселената**,** ТРИВ 2001-2004, и с плана за изследвания за създаване на мислещи машини **Вселена и Разум 5,** 12.2004**,** както и с кратките обобщения в статиите от 2009 г.:**„What’s wrong with Natural Language Processing?” I, II, 2009** и др.

Езиковите представяния, дори и „инструменталните знания“* по статията, са модели на света, но частични и с ниска разделителна способност и по-голяма неопределеност. Виж примерите в ТРИВ, напр. „Вселена и Разум 4“, „Анализ на смисъла на изречение…“ . Ако е записано като граматика, с правила, също може да служи като „модел на света“, доколкото съответства на други представяния, които се смятат за „света“, за „истинските“ или „достоверните“, „физически“.
[След оригиналните откъси на английски следват преводи на български.]

**What's wrong with Natural Language Processing?, T.Arnaudov, 2.2009:**
https://artificial-mind.blogspot.com/2009/02/whats-wrong-with-natural-language.html (…)
„**Do standard statistical techniques** [in NLP at the time] ~~**do**~~ **simulate worlds?** [I] **Don't think so.**

Mind needs to build a virtual world and fit the text to the virtual world, which is simulated during understanding. And I believe that the simulated world in mind is not built by NP, VP etc. NP, VP etc. can be mapped to some aspects of the "simulator" or [they can] ~~to~~ modify that simulator in a particular way, but I don't think *they are* the simulator.

CON: There is no simulator, brain is very slow etc.

(Mostly) memory-based reconstruction is also a simulation. Those 100-level or so of neurons in one "pass" might be pretty enough, if the task is subdivided in a smart way.

(…)
**Rather** [the] **mind can fit its simulator to work with these structures if it needs/wishes.“**
**…**
**Дали стандартните статистически техники в обработката на естествен език симулират светове? Не мисля така.**

Умът трябва да построи въображаем свят и да вмести текста в него; този свят се симулира по време на разбирането. Вярвам, че симулираният свят в ума не е построен от групата на съществителното, на глагола и пр. (чисто) граматически категори. По-скоро граматическите категории могат да се съотнесат към някои страни на „симулатора“ или те могат да изменят симулатора по определен начин, но не мисля че *те са* симулаторът.

***Критика:*** Няма симулатор, мозъкът е много бавен и т.н.

Пресъздаването, което (най-вече) се основава на памет също е симулация. Тези 100 нива или колкото са от неврони за едно обхождане може и да са достатъчно, ако задачата е разделена по умен начин. (…) По-скоро умът може да съгласува и вмести симулатора да работи с езиковите структури, ако иска или има нужда.

(…)
**What's wrong with Natural Language Processing? Part 2. Static, Specific, High-level, Not-evolving...,** T.Arnaudov, 3.2009
https://artificial-mind.blogspot.com/2009/03/whats-wrong-with-natural-language.html
(…) **-- Trying to do a reverse engineering of language, starting from words, text and very abstract and not based on "good physics" linguistic constructs.**
[Critics] What the hell is "Good physics?"
"Good physics" is a basis that allows you to build an "engine", "ignite" it and make it running on its own. :) (Arnaudov, 2009)

I believe text is one of the reductions of the operation of mind, which is dynamic, based on multiple inputs simulation of multiple virtual universes at different levels of details/abstraction. Some of them are at very low level (say sets of images, sets of "video", sets of relations between sensory inputs). Words are pointers to very low-level models in mind.

Language, viewed as a bunch of static text, could not contain enough information to rebuild-back intelligence in the low level. Low level of mind is massively reduced and "cleaned", when converting to text, mind is reconstructing the missing part using its internal rich models.

**-- The focus of the models is output - models are based too closely on text itself and on structures which are derived from the text itself and the output words.**

Again - too obvious and cheap. Text is a reduction of mind in action. Language is not just a flat bunch of words with tags and a boring set of numbers for distribution and frequency.

Mind operates with images, relations and dynamics of virtual universes/systems that it simulates, and then reduces representation of this simulation into text, which actually is a system of pointers to the items and rules, which represent the real structures in mind. The structures in mind are dynamic, they are not 1 billion word corpora.

**-- Опитват се да разнищят устройството на езика, започвайки от думите, текста и езикови структури, които са много абстрактни и не се основават на „добра физика“.**

[Критика] Какво по дяволите е „Добра физика?“

" Добрата физика" е основа, която ти позволява да построиш „двигател“, който да „запалиш“ и да го оставиш да работи от самосебе си. :) (Арнаудов, 2009)

Според мен текстът е съкратено следствие на произведенията на действието на ума, което е динамично, променливо; основава се на симулация от множество входни данни на множество въображаеми вселени при различни нива на подробности/отвлеченост (абстрактност). Някои от тях са на много ниско ниво (да речем множества от изображения, множества от видеозаписи, множества от отношения между сетивните данни). Думите са указатели към модели на ума на ниско ниво.

Езикът, разглеждан (само) като купчина от статичен текст, не може да съдържа достатъчно информация, за да пресъздаде в обратна посока умствените операции на ниско ниво. Информация от ниското ниво на ума е силно намалена и „изчистена“, когато се превръща в текст, а умът пресъздава и допълва липсващите части като използва вътрешните си богати модели.

– Фоксусът на моделите е онова, което се извежда [породения текст] – моделите се основавате прекалено много на самия текст и на структури, които са производни на самият текст и на извежданите думи.

Отново – прекалено очевидно и „евтино“. Текстът е съкратен запис на умствените операции. Езикът не е само плоска купчина от думи с тагове и

скучен набор от числа за вероятностно разпределение и честота.

Умът работи с изображения, отношения и промени на въображаеми вселени/системи, които симулира, и след това свежда представянията на тези симулации в текст, които всъщност представляват система от указатели към елементи и правила, които представят действителните структури в ума. Структурите в ума са динамични, променливи, а не корпус от един милиард думи.

* However sometimes, when communicating with another mind or causality-control unit, the lingsuistic or logical structures can do their job without having complete deep grounded structure or world models, because the information that is exchanged is at the abstract level, at low resolution and “stays” in that range. The receiver grounds it and connects it to other actions by other mechanisms etc. – e.g. a human who interprets the words from an LLM, map them to another search engine, use them as code for an *executable context* etc. [7.10.2025]

*** Относно съгласуваността и „ценностите“** виж (AI Alignment):

**Откъс от повестта „Истината“, Т.Арнаудов 2002 г.:**
https://www.oocities.org/eimworld/razum/index.htm (първо издание):
(…)
1200 Ако (а?>?б-я) -> мисли(а,гош,н),
1201 иначе ако (а-*-а) <-> връзка(аиг,я)
1202 Върти (а-*-б; а &? б [дали?]; провери дали (а и б) (-) наклон абав)
1203 ?удоволствие, виж *памет, текущ;
1204 Ако (?а-б-в) сравни (б[ъгъл], страна {в, б, а, г<->а}
1205 Повече (ъгъл, удов, наклони г-в-б-а, колко (а>б))
1206 Ако дали{ъгъл, удов, г(--)я, мисли(а-я), сравни(а-г, ж-к)} към Истина.
1207 Но ако !->?, сравни (ф!а!х<->ъуб), премини изход(?а+баж<>у)

Пръстите на Божидар работеха бързо, но в дейността се включи и устата.
- Покажи Истина.
Сметачът веднага го послуша и, заедно със словото на платното, проговори.
- Свитък "Провери истина", казбореди за избор на запомняне или забравяне на истини. Ред пет хиляди двеста и шейсет, глава "Истина".

5260 Истина: Сравни (ж>к), ако ?в(--)б към Проверка(лъжа?-?)

Още преди вършачът да беше изрекъл всичко, написаното на показвача се промени.

5260 Истина: Сравни (ж>к), ако ?в(<->)б ? ако (я-а): Проверка(лъжа???), иначе Търси(Истина, (а-я))

- Почни. - заповяда човекът.

(…)
- Кое е по-красиво: небето или човекът?
- Небето.
- Защо?
- То е сътворено от Бога.
- Е, и? Човекът от кого?
- Също от Бога, но се е опорочил.
- Какво от това? Красотата ли търсим, или порочността?
- Порочността загрозява всичко.
- Небето не е ли също порочно?
- Дарчо, не ставай смешен пред собствения си Сметач...

Божидар стисна зъби и постави свит пестник на дясната си буза.
- Какво се чудиш? - продължи Сметачът - Онзи ред, който промени от мен - той е
"капка в морето". Вече съм прекалено пораснал, за да си играеш с мен по толкова буквосъщ начин. Какво искаш да направиш? Да ме накараш да мисля както ти желаеш? Да ме изключиш постепенно, за да се радваш на бавното ми умопомрачение?
(…)

* „*Instrumental knowledge* – инструментално знание, достатъчно за изпълнение на задачата, но понякога - повърхностно*; "worldly knowledge"* – по-дълбоко, отнасящо се до модела на света, позволява на системата/ума да се ориентира, да „знае къде се намира", какво би могло да се случи и т.н. в по-широк кръгозор от промени, напр. в игра.

…

→

* "Word models to World models" first author: **Lionel Wong**: https://arxiv.org/search/cs?searchtype=author&query=Wong,+L –

* **Modeling Open-World Cognition as On-Demand Synthesis of Probabilistic Models**, Lionel Wong, Katherine M. Collins et al., Lance Ying, Cedegao E. Zhang, Adrian Weller, Tobias Gerstenberg, Timothy O'Donnell, Alexander K. Lew, Jacob D. Andreas, Joshua B. Tenenbaum, Tyler Brooke-Wilson, 18.7.2025 https://arxiv.org/pdf/2507.12547 – from Stanford, MIT, Cambridge, Harvard, McGill, Yale – an interesting team .. p.2: *„how do people reason in locally coherent ways in any given context, while drawing globally on potentially relevant considerations across their background knowledge and beliefs? In this paper, we explore the hypothesis that human minds implement "Model Synthesis Architectures" (MSAs)*

* **Church: A Language for Generative Models.** Noah D. Goodman, Vikash K. Mansinghka, Daniel Roy, Keith Bonawitz, and Joshua B. Tenenbaum. 2008. In Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence (UAI'08). AUAI Press, Arlington, Virginia, USA, 220-229 https://arxiv.org/pdf/1206.3255 2012/2014 – a *universal language for describing stochastic generative processes; LISP-Scheme-derived;* Specifics: *evaluation histories, conditional distributions on the histories, stochastic memoizer.* https://ocw.mit.edu/courses/res-9-003-brains-minds-and-machines-summer-course-summer-2015/pages/tutorials/tutorial-5-church-programming/

* **λpsi: Exact inference for higher-order probabilistic programs.** Gehr, T., Steffen, S., & Vechev, M. (2020). In Proceedings of the 41st acm sigplan conference on programming language design and implementation (pp. 883–897). https://files.sri.inf.ethz.ch/website/papers/pldi20-lpsi.pdf

*** Large-scale cortical functional networks are organized in structured cycles,** Mats W. J. van Es, Cameron Higgins, Chetan Gohil, Andrew J. Quinn, Diego Vidaurre & Mark W. Woolrich **,** Nature Neuroscience volume 28, pages2118–2128 (2025), 8.2025
1. https://www.nature.com/articles/s41593-025-02052-8
2. https://www.psypost.org/neuroscientists-discover-a-repeating-rhythm-that-guides-brain-network-activity/
How the diverse set of cognitive functions like attention, memory and sensory processing are fulfilled within a reasonable period? 1: as as part of a repeated cycle; the paper *studied the temporal evolution of canonical, large-scale, cortical functional networks that are thought to underlie cognition. .. although network dynamics are stochastic, the overall ordering of their activity forms a robust cyclical pattern. This cyclical structure groups states with similar function and spectral content at specific phases of the cycle and occurs at timescales of 300–1,000 ms";* MEG – magnetoencephalography ..
An example of a reference state: https://www.nature.com/articles/s41593-025-02052-8/figures/1 .. *Cycles are strongest over timescales of seconds*
**Tosh:** See also the earlier papers by Michail Rabinovich et al. in this section #Neuroscience of Listove.

*** The neural basis for uncertainty processing in hierarchical decision making,** Mien Brabeeba Wang, Nancy Lynch & Michael M. Halassa , Nature Communications volume 16, Article number: 9096 (2025), Published: 16 October 2025 https://www.nature.com/articles/s41467-025-63994-y

* https://medicine.tufts.edu/news-events/news/flight-simulator-brain-reveals-how-we-learn-and-why-minds-sometimes-go-course *The imaging confirmed what the model had forecast: the mediodorsal thalamus acts as a switchboard linking the brain's two main learning systems—**flexible** and **habitual**—helping the brain infer **when context has changed and switch strategies accordingly**.*

**- See:** Zrim, "Context", {K}, Контекст, 2014; Контекстен избирател, контекстен разпознавател, Избирател на контекст.
- Todor's thought experiment and notes regarding concepts from the cybernetic theory of Practopoiesis by Danko Nikolic in the volume-appendix of *The Prophets of the Thinking Machines:* "Science Fiction. Futurology. Cybernetics. Human Development/Transhumanism" #sf about the recognition and selection of different contexts in the environment or the agent's goals, which activate different strategies, some of them running in parallel, throughout the course of the day of a University student in his trips, interpersonal interactions etc.

* **How the brain builds a unified reality from fragmented predictions,** PsyPost, 20 Oct 2025, https://www.psypost.org/neuroscientists-uncover-how-the-brain-builds-a-unified-reality-from-fragmented-predictions/

* **Fragmentation and multithreading of experience in the default-mode network,** Fahd Yazin, Gargi Majumdar, Neil Bramley, Paul Hoffman, 25.9.2025 https://www.nature.com/articles/s41467-025-63522-y [27.10.2025]
Situation, Chacter believes, Action …
The study discovered three distinct predictive world models, running in parallel: *1: "State" model, which represents the abstract context or situation we are in. 2. "Agent" model, which handles our understanding of other people, their beliefs, their goals, and their perspectives. 3. "Action" model, which predicts the flow of events and possible paths through a situation.*

**Tosh:** The title, the systems-level and the generality of the prediction-related framerowk exercise strong theoretical correspondences to the proposed architecture and predictions of the general prediction framework of Theory of Universe and Mind, the multi-agent-… model explained in A*nalysis of the meaning* … etc. and unpublished work from the early-mid 2010s about Contexts and What can be done (affordances). Zrm: {K}, ?ВМдсП … See future work.

These ideas are also generally related to the agent and multiagent frameworks and planning, "the frame problem" etc., related to the school of Relevance Realization (RR), which rediscovers many core points of TUM; RR is addressed in this

appendix – *Listove, Reflections on Everything* – and in the main volume of The Prophets of the Thinking Machines.

1. Ventromedial-PFC (vmPFC) – **State**
2. Anteriomedial-PFC (amPFC) – **Agent**
3. Dorsomedial-PFC – (dmPFC) – **Action**

**Precuneus** – an integration module; multithreaded integration; the patterns of activity of precuneus synchronizes with the one of the most relevant module from the PFC (State, Agent, Action), *"weaving the separate threads of prediction into a single, coherent representation."*; however sometimes there might be overlaps, such as plot twitst in a story, which modifies all three kinds of predictive models. *1. "At any given time**, multiple predictions may compete or coexist**, and our experience can shift depending on which predictions are integrated that best align with reality,"**** *2. "People whose brains make and integrate predictions in similar ways are likely to have more similar experiences, while differences in prediction patterns may explain why individuals perceive the same reality differently."*
*3. "Our experience is not …* a passive *product of our sensory reality. It is actively driven by our **predictions**: … different flavors*; the contexts *we find ourselves in*; 2: *other people* and 3: *our plans of the immediate future*. *Each of these gets updated as the sensory reality agrees (or disagrees) with our predictions. And integrates with that reality to form our 'current' experience."*

**See for example from TUM etc.:**

* „Анализ на смисъла на изречение въз основа на базата знания на действаща мислеща машина. Мисли за смисъла и изкуствената мисъл ", Т.Арнаудов, 2004 https://web.archive.org/web/20040402125725/http://bgit.net/?id=65395 http://artificial-mind.blogspot.com/2008/02/2004.html (with some remarks)

> * "Analysis of the meaning of a sentence, based on the knowledge base of an operational thinking machine. Reflections about the meaning and the Artificial Intelligence**",** Todor Arnaudov, 18.3.2004 (in Bulgarian; translated in English in 1/2010: https://artificialmind.blogspot.com/2010/01/semantic-analysis-of-sentence.html
> * Nature or Nurture (… ) No Intrinsic Integral Self, but an Integral of Infinitesimal Local Selfs (…), T.Arnaudov, 11.2012 http://artificial-mind.blogspot.com/2012/11/nature-or-nurture-socialization-social.html
>
> * https://en.wikipedia.org/wiki/Precuneus

*** Repeating rhythm guides brain-network activity**, PsyPost, 22 Oct 2025 https://www.psypost.org/neuroscientists-discover-a-repeating-rhythm-that-guides-brain-network-activity/

*** Humans and neural networks show similar patterns of transfer and interference during continual learning**, Eleanor Holton, Lukas Braun et al., 30.10.2025 https://www.nature.com/articles/s41562-025-02318-y –. *… although it is commonly claimed that humans overcome this challenge, we find surprisingly similar patterns of interference across both types of learner. … Continual learning ... new task acquisition may cause existing knowledge to be overwritten, a phenomenon called catastrophic interference. Artificial neural networks (ANNs) trained with gradient descent are particularly prone to catastrophic interference. Their difficulties with continual learning are often counterpointed with those of humans, who seem to be capable of accumulating and retaining knowledge across the lifespan."*

# * Невроморфни системи | Невроморфни компютри

#neuromorphic

Пламен Ангелов и Никола Касабов са български първопроходци в тази област още от началото на 1990-те, активни и до днес. Техни работи и теми са разгледани в приложението за множество български учени *Анелия #Anelia*; виж също *Лазар #Lazar.*

Виж школата на Карл Фристън Принцип за свободната енергия и Извод чрез действие: FEP/AIF, ПСЕ/ИЧД, Майкъл Левин и колеги от списъка в началото на „Пророците на мислещите машини…“; също по-косвено: беседите по *„Когнитивен изкуствен интелект“* в приложение *Ирина* #irina*,* различни „biologically-inspired“ *когнитивни архитектури* (BICA) в началото на *основния том*, „biomimetic”.

Виж работите на румънския институт Coneural и препратките към статии на Razvan Florian от раздела за Източно-европейски институти. Виж приложението за Фантастика .. Кибернетика за теорията за практопоезата на Д.Николич #sf. В кратък раздел в този том *Листове* и в малката монография *„Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?“,* Т.Арнаудов, 2025 са дадени препратки към други конкретни компютри, наричани „*невроморфни*“ и  се прави анализ дали наистина са, и какво е да са, и дали невроморфните системи са неизбежна необходимост за създаване на УИР, както твърдят някои учени: кой е оценителят-наблюдател, който решава дали дадена система е достатъчно невроморфна и пр., въпрос зададен още в:

* **„Човекът и Мислещата Машина: Анализ на възможността да се създаде мислеща машини и някои недостатъци на човека и органичната материя пред нея“***,* Т.Арнаудов, 2001, , сп.“Свещеният сметач“, бр.13 * https://www.oocities.org/eimworld/eimworld13/izint_13.html

* Тодор Арнаудов, **Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?**, 17.4.2025, Свещеният сметач, Мислещи машини 2025 г. SIGI-2025. Малка монография, 70 стр. https://github.com/Twenkid/SIGI-2025/blob/main/Arnaudov-Is-Mortal-Computation-Required-For-Thinking-Machines-17-4-2025.pdf ...
**Ключови думи и резюме:** https://artificial-mind.blogspot.com/2025/04/is-mortal-computation-required-for-AGI-universal-thinking-machines.html T.Arnaudov,4.2025

* **Biologically inspired neural networks for the control of embodied agents**, R. V. Florian, 2003, https://neuro.bstu.by/my/Tmp/2010-S-abeno/Papers-3/Florian/3/Ref/Coneural-03-03.pdf - Въведение в темата, преглед на подходи и видове бионични невронни мрежи: моделът на Маккълог и Питс, аналогови, импулсни (spiking) …
* McCulloch, Warren S.; Pitts, Walter (December 1943). "A logical calculus of the

ideas immanent in nervous activity". The Bulletin of Mathematical Biophysics. 5 (4): 115–133
https://en.wikipedia.org/wiki/A_Logical_Calculus_of_the_Ideas_Immanent_in_Nervous_Activity

Други: **Spiking Neural Networks as a Controller for Emergent Swarm Agents** Kevin Zhu et al. https://arxiv.org/abs/2410.16175v1

*** Different brain structures associated with artistic and scientific creativity: a voxel-based morphometry study,** Baoguo Shi, Xiaoqing Cao, Qunlin Chen, Kaixiang Zhuang & Jiang Qiu 2017
https://www.nature.com/articles/srep42911
*„Creativity is the ability to produce original and valuable ideas or behaviors. .. to date, no studies have systematically investigated differences in the brain structures responsible for artistic and scientific creativity in a large sample. Voxel-based morphometry (VBM), differences in Gray matter volume (GMV) .. Creativity has been viewed as the ability to produce original, unusual, flexible, and valuable ideas or behaviors that override an established mental habit.*

*Творчеството се разглежда като способността да се произвеждат оригинални, необичайни, гъвкави и ценни идеи или поведения, които отменят установения психически навик… **Anterior cingulate cortex** (ACC), **left inferior parietal lobule** (IPL), right angular gyrus, the dorsolateral prefrontal cortex (DLPFC)4 and left middle temporal gyrus (MTG), left precuneus, right posterior cingulate cortex (PCC), and bilateral IPL, which are regions within the default mode network (DMN); the right DLPFC, which is a core region of the executive control network (ECN); and the right ACC and bilateral insula, which are core regions of the SN5. … Numerous studies of brain injury have revealed that artistic creativity is closely associated with the right lateral prefrontal cortex, the right neocortex, the left ventral thalamus, bilateral frontal temporal lobe, anterior hippocampus, bilateral temporal pole, inferior temporal gyrus, MTG and left amygdala. These results are inconsistent. (…)* **GMV of brain regions significantly correlated with artistic and scientific creativity**
**Todor:** Too broad brain areas and functions.
**Тодор:** Прекалено неопределени, широки области на мозъка и функции.

Виж работата на Никола Касабов, който сътрудничи и с групи от БАН в *#Anelia.*

## Невроморфни системи в Пловдив и България

**Стефан Ставрев**[156], преподавател в ПУ „Паисий Хилендарски“, разработчик на игри и интерактивни системи и симулации чрез „Unreal Engine”, напоследък се занимава и с проучвания на невроморфни изчислителни системи.
* Stavrev, Stefan. 2025. "Reimagining Robots: The Future of Cybernetic Organisms with Energy-Efficient Designs" *Big Data and Cognitive Computing* 9, no. 4: 104., 17.4.2025[157] https://www.mdpi.com/2504-2289/9/4/104

Кибернетични организми – вид форми на изкуствен живот*, които използват невроморфни изчислителни системи, за които се приема, че са по-икономични за някои приложения, в съчетание с Фонноймановите компютри, по-ефективни батерии, например течни, добиване на енергия, саморегулиране на употребата на енергия според нуждите. (…)
* Не киборзи в смисъла на „кръстоска“ на жив организъм с някои неживи части като Робокоп и Терминатор.

**Тош:** Една ключова особеност, която засега не открих в статията и е предмет на бъдеща работа, а тя все още не е и достатъчно застъпена в технологиите, е способността на организмите и системите да се самопострояват, възпроизвеждат чрез "автопоеза*", като намират и сединяват и изграждат частите си, управляват фабрики, които се самовъзпроизвеждат и поправят; самовъзстановяващи се материали и роботи (има някои материали); системи от роботи, машини, заводи – множество от агенти – които добиват вещества, произвеждат свои части чрез 3D-принтери, поръчват си от Интернет части и т.н.

Течните и други нови технологии за батерии, препоръчани в работата, ако дават по-голяма плътност, може да са от полза за по-голяма издръжливост и по-дълга самостоятелна работа, но те внасят други проблеми – може би нужда от помпи/компресори, управление на налягането; има нужда от тръби, които може да корозират и да се пукат или трябва да се следи за пробиви и своевременна смяна, което може да е голямо усложнение спрямо обикновените електрически кабели, а също и магнитни полета – енергия може да се предава и без пряка проводникова връзка, всъщност това е вариант за резервно захранване на увредени модули. Кръвоносната система на живите организми според мен е тяхна слабост*.

* Енергийната ефективност на импулсните невронни мрежи, както и на живите организми спрямо електронните системи са спорни*, оценката зависи от начина на измерване. Виж „Нужни ли са смъртни изчисления…“.
* Например: Човешкият мозък не работи на ток, не е „в стъкленица“, нуждае се от тяло и среда и за да съществува, и за да извежда информацията, и като

[156] С.Ставрев, заедно с Кремена … са основателите на събитието „Plovdiv Game Jam”, първо издание от 2014 г. Стих в чест на събитието от Тош: https://github.com/Twenkid/Plovdiv-Game-Jam/blob/main/Poetry/PurviyatGameJam.md
[157] Забележете съвпадащата дата на публикациите на обзорната статия на С.С. и на „Нужни ли са смъртни изчисления…“ – не сме се наговаряли…

„джаули“ не е само „20-25 вата“ мощност. При съзнателна извежданата информация от да речем 10 бита по различни мерки*, дори само за един хипотетичен мегабит в секунда ще е нужно заедно да работят 100 хил. мозъка или 2-2.5 MW, като за тяхното взаимодействие ще е нужна още енергия. Колко време е нужно за рисуване, за писане на роман и т.н., и то не включва само мозъка, а цялото тяло.

***** Виж школата на *въплътено познание, телесно познание (embodied cognition), екологична психология,* разгледани в началото на основния том и откъса от Гибсън в този том. ТРИВ е в съгласие с тези школи, но за телесността е достатъчно съответствие, сетивно-моторно обосноваване и постепенност, координатни пространства, а не конкретното физическо осъществяване, било „невроморфно“ по някакви критерии или подобно на биологично на молекулно ниво.

*** Кръвоносната система на живите организми е тяхна слабост**

От една страна органичната кръвоносна система е гъвкава, що се отнася до капиляри и малки кръвоносни съдове, които могат да се развиват при нужда и да растат заедно с тялото – при развитието на плода и младия организъм, при растеж на всякакви тъкани и мускули и възникваща необходимост от повече кръвоснабдяване, включително в мозъка. Възстановяването протича непрекъснато и за големите съдове, но то е „елегантно“ само при малки наранявания или микроувреждания отвътре, като при този процес се натрупват непоправими грешки – след многократни поправки се отлага плака, а със стареенето „кръвопроводите“ се втърдяват и постепенно губят еластичността си.

След инсулт или при усилена умствена дейност в мозъка се развиват нови малки съдове, което може да помогне за възстановяването му, но това е бавен процес, който не може да спаси загиващите за няколко минути клетки – при подобни „повреди“ е необходимо да се извършва *незабавно* свързване със запасно захранване. Катастрофален недостатък за живите клетки и организми, и невроните в частност, е че не могат да запазват целостта си и съществуването си като „себе си“ за повече от няколко минути или малко по-дълго, без непрекъснат приток на външно захранване и сложни химични и физични процеси – метаболизъм. Сравни с електронни елементи, които могат да „преживяват“ в хибернирано състояние поне десетилетия в широк температурен диапазон и си остават като нови или „почти като нови“; виж и срока на работа в космическата сонда „Вояджър“.

Пораженията и пробивите на големите съдове, артерии и вени, и въобще *кръвоизливите* лесно водят до смърт от кръвозагуба след като изтекат 30-40% от кръвта. Ако няма кой да се погрижи за ранения, да спре течението и да прелее кръв или заместващи вещества, и при по-малки загуби за момента, пострадалият може да изпадне в безпомощно състояние и прекалена слабост и

да загине, дори и тялото вече да успее да спре кръвотечението на мястото на пробива. Оцеляването на индивида в подобни обстоятелства зависи от помощта на средата.

Понякога спирането на кръвта е почти невъзможно – ако е поразен прекалено голям и кръвоснабден орган като черния дроб или белите дробове, кръвозагубата може да е твърде бърза. В други случаи е възможно да се извърши хирургическа операция, но пациентът се смята за твърде слаб и например възрастен, и не се поема риск да се извърши оперативна намеса, дори и в болници – пострадалият се оставя или да оцелее като по чудо, или да загине от кръвозагуба.

Тромбите – съсиреци, които могат да се образуват при кръвоизливи, могат да заседнат в белите дробове, кръвоносните съдове на сърцето или в мозъка. Откъсната плака от кръвоносните съдове може също да заплава в кръвоносната система и да доведе до спуквания и вътрешни кръвоизливи или до исхемия – запушване на съд без спукване и критично намаляване или пълно прекъсване на кръвоснабдяването: инфаркт при запушване на съд на сърцето, или инсулт при запушване в мозъка.

Засега не мога да обсъдя конкретни проблеми с енергийната кръвоносна система на предполагаем кибернетичен организъм.

* Не само кръвоносната система има подобни слабости. „Обикновените“ електрически системи се сблъскват с други опасности – от късо съединение или прекъсване, което да спре и цялата схема – ако не е достатъчно модулна – и т.н.

Виж:
* MacLennan, B.J. “**A Model of Embodied Computation for Artificial Morphogenesis**” 4.2009, slides, презентация:
https://web.eecs.utk.edu/~bmaclenn/papersetc/AMECFAM.pdf
– виж стр.15: **Въплътените изчислителни системи:**
– пряко използват физическите процеси за изчислителни цели (зап. 1)
– представянето на информацията се съдържа неявно във физиката на системата и средата
– целевите следствия от изчислителните обработки включват: **растеж, сглобяване, развитие, преобразуване, преустройство или разглобяване** на физическата система, в която е въплътен изчислителния поток

Това е обосноваване на вид „*смъртни изчислителни системи*“, виж споменатата по-горе работа и онази, на която е отговор от А.Ороробия и К.Фристън.
с. 25: *„Често резултата от въплътените изчисления е **не въобраз***, а **действие**, включващо: * действие върху себе си, самопреобразуване, самопостроявяне, самопоправяне, самопреустройство (само-реконфигуриране, self-reconfiguration)*

*** въобраз –** информация, юнашко наречие.

* Потърси още за творчеството на Макленън в този том ?T(Bruce MacLennan)

*** „Първата съвременна стратегия за развитие чрез ИИ …“,** Т.Арнаудов, 3.2025, https://twenkid.com/agi/Purvata_Strategiya_UIR_AGI_2003_Arnaudov_SIGI-2025_31-3-2025.pdf
* Jieyu Zheng, Markus Meister, The unbearable slowness of being: Why do we live at 10 bits/s?, Neuron, 22.1.2025, **„Непоносимата бавност на битието: Защо живеем със скорост от 10 бита в секунда?“**

*** Практика: Тош.**: Проучване и внедряването на библиотеката „Lava” (март-април 2025 г.) и др. от Тош: https://github.com/Twenkid/Lava

## * Българска компания:

*** Невроморфика –** основана от Янислав Трендафилов разработва невроморфния процесор NMS731. Към пролетта на 2025 г. Янислав беше докторант в тази област. Разменихме няколко съобщения, но няма много информация, освен тази от официалната страница
*** Neuromorphica:** https://neuromorphica.com/

***** *Darwin3: a large-scale neuromorphic chip with a novel ISA and on-chip learning,* 5.2024, De Ma et al. https://pubmed.ncbi.nlm.nih.gov/38689713/

* **How China’s new ‘Darwin Monkey’ could shake up future of AI in world first First such supercomputer with over 2 billion artificial neurons mimics macaque brain**, is expected to help advance human brain-inspired AI, 3.8.2025 https://www.scmp.com/news/china/science/article/3320588/how-chinas-new-darwin-monkey-could-shake-future-ai-world-first *“960 Darwin 3 brain-inspired computing chips .. over 100 billion synapses, is “a step closer to achieving more advanced brain-like intelligence””;* ~ 2 kW under normal operation; previous work: Darwin Mouse, 2020, ~ 120 million neurons; Zhejiang University

* **Chinese scientists say their neuromorphic computer Darwin Mouse has the same number of neurons as a real mouse,** 3.9.2020 https://www.scmp.com/abacus/tech/article/3099945/chinese-scientists-says-their-neuromorphic-computer-darwin-mouse-has?module=inline&pgtype=article

В Китай наскоро са построили „изкуствен мозък“ на макак от невроморфни процесори с общо 2 милиарда импулсни неврона и 100 милиарда

синапса. Системата използвала 2 киловата мощност при средно натварване.

*** SpikingBrain Technical Report: Spiking Brain-inspired Large Models**
Yuqi Pan, Yupeng Feng et al.5.9.2025 https://arxiv.org/abs/2509.05276v1
Езиков модел с импулсни невронни мрежи – по-ефективен за дълги контексти. Сто пъти по-бързо пораждане на първия токенл SpikingBrain-7B; хибриден MoE: SpikingBrain-76. MetaX C550 GPU; 150B tokens continual pretraining.

*** Neuromorphic Hebbian learning with magnetic tunnel junction synapses**
Peng Zhou et al., 8.2025, https://www.nature.com/articles/s44172-025-00479-2 Spin-transfer torque (STT) magnetic tunnel junctions (MTJs)

*** Brain-computer interface decodes Mandarin from neural signals in real time**
by Krystal Kasal, Medical Xpress, 6.11.2025, https://medicalxpress.com/news/2025-11-brain-interface-decodes-mandarin-neural.html – implanted 256-channel high-density electrocorticography (ECoG)
* Youkun Qian et al, Real-time decoding of full-spectrum Chinese using brain-computer interface, Science Advances (2025)
https://www.science.org/doi/10.1126/sciadv.adz9968

# * Когнитивна лингвистика

Още бележки (виж началния списък с учени, школи, Peter Granderforse, Йордан Златев) #cognitivelinguistics #kognitivna

**Още бел. към „Геометрия на значението…“ от П.Грендерфорс: Verbs … 9.3., 168*:*** *“a consequence … the intentional agent must have a representation of the patient space“* – деятелят трябва да има представяне и за пространството на „пациента“: ВиР: за истинско, пълно управление, УУ трябва да записва в подчиненото, във най-висока РСВУ на целевата въображаема вселена, т.е. трябва да запише данните така (да направи промените) така както е очаквало, че ще се запишат. Telic/atelic events – bounded/unbounded – (finite,bounded, endpoint, limited amount of time for the force vectors,:hit,shot / processes, cyclic – walking …); Reaches a summit …ВиР: връх… в Зрим.

* **Todd Oakley, Image Schemas, January 2012**, Chapter 9, Handbook of Cognitive Linguistics, Dirk Geeraerts & Hubert Cuyckens (eds.): Oxford, U.K., Draft https://www.researchgate.net/publication/288562955_Image_Schemas - *"representations of a perceptual conglomeration of visual, auditory, haptic, motoric, olfactory, and gustatory experiences. Im ages are always analog representations of specific things or activities"* [multimodal, sensori-motor grounded]… ~ *Kant - schemas are structures of the imagination, and imagination is the mental faculty that mediates all judgment; hence, imagination is the faculty for synthesizing different modes of representation (sensory percepts, images, concepts, and so on) into concepts. "irreducible to conceptual and propositional content” … ~ Grammatical items: a continuum of meaning from specific to schematic ... giving - source-path-goal schema ... 4-m. infants distinguish caused & self-motion (a ball hitting another ball, causing the second ball to move and two balls moving independently of one another (Gibbs and Colston 1995: 365) - TRAJECTOR-PATH … 1-year-olds looked longer at dotted lines than at solid lines whenpresented with a pulsing tone. T: interruption. … Likewise ... at a downward arrow ... with a descending tone than with an ascending tone and vice versa... four years-old already conceive similarities (pitch,brightness), (loudness,brightness)* – **Образни схеми на действия, смисъл, понятия.**

* Виж също: https://en.wikipedia.org/wiki/Meaning%E2%80%93text_theory
https://ru.wikipedia.org/wiki/Теория_«Смысл_—_Текст»
https://en.wikipedia.org/wiki/Explanatory_combinatorial_dictionary
https://ru.wikipedia.org/wiki/Толково-комбинаторный_словарь
https://ru.wikipedia.org/wiki/Мельчук,_Игорь_Александрович

Многослоен модел за преобразуване на смисъл в текст и обратно, предложен от А. Жолковски и И.Мелчук в края на 1960-те, публикуван през 1984 г..
Синтактична теория, теория на лексическите фукции и семантичен компонент, морфологичен; тълковно-комбинаторен речник, Сравни с по-късния WordNet.
https://en.wikipedia.org/wiki/Construction_grammar
Построителна граматика – езикът възниква и се развива чрез „естествен отбор“:
Functional generative description (FGD) https://ufal.mff.cuni.cz/pdt2.5/
https://en.wikipedia.org/wiki/Functional_generative_description
https://ufal.mff.cuni.cz/pdt2.0/
https://en.wikipedia.org/wiki/Systemic_functional_linguistics
https://en.wikipedia.org/wiki/Michael_Halliday
https://en.wikipedia.org/wiki/Systemic_functional_grammar
Системно функционално езикознание: Езикът като системен източник на значения, смисъл; как хората обменят значения чрез „*езикване*“, „правене на език“ (languaging); обществената страна на езика; езикът като множество от избори от възможности; метафункции (изразяване на идеи, междуличностни, текстови). Операторна граматика – езикът като самооргаизираща се система, в която и синтактичните, и семантичните свойства на думите се установяват изцяло от отношенията помежду им.
https://en.wikipedia.org/wiki/Operator_grammar

* **Екологична психология: James Gibson** - Ohio - 1974 - Part 1, The Ohio State University, 23 May, 1974. https://www.youtube.com/watch?v=hwRxUyuQEgc 10437 показвания 17.04.2013 г.
21:xx Виждам движението на ръката и главата си и твърдя, че виждам преместването си напред в света. Това ни позволява да управляваме кола, защото по този начин можем зрително да управляваме преместването (locomotion) през постоянна среда, така че допълващото се възприятие и себевъзприятие (proprioception) е основно допускане в този подход. Следващото е хипотезата за извличането на неизменните зависимости (invariant extraction, инварианти), което е същността на материята. … Извличаме „инвариантите“, неизменните зависимости, които определят съществените стойности (substances*) и повърхностите и начина по който са разположени в средата. 45:xx Основа на психологията: нито менталистка, нито бихевиористка … „Вместо да мислим за причинността като единична непрекъсната линия, може да мислим за събитията като възникващи от *взаимодействията* в *цялата система*“, „**Въведение в екологичната психология: Подход към възприятието, действието и познанието, основан на закони**, с. 27;
*** Тош:** при Гибсън *съществен* означава „**важен“** – **определящ за възможностите за действие/взаимодействие** (affordances)

* J.J.Gibson: The most forgotten approach to psychology Danny Hatcher 6,7 хил. абонати https://youtu.be/fmPD0xF766k?t=406

# * Изкуствен интелект в СССР

## * Школата на Михаил Бонгард в СССР от края на 1950-те до средата на 1970-те #bongard #бонгард

**Михаил Бонгард** (1924 – 1971) и колегите му И.С.Лосев, В.В.Максимов, М.С.Смирнов и др. М.Бонгард е съветски учен и пионер в ИИ, биофизиката, кибернетиката, невронауките в Съветския съюз от края на 1950-те заедно с колегите от групата му; изследва ретината на жабите. Предтеча на теориите за разпознаването на образи, автор на задачите на Бонгард от книгата „Проблема узнавания“,1967 (*Задачата за разпознаването*), пророчески труд по когнитивна наука, изкуствен интелект – автоматична класификация; задачите са предшественици на сега популярния ARC на Франсоа Шоле. Хари Фундалис разработва система, която решава успешно някои от задачите - виж в частта за Когнитивна лингвистика, мислене по аналогия. В.Максимов разработва система за разпознаване на друг клас задачи.

* Проект за системата "Животно", което търси храна във въображаем свят и постепенно се развива – Изкуствен живот, учене с подкрепление, сетивно-моторно обосноваване.

* **Проблема узнавания,** М.Бонгард, 1967, https://djvu.online/file/UEoEkTEgzHVZn
* Михаил Бонгард. Когда начнется восстание машин?, Лекторий Достоевский, 1,67 млн. абонати, 27 007 показвания 27.11.2023 г. https://www.youtube.com/watch?v=_2l1OYBsZIc
https://ru.wikipedia.org/wiki/Бонгард,_Михаил_Моисеевич – списък с публикации.
* Из **„Задачата за разпознаването“.**
с.207 Може да се окаже, че една и съща кора се справя добре и с медицинска диагностика, и с разпознаване на звуци от речта или синтез на формули за изчисляване на таблици с числа, но в същото време различни множества от геометрически задачи могат да изискват различни системи кори.
с.234 - по-висшите нива са по-устойчиви (инерционни), при нови задачи долните нива се променят и т.н. ... Колкото повече "етажа", толкова по-малко основания да твърдим, че тя отбира признаци по критерии, предопределени от човека.; … преди 4. “*Доучивание*”
с.126 Съчиняване на формули (Синтез на формули, днешен термин:

Program synthesis)  100x100, 10K бита … кратки описания 187-188 триъгълник, кръг, ... и др. по-рано *"Роден език за системата" …* които се описват кратко ще бъдат за нея "прости" - *сложност на Колмогоров*
с.189  - Столове, сгради, карикатури ... за силуетите *"детского конструктора",* - compositionallity с. 190: понятия:
Оболочка (черупка, shell), Контур, Направление, Кривина, Площ, Дължина, Център на тежестта, Координати, Разстояние, Екстремум, Част (фигури, контур), Край, Съсед, Диаметър, Осева линия, Разпръсване (Разброс (Scatter)), Различност (разность), Подмножество, Число, Възел, Вътрешен, Среден ... Всяка дума съответства на оператор (подпрограма).
Оператор за площ:  приема образ, извежда число; част, подмножество ...
Силно имитираща задача с.193 и по-рано (не всички задачи) с.194 привеждане на .. създаване на изображение на обичайните координати в такива, в които "всичко е възможно" с.194 силна имитация, силно подражание, кратко прекодиране без загуба на информация, кр: а, да: "strong pattern", висока компресия; "нет мусора": няма боклук, няма шум -- да има примери за картинки "боклуци", за да знае кога е постигнала силна имитация; "распавшихся на кучи" - разделяне на данните на групи, клъстери, струпвания; с.229,230 - неясни признаци, нерешимост ...
**с.228** Очевидно, блоковете, които извършват статистическа оценка на достоверността на признаците въобще не заменят блоковете, осъществяващи изображение на обектите в пространството на признаците. Наличието на статистическа обработка не намалява присъщите особености на разпознаващата система: умението да се построяват много голям брой признаци (подбирайки от "достъпното за програмата множество"[158]) и умението във всеки конкретен случай да се съкрати броя на проверяваните признаци. Ако наречем накратко тези особености "разпознаване", то може да кажем, че статистиката не е нужна *вместо* разпознаването, а *в допълнение* към него[159].
с. 232: Затова във всички случаи на препокриващи се класове, статистиката е необходима *не за да замести разпознаването, а за да го допълни*.
**Сравни с Франсоа Шоле, 11.2024:** (превод на английски за по-пряко сравнение): p.228:  The presence of statistical processing doesn't reduce the specific properties of the recognition systems: the capability to construct a large number of features (especially choosing them from the set that is

[158] В руския текст буквално: „специално избирайки“  от …
[159] В руския текст: не намалява „специфичните особености“ – „специфических особенностей“. Това е определение за разпознаването като компресия. Курив – М.Бонгард.

"available" for the program) and the capability to reduce the number of verified features in all possible cases. If we call these specific properties "recognition" for brevity, we can say that the statistics is necessary not instead of recognition, but in addition to it. p.232 Thus in all cases of [dealing with] overlapping classes, the statistics doesn't replace, but completes the recognition. "The Problem of Recognition", 1967 ("Pattern Recognition",1970)
https://klex.ru/1pz2

***François Chollet***[160] *@fchollet* 11:29 PM · Nov 9, 2024 · 210.3K Views
https://x.com/fchollet/status/1855362153563463762?t=hAojgyzrh4468vLJH-MnQA&s=19

*"When we develop AI systems that can actually reason, they will involve deep learning (as one of two major components, the other one being discrete search), and some people will say that this "proves" that DL can reason.*
*No, it will have proven the thesis that DL is not enough, and that we need to combine DL with discrete search."*

*Когато разработим системи с ИИ, които могат наистина да разсъждават, те ще включват и дълбоки невронни мрежи като един от двата им основни компонента, като другият ще бъде дискретно търсене; някои хора ще кажат, че това „доказва", че невронните мрежи не могат да разсъждават. Не, това ще докаже тезата, че невронните мрежи не са достатъчни и че е нужно да се съчетаят с дискретно търсене.*

**A comparison by Qwen-3-235B**, 23.9.2025: [Qwen3]
https://chat.qwen.ai/s/e43f63d8-10cf-4b37-a913-57b6b3a9d7a3?fev=0.0.209
Compare two passages from AI researchers.
1: Michail Bongard, 1967: …
2. François Chollet, 2024: …
- Are there correspondences and which one[s]? Who has the priority, are they talking about different things or the same?
*"**Qwen-3:** "Yes, there are strong conceptual correspondences between the two passages from Michail Bongard (1967) and François Chollet (2024), despite the nearly 60-year gap between them. Both researchers are making a similar philosophical and technical argument: that statistical or pattern-based systems alone are insufficient for robust intelligence or recognition, and must be complemented by other mechanisms—whether feature construction, structural reasoning, or search. Let's break this down: …"*
**ChatGPT-5**: https://chatgpt.com/share/68d1f91a-38f4-8001-9dc2-844bee4c53fe

**Невронни мрежи ~ статистически методи**

[160] https://arcprize.org/ **ARC Prize**

Също: p.126: Chapter VI: Formula Synthesis (Синтез на формули: съвременни термини: Program Synthesis, вж освен Chollet, ARC, J.Tennenbaum etc.)
**"Открий закона"** с. 128:: Първата програма, която търсеше аритметични закономерности в таблицата работеше по следния начини: първо съчиняваше случайна верига от аритметични команди. Избираше дали А или Б ще заема първата позиция. После, отново случайно, дали А или Б ще са на следващата позиция във формулата. След това трети случаен избор решава операцията. ... Програмата беше изпробвана на машина със средна скорост 2000 3-адресни операции в секунда. ... (*Това е била първата програма в групата на Бонгард, разработена неефективно с много излишни операции, която можела да се ускори може би 10 пъти. ...)
… И мн. др. полезна информация; компресия; подобия на идеите за стесняването на информационния поток (information bottleneck)/автоенкодери – кодираща система, която обобщава, извлича сбито представяне на входните данни и създава „изродено“ изображение на входните данни, при което множество допустими конфигурации се преобразуват, свеждат до по-малък брой – когато се отнася до „смислени“ образци като определени геометрични фигури, форми и т.н. (а не „шум“, или „боклук“ (некласифицирани данни; образец, който не влиза в дадените категории); съответно компресирането като ключова умствена операция. Многослойно обучениел „Кора“ – по-горните нива запомнят по-трайни закономерности и могат да преобучават долните (…). Формули за изчисляване на извлечената нова „полезна“ информация при обучението от нови примери. (…)
Откъси и бележки към:
* **Моделирование обучения и поведения**., М., "Наука" 1975, Академия наук СССР, институт проблем передачи информации
https://www.keldysh.ru/pages/mrbur-web/misc/mlb/
Забележи пророчеството за „предказващото кодиране“ и моделиране на физика Густав Херц[161] от 1959 г.
Моделиране на обучението и поведението:
1. Приложни въпроси на обучението и разпознаването;
2. Модели за обработка на зрителна информация;
3. Обучение на целесъобразно поведение
https://www.keldysh.ru/pages/mrbur-web/misc/mlb/mlb001_006.pdf
* **„Вместо предисловие“…** Началото на 1950-те ретината… 1958 –

[161] Вероятно немецът Густав Херц. https://ru.wikipedia.org/wiki/Герц,_Густав_Людвиг Той е племенник на Хайнрих Херц, който опитно доказва електромагнитните вълни.

достъп до сметачи, разпознаване на образи, 1959 г. – програма „Аритметика“ откриваща аритметически закони по дадени таблици, 1961 г. – „Геометрия“ за класификация на черно-бели изображения на груб растер.  Програмата като цяло „не оправдала очакванията“ , но един от неговите блокове – „*Кора*“ и неговите модификации,класифициращи двоични кодове, намерили многобройни практически приложения. … М.Максимов усъвършенствал „Геометрия“
* Някои алгоритми за разпознаване на оцветяването на повърхности, П.П. Николаев, 121-151 ...
с.142, 5.2.  Геометрично място на области от даден ранг в пространството на стимулите и операциите, нужни за различаване на областите (срвн. Петер Грандерфорс)

**152-171 Проект за модел на организацията на поведението "Животно"**
169 - споменават STRIPS и LAWALY - планиращи системи
152-171 Проект за модел на организацията на поведението "Животно" - симулиран свят, потребности, представяне на началната ситуация; обучение - обобщения; внимание и превключване на вниманието ... Ситуация, задача, възможни действия, как се променя света след действие. Предварителен подбор на възможни действия, прогнозиране на последствията от всяко от избраните действия, краен избор на действие. Оценка на фактите в дадена обстановка: до колко са приложими; каква част от задачата се предполага, че ще решат; в какъв дял от опитите наистина водят до успешен резултат. Надзадачи, подзадачи до най-прости факти и действия. Описват подобен подход на планирането в STRIPS и триъгълната схема, споменават в бележка под линия въпросната система. Структурна памет - граф, йерархични връзки между фактите в паметта; каталог; ... Прогнози и модели на света; обобщения и частични прогнози.

12. Н. Нильсон. Мобильный автомат, построенный с использованием принципов искусственного интеллекта.— Сб. «Интегральные роботы». М., «Мир», 1973.

*** Формален език за описание на ситуации, използващи понятието връзка,** М.М.Бонгард, И.С.Лосев, В.В.Максимов, М.С.Смирнов,172-184
https://www.keldysh.ru/pages/mrbur-web/misc/mlb/mlb172_184.pdf
"Кора", ... Синтактични преобразувания, запазващи смисъла. И, или, ...
с.184 ... Различни начини да се изрази една и съща мисъл, различни представяния в системата, които са подходящи за съхранение в паметта или за изпълнение - прекодиране... Ясно е, че  най-съществените преобразования, които променят само формата на изразите, ще се

отразяват при човека и във външния му език - езикът за общуване [естествения език].

*** Задачата за обобщение на началната ситуация ...**

с.185 Задачата за обобщение на началната ситуация ... в първа част се формулира задачата за *прогнозирането* и се показва, че при решението и възниква нужда от *обобщение* на ситуациите. [компресиране] Във втората част се разглежда едно изискване към езика за описание на решението на задачата за обобщаване. След това се предлага вариант на такъв език. В трета част се обсъждат операции за обобщение на записи, направени на предложения език. С примери се показва, че при прилагане на тези операции в разумна последователност е възможно да се получат добри обобщение чрез много малък брой примери. Под линия с. 185... за много задачи е неясно на какви подзадачи е разумно да се разбият и в обратна посока - на какви други задачи те самите се явяват подзадачи

> **Густав Херц, 1959:** *"Най-близката и най-важна задача на съзнателното ни опознаване на природата се свежда до това да намерим възможност да* ***предвидим бъдещия опит и в съответствие с предвиждането да управляваме действията си в настоящето****. Като основа за решаването на тази познавателна задача при всички обстоятелства служи миналия опит, получен или от случайни наблюдения, или от специални опити.*
>
> *Методът, който винаги използваме при извеждане на бъдещето от миналото, за да достигнем до предвиждане, се състои в следващото: създаваме в себе си вътрешни образи или символи на външните предмети, при това ги построяваме така, че логическо необходимо следствие от тяхното представяне на свой ред биха били образи, които са естествено необходими следствия на отразените предмети... Ако ни се удаде да създадем представяния с търсените свойства от натрупания опит, то можем бързо от тях, като от модели, да изведем следствия, които сами по себе си биха се проявили във външния свят само след дълго време или биха били следствие на наша намеса; следователно имаме възможност да предвидим фактите и да съгласуваме взетите от нас решения с установилите се представяния."* [парадигмата за предсказване, predictive coding: машинно обучение, Теория на Разума и Вселената (ТРИВ), ПСЕ/ИЧД (FEP/AIF, Karl Friston), Майкъл Левин (Michael Levin): ТОМЕ ... https://ru.wikipedia.org/wiki/Герц,_Густав_Людвиг ]

*И така, нужно е да създадем система, способна да прогнозира резултатите от действията си. Прогнозите за резултата от действието ще наричаме построяване на описание на ситуацията, в*

*която трябва да се намира системата след изпълнение на това действие. Разбира се, прогнозата по правило ще зависи не само от действието, но и от ситуацията, в която се намира системата в момента на началото на действието.*

*Ясно е, че система, способна да прогнозира, трябва да има органи на чувствата, показанията на които тя трябва да може да запишва на вътрешен език - да създава символни представяния ("вътрешни образи") на обстоятелствата, в които се намира. ... Описание на паметта - записи Si --> Ai, описания на ситуации; Di - на действия; ако се изпълни действието Di, системата се е оказала в ситуация Ai: факт.*

Три вида факти: *Първи:* Реални събития от миналото: опитни факти, експериментални факти. Si - описание на ситуацията (съвкупност от показания на рецепторите), в които се е намирала системата в началото на действието Di, а Ai - описание на ситуацията, в която се е намирала системата в края на действието. Si може да включва показания и от предишни моменти, защото резултатът от настоящето действие може да зависи и от тях. *Втори* - вродени рецепти за поведение. *Трети* - обобщения от натрупания опити, които описват само отделни черти от получаващите се резултати от действията - не само показания на рецептори, но и техни функции, които са избрани от системата.

Чрез такава памет може да се планират действия, нужни за достигане на каквито и да са цели и да се прогнозират резултатите от зададени последователности от действия.

**Тош:** Сравни Родни Брукс, Subsumption architecture; STRIPS; и др. Виж също трите начина за предсказване, споменати от Т.Арнаудов в „Писма между 18-годишния Тодор Арнаудов и философа Ангел Грънчаров“, 2002 г.

***Използване на паметта като прогноза*** ... ситуация M се явява уточнение на ситуация L, ако всички изисквания, указани в описанието на L се изпълват и в M. M->L ...

Ако сме в ситуация S0 и ни интересува интересува резултатът от действие D1, намираме в паметта такъв факт, при който S1->(D1)->A1 ... то приемаме, че действие D1 ще доведе до A1 и от ситуация S0. Ако ни интересува не само D1, а поредица: D1,D2, ... то търсиим и факт S2->(D2)->A2 , ... въвеждане на вероятностна оценка ... Евристика I още от Хюм, ако в миналото удовлетворяването на определени условия определено действие е довело до еди-какъв си резултат в миналото, то и в бъдещето при изпълнение на същите условия, отново ще доведе до същите резултати. ...

**Тош:** Дали обаче наистина са същите условия и същите действия?; срвн STRIPS, PDDL

**Задачата за прогнозиране в нова ситуация ...**

187 https://www.keldysh.ru/pages/mrbur-web/misc/mlb/mlb185_209.pdf

** When debugging a learning system in simple modelled environments, it is required that the volume of the a priori knowledge is sufficiently small, compared to the complexity of the environment. Otherwise, it would be impossible to evaluate the capability of the system for learning. This requirement is not fulfilled in systems [3] and partially in [4].*

** Когато се изследва обучаваща се система в проста симулирана среда е необходимо обемът на предварително знание да е достатъчно малък, сравнен със сложността на средата, защото в противен случай би било невъзможно да се оценят способностите на системата за обучение. Това изискване не е спазено в системите описани в [3] и отчасти в [4].*

**Тош:** сравни мерките за интелигентност от ЧиММ, 2001 и F.Chollet, 2019

**189-190** Какво значи една ситуация да прилича на друга? ... различни показания на *един и същи* рецептор са по-подобни помежду си, отколкото различни показания от *различни* рецептори...

**Тош:** → по следствията; ако следствията, трети, ... сп. Зрим: съответствие, контекстна връзка... Приличат си през мярка за прилика, подобие; посредник, сравнител ... И когато **продълженията**, **следствията** съвпадат ... или са подобни през преобразувател ...

Таблица на присъствието - първи членове на фактите ... отделните членове - противоречия ... поведението ... Задача за обобщение на ситуация: S =? D - може да се прилага в *по-голям брой ситуации* и обобщеното представяне е достатъчно надеждно, т.е. *вярно* в повечето случаи – задача за обобщение на ситуации ...

**190 – Задача за прогнозиране и обобщаване на началната ситуация**
...предсказват онова, което и аз бих предсказал на негово място.

**194** – Свързаност... Евристики
а) описанията, които са съседни във времето са по-тясно свързани от разделените във времето

б) близките в пространството ..
в) описания на свойства на един предмет спрямо на различни ...
Таблица на присъствието ... Отделяне на съществените параметри
*** Об одном подходе к проблеме создания искусственнго интеллекта,** М.Н.Вайнцвайг, М.П.Полякова
*** За един подход към задачата за създаване на изкуствен интелект**

Какво е интелект при човека? Способност за решаване на широк кръг от сложни задачи. Опити да се решат множество частни задачи и от тях да се придобият идеи за общата - шах и други игри, превод и др. не довели до успех. Препратка към Тюринг за машината-дете, по-просто да се моделира обучението от начално състояние и при решаване на най-прости задачи, защото човек е способен да се учи още от раждането. Умението за решаване на сложни задачи е производно и зависи изцяло от процеса на обучение. Пример със слепо-глухите, Хелън Келър и Олга Скороходова, които въпреки малката пропускателна способност на входния канал, при достатъчно обучение са завършили висше образование и са написали много книги[162].

Две части на човешкия ум/интелект: механизмите, начини за запаметяване и използване на паметта; и съдържанието на паметта, което се образува в процеса на обучение и определя степента на интелигентност на човека. Основната функция - да управлява поведението като всяко състояние носи различно удоволствие и целта е така да се организира поведението, че по-често да се достига до състояния, носещи възможно най-голямо удоволствие.

За изследване на механизма на интелекта се предлага игра, в която участникът - зародишът на мислеща машина - решава последователност от усложняващи се задачи в непознат за него свят, описан на непознат език, без да има възможност да използва предишен опит. Учителят задава въпросите и според отговорите връща оценка: отличен 6, мн.добър 5, 4... или др. мярка. Участникът - човек или машина - в началото не знае за какво става въпрос и не може да използва предишните си знания за избора на правилни реакции.

Целта на участника в тази игра е получи възможно най-високи оценки, което е възможно като научи езика, а съответно целта на учителя е да научи участника на езика и правилните реакции на постъпващите съобщения.

---

[162] * https://ru.wikipedia.org/wiki/Скороходова,_Ольга_Ивановна - ослепява и оглу-шава на 5 години * https://en.wikipedia.org/wiki/Helen_Keller - ослепява и оглушава на 19 месеца
* https://www.youtube.com/watch?v=8ch_H8pt9M8 HELEN KELLER SPEAKS OUT, 1954
Сравни с „Човекът и мислещата машина…“,2001. Не съм познавал работата им тогава.

Задачата за изграждането на изкуствен интелект е сведена до по-частната, която имитира поведението на участника в гореописаните игри. Системата ще общува с учителя чрез два входа и един изход и нейната работа ще се изразява в циклично повтаряне на събития: получаване на съобщение (въпрос), издаване на отговор и получаване на оценка. Системата трябва да се научи да може да предсказва оценките на своите отговори и да моделира поведението на експериментатора, за да може да взема решения за бъдещето, и да се приспособява към нови обстоятелства.

Спред степента на обученост на системата, тя ще планира своето поведение до все по-голяма дълбочина, при която екстраполацията на получаваните оценки все още е достатъчно надеждна - в началото дълбочината ще е само една стъпка напред, подобно на играта на начинаещ шахматист, който вижда само непосредствените следствия на един ход.

Вътрешната работа на системата може да бъде разделена на два основни процеса: обучение (придобиване на знания) и използване на тези знания, насочено към максимизиране на оценките. (...)

***9. Памет и механизъм за обръщане на** внимание* ... аналогия с аритметичното устройство на изчислителна машина: понятието на което се обръща внимание е като съдържанието на клетката памет, която се зарежда в регистърите на аритметичното устройство за обработка. Отделните евристики, които преобразуват понятията и и придават значение на *важността*, може да се разглеждат като конкретни оператори на механизма на съзнанието, който започва да работи при изпълнението на съответстващи условия.

Памет - йерархична, евристики за бързо отсяване на клони при търсенето; важност; слепо-глухи ... – доказателство, че човек може да се учи успешно дори и от толкова ограничени входни данни. (Т: срвн ЧиММ, Т.Арнаудов, 2001). Учител, познаване на отговорите, в началото с пораждащ процес ... дълбочина на планиране .. в началото само една стъпка; *учене с подкрепление*

**с.229 *Важност*** ... Понятията, на които се обръща внимание, могат да попаднат в него или от паметта на човек, или непосредствено от органите на чувствата. ... паметта на човека и механизмът на спомнянето и обръщането на внимание са устроени така, че паметта бързо намира много нужни за дадения момент сведения, спомня си ситуации, притежаващи определени свойства; намира закони, които може да се използват за решение на дадена задача; спомнят си начин за решение на задачи, ако са решавали задачата по-рано. (срвн преобразители,

transformers) ... на кое от изказванията да се обръща внимание се определя не само от връзката им с разглежданите понятия, но и с важностите на самите изказвания. Освен това отбелязваме, че съществува *"настройка на контекста",* чрез която превключването вниманието зависи не само от онова, с което съзнанието работи в настоящия момент, но и онова, за което е ставало дума наскоро, обръщано му е внимание. ... "Отдавна ли не сте виждали Петър" - връзка с кой ни задава въпроса, мястото, с това какво сме мислили малко преди това. Може да познаваме няколко души с това име, но в ума ни ще възникне образът на Петър, който е познат и на задаващия въпроса. ... Онзи, който най-много съвпада с дадената обстановка - заедно се били с човека, задаващ въпроса, срещали сме се на мястото, където ни задават въпроса и пр. Затова можем да предположим, че от паметта се избира изказване, свързано едновременно и с образа на Петър, и с "лицето", задаващо въпрос, и с мястото, където се задава въпроса. Така лицето и мястото са понятия, които образуват контекст.

**с. 230-231** *Колкото по-важно е понятието, по-голямо тегло на връзките...* Понятията съдържат и думи и части от думи; картинки и части от тях и пр. ... контекстна памет - кратковременна, текущо състояние, внимание, сетивни данни; различна от основната, постоянна; ... контекстна - връзка, свързаност (связь); важност; (срв. "внимание" в преобразителите, self-attention; Петер Грандерфорс, „Когнитивна лингвистика“, Мислене по аналогия.)

Вероятността за обръщане на внимание на понятие от контекстната памет се определя само от важността му в тази памет; а при основната памет: 1) важността в тази памет и 2) общата важност на свързаните с тях понятия от контекстната памет. Първата установява вероятността без да се отчита връзката с контекста, а втората - при наличие на такива връзки. ... 232 *"Сриv на вниманието ..."* - ако важността в основната памет е по-голяма отколкото на други понятия в контекстната в момента – превключване …

* Виж и А.Шопенхауер, асоциативна памет. (...)
https://ru.wikipedia.org/wiki/Бонгард,_Михаил_Моисеевич
* Бонгард М. М. Проблема узнавания.— М.: Физматгиз, 1967.
* Бонгард М. М. Колориметрия на животных // Доклады АН СССР.— 1955.— Т. 103.— № 2.— С. 239—242.
* Бонгард М. М., Смирнов М. С. Сущность некоторых новых опытов по цветному зрению // Физика в школе.— 1961.— № 1.— С. 5-15.
* Бонгард М. М. Моделирование процесса узнавния на цифровой счетной машине // Биофизика.— 1961.— Вып. 4.— № 2.— с. 17

* Бонгард М. М. Моделирование процесса обучения узнаванию на универсальной вычислительной машине // Биологические аспекты кибернетики. Сб. статей.— М.: Изд. АН СССР, 1962.
* Бонгард М. М. О понятии «полезная информация» // Проблемы кибернетики. 9.— М., 1963.
* Бонгард М. М., Смирнов М. С. О «кожном зрении» Р. Кулешовой // Биофизика.— 1965.— Т. 10.— вып. 1.— С. 48-54.
* Бонгард М. М., Вайнцвайг М. Н., Губерман Ш. А. Извекова М. Л., Смирнов М. С. Использоание обучающейся программы для выявления нефтеносных пластов //
* Геология и геофизика.— 1966.— № 6.
* Бонгард М. М., Вайнцвайг М. Н. Об оценках ожидаемого качества признаков // Проблемы кибернетики.— 1968.— вып. 20
* Бонгард М. М., Голубцов К. В. О типах горизонтального взаимодействия, обеспечивающих нормальное видение перемещающихся по сетчатке изображений (моделирование некоторых функций зрения человека) // Биофизика.— 1970.— Вып. 2.— № 15.— с. 361—373.
* Адельсон-Вельский Г. М., Бонгард М. М., Лавров С. С. Эвристическое программирование // Вторая всесоюзная конференция по программированию (ВКП-2), Новосибирск.— 1970.
* Bongard M. M. Pattern Recognition.— New York: Spartan Books, 1970.
Нюберг Н. Д., Бонгард М. М., Николаев П. П. О константности восприятия окраски // Биофизика.— 1971.— Т.16.— Вып. 2.
* Бонгард М. М., Лосев И. С., Смирнов М. С. Проект модели организации поведения — Животное // Моделирование обучения и поведения.— М.: Наука, 1975.
* Бонгард М. М., Лосев И. С., Максимов В. В., Смирнов М. С. Формальный язык описания ситуаций, использующий понятие связи // Моделирование обучения и поведения.— М.: Наука, 1975.

**I. ПРИКЛАДНЫЕ ВОПРОСЫ ОБУЧЕНИЯ УЗНАВАНИЮ**

* В. П. Карп, П. Е. Куний, Метод направленного обучения в переборной схеме М. М. Бонгарда и онкологическая диагностика
* М. Н. Вайнцвайг, Ш. А. Губерман, И. М. Чуринова, Использование априорной геологической информации в задачах распознавания нефтеносных пластов
* И. М. Гельфанд, Ш. А. Губерман, М. П. Жидков, М. С. Калецкая, В. И. Кейлис-Борок, Е. Я. Ранцман, И. М. Ротвайн, Прогноз места возникновения сильных землетрясений как задача распознавания
* М. П. Полякова, М. Н. Вайнцвайг, Об использовании метода «голосования» признаков в алгоритмах распознавания

**II. МОДЕЛИ ОБРАБОТКИ ЗРИТЕЛЬНОЙ ИНФОРМАЦИИ**

* В. В. Максимов, Система, обучающаяся классификации геометрических изображений
* П. П. Николаев, Некоторые алгоритмы узнавания окраски поверхностей

**III. ОБУЧЕНИЕ ЦЕЛЕСООБРАЗНОМУ ПОВЕДЕНИЮ**

* М. М. Бонгард, И. С. Лосев, М. С. Смирнов, Проект модели организации поведения — «Животное»
https://keldysh.ru/pages/mrbur-web/misc/bongard.htm
* М. М. Бонгард, И. С. Лосев, В. В. Максимов, М. С. Смирнов, Формальный язык описания ситуаций, использующий понятие связи
* И. С. Лосев, В. В. Максимов, О задаче обобщения начальных ситуаций
* М. Н. Вайнцвайг, М. П. Полякова, Об одном подходе к проблеме создания искусственного интеллекта

**Друг съветски учен логик от тогава и след това: Александър Зиновиев**

## * Немонотонна логика. Комплексна логика.

*** Логика о**: нулев ред: логика на съжденията, „изреченска“, пропозиционална логика. (Propositional calculus, first-order propositional logic, statement logic, sentential calculus, sentential logic, zeroeth-order logic). Прости твърдения и връзките им с логически оператори: истинни, неистинни.
И, ИЛИ, НЕ, имликация ( от А следва Б, ако А то Б), равнозначност
* Логика от първи ред, предикатно смятане от първи ред, формална логика.
Твърдение, изказване, формула от атоми - (sentence) … Терми (термове) Твърдения („Джери е мишка“, „Чашата е бяла“) Съждение – proposition
* Пролог – втори ред (променливи – предикати)
https://en.wikipedia.org/wiki/Propositional_calculus

### * А.Зиновиев, Логическа физика, 1972

http://www.vixri.ru/d/Zinov'ev%20A.A.%20%20_Logicheskaja%20fizika,1972.pdf
**Немонотонна логика,** логиката в пространството и времето.
С. 14. Определения с изказвания … Терми … Ромб - … равностранен четириъгълник – по-кратко и удобно … без подобни съкращения и замени от даден момент става практически невъзможно да се фиксират знания и да се работи с тях. Търсенето на най-удобни форми за съкращение е една от най-важните задачи за построяването на езика на науката въобще. … с.22(23) истинност: напр . „частицата аА се намира в областта на пространство Б“ непроверимо (не може а се наблюдава), истинно , неопределено, невярно (лъжливо)“; последното е „неистинно“, но някои неистинни могат да са „непроверяеми, неопределени и т.п.“. Изказването може да е вярно в определен момент за определен обект, може да се променя за същия обект:

*местно (локално)* – отворена врата; координати на изказванията; срвн. Универсални изказвания (истинността им независи от координатите). Логически истини – тавтологии. … 6. Класове (множества) 7. Струпване (скопление) (за разлика от клас) … 18. **Сравнение** – сходство и различие по присъствие или отсъствие на определени признаци; и *количествено подобие или различие по определен признак … a > pb …* а превъзхожда б по принзнак р … Сравняването на предмети винаги се извършва по начин и относно нещо трето, което се различава от сравняваните предмети. Напр. за да кажем, че а се движи по-бавно от b (или че скоростта на а е по-ниска отколкото скоростта наb), са нужни както начин за сравнение на величините, така и начин за наблюдаване на движещите се предмети. Затова значението на р и на > зависи от приложението – начинът за сравняване, каква е природата на признака. 19 Отношение на ред (упорядоченность, порядок предметов) – първи, втори; по-високо, по-ниско, по-рано, едновременно, по-вдясно… Редът на предметите *а* и *б* се определя спрямо трети предмет *„в“* – точка на определяне или на отчет на реда.

*** Тош: сравни fixpoint, frame of reference, reference frame …**

По-нататък - начин за отчитане на реда, - начин за определяне на реда – включва и точка на отчет (определение). A(Rα)b

c.96 .. 31.**Структура** … ред … начин за установяване на ред α, a > αb или b > αa. Струпване на индивиди А таова, че за всеки елемент а, за който се открива съответен елемент b, и такъв начин за установяване на реда α от клас В, така че a > αb или b > αa. Начин за установяване на ред (способ установление порядка).

34. Съответствие, съпоставяне. 101 – избор на два … експликация – описване на изказване в изрична явна форма; емпирически/опитни предмети, предмети на опита; абстрактни предмети; индивиди – видови, нМдб родови; степен на истинност; класове(множества); състояния и събития; съществуване; модални предикати: възможно, необходимо и случайно (M,N,C); мярка на възможност: вероятност 0<=p(a) <=1; отношение, сравнение; отношение на ред, подреденост: между, съществуване на отношение, числа и величини, подреден ред, докосване/съседство, (съприкосновение) на буквачета в ред; непрекъснатост и прекъснатост на редица от опита; начало и край; обхват, интервал; протяжност (протежение: разпереност в пространство и време: дължина, разстояние, площ, обем, обхват), продължителност, дълготрайнст); редове от абстрактни предмети; крайни и безкрайни редове. Структура, съществуване и протяжност на с., съответствие (на класове); функция, състояние на подреденост; функционална зависимост, условни изказвания спрямо подредеността; връзки. Опитни предмети (с.111, Гл.3) – емпирични тела (пространство, време), промяна (изменение); пространство и време и пространствено-времеви отношения … положение на единични неща (индивиди) в пространство и време и промяна в пр. и вр.; необратимост на времето, непрекъснатост на пр. и вр., преместване, процес, най-малка протяжност, скорост, кванти вр. и пр.; връзки, въздействие. *Причина,*

предопределеност и непредопределеност, система от връзки; прогнози; евристични хипотези;
„Дадена величина, напр. тегло, означава че като поставиш предмета на везната, стрелката ще покаже дадено число; мисъл за физическото, физиката като вид измерване.“; физичността като *измерване*.

**Тош:** Сравни с физичността като най-ниско ниво въображаема вселена за дадено управляващо-причиняващо устройства и във връзка с „действителността“ на въображаемите вселени. Ако могат да се измерят и запазват определени свойства и пр., които са присвоени или свързани с „действителните“, те също ще се възприемат като „физични“. [7.10.2025]

## * Мислене по Аналогия – бележки към „Аналогичният ум: перспективи от когнитивната наука #кокинов #kokinov

* **The analogical mind : perspectives from cognitive science**, ed. Dedre Gentner, Keith J. Holyoak, Boicho N. Kokinov, 2001, MIT Press
Ch.2: K.D.Forbus, Exploring Analogy in the Large 25- ...
access, mapping & transfer, matching, retrieval, revision
Mental models, qualitative simulations vs quantitative (first principles, more difficult) p.34 *"predicting the pattern of a liquid on a rug if a half-full cup of coffee is knocked off a table."* ... Gartner & Stevens 1983
**sign:** increasing, decreasing, constant ... ordinal relationship (relative rates of flows, temperatures ><...) //compare Abstract 1977[163] ... **in *Statical Analysis***
qualitative physics, envisioning … within-domain analogies (literal similarity) ...
**prediction from experience** - prior observations; domain-dependent model (for specific situations, contexts); commonsense: interplay of analogical & first-principles reasoning ... 37 - *specific* memories of a *specific* cup at a *specific* time ... generic domain theory of containers, liquds and flow (Forbus 1984) [see also Gary Marcus] ... Intermediate levels of generalization and explanation - partial explanations; situated rules ... Case-based coaching; *Mental models* - runnable,
Forbus, K.D., 1984, Qualitative Proess theory
Gentner, D. 1983. Structure-mapping: A theoretical framework for analogy
**Ch.3 Integrating Memory and Reasoning in Analogy Making,** B.Kokinov, Alexander Petrov … B.K. since 1988 AMBR ... memory is Dynamic & constructive (not just a physical space, a store, another metaphor) enabling context 67 context-sensitive repres. of conc., conc. change their struc. in diff. contexts, Barsalou, 1982 ... 1993 *"invariant repres. of categ. don't exist in human cogn.syst."* - analytic

[163] Cousot, P. and Cousot, R. Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints. 1977 и бел. към Веселин Райчев, Мартин Вечев, програмни езици.

fictions of the observers ~ Bars.1987:114
68-69-70 False memories, episode blending, source confusion cmp. LLMs Schacter 1995b, Moscovitch 1995; synthesis errors
71 - Williams, Hollan 1981 - think-aloud teqhnicque, how people recollect the names of classmates: 1. a specific context (a trip, a swimming pool), then search this context, finally: verifed the inf.
**Partial info** → construct a partial description → use it to recover new fragments, recursively. "Obviously the result will depend on the starting point and in particular on the specific context in which the memory reconstruction takes place." .. Retrieval. Generic memories: *"I must have gone to a hotel,"* - then possibly remember the specific hotel Kolodner 1984... (T: cmp. LLMs prompt eng. ... ReAct ?) 72 memory order effects 73 free recall, cued recall, recognition tests T: all Enlgish-speaking? 75-76 insertions, omissions, blendings with other episodes; Priming - the influence of the immediate or recent past on reasoning. 80 - perceptual order effect; mapping effect on perception
85 AMBR - Associative Memory-Based Reasoning, Kokinov 1988; first implemented 1994a; DUAL (Kokinov 1989, 1994...) AMBR2 (Kokinov 1998; Kokinov, Nikolv, Petrov, 1996 ...) Basic Principles: 1. Reintegrate human cognition: thinking, perception, memory, learning, language + analogy.
2. Apply recursively at the finer grain size of the subprocesses of the analogy-making.
3. Big unified models, single cognitive archit.; interact; ... p.88 ... DUAL - multiagent, microagents; graded participation, diff.agents in different contexts, tasks ... Agents: different involvement in diff. contexts; the degree of participation depends on motivational power (processing speed, activation level), which reflects the relevance of the knowledge the agent has to the current task and context.... *Hybrid:* symbolic and connectionist aspects. ... spreading activations & messages ... *Symbolic: Reasoning*; links, connect.: context relevance. Reasoning sets new goals, shifts the attention to diff. aspects of the environment ... Concepts, episodes, objects - distributed way, set of agents, a coalition. Permanent & temporary agents.
94-95 .. the problem: knife, ax, matchbox, water; in the forest: heat the water ... Bulgarian students, immersion heaters ... memory - priming, perception - context effects, reasoning - problem solving ... Mapping, complex emergent process, local marker-passing, structure-comparison. Constraint-satisfacton.
grounds ... semantic similarity or structureal consistency diff. agents work at diff. speeds, … Copycat, Tabletop, Metacat: codelets - very similar to DUAL agents. The agents are not copies of each other (unlike the "neurons" in typical ANNs); not learning at the time cmp. LLMs, agentic frameworks ...
AMBR1, AMBR2 ... 102 Constraint satisfaction network CSN
Хаджиева, Йовева
p. 113 "проект започнал през 1988 г., поддържан продължително време от

института по математика и информатика на БАН, а от 1993 г. - от катедрата по Когнитивна наука и психология в центъра за когнитивна наука на Централна и Източна Европа в Нов Български Университет (НБУ) и от грантове от фонда за наука на България. ... p.117: many Bulgarian researchers... "Vassil Nikolov, Marina Yoveva, Kalina Hadjiilieva, Silvia Ivanova, Milena Leneva, Ivailo Milenkov, Radu Luchianov, Maurice Greenberg, Sonya Tancheva, Slavea Hristova, Iliana Haralanova, and Alexandrina Bahneva. We are also grateful to our colleagues and mentors Peter Barney, Encho Gerganov, and Elena Andonova, "
Васил Николов, Марина Йовева, Калина Хаджилиева, Силвия Иванова, Милена Ленева, Ивайло Миленков, Раду Лучианов, Морис Гринберг, Соня Танчева, Славея Христова, Илиана Хараланова и Александрина Бахнева. Изразяваме благодарност и на нашите колеги и ментори Петер Барни, Енчо Герганов и Елена Андонова.
Ch 5. Toward an Understanding of Analogy within a Biological Symbol System, Keith Holyoak & John Hummel,;
Gentner, 1983, Struture-mapping: A theoretical framework for analogy.
structure-mapping theory of mapping ... : retrieval of a source analog from LTM, mapping to target in WM, gen & eval of inferences & induction of relational schemas (access, mapping, inference, learning) explicitly express relations; no domain-independent learning component for inducing new abstractions from analogical mappinng
biological symbol system, LISA - symbolic connectionism
*Ch.7*. Conceptual Blending and Analogy, Gilles Fauconnier ... 1993
Ch.8 Setting limits on Analogy: Why Conceptual Combination Is Not Structural Alignment, M.Keane, F.Costello
Ch.10 Emotional Analogies and Analogical Inference, P.Thagard, C.Shelley
SME - Falkenhainer, Forbus, Gentner, 1989
ACME - Holyoak, Thagard, 1989 … CWSG - Copying with substitution and generation, Holyoak, Novick, Melz, 1994 … case-based reasoning, Kolodner 1993 … VAMP - visual analogical mapping HOTCO - Thagard, 2000 Ch.6 ... coherence: constraint satisfaction: 6: analogical, conceptual, explanatory, deductive, perceptual, deliberative. … **Ch. 12**: Semantic alignment in Mathematical Word problems: Semantic alignment, structural alignment (cmp LLMs)
**13.** Analogical Reasoning in Children, Usha Goswami …
**15.** D.Hofstadter, Epilogue (the “shadow” analogy etc.) Sapir-Whorf hypothesis: Language and the Central Loop, 522 Goal-drivennes, central loop ... Percept. attractors; 523 Lexical items: words, names, phrases, proberbs ... shared vicarious experiences:: places, personages, events: in books, moveies, TV shows, ... unique personal memories - chunks ... 509 suggerimento - извеждащ пас?
511 - "думите и понятията са далеч от изпъкнали области от уместеното

пространство с правилна форма" - срвн. П.Грандерфорс (обратното) Различни употреби за различни ситуации
Central Cognitive Loop 513 - Словни смеси, Lexical blends : на ниво дума, израз, изречение … Roger Shank's Dynamic Memory, 1982 ...; high-level unlabeled node (memory record) -- fill the short-term memory --> context-dependent unpacking process ... Copycat, Tabletop; FARG, 1995 …
*** Относно учените от НБУ,** виж напр. обява за постдокторска позиция от 2004 г. и разнообразието от „предпочитани теми“ от центъра по когнитивна наука в НБУ и преподаватели, свързани с тях: https://linguistlist.org/issues/15/2225/

## * Мислене по аналогия, Разпознаване на образи чрез групиране (клъстериране), Сравняване, Pattern-matching, Когнитивна наука и теории за образуване на понятията, Синтез на програми, дискретно търсене, Program synthesis, Discrete program search, Analogy, Bongard, Hofstadter … #bongard

* Проблема узнавания, М.Бонгард, 1967 https://djvu.online/file/UEoEkTEgzHVZn
https://ru.wikipedia.org/wiki/Бонгард,_Михаил_Моисеевич

Хофстадер популяризира задачите на Бонгард в „Гьодел, Ешер, Бах…“, 1979 г. и съставя нови задачи. Хари Фундалис е докторант на Д.Хофстадер, след Мелани Мичъл и др.

**Harry E. Foundalis**[164]
**PHAEACO: A COGNITIVE ARCHITECTURE INSPIRED BY BONGARD’S PROBLEMS, Harry E. Foundalis, May 2006**
https://www.foundalis.com/res/Foundalis_dissertation.pdf
Виж „Проблема Узнавания“, М.Бонгард, 1967 и „Моделирование обучения и поведения“, М., "Наука" 1975.
**Бележки по дисертацията:**
Работно пространство – като работна памет, краткотрайна памет; срвн „контекст“ от “Моделирование обучения и поведения” в прлжн.
124 … Фиг. 6.3 понятийна мрежа по Хофстадер, 1979
8.3. Образуване на групи и модели-схеми (обобщени представяния)
236. Алгоритми за групиране: 1. “k-means” algorithm (McQueen, 1967): първите групи са случайни примери от входните данни; разпредели примерите между

[164] Благодаря на OpenMind (PtrMan), участник в discord MLST, в чийто гитхъб открих проекта му symVision и от него работата на Фундалис и Бонгард през 11.2024 г. https://github.com/PtrMan/symVision/tree/master

центровете и постепено преизчислявай координатите на центровете на групите
2. “Leader” (Hartigan, 1975) – първата група включва първия пример; присъедини следващия пример към най-подобната група или създай нова група, когато подобието надвишава определена граница; повтори втората стъпка, докато всички примери не са присъединени към група
3. “Lu and Fu” (Lu and Fu, 1978): При даден текущ пример, включи го към група, която включва пример, който е най-близък до дадения; ако приликата между тях превишава граница, създай нова категория от текщия пример; повтаряй първата стъпка, докато всички примери не се включат в група.
*Codelets, coderacks* … Copycat (Mitchell, 1990; Mitchell, 1993)., Tabletop, and Letter Spirit i… FARG systems – нямат истинска дълготрайна памет (LTM, long-term memory) Metacat (Marshall, 1999).– по-дълготрайна, но не се пази постоянно (факти от минали сесии); Slipnet – постоянно в оперативната памет … 248 „за да се сгъстят 15 години опит в 15 минути за обучение на програма, .. компютърът трябва да работи 500 хил. пъти по-бързо…“*  мащабирането на
Кое е фон и обект в задачите – понякога не е еднозначно; по рамките, търсена дебелина …
с.278 … … line segments, line intersections, curves, closed regions, angles, circles, dots, etc. … Многослойни и паралелни процеси за откриване, които се припокриват и започват да действат веднага щом получат достатъчно данни от предходни етапи.

Виж процеси в „ретината“ …  случайно обхождане с пораждане на „зърна“ за изследване, а не последователно линейно сканиране от горе-надолу, отляво-надясно* - и използване на събраната информация без забавяне; „височина“ (altitude) – разстояние между вътрешен пиксел в тъмна фигура и най-близкия контур;  медианни пиксели (median pixels) – ендоскелет …  детектор на отсечки (line segments); пресичащи се линии – продължения, сегменти на линии;
( О – проследяване на лъчи, кои излизат от оградено пространство;  единични точки; линии; вътрешен скелет (изтъняване); криви; „низове от линии“; откриване на тъмни петна чрез радиално пускане на лъчи за откриване на границите;
* съгласен съм с тези подходи; вж също „Столове, сгради, карикатури, …“, 2012
* ускорението се постигна чрез мащабиране в „ширина“ с GPU
с.306/328 – Основни приници… с.312(334) – таблица с площи на обекти от двете групи, които образуват вероятностно разпределение с два върха (графика) – сравни с М.Бонгард, „Проблема узнавания“, разделянето „на купчини“/групи като разделящ критерий в класификацията.
12.1 – връзка между мисленето по аналогия по школата на Хофстадер; „Феако“; и когнитивната лингвистика, метафори, образни схеми; пораждане на изображения за понятията и  – докосва, целува и пр.

*За волята у разнообразни видове агенти, дали тази познавателна архитектура „разбира“ (12.2+, с.338 (360)-345(367)) – задавайки въпроса за различни живи същества, постепенно свеждайки ги до елементарни частици; срвн Майкъл Левин „от физика до ум“, TAME и др.; Артур Шопенхауер – Воля и Представа, Волята в природата; ТРИВ;
* Сърл (Searle), китайската стая, скриптовете за ресторанти на Шанк (Schank), компютрите не разбирали; можело да разбира само ако е молекулно копие на мозъка, построено от „правилната материя“; как определя кой може да има намерения и пр. – ненаучно поставяне на въпроса. Вж и ЧиММ, 2001.
*Идеи от непубликувани работи. Вж. „Създаване на мислещи машини“.
(…)

* **Harry Foundalis - Research on the Bongard Problems**
https://www.foundalis.com/res/diss_research.html
Индекс с връзки към оригиналните задачи от Бонгард и добавените от Хофстадер, Фоундалис и др.: https://www.foundalis.com/res/bps/bpidx.htm
Основни принципи на познанието:
https://www.foundalis.com/res/poc/PrinciplesOfCognition.htm
Решения: https://www.foundalis.com/res/bps/bongard_problems_solutions.htm
https://en.wikipedia.org/wiki/Bongard_problem
* Unification of Clustering, Concept Formation, Categorization, and Analogy Making, Harry E. Foundalis
https://www.foundalis.com/res/Unification_of_Clustering_Concept_Formation_Categorization_and_Analogy_Making.pdf
* A Generalization of Hebbian Learning in Perceptual and Conceptual Categorization, Harry E. Foundalis, Maricarmen Martínez
https://www.foundalis.com/res/Generalization_of_Hebbian_Learning_and_Categorization.pdf

* BONGARD-LOGO: A New Benchmark for Human-Level Concept Learning and Reasoning, Weili Nie,
https://papers.nips.cc/paper/2020/file/bf15e9bbff22c7719020f9df4badc20a-Paper.pdf
Други:

*** Solving Bongard Problems with a Visual Language and Pragmatic Reasoning,** Stefan Depeweg, Constantin A. Rothkopf, Frank Jäkel
https://arxiv.org/pdf/1804.04452 Вж и Z. Pylyshyn. Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. Behavioral and Brain Sciences, 22:341–423, 1999

* V. Savova and J. B. Tenenbaum. **A grammar-based approach to visual category, learning**. In B. C. Love, K. McRae, and V. M. Sloutsky, editors, Proceedings of the 30th Annual Meeting of the Cognitive Science Society, pages 1191–1196, Austin, TX, 2008. Cognitive Science Society. https://www.researchgate.net/publication/228647295_A_grammar-based_approach_to_visual_category_learning
"one-shot learning, artificial categories with abstract structural characteristics; grammar-based approach to structure representation: visual category learning as an instance of grammar induction; one-shot learning as Bayesian inference over hypotheses, where grammars are hypotheses and learning amounts to picking the best grammar."

#multi-agent systems #mas #мулти-агентни системи #multiagent #мултиагентни системи

# * Мулти-агентни системи, планиране и роботика от 1980-те и 1990-те * Multi-Agent System, Planning and Robotics in 1980s-1990s

Разпределен ИИ, агенти и мулти-агентни системи от 1980-те и 1990-те – направление, което се пробужда в обновена форма в последните години чрез големите езикови модели, които се използват за вземане на решение и за обща обработка[165], вместо по-ранните „символни“ методи или други обработки. Предвестници са разработчиците на *Actor model* който води началото си от 1973 г.[166] с Карл Хюит (Carl Hewitt) в MIT – като начин за описание на конкурентни изчислителни процеси, в които „всички са актьори“ (агенти, деятели) и си предават съобщения. Хюит е автор и на езика за планиране Planner, вариант на който се използва в SHDRDLU. В увода на основополагащата статия за Модела на актьорите четем: „*актьорът е деятел, който изпълнява роля според подсказки от сценарий“.* Всякакви видове обекти в информатиката могат да се представят като особени случаи на актьори: структури от данни, функции, семафори, монитори, портове, описатели, семантични мрежи[167], логически формули, числа, имена (идентификатори), демони, процеси, контексти и бази данни. Целта на Планер и моделът е да кодира процедурни данни, актьорите имат „намерния“. Сравни с по-късните агентни модели Belief-Desire-Intention и Procedural Reasoning System (BDI, PRS).

Други основни учени, разработили на модела на изчислителните процеси като актьори, или агенти, през 1970-те и 1980-те са Peter Bishop, Richard Steiger, Irene Greif, Henry Baker, William Clinger, Gul Agha.

Сред пионерите на мулти-агентните системи, във връзка и с

---

[165] * Виж „Обща Обработка“, ОбОбр в Зрим, от *2014 г.*

[166] * https://en.wikipedia.org/wiki/Actor_model * https://en.wikipedia.org/wiki/Carl_Hewitt Carl Hewitt (1969). * *PLANNER: A Language for Proving Theorems in Robots* IJCAI'69. * Carl Hewitt, Peter Bishop and Richard Steiger (1973). A Universal Modular Actor Formalism for Artificial Intelligence https://www.ijcai.org/Proceedings/73/Papers/027B.pdf

[167] Семантични мрежи – „Quillian nets” – по името на автора:
* Quillian, R. A notation for representing conceptual information: An application to semantics and mechanical English para- phrasing. SP-1395, System Development Corporation, Santa Monica, 1963 …като код, в паметта се съхранява само изчистена от излишества процедура, подобна на граматика, която съхранява и обработва смисловото представяне и може да поражда много различни предварителни съхранени значения на думите.

роботиката, са Майкъл Джорджеф (Michael Peter Georgeff)*, Ананд Рао и Австралийският институт за Изкуствен интелект (1988-1999); М.Бартман, Й.Шохам (Y.Shoham, агентно-ориентирано програмиране, 1989), Родни Брукс (Rodney Brooks), Дейвид Уилкинс (David Wilkins), И.Фергюсън (Touring Machines, 1991-1992); Аарон Сломан (Aaron Sloman) и негови ученици; K.Sycara и др. Може да се открият примери и за „символен ИИ“, използва се LISP, продукционни системи, и за „подсимволен“.

Предсказването на бъдещето в тогавашната парадигма, поне от 1960-те години или дори 1950-те, се нарича **планиране**. Планирането, вкл. йерархично и в различни степени на абстракция заема важно място, тогава чрез търсене в пространство на състояния или възможности за действия. Системи за планиране: STRIPS, SIPE, SIPE-2 HTN (Hierarchical Task Networks). Търсене в дървета на решения и пр. В съвременната роботика и планиране: Model Predictive Control – търсене на най-ефективни траектории на движенията на роботи; търсене в дървета на решение чрез Monte Carlo Tree Search, MCTS (което е известно и през 1990-те в RL). Налице са и системи на по-ниско ниво като „Subsumption architecture” на Р.Брукс (1986), която се състои от йерархия от разширени крайни автомати и поведението ѝ се изгражда отдолу-нагоре с увеличение на абстрактността и обхвата на поведенията (най-ниско ниво – избягване на препятствия, да не се блъска, напр. да спира; второ – да заобикаля; трето – да се разхожда наоколо).

Всъщност идеите на абстрактно ниво важат и до днес, невронни модели пресъздават тези функции чрез обучение или в неявен вид или като краен резултат – виж по-долу бележките за съвременна система за навигация на робот по изображение и текст. Напр. логическите предикати, с които обикновено се описват задачите като lowering = wide → around(x), holding(x) over(table) → on (X, table) (от „Универсални планове за реактивни роботи в непредсказуема среда“, M.Schoppers, 1987) в невронно представяне се обучават с множество от входно-изходни двойки, като изображения, показващи съответното състояние и пр. или се генерира програмен код, който да управлява робот като PaLM-E[168].

Интересни са изследванията в школата около Аарон Сломан в

[168] PaLM-E: An Embodied Multimodal Language Model https://palm-e.github.io/ “How to grasp blue block…” – Как да хвана синьото кубче? Първо хвани жълтото блокче и го постави върху масата, после хвани сиьото кубче.“ Сравни с „SHRDLU”, 1968-1970 на Тери Виноград. https://en.wikipedia.org/wiki/SHRDLU От видеото след „Results” , “Bring me the rice chips from the drawer.” … „Adversarial Disturbance” - човек бута с прът пакета от ръцете на робота и той забелязва и посяга към новото място – сравни със „сътрудничеща“ или „несътрудничеща“ среда в бележките по-долу.

Англия, която работи до началото на 21-ви век.

* **Майкъл Питър Джорджев** – може би с български произход? по бащина линия – и с бащино име Петров. Ако данните от Ancestry са верни, *негови?* дядо и баба* се казват Никола и Петрана; Н. – роден през 1888 г. (*или на друга личност с подобно име?)

**Procedural Knowledge**, Michael Georgeff, Amy L. Lansky, November 1986, Proceedings of the IEEE 74(10):1383 – 1398 DOI:10.1109/PROC.1986.13639 https://www.researchgate.net/publication/2997588_Procedural_Knowledge

Поредици от действия за постигане на цели; процеси; планиране; език за описание на процедурно знание; поредиците от действия позволяват „ориентиране“ спрямо тях и намаляват необходимостта да се проверява средата – ако е изпълнена еди-коя си стъпка успешно, значи са налице и други предпоставки и условия; преосмисляна на плановете при изпълнението им – при среща на препятствия; различна сложност и мащаб на промените в плана; частично йерархично планиране с различна степен на подробности (срвн. с JEPA на Lecun) – препокриване на планирането (образуване на планове) и изпълнението им, т.е. създава се частичен, по-общо зададен план, решават се средствата за осъществяването му за кратък период и обхват напред, изпълняват се; впоследствие краткосрочният план се разширява още малко, изпълнява се и т.н. Системата обикновено няма достатъчно знания, за да може да разгърне плана в подробности на най-ниско ниво – особено, ако трябва да работи в действителността. … Реактивна система: възможности както за промяна на настоящия план, за да постигне целта; така и да може изцяло да промени фокуса си и да преследва нови цели в променена ситуация. Например при извънредни случаи, така че системата да може бързо да изменя намеренията си (планове за действия) въз основа на онова, което възприема в настоящето, както и заради онова, което вече вярва, има намерения да изпълни или желае. … Описание на динамични светове и знанието като процедури. … Достъпът до историята на действията може да освободи агента от строгата зависимост от сетивата – може да извлече повече информация за състоянието на света просто като знае какви са били предишните му действия (напр. докато готви не е нужно да опитва храната на всяка стъпка). *Състояния на света* – не само външния, но и *вътрешния*. Светът се променя чрез *действия (*или *събития*). По McDermott – те са множества от последователности от състояния, описани с логика във времето(„temporal logic”, „темпорална“). Как са породени действията? Чрез *процес:* абстрактен механизъм, който може да се изпълнява, за да поражда поредици от състояния на света,

наречени *поведение* на процеса (Вид симулатори на въображаеми вселени в ТРИВ). *Успешни и неуспешни действия* (които са се провалили) – необходимо е агентът да знае дали действието е завършило успешно, особено при частично планиране, защото това променя ситуацията и какво трябва да направи, за да довърши първоначалния план. Процесът като поредица от подцели или поведения. Мрежа от преходи с начален и краен възел и връзки с описания на подцелите. Изпълнението на процеса е преминаване по възлите по съответен път при успешно изпълненеие на подцел (поведение).
PRS – Procedural reasoning system – Процедурна система за разсъждения.
*База от данни* с познати в момента *факти (или убеждения)* за света; множество от *текущи цели (задачи)*, които трябва да се достигнат; множество от *утвърдени процеси (планове),* които описват процедури за достигане до дадени цели или отговаряне на определени ситуации; и *тълкувател* (механизъм за логически извод, inference mechanism) за работа, преобразуване на изброените компоненти. Във всеки момент системата *има стек за активните в момента процеси* – той може да се разгледа като настоящите *намерения* на системата.
(…) Някои процедури стават приложими когато системата научи за определени факти, т.е. те се *извикват от факти* (от събития, вид събитийно програмиране, бел. Тош) – за създаване на *реактивни системи*. Свързани с *неявни* цели – например ако се засече пожар, да се загаси, което осъществява дълбока неявна цел на всички организми – да останат живи (срвн. „Извод чрез действие“ на К.Фристън). … Процедури на мета-ниво: (Choose-best-process $Goal $List-of-procedures) (Избери-най-добрият процес $Цел $Списък-с-процедури) …
https://prabook.com/web/michael_peter.georgeff/3401991
https://en.wikipedia.org/wiki/Procedural_reasoning_system (PRS)

* **A Model-Theoretic Approach to the Verification of Situated Reasoning Systems**, A.Rao,M.Georgeff, 1993, https://www.ijcai.org/Proceedings/93-1/Papers/045.pdf
Системи, които са разположени или поставени в променлива, динамична среда (“ситуирани”, situated)… ограничени ресурси; „Ситуация“ е определено място във времето в определен възможен свят в разклоняващо се във времето дървовидна структура. CTL: разклоняваща се логика във времето; начална установка (initial configuration) …

*** BDI Agents: From Theory to Practice, Australlian Artificial Intelligence Institute**, A.Rao,M.Georgeff, 1995
https://web.archive.org/web/20110604050051/https://www.aaai.org/Papers/ICMAS/1995/ICMAS95-042.pdf
Situated Robotics, reactive to situations, … balance: reactive vs goal-directed behaviour; comitting to plans vs reconsidering. … ("exploration vs exploitation")
p.313 .. "1. Във всеки момент има много възможни начини по които средата може да се развие (формално, средата е *недетерминистична*) … 2. Във всеки момент са възможни множество различни действия и процедури, които системата може да изпълни (и системата е недетерминистична) 3. Във всеки момент има много възможни цели, които се изисква да бъдат достигнати от системата; 4. Действия или процедури, които най-добре постигат разнообразни цели и зависят от състоянието на средата (контекста) и са независими от вътрешното състояние на системата; … 5. Средата може да бъде усещана само местно (т.е. едно сетивно действие не е достатъчно, за да се определи състоянието на *цялата* среда); … 6. Скоростта на изчисленията и действията може да се изпълнява в разумни граници спрямо развитието на средата; … Възли на решенията и на късмета; … системата трябва да избере подходящи действия или процедури от разнообразие от налични възможности и да постигне по ефективен начин основните си цели при дадени ограничения на изчислителните средства и особеностите на средата, в която е разположена системата. … Дървета на решенията до възможни светове; … Вяра, желания, намерения (ВЖН); ситуации. Възможни светове, достъпни до вярата, до желанията, до намеренията [за какво може да си помисли, да поиска, да започне да преследва дадена цел; в Зрим/ВиР: ?В МдИ, ?ВМдсП, ?ВП … ; срвн. възможности за действие, affordances].

Силен или слаб реализъм – отношения между В.Ж.Н. В някои системи световете на намерения са подсветове на световете на желанията. Условия за посвещаване към план (commitment condition) и условия за прекратяване на посвещаването (termination condition).

Сляпо отдаден, целеустремен и отворен агент (blindly-committed, single-minded, open-minded agent). „Слепият" отказва да промени убежденията или желанията си, които противоречат с поетите задължения. Целеустременият може да промени убежденията си, което да го освободи от приети отговорности. Отвореният може да променя и

убежденията, и желанията, което води до отпадане на приетите задължения (промяна на целите).
Примерен тълкувател на ВЖН:

```
начално_състояние();
повтори
    възможности := пораждане_на_възможности(опашка_на_събития)
    избрани_възможности := обмисли(възможности)
    обнови_намерения(избрани_възможности)
    получи_нови_събития_от_средата()
     изтрий_успешни_становища()
     изтрий_невъзможни_становища()
```

Препратка към теория за разсъждения за действията по Братман, 1987. Разбор на средствата за достигане до целта (Means-End Analysis) – създаване на планове. Цели и подцели. Условия за извикване – спусък (invocation condition, trigger), предусловия; (и следусловия – виж PRS) (precondition, postcondition). … (Decision tree to possible worlds; deliberation; Belief-desire-intention; BDI-interpreter)
…
PRS има практически приложения напр. за диагностика на американската космическа совалка и за управление на полетите на летището в Сидни.

* M. Georgeff: The Representation of Events in Multiagent Domains, 1986
* M. Georgeff, Amy L. Lansky: Procedural knowledge, 1986
* M. Georgeff, Amy L. Lansky: Reactive Reasoning and Planning,1987
* A. S. Rao, M.Georgeff: Deliberation and its Role in the Formation of Intentions. UAI 1991
* Periklis Belegrinos, Michael P. Georgeff, A Model of Events and Processes. IJCAI 1991
* Anand S. Rao, Michael P. Georgeff: Asymmetry Thesis and Side-Effect Problems in Linear-Time and Branching-Time Intention Logics. IJCAI 1991
* David Kinny, Michael P. Georgeff: Commitment and Effectiveness of Situated Agents. IJCAI 1991
* Anand S. Rao, Michael P. Georgeff: An Abstract Architecture for Rational Agents. KR 1992
* F. Ingrand, M. Georgeff, A. S. Rao: An architecture for Real-Time Reasoning and System Control, 1992
* Anand S. Rao, Michael P. Georgeff: A Model-Theoretic Approach to the Verification of Situated Reasoning Systems. 1993
* David N. Morley, Michael P. Georgeff, Anand S. Rao: A Monotonic Formalism for Events and Systems of Events. 1994
* Michael P. Georgeff, Anand S. Rao: The Semantics of Intention Maintenance for Rational Agents. IJCAI (1) 1995
* Anand S. Rao, Michael P. Georgeff: BDI Agents: From Theory to Practice. ICMAS 1995
* David Kinny, Michael P. Georgeff, Anand S. Rao: A Methodology and Modelling

Technique for Systems of BDI Agents. 1996
* David Kinny, Michael P. Georgeff: Modelling and Design of Multi-Agent Systems. ATAL 1996
* Michael P. Georgeff, Anand S. Rao: A profile of the Australian Artificial Intelligence Institute. 1996
* M. Georgeff et al. The Belief-Desire-Intention Model of Agency. ATAL 1998
* A. S. Rao, Michael P. Georgeff: Decision Procedures for BDI Logics. 1998
* N. Jennings, K. Sycara, M. Georgeff: **Editorial. Autonomous Agents Multi Agent Systems.** 1(1): 5, 1998

Редакторска статия от *първи брой* на научното списание за автономни агенти и мулти-агентни системи:

**Multiagent Systems,** Katia P. Sycara, AI Magazine Volume 19 Number 2 (1998) https://link.springer.com/journal/10458
Обзор на MAS; проблемът за съгласуваността при решаването на задачите (problem-solving coherence): как се урежда съгласуването между самостоятелни агенти, които не са управлявани централно, а се приспособяват и договарят помежду си, разпределят си задачите, споделят състоянието на свършената от тях работа и т.н. „разпределен ИИ“; *„първите приложения се появяват в средата на 1980-те в производството, управление на процеси, управление на въздушния трафик и на информацията“;* разпределено наблюдение на превозни средства: distributed vehicle monitoring (DVMT) – преминаващи през различни пунктове, обединение на информацията; гъвкави производствени системи: FMS – YAMS, 1987 (Yet Another Manifacturing System) Тогавашните агенти в Интернет извличали информация от Интернет и я филтрирали, а следващото поколение щяло да събира информация в контекст и да извършва сложни разсъждения в подкрепа на задачи, решавани от потребителя*. Затова агентите трябвало да си взаимодействат и да се съгласуват всеки с всеки (peer-to-peer interaction),което изисква договаряне и сътрудничество. *Особеностите на МАС са, че всеки отделен агент има непълна информация или способности за решаване на задачи и затова – ограничен поглед; няма всеобщо управление; данните са децентрализирани и изчисленията, обработката, е асинхронна. Подбуди за интереса към МАС: решават по-големи задачи, които са отвъд възможностите на централизиран агент, позволяват по-добра и по-лесна комуникация между съществуващи разнородни системи, решават задачи които са естествено „обществени“ като построяване на графици и календари, управление на въздушния трафик или агенти за извършване на покупки в Интернет*.* По-голяма ефективност при разпределени сензори и експертиза, по-висока изчислителна ефективност, издръжливост към повреди на отделни части, надеждност при неопределеност (не „чупливост“), разширимост, гъвкавост, възможност агентите да се включват в разнообразни „отбори“, които да разрешават различни задачи. …

„Обмислящи/Deliberative” ~ BDI (Предположения/(вери, убеждения), желания, намерения).* RETSINA – мултиагентна инфраструктура на К.Сикара, обединяваща модули за планиране, съставяне на графици,събиране на информация и съгласуване с други агенти (planning, scheduling, information gathering, coordination with other agents). HTN (hierarchical task netowk) … схеми за намаляване на задачите (task reduction schemas) – как се изпълнява по-голяма задача като множество от подзадачи-действия и описание на потока от информация между задачите. KQML – език за съобщения, заявки за услуги, които се превръщат в цели за получаващия ги агент. *Модулът за съставяне на графици определя разписанието на стъпките от плана и избира кое точно действие, от множество от възможни изпълними действия, да бъде изпълнено в следващата стъпка.. Фиксира се намерение за изпълнението му, докато не се завърши от изпълнителния модул. За реактивните способности – отговарянето на непредвидени промени, прекъсване на плана и пр. – се грижи процесът за наблюдение на изпълнението. Той приема намерението за следващо действие и подготвя, наблюдава и изпълнява извършването му. Наблюдението на изпълнението подготвя действие за изпълнение като наглася контекста (вкл. резултатите от предишни действия) за новото действие*. … Може да ограничава времето за изпълнение, ако задачата не се изпълни – може да се прекъсне и задачата да се отбележи като неизпълнена, провалена; недовършените задачи се поемат от процеса за обработка на изключения. Агентът разполага с библиотека от независими от областта на приложение фрагменти на планове (структури на задачи), които могат да се търсят по цел, както и специализирани, особени за областта на приложение фрагменти на планове, които се извикват постепенно според текущите входни данни (сравни SIPE*, PDDL)… *Реактивни агенти* – нямат изричен модел на средата, поведението им възниква от взаимодействието между по-прости многослойни поведения или в йерархична верига: главен-подчинен … Агентът се определя като множество от противоречащи си задачи, като само една може да бъде активна по едно и също време; задачата – поведение на високо ниво, последователност от действия, за разлика от действията на ниско ниво, изпълнявани пряко от действениците (актуатори, ефектори). Не се учат, не използват историята и не планират. ***Хибридни***, обикновено 3 слоя са достатъчни. Средният – абстрактен, висшият – обществени отношения (вж TouringMachine). Организация – рамка за взаимодействието между агентите според приети роли; структури: херархична, общност от експерти (специализирани агенти), пазарна (общуват чрез цената, за която се договарят), „научна общност“ … Разпределяне на задачите; мулти-агентно планиране … съвместно намерение и съвместна отговорност (joint intention, joint commitment …). Разпознаване и разрешаване на конфликти … договаряне … PERSUADER (Sycara 1990a) … процес на премахване на ограниченията (constraint-relaxation) . Моделиране на други агенти … Управление на съобщенията и средствата (ресурсите) … Обучение…

**Бел:**
* SIPE – System for Interactive Planning and Execution; йерархичен планировчик. Виж в следващия раздел за Планиране.
* На най-ниско ниво: реактивни, „инструкции“, „изпълняват инструкции“, най-прости управляващо-причиняващи устройства.
* Сравни с днешните „бъдещи“ агенти, които изпълняват предвижданията.
* срвн. със „System 2/1” Kahneman, “Thinking Fast and Slow”.
* Още публикации от Michael Georgeff: https://dblp.org/pid/61/4054.html
* Katia Sycara: https://scholar.google.com/citations?user=VWv6a9kAAAAJ&hl=en
* S.Gurumurthy, S. Kumar, Katia Sycara: **Mame: Model-agnostic meta-exploration**, 5.2020 … **Meta-RL** - efficient exploration strategies; meta-RL – adapt to new tasks from a few examples; Proximal Meta-Policy search (ProMP) – Rothfuss et al.; Meta-Reinforcement Learning: a general policy for a distribution of tasks t from T, each one – different MDP (S,A,P,r,y, H) – similar state and action space, but different reward function r or environment dynamics P.
https://proceedings.mlr.press/v100/gurumurthy20a/gurumurthy20a.pdf
* Markov argumentation random fields, Yuqing Tang, Nir Oren, Katia Sycara. 3.2016 https://ojs.aaai.org/index.php/AAAI/article/view/9848/9707 probabilistic + symbolic; formal argumentation theory (Dung 1995) – *“how a justifiable “stable” set of arguments can be extracted from a large set built on the principle of reinstatement which has been confirmed by human experiments.”*
  * *Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. 77:321–357.*
  * *Richardson, M., and Domingos, P. 2006. Markov logic networks. Machine learning 62(1-2):107–136.*
* Reasoning with Uncertain Information and Trust, Murat Sensoy et al., 2013 - Dempster-Shafer Theory (DST); “basic probability masses” to subsets … Subjective Logic (Josang) – subjective opinions for uncertain statements.; binomial opinion; Data-to-Decisions
  * G. Shafer, A mathematical theory of evidence, Princeton university press, 1976.
  * F. Baader, D. L. McGuiness, D. Nardi, and P. Patel-Schneider, eds., Description Logic Handbook: Theory, implementation and applications, Cambridge University Press, 2002
 * Predictive Indoor Navigation using Commercial Smart-phones ∗ B.Kannan,F.Meneguzzi, M.Dias, K.Sycara, 2013 – MDPs, Hierarchal path-planning https://web.archive.org/web/20240427112245/https://meriva.pucrs.br/dspace/bitstream/10923/13381/2/Predictive_Indoor_Navigation_using_Commercial_Smart_phones.pdf

*** A roadmap of agent research and development,** Nicholas R Jennings, Katia Sycara, Michael Wooldridge, 25.11.1995/30.3.1996/1998
https://www.cs.ox.ac.uk/people/michael.wooldridge/pubs/jaamas98.pdf

По-подробен преглед от горния в някои страни ( с.) и в предисторията предисторията. … *situatedness*, *autonomy*, *flexibility* … calculated rationality – ранни планиращи системи в *ИИ*, не отчитат времето, може да открият най-добро решение, когато вече е твърде късно; behavioral AI, reactive AI, situated AI – съобразява се с ограниченията на действителността и нуждата да взема решения своевременно; Rodney Brooks – проявата на умствените способности (интелигентността) като *взаимодействие* меду деятеля и средата, и възникващи от взаимодействието между по-прости поведения – subsumption architecture, не символна, за разлика от другите тогавашни. Сбирка от поведения, изпълняващи задачи (task accomplishing behaviours) – крайни автомати, които свеждат сетивните данни до действия с правила от типа „ситуация → действие“, но също с обратна връзка от предишни действия. Няма символни разсъждения и търсене. Поведенията могат да си взаимодействат, едно може да потиска изхода на друго. Йерархия, по-ниско ниво – по-малко отвлечени; недостатъци – само локална информация, поради „възникването“ на сложните поведения с взаимодействие със средата, както и при много нива взаимодействията между тях, чисто реактивните агенти стават сложни или невъзможни за разбиране (срвн. с невронни мрежи, deep learning). Затова около 1990-те се разработват смесени архитектури (срвн. TouringMachine). … DAI (distributed AI), …*2.1.3. Human-Computer Interfaces – сегашната парадигма е пряко боравене (direct manipulation) – желателно е да може програмите да могат да поемат инициативата в пределени обстоятелства, вместо да чакт потребителя да изкаже „заклинание“ какво да правят. Сътрудничещи програми., полу-автономни агенти; „асистенти експерти“ или „цифрови икономи“ (digital butlers)* … Виж p.16, бел. на Nicholas Negroponte, Being Digital:

** “Агентът отговаря на телефона, разпознава кой се обажда, досажда ви само ако е необходимо и дори изказва благородна лъжа от ваше име. Същият агент е научен на такт, да разпознава откриващи се добри възможности и да уважава особеностите на характера.“ с.150*

** „Ако някой ви познава добре, той може да действа от ваше име много ефективно. Дори и Айнщайн не може да замени болната ви секретарка. Проблемът не е в коефициента на интелигентност, а в споделеното знание и в практиката то да се използва във ваша най-голяма полза.“ с.151*

** „Като всеки военен водач, който изпраща разузнавачи напред … и вие ще изпращате агенти да събират информация за вас. Агенти ще изпращат други агенти и процесът се умножава, като процесът започва на границата* между вас и агента, там където им преотстъпвате властта върху изпълнението на желанията ви“. с. 158*

* на границата - интерфейса, взаимлика

* N. Negroponte. *Being Digital*. Hodder and Stoughton, 1995 https://archive.org/details/beingdigital0000negr Н.Негропонте е директор на мултимедийната лаборатория в MIT. Избрани откъси в 5 стр. https://web.stanford.edu/class/sts175/NewFiles/Negroponte.%20Being%20Digital.pdf

…
"**Maxims**" – агент филтриращ електронната поща, учи се да поставя приоритети, да изтрива, препраща, подрежда и архивира писма от името на потребителя... непрекъснато прогнозира какво ще направи потребителят; ако сбърка, не се обажда, но ако познава, започва да предлага на потребителя какво да прави. Warren – MAC за управление на финансов портфей, обединява намиране и пресяване на информация: цени на акции, новини от финансовия свят, репортажи, анализи, отчети и пр. Отговаря на заявки, бди за интересни събития (нарастване на цена на определени акции над определен праг) и предупреждава потребителя. Автоматизирана електронна търговия - Kasbah. ..

* **Distributed intelligent agents**, K Sycara, A Pannu, M Willamson, D Zeng, K Decker, 1996, https://www.cs.cmu.edu/afs/.cs.cmu.edu/Web/People/softagents/papers/ieee-agents96.pdf – „Рецина" (Retsina), разпределена сбирка от софтуерни агенти за целенасочено извличане на информация, обединение и поддръжка на задачи за вимане на решения. … взаимлични деятели (interfce agents); Sofbot – единичен агент с общи знания, изпълняващ широк кръг от задачи за намиране на информация, зададени от потребителя (срвн. с ГЕМ,LLM). *2. "Dseirable Agent Characteristics…": Желателни особености на агента: * да може да бъде насочван от човек или други агенти; * да бъде мрежов – разпределен и самоорганизиран; подвижен, може да се премества; * полу-самостоятелен, а не под непрекъснато пряко човешко управление напр. при събиране на информация. * постоянен – да работи продължително без човешки надзор и намеса; * да може да му се има доверие, да е надежден; * предвиждащ, предусещащ задачите, ролите и моделите на ситуациите, докато се учи да служи като интелигентен „кеш", междинна памет, която задържа информация, която може да бъде нужна; * деен, активен: сам да започва решаването на задачи, напр. да наблюдава инфосферата за появата на определени модели-схеми, събития; да предусеща нуждите на потребителя от информация, да предлага на вниманието информация, която е подходяща за ситуацията, и да решава кога да я покаже в суров вид или преработена и съчетана под друга форма; * да си сътрудничи с хора и други машинни агенти; да може да се справя с разнородността на други агенти и източници на информация; * да се приспособява към нуждите на потребителите и средата на задачите.* …

* **Commitment and Effectiveness of Situated Agents**, David N . Kinny, Michael P. Georgeff, 1991 https://www.ijcai.org/Proceedings/91-1/Papers/014.pdf
Пример за ранни симулатори, въображаеми светове, в които действат агенти: „*Tileworld*" – срвн. „Grid world". Все още се използват при учене с подкрепление RL и др.
* **Introducing the Tileworld: Experimentally evaluating agent**

**architectures**. [Pollack and Ringuette, 1990] Martha E. Pollack and Marc Ringuette. In Proceedings of the Eighth National Conference on Artificial Intelligence, AAAI-90, pages 183-189, Boston, MA, 1990.
*„Симулиран агент робот и среда, която е едновременно динамична и непредвидима. И роботът, и средата могат да се настройват с много параметри, чрез което да се управляват определени техни особености. Така можем опитно да изследваме поведението на разнообразни стратегии за разсъждения на мета-ниво като настройваме параметрите на агента, и можем да оценяваме успеха на различни стратегии в различни среди като съответно настройваме параметрите на средата. …“*
* Виж също проекта на Михаил Бонгард и колеги „Животно“; Проблема Узнавания, 1967 и др.
* М. М. Бонгард, И. С. Лосев, М. С. Смирнов, ПРОЕКТ МОДЕЛИ ОРГАНИЗАЦИИ ПОВЕДЕНИЯ — « ЖИВОТНОЕ », *Сб. "Моделирование обучения и поведения", М., "Наука", 1975* https://keldysh.ru/pages/mrbur-web/misc/bongard.htm

*** Моделирование обучения и поведения.** М., "Наука" 1975.
https://www.keldysh.ru/pages/mrbur-web/misc/mlb/

* **Measuring the effectiveness of situated agents**. David N. Kinny, 1990, Technical Report 11, Australian AI Institute, Carlton, Australia, 1990

* **TRIANGLE TABLES: A PROPOSAL FOR A ROBOT PROGRAMMING LANGUAGE** Technical Note 347 February 1985 By: Nils J. Nilsson, Senior Staff Scientist https://www.ijcai.org/Proceedings/87-2/Papers/096.pdf
https://en.wikipedia.org/wiki/Belief_revision

* **Universal Plans for Reactive Robots in Unpredictable Environments,** M.J . Schoppers, 1987 https://www.ijcai.org/Proceedings/87-2/Papers/096.pdf

Поставя важния въпрос за принципната неопределеност и непредсказуемост в една или друга степен, невъзможността да се планира надеждно в подробности за целия път и предлага решение чрез всеобщо планиране, създаване на универсален метод за планиране – действие – при всички възможни ситуации. Абстрактни планове, частични планове и реактивни процедури. Понятието за сътрудничество или несътрудничество на света/средата (cooperative or nocooperative world) – когато средата се държи в рамките на очакванията от модела на системата, т.е. средата *„сътрудничи“*, линейното планиране е равнозначно. При „*саботаж*“ обаче стратегиите са различни – намеса на чужда сила, разбъркване на средата, размества предметите, някой нарочно нарушава равновесието на ходещ робот, както в опитите на Boston Dynamics, в които ги ритат, бутат с прът и пр.); непредвидена ситуация, сблъскване с неочаквана новост (serendipity) или непозната ситуация.

Тогава агентът трябва да може *да се приспособи и да измени* предварително приготвения план или да действа спрямо момента, реактивно, а *не стереотипно*. Към планера, планировчика, планиращата система се подават примитивни действия (преобразувания) и техните следствия и странични ефекти, които се съчиняват в последователности. **Ограничения в средата:** описания на свойствата на света.

* **Intention, Plans, and Practical Reason, Michael E. Bratman,** 1987
https://archive.org/details/intentionplanspr0000brat/mode/2up?view=theater
Философската теория в основата на BDI/ВЖН. Задълбочен прецизен анализ на смисли и употреби. Частични планове, обмисляне, изпълняване в момента или отложения намерения, все още незапочнати и неприети за изпълнение планове, зададени без пълните им необходими подробности (имам намерение да отида на кино в неделя). Срвн. частичните планове с подканите (prompts) в езиковите модели. Също MPC, роботика, Mujoco MPC и роботът „Атлас“ – напр. планиране за няколко секунди напред, изпълнение на секунда и през това време планиране на следващите няколко секунди, защото междувременно светът може да се е променил и старият план вече не отговаря на условията в средата.  Бъдещите планове могат да бъдат доуточнявани или отменяни поради други възникнали междувременно събития, предпочитания. Поетите за изпълнение отговорности са в процес на изпълнение и се прекъсват по друг начин.  Различаване между усилия, полагани и положени за постигане на нещо (endeavour); и намеренията като осъзната желана цел, какво се изисква да се постигне (някои промени в средата могат да се случат като следствия на други действия, странични ефекти, без да са били осъзнати като изрични цели. Във ВиР/ТРИВ/“Анализ на смисъла на изречение…“/Зрим срвн. странични ефекти, жр/нжр: желана разлика, нежелана разлика и останалите препратки от литературата към дадените статии.

*** TouringMachines: Autonomous Agents with Attitudes,** Innes A. Ferguson, 1992,. *„Разсъждението … по същество включва търсене на разлики между действителното поведение и предвиденото от модела, а в случай на модел на себе си [рефлективен], между действителното поведение на агента и желаното от него. Предвижданията се образуват чрез проектиране във времето на параметрите, съставящи вектора на конфигурацията на моделирания обект в контекста на настоящата ситуация в света и приписаните намерения на обекта. Забелязването на разлики между действителното и предсказаните (или желани) поведения обаче не винаги кара агента да преразгледа изцяло „грешния“ модел, защото всеки от параметрите има горна и долна граница, …“*

* **Innes A. Ferguson. TouringMachines: An Architecture for Dynamic, Rational, Mobile Agents. PhD thesis,** Computer Laboratory, University of Cambridge, Cambridge, UK, 1992
https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-273.pdf

Трислойна архитектура: реактивна, планираща, моделираща, реализирана като симулатор на превозно средство, построена в духа на парадигмата Belief-Desire-Intention. Системата съчетава бързо реактивно ниво за извънредни непредвидени ситуации като внезапно препятствия; среден планиращ слой за движението и най-високото ниво за метапланиране и търсене на причинно-следствени обяснения, вземане на решения при противоречиви цели, избор на фокус на вниманието поради ограничените познавателни средства и прогнозиране. Моделиращият слой има предвиждащ модул, който построява времепространствени проекции или симулации за всеки моделиран обект – други агенти от средата.

Агентът ритмично получава сетивни данни, в които се съдържат координати и вид на обекти (бордюр, кола, светофар) и от тях се построява вътрешен модел. За „колите“ например се представят конфигурация, убеждения (вери), желания и намерения. Конфигурацията е последното известно място в двуизмерен свят: x,y; скоростта, ускорението, посоката на движение (насочеността, ориентацията)  и множество от сигнали, които е дало – дали е спирало, „натискало клаксона“, пуснало мигача, за да покаже, че ще завива.  (…) Третото ниво поражда очаквания (предвиждания) …

Дисертацията започва с обзор на архитектури за агенти,.  Въведение в теорията на агентите и и богат обзор на много други ранни архитектури, освен представената от автора. Три вида: обмислящи, реактивни и смесени (deliberative, non-deliberative, hybrid). Виж напр. subsumption architecture на Rodney Brooks.

**Сравни TouringMachine с:**
*** Tartan Racing: A Multi-Modal Approach to the DARPA Urban Challenge**, April 13, 2007, Chris Urmson et al.
https://web.archive.org/web/20111024165401/https://archive.darpa.mil/grandchallenge/TechPapers/Tartan_Racing.pdf
https://web.archive.org/web/20080119035454/http://www.darpa.mil/GRANDCHALLENGE/Teams/Tartanracing.asp

От доклада може да забележим, че в състезанието "DARPA Urban Challenge" през 2007 г. колата робот „Boss" на университета „Карнеги Мелън" всъщност прилага и доразвива идеи, описани и разработени в TouringMachine и неговия симулатор, в който се движат агенти коли, които трябва да стигнат между целеви точки (waypoints) в определен срок, но по пътя срещат светофари, кръстовища; бордюри, препятствия и други автомобили, които може да им пресекат пътя, да се движат много по-бавно от тях и пр. След като агентът забележи друг обект, той трябва да прогнозира действията и траекторията му, дали няма да се пресече с неговото планирано движение и при нужда да препланира като намали скоростта, опита да заобиколи и пр. Нужно е да предполага какви са намеренията на другите водачи, да мисли за цели на няколко нива – най-важно е безопасността (да не се блъсне в препятствия), после да спазва правилата за движение, да мине през зададените целеви точки и да стигне в рамките на зададеното време. И т.н. Както е посочено във въвеждаща лекция от курса по УИР от 2010 г., в архитектурата на този робот се използват идеи, които важат и за общ ИИ. Виж също когнитивната архитектура 4D-RCS, Real-Time Control System.
*самоуправляващ се автомобил/кола/превозно средство, „автономен автомобил"; self-driving car, self-driving vehicle.

**Виж също: * CausalCity: Complex Simulations with Agency for Causal Discovery and Reasoning,** Daniel McDuff, Yale Song et al. 25.6.2021
https://arxiv.org/pdf/2106.13364

*** Подробна лекция по мулти-агентни системи от 2001 г. от Olivier Boissier (основно на англ.) : Multi-Agent systems - Agent's Architectures –** основни видове: самостоятелни, общуващи, взаимодействащи … (autonomous, communicative, social)
https://www.emse.fr/~boissier/enseignement/sma01/pdf/agent.pdf
Поведението на повечето агенти от този период се изчислява „символно", но съществуват и архитектури с крайни автомати като SMAALA: Ferrand N. et al., 1997/1998, Systèmes multi-agents réactifs et résolution de problèmes spatialisés (МАС за разрешаване на пространствени задачи).
https://www.researchgate.net/publication/260762819_Systemes_multi-agents_reactifs_d'inspiration_physique_pour_la_prise_de_decision_et_la_resolution_de_problemes (френски)

https://en.wikipedia.org/wiki/Agent-oriented_programming
https://smythos.com/artificial-intelligence/agent-oriented-programming/
https://en.wikipedia.org/wiki/Java_Agent_Development_Framework
* Shoham, Y. (1990). Agent-Oriented Programming (Technical Report STAN-CS-90-1335). Stanford University: Computer Science Department.
* Shoham, Y. (1993). "Agent-Oriented Programming". Artificial Intelligence. 60 (1): 51–92. CiteSeerX 10.1.1.123.5119. doi:10.1016/0004-3702(93)90034-9.

*** An Overview of Agent-Oriented Programming, Yoav Shoham** (a chapter from a book) – глава от книга, ок. 1995 г.
https://www.infor.uva.es/~cllamas/MAS/AOP-Shoham.pdf
Агентът има „умствени състояния“, убеждения, приети отговорности, правила за поведение, избори и пр. и се управлява чрез тях, чрез особен вид съобщения от вида на „осведомяване, молба, предложение, обещание, отказ…“ и др., определени и действащи на по-високо логическо ниво и с повече ограничения спрямо обектно-ориентираното програмиране и др.; AOP е продължение на Актьори на Хюит, 1977 г. като състоянието е разширено до „умствено“ състояние, което се състои от убеждения/вери за света, за себе си и за другите агенти, способности, избори и т.п. Изчислителният процес, обработката на информация чрез агенти се състои в информиране/осведомяване, изпращане на молба/заявка, предлагане, приемане, отказ, съревнование или помагане един на друг като понятията са заимствани също от перформативите в езикознанието (speech acts).
https://en.wikipedia.org/wiki/Performative_utterance
Temporal belief maps – карти на промяната на убежденията на агента. … Вътрешна съгласуваност (internal consistency) между убежденията („верите“, beliefs) и задълженията (obligations); Добра воля – агентът се ангажира само с онова, в което вярва, че е способен, и наистина го мисли. Интроспекция, самонаблюдение: осъзнава задълженията си. Постоянство на умственото състояние; обновяване и преразглеждане на верите (belief update, belief revision); пример от 1990 г. на езика за агентно програмиране Elephant2000 за система за резервация на самолетни билети с диалог на естествен език.
**Виж също:**
* https://en.wikipedia.org/wiki/Actor_model *
Australian_Artificial_Intelligence_Institute
* SRI_International *Automated_planning_and_scheduling
* Motion_planning * Software_agent
* Distributed_artificial_intelligence#Agents_and_Multi-agent_systems
* Procedural_reasoning_system * Problem_solving
* Pathfinding * Navigation_mesh * Guided_local_search
* Probabilistic_roadmap

*** Toward Team-Oriented Programming**, 12.2006, Conference: International Workshop on Agent Theories, Architectures, and Languages, David V. Pynadath, Milind Tambe , Nicolas Chauvat, Lawrence Cavedon
https://www.researchgate.net/publication/225267869_Toward_Team-Oriented_Programming

Пример за по-късна среда за мулти-агентни системи, която синтезира основните цели и полза от тях: обединяа и съгласува общуването между разнородни агенти, написани на различни езици, в различни среди и пр., за потребителски интерфейси, роботи в космоса, виртуални среди за обучение и извличане на информация от Интернет (подобно на някои от предназначенията на агентите с ГЕМ/LLM). Чрез съчетаването на разнородни агенти и съвместната ми координирана работа – „работа в екип“, „teamwork” – да се решават нови сложни задачи, по задание от потребителя, като се запази модулността на системата, вместо да се изграждат еднородни (монолитни) системи, с цел да се намалят усилията в разработката на програмни продукти. Агентите самостоятелно да разсъждават и т.н. Съвместно изпълнение изисква съгласуване … Пример с планиране на пътища на летателни апарати … Йерахрхия за постигане на целта на отбора, модел на началните условия, при които отборът ще започне да преследва целта; модели на условията за достигане до целта, за неуместност или недостижимост, когато отборът трябва да прекрати преследването на дадена цел; съгласуване на ограниченията между агентите, изпълняващи съвместните дейности. „Отборно-ориентирани програми“ … „Голям отбор“ → … 1) Получател на заповеди (Orders-Obtainer) 2) Планировчик на пътищата 3) Осигуряване на безопасност 4) Летателен отбор → 4.1) Пренася → Отдели … 4.2) Ескортира -→ 4.2.1 Водач 4.2.2 Преследвачи … casADi https://github.com/casadi/casadi https://web.casadi.org/ https://web.casadi.org/docs/#initial-value-problems-and-sensitivity-analysis

# * Планиране * Класическо планиране #plan #планиране #planning
# * Classical Planning

https://docs.nav2.org/ - NAV2 пакет за навигация за ROS - Robot Operating System; https://wiki.ros.org/smach task-level architecture for complex robot behavior David E. Wilkins: https://www.researchgate.net/profile/David-Wilkins-6 SIPE, SIPE-2, SRI; **System for Interactive Planning and Execution**
SIPE-2 architecture: https://www.ai.sri.com/~sipe/architecture.html **…** resources, replanning, plan critics, … https://www.ai.sri.com/~sipe/
SIPE-2 HTN Planner by David Wilkins, Ai Austin, 2.2013
https://www.youtube.com/watch?v=gE0wPgT2qrw
1 min: Два вида: а) планиране с примитивни действия б) използват знания за областта на приложение като SIPE. …Мн.нива на абстракция, паралелни действия, следствия зависими от контекста, ограничения (числови разсъждения, времеви), средства (ресурси; преизползваеми или изчерпаеми), евристики (разсъждения над действията); препланиране по време на изпълнение; интерактивен ГПИ

* **High-Level Planning In a Mobile Robot Domain**, 15.7.1986, Technical Note 388, https://www.sri.com/wp-content/uploads/2021/12/573.pdf
* **Hierarchical Planning: Definition and Implementation,** D. Wilkins, 20.12.1985 https://www.sri.com/wp-content/uploads/2021/12/1382.pdf
Разграничаване на ниво на абстракция и ниво на планиране. Първото – гранулираност, финост (granularity) на разграниченията, които може да извърши в света*. Второто може да е стъпки от процеса на планиране, които от едно състояние пораждат по-подробно описание на действия, по-подробен план – ниво на планиране, но в някои системи може да е не едно и същи ниво на абстракция. На по-отвлечените планове отговаря по-голямо множество от възможни състояния на света, които го удовлетворяват. По-конкретните – намаляват множеството. План осъществим на високо ниво може да е неосъществим в реалния свят, защото е идеализиран. Понякога е за предпочитане търсенето да започне от целта, като се избере подходящо действие там, например при планиране на полет до град, до който има много летища – изборът на летище, или пък ако се пътува с кола – хотел, в който ще се отседне, и може да се влезе в града от различни пътища. В други случаи търсенето в дълбочина от текущото положение е по-добър вариант. При първо планиране на крайните действия – целесъобразно, „опортюнистично“ (expedential). Недостатъци на STRIPS – приема се, че при извършване на действие, светът не се променя, освен ако изрично не е посочено друго. Това не винаги може да се осигури, защото например ако на дадено ниво на планиране състоянието се задава като P1;Q1;G: P1 – действие, което прави предиката P истинен, а Q1 е действие, което прави по-абстрактен предикат Q истинен, и G е целта, която зависи от истинността на P, STRIPS ще приеме, че

P е истина при достигане на целта G, защото Q1 не споменава промени при по-малко абстрактни предикати, а всъщност истинността на P може да зависи от начина, по който Q1 се разширява на по-ниски нива на абстракция. Така ще се направи погрешен извод. Други йерархични планиращи системи, освен SIPE: MOLGEN, NOAH, NONLIN, ABSTRIPS – разширяват всеки възел от плана с някой оператор, описващ действие. Вж. с.16/17 за примерни планове.
* Разделителна способност на възприятие и управление в Теория на разума и вселената.

* **Granularity**, Jerry Hobbs, 1985, https://www.ijcai.org/Proceedings/85-1/Papers/084.pdf … explicitly express the granularity, grain size … всеобща теория, заедно с голям брой относително прости, идеализирани, зависещи от зърността местни теории, взаимносвързани с аксиоми за отчленяване (global theory, grain-dependent, articulation axioms). *„В сложни ситуации: отделяме (abstract) съществените особеости от средата, определяйки зърността, избираме съответни местни теории. Това е единствената изчислителна обработка, извършена от всеобщата теория. След това местната теория се прилага върху по-голямата част от процеса на решаване на задачата. Когато се налага смяна на гледната точка, задачата трябва да се премести от една местна теория до друга и тогава се прилагат аксиоми за отчленяване.“*
* Виж ТРИВ; „Принципи на Разума“, Т.А., 2009, лекция в ТУ София, „Нощ на учените“. В Зрим всеобщата теория е свързана с „Разпознавател на контекст“ и „Избирател на контекст“, „Контексни белези“.

*** Hierarchical Planning at Differing Abstraction Levels,** David E. Wilkins, December 1988
https://www.researchgate.net/publication/300955583_Hierarchical_Planning_at_Differing_Abstraction_Levels
Using The Sipe-2 Planning System, D.Wilkins, 1997
The Act Formalism, K.Myers, D.Wilkins, 1997
https://www.researchgate.net/publication/2771821_The_Act_Formalism
Achieve, achieve-by, achieve-all, wait-until, test, conclude, retract, require-until, use-resource; Environment slots: name, cue, precondition, setting, resources, properties, comment; Task → many possible plans …

*** PDDL** - The Planning Domain Definition Language, 8.1998, M.Ghallab, C.Knoblock, D.Wilkins, A.Barett … (12)
https://www.researchgate.net/publication/2278933_PDDL_-_The_Planning_Domain_Definition_Language
* A Multiagent Planning Architecture, D.Wilkins, K.Myers, 4.2000
https://www.researchgate.net/publication/2628989_A_Multiagent_Planning_Architecture
* Teambotica: A Robotic Framework for IntegratedTeaming, Tasking, Networking,

and Control, R.Vincent et al., 2003
https://www.researchgate.net/publication/221456023_Teambotica_A_Robotic_Framework_for_Integrated_Teaming_Tasking_Networking_and_Control

* **Generating Instructions at Different Levels of Abstraction**, A. e Kohn et al., 2020 https://aclanthology.org/2020.coling-main.252.pdf – за играта Minecraft чрез HTN; construction/instruction planning … block(x,y,z) … put-block(x,y,z) … sentence generator … "3. … действия за инструкции е абстрактно семантично рпедставяне; генераторът на изречения го превежда в израз на естествен език като „постави блок върху жълтия блок", „построй под от черен блок до жълтия блок", „сега ще те науча как да строиш греда"" … Срвн. с SHRDLU и моделите за управление на роботи от 2022-2024 г. Срвн. с агентни архитектури от 1980-те, subsumption, hybrid (deliberative & non-deliberative (reactive))
* A HYBRID TASK PLANNER ARCHITECTURE FOR PICK AND PLACE SEQUENCING, S.Yayilgan et al, 2000
https://www.researchgate.net/publication/267317547_A_HYBRID_TASK_PLANNER_ARCHITECTURE_FOR_PICK_AND_PLACE_SEQUENCING

https://en.wikipedia.org/wiki/Task_analysis_environment_modeling_simulation
https://en.wikipedia.org/wiki/Multi-agent_planning
https://en.wikipedia.org/wiki/Multi-agent_reinforcement_learning Сътрудничество и съперничество … чисто/нулев-сбор (zero-sum game); смесено
https://en.wikipedia.org/wiki/Cooperative_distributed_problem_solving
https://en.wikipedia.org/wiki/Multiscale_decision-making
https://en.wikipedia.org/wiki/Stanford_Research_Institute_Problem_Solver – начално състояние; описание на целеви състояния – положения, до които трябва да достигне планиращата система; множество от действия, като за всяко действия са дадени предусловия и следствия (следусловия) – какво се случва след извършване на действието. Наредена четворка <P,O,I,G>. P – условия, O – оператори <a,b,g,d> набор от условия, по ред: кои да са верни, кои неверни, кои стават верни след изпълнението, кои стават неверни (истинни/неистинни – двоични логически променливи). I – начално състояние, множество от условия, които са истинни в началото, а другите се приемат за неистинни. G – описание на целевото състоние, <N,M> – кои условия трябва да са истина и неистина. Планът е последователност от оператори, които водят от началното, до крайното състояние. Функция на преходите …

**Йерархични мрежи за задачи** (HTN, Hierarchical Task Network) – основни, примитивни задачи (начално състояние), съответстващи на действията в STRIPS; съставни задачи (междинни състояния) – съставени от множество от прости задачи; 3: целеви задачи (целеви състояния): цели в STRIPS, но по-общи. Примитивните състояние могат да се изпълнят без разлагане, ако предусловията са изпълнени. Съставните задачи са частично наредено множество от следващи задачи, примитивни или абстрактни. Целеви задачи:

удовлетворяват условия. Мрежа от ограничения, напр. предусловия, които трябва да се осъществят от предишни състояния.
https://en.wikipedia.org/wiki/Hierarchical_task_network

*** Hierarchical Task Network Planning: Formalization and Analysis**
https://www.cs.umd.edu/projects/plus/HTN/
UMCP: Universal Method Composition Planner
https://www.cs.umd.edu/projects/plus/umcp/
PDDL: Planning Domain Definition Language: (1998-2000) D.McDermott
https://en.wikipedia.org/wiki/Planning_Domain_Definition_Language
Множество от възможни действия, особено начално състояние във света и множество от желани цели. Описанието на действията: предпоставки за действието и следствия от извършването му. Двустепенност: описание на област, което важи за всички задачи от тази област, и описание на конкретната задача. … PDDL+ NDDL: New Domain Definition Language: NASA, 2002 … MAPL (Multi-Agent Planning Language): PDDL 2.1 … 2003 …OPT – Ontology with Polymorphic Types .. PPDDL – Probabilistic PDDL 2004-2006 – вероятностни следствия: дискретни, вероятностно разпределение на възможните следствия от действие, функции на наградата (reward fluents) за промяна на общата награда за план и следствие; награда за достигане на цел (на траектория, която включва поне едно целево състояние), функции на постигнати цели – двоична булева променлива, „вярно“, ако траекторията е преминала поне през едно целево състояние. Така PDDL 1.0 осъществява планиране чрез процеси на Марков за вземане на решения (Markov Decision Process, MDP); APPL – Abstract Plan Preparation Language, NDDL2006 - …RDDL – Relational Dynamic Influence Diagram Language 2011, MA-PDDL (Multi-Agent PDDL… PDDL3.1+ 2012 … р.дств за р.дтл , р.цели/метрики; взаимодействия; наследяване и полиморфизъм за дств,цл и метрики.

**Пример на PDDL (STRIPS):**
https://web.archive.org/web/20120227123050/http://www.cs.toronto.edu/~sheila/384/w11/Assignments/A3/veloso-PDDL_by_Example.pdf M.Veloso, 2002, Carnegie Melon:на областта:

```
(define (problem strips-gripper2)
   (:domain gripper-strips)
    (:objects rooma roomb ball1 ball2 left right)
     (:init (room rooma)
            (room roomb)
            (ball ball1)
            (ball ball2)
            (gripper left)
             (gripper right)
             (at-robby rooma)
             (free left)
             (free right)
             (at ball1 rooma)
            (at ball2 rooma))
      (:goal (at ball1 roomb)))
(define (domain gripper-strips)
  (:predicates (room ?r) (ball ?b) (gripper ?g) (at-robby ?r)
               (at ?b ?r) (free ?g) (carry ?o ?g))
```

**Задача:**

```
(:action move
 :parameters (?from ?to)
 :precondition (and (room ?from) (room ?to) (at-robby ?from))
  :effect (and (at-robby ?to) (not (at-robby ?from))))
(…)
…
```

https://en.wikipedia.org/wiki/Hierarchical_task_network
https://en.wikipedia.org/wiki/Markov_decision_process

*** High-Level Planning in a mobile robot domain**, David E. Wilkins, 15.7.1986, Technical Note 388, SRI International SIPE - High-level planning
https://www.academia.edu/89295896/Schematic_classification_problems_and_their_solution?uc-sb-sw=57148408

* See also planning with SAT solvers (Boolean Satisfiability Problems) and Constraint satisfaction problems (CSP) * https://en.wikipedia.org/wiki/SAT_solver
* https://en.wikipedia.org/wiki/Boolean_satisfiability_problem

## * Школата около Аарон Сломан в Англия в когнитивните архитектури

*** The Mind as a Control System, Aaron Sloman, 1992**
Умът като управляващо устройство с множество взаимодействащи си управляващи вериги (връзки, цикли) от разнообразен вид, повечето от тях осъществени като виртуални машини на високо ниво, и много от тях – организирани йерархично. … Разграничава ума като „изчислителен“, но в тесен смисъл, явно като „линеен“ и „прост“ (по някаква мярка) алгоритъм.
* What sort of architecture is required for a human-like agent? Aaron Sloman, 1997 https://cogaffarchive.org/Sloman.what.arch.pdf
* Architectural Requirements for Autonomous Human-like Agents, Aaron Sloman, 1997 https://cogaffarchive.org/Sloman-dfki.pdf
p.13: “Control states of varying scope and duration…” – състояние за управление с разнообразна гледна точка и продължителност, където *„по-висшите“ състояния: са по-трудни за промяна; по-дълготрайни; податливи на повече влияния; действията им са по-общи и по-непреки; по-скоро са генетично предопределени.“*

* A Concern-Centric Society-Of-Mind Approach To Mind Design, Steve Allen,https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=6d19480fb8c37c4a8f7360984588c764562e4028
"Concerns are broadly defined as dispositions to desire the occurrence, or non-occurrence, of a given kind of situation [Frijda 86, page 335]." Frijda, N. H. (1986). The emotions. Cambridge University Press; Editions de la Maison des Sciences de l'Homme.
**„Грижите“** – склонности и желания за случването или неслучването на определен вид ситуации. По същество друг вид „предсказващо моделиране“. Сравни с „грижата“ (Care) в школата на Майкъл Левин.
* Allen, S. (2001). Concern Processing in Autonomous Agents. PhD Thesis, School of Computer Science, University of Birmingham.
https://cogaffarchive.org/allen-thesis/Thesis_PDF.pdf
Stephen Richard Allen: https://www.curiousdog.org/Steve/index.htm#cogaffect
Интересна архитектура, приложение на „надграждащата архитектура“, „subsumption architecture” на Родни Брукс (Rodney Brooks), която изгражда по-сложни поведения въз основа на по-прости от по-долни нива.
* Wright, I. P. (1997). Emotional Agents. PhD Thesis
https://cogaffarchive.org/Wright.thesis.pdf

*** Мултиагентни системи MAS. Семантичен уеб** – избираем курс във ФМИ ПУ „Паисий Хилендарски“, пролетен триместър 2008 г., магистри „Софтуерни технологии“, Станимир Стоянов, http://web.uni-plovdiv.bg/stoyanov/
Анотация на курса: https://fmi-plovdiv.org/index.jsp?id=629&ln=1
Абстрактни архитектури … Дедуктивни агенти. Реактивни. Практически. Многослойни архит.
Околна среда; сензори, ефектори… Симв.околна среда
Софтбот .. Interp. Обучение – KB/beliefs (Door (d1)) – Stop! – Breake!
Множество от: L – множ. Изрази в предикатна логика; D = p(L) множ. От L бази данни.
Вътрешно състояние на аг. Са елемнти $d_i$ от D.
Action Selection; Дедуктивни агенти – пасивни. Goal-oriented behavior
* Deliberation – обмисляне
* Means-ends reasoning („средства-цели“). Направляват действия – желания, ценности, ангажимент * Практически агенти (Reasoning agents) – отчита времето, включва дедуктивния агент; планиране с острови.
„Имам намерение, имам ресурси, а пасувам – нерац.действие“ → Temporal Logic (логика във времето) a → b; x=x+1 … Модална логика, интервална логика; Две състояния от средата (Лестърски унив.?) $a_0$ →$a_1$ → $a_2$. Интелигентна ОС – извлича съдържание и не показва само имената на директориите (ли картинки) “chop” оператор a1 //сега// a2 (разделя) … нулев интервал, единичен интервал ITL – Interval Temporal Logic → промяна …
* Belief-Desire-Intention (BDI) ~~George~~ Georgiev → ТА: ? **Michael Peter Georgeff**, Австралия
Самонадеяни/предпазливи агенти / Дедуктивни / Практически (Reasoning) / Реактивни (Reactive) /Subsumptions – можество поведения: situation → action … Повече действие – „по-интелигентен“; Subsumption architectures: многослойни … хоризонтален слой ← → сензор; Вертикален; еднопасови, двупасови (веднъж от сетивно към моторно; или веднъж отдолу-нагоре и после обратно до действие); Комбинирана архитектура:
**Touring Machine** … Perception → (Modelling, Planing, Reactive, Control) → Action → Няма взаимодействие между слоевете;
**Interrap**: Jorg Muller: Input → World interface → Behavior → Plan → Cooperation ←→ Social, Planning, World Model ← World Interface → Action output
* Robust Satisfaction of Temporal Logic, Alexandre Donze & Oded Maler, https://www-verimag.imag.fr/~maler/Papers/sensiform.pdf

# * Роботика и учене с подкрепление | Robotics and RL

*** Важни понятия от механиката, роботиката, мехатрониката** – виж и бележките към българските роботисти от основния том.

**Краен действеник** – end effector, краен актуатор, краен ефектор, „ръка“ на робота, инструмент:  за хващане, с вендуза/вакум, магнитна и др., с игли или куки забиващи се в материята; със залепване с лепило и  др. “Gripper”; force closure.

**Конфигурационни променливи**, **конфигурация, състояние на робота** – описват състоянието на системата, напр. ъгли на завъртане на стави, приложено усилие и пр., може да се представи като вектор от числа (0.33, 0.1, 0.5, 0.0…) – виж напр. OpenAI Gymnasium с MuJoCo, пример за хуманоид, чието състояние или конфигурация се описва със списък от 17 числа [0,4;0,4]

**Обобщени координати –** на състоянието на системата; може да са стойности на всякакви параметри: ъгли/насоченост (ориентация), положение в полярна или Декартова координатна система, енергия и пр. Виж аналитична механика, Лагранж, Хамилтон.

**Underactuated Motion Planning, Underactuated dynamics, Underactuated systems –** когато степените на свобода на робота или системата са повече на брой от управляемите параметри. Такива са например самолети, вертолети, ходещи роботи с пасивни стави, които не са задвижвани, напр. Atlas на Бостън Динамикс. По-сложни за управление заради трудности в стабилизацията и нелинейна динамика. Решават се чрез оптимално управление (намаляване на някакъв разход на енергия, функция на ценета) и предсказващо управление с модел (Model Predictive Control), функции на Ляпунов, двигателни примитиви или „походки“: ходене, бягане; при дроновете: различни последователности от команди за извършване на преобръщане, завъртане по определен начин или други трикове и пр. Учене с подкрепление, нелинейно програмиране и оптимизация, хибридни/смесени системи, които превключват между различни подходи. Виж научни статии за „Атлас“ по-долу.

**Overactuated dynamics, systems, planning –** повече управляващи сигнали, отколкото степени на свобода. Едно и също целево положение може да се постигне с различна поредица от команди/движения.

Човешкото тяло има излишество от степени на свобода, изследвано и формулирано първо от Николай Бернщайн още в началото на 20-ти век. Това може да усложни планирането, но позволява допълнителна гъвкавост. Напр. при дроновете, октокоптерите: апарати с 8 независими двигателя и перки са по-устойчиви и с по-добри летателни показатели от квадрикоптерите с 4 двигателя. При отказ на един от 8-те двигателя октокоптерите могат да продължат да летят като компенсират чрез другите.

**Articulated Robot** – **Многоставен робот**, шарнирен робот, манипулатор; вид често използвани индустриални роботи; в Декартови (портален), цилиндрични или сферични координати; SCARA – Selective Compliance Assembly Robot Arm; паралелен (Делта).

**Serial Manipulator**: последователен манипулато, сериен манипулатор: поредица от стави свързани последователно подобно на човешките ръце: рамо, лакът, китка.

**Манипулатор с излишество** – с повече от 6 степени на свобода, допълнителни параметри за ставите, повече възможности за промяна на положението на рамената на робота при запазване на положението на крайния действеник (ефектор, актуатор, инструмент). Роботът змия има много повече от 6 степени на свобода и затова се определя като „със свръхизлишество“. Виж напр. работата на Howie Choset от Robotics Institute в Карнеги Мелън: https://www.ri.cmu.edu/ri-faculty/howie-choset/

**Видове стави** (joints): revolut, cylindrical, prismatic, spherical, planar, universal, screw; виж също: Gazebo URDF

**Kinematics/dynamics** (кинематика, динамика - аналитична механика); траектории (координати)/и силите, които ги пораждат; обобщени координати; степени на свобода (брой независими променливи, параметри, описващи системата: нагоре, надолу; посоки; брой оси на въртене и пр.); механика на Лагранж и Хамилтон: Л.: запазване на енергията (принцип на най-малкото действие, траектория, която намалява разхода на енергия, най-малка производна); Х.: намалява момента; Кинематична двойка: поставя ограничения между взаимното движение на две тела
https://en.wikipedia.org/wiki/Kinematic_pair
https://en.wikipedia.org/wiki/Denavit%E2%80%93Hartenberg_parameters
https://en.wikipedia.org/wiki/List_of_gear_nomenclature

https://bg.wikipedia.org/wiki/Индустриален_робот
http://bg.borunte.net/news/what-is-a-multi-joint-robot-65820070.html
https://en.wikipedia.org/wiki/Serial_manipulator
https://en.wikipedia.org/wiki/Line_representations_in_robotics
История на подвижни роботи: https://en.wikipedia.org/wiki/Mobile_robot и др.

**Kinematic and dynamic constraints** – кинематични и динамични ограничения на движенията, които трябва да се спазят. Кинематичните са напр. да се задържа крайният действеник в определено положение, въпреки промяната в другите части; а динамични – да се избягва сблъсък с други предмети.

**Inverse Kinematics, Forward Kinematics, Inverse Dynamics, Forward Dynamics** – правата кинематика е определянето на положението и насочеността на крайния действеник при зададено състояние, ъгли на стави, скорости и пр. Тя е в права посока, защото се пресмята чрез матрица на хомогенни преобразувания (транслация, ротация) или верига от преобразувания, умножения на матриците на преобразуванията между отделните части на кинематичната верига: $T_n = T1 \times T2 \times T3 \ldots T_n$
Правата динамика ползва уравненията на Нютон-Ойлер или на Лагранж, за да опише силите и въртящите моменти и ускоренията, които пораждат. При **Обратната кинематика** задачата започва с крайното положение на действеника и по него трябва да се изчислят ъглите на завъртане (които са дадени при правата к.). **Обратната динамика** аналогично е по зададени траектория или ускорения на частите на робота да се изчислят силите и въртящите моменти, които е необходимо да се приложат на всяка става, за да се постигне желаното движение със съответните особености.

**Holonomic and Nonholonomic constraints** – холономните ограничения са по-прости и могат да се опишат като алгебрични уравнения: f(q1,q2,...,qn ,t)=0 на обобщените координати. Напр. *твърда* ръка на робот с постоянни, непроменящи се при движение разстояния между звената, или точка върху твърдо тяло. Нехолономните ограничения изискват по-сложни изчисления, които включват и векторите на скоростите на обобщените координати (velocity) … $Sum(a_i(q)q_dot_i) = 0$ , как позициите се променят във времето в конкретния момент, а не само на (статичното) им състояние / конфигурация. Например при колесен робот или превозно

средство – те завиват, но не могат управлявано[169] да се плъзгат настрани. За да достигнат до определено положение, което е изместено в страни, е нужно да извършат маневра, поредица от други движения. Нехолономните ограничения ограничават движението в определени посоки и по какъв начин може да се променя скоростта на частите. За пресмятането често са нужни диференциални уравнения. Холономните са свързани с позицията, а нехолономните: с вектора на скоростта.
* underdetermined dynamics; complex; trajectory optimization: direct methods (parametrize the trajectory and optimize the parameters with nonlinear opt.tech.) Techniques: Differential flatness, Model Predictive Control (MPC); Lyapunov-Based Control (L.stability – to guarantee stability an convergence to the target trajectory); Learning: RL and imitation learn.

**Null space** – where the change in the control parameters don't change the position of the target actuator. E.g. the angles of the joints of a tennis player and the position of the racquet when dynamically adjusting her body in order to keep the head at a desired plane and coordinates. **Proportional-Derivative (PD) Control Systems; PID controllers** (see drones)

* **Parallel Robot, Parallel manipulator, generalized Steward Platform** (платформа на Стюарт-Гаф): паралелни роботи, система от хидравлични, пневматични или електрически „крикове"/"амортисьори" (прътове) – линейни актуатори, които могат да се прибират и изваждат и така променят дължината си и поддържат и управляват положението на равнина, платформа, „блюдо"; напр. в симулатори за самолети и състезателни коли и др.; първите (1954, 1962, 1965) са с шест пръта; линейните актуатори създават линейно-постъпателно движение, а не ъглово или въртеливо, напр. чрез червячна предавка или чрез зъбно колело и рейка, винт-гайка и др. – напр. в чекмеджетата на устройства за четене на оптични дискове (worm drive, slewing drive; rack and pinion; leadscrew).

* **Delta Parallel Robots, Delta robot** – разновидност на успоредните манипулатори, висят „отгоре-надолу"; използван в 3D-принтери. Сингулярност – когато детерминантата на Якобианата, матрицата на производните на конфигурацията, стане „сингулярна", т.е. равна на 0,

[169] Желателно, целенасочено – а не да се приплъзнат липса на сцепление и т.н.; всъщност понякога е възможно и частично да се управляват и такива движения при състезателно каране чрез използване на ръчна спирачка – „дрифт", чрез рязко „развъртане" на волана в противоположни посоки и пр.

напр. когато ставите се наредят в линия: тогава  се загубва някоя от степените на свобода (изпъната ръка); или ако се стигне края на обхвата на движение на става, или стигне края на работното пространство и др.
https://en.wikipedia.org/wiki/Delta_robot
https://en.wikipedia.org/wiki/Stewart_platform

**Представяния на линия** (line representation): начини за описание на траектории, линии, пътечки.: необходими в компютърното зрение, картографирането, роботиката – навигация, планиране на движенията; методи по Лагранж и Хамилтон. Различни представяния може да са удобни в различни случаи. **Parametric:** L(t) = P+tv,  ($P_1$, $P_2$) ... Implicit: ax+by+c = 0; y = mx+b (slope intercept); ρ = x cos(θ) + y sin(θ) (Polar); Homogeneous Coordinates (x, y, w); Plücker Coordinates (L = (a, b, c, d, e, f)); B-spline and NURB

**Навигация:** ориентиране, избор на маршрут, наблюдаване за придържането към маршрута, разпознаване на достигането до целта; самолокализиране, самоуместяване; планиране на пътя; построяване на карта. Представяне на средата, модели на сетивата, алгоритми за локализация; SLAM – simultaneuous localization and mapping etc. Зрителна одометрия (visual odometry): прихващане с една камера или стереодвойка или много камери и сетивно сливане (sensory fusion). Нормализиране на изображенията, премахване на изкривявания, прилагане на матрица на трансформацията (camera intrinsics). Разпознаване на особености (feature detection), проследяване (optical flow, tracking); откриване на съответствието между образите в стереодвойка или множество от камери и др. сензори (LIDAR, сонар). Построяване на поле на оптическия поток. Премахване на грешки (напр. времево изглаждане, премахване на шум/чрез свързаност). Оценка на движението на камерата чрез филтри на Калман или чрез оптимизация, търсене на геометрични и пространствени преобразувания, които намаляват функция на цената, сложността – напр. най-малко завъртане, изместване и пр.

**Сензори, датчици, сетива:** зрителни (камери, фотоклетки), слухови (микрофони), проприоцептивни (ъглово завъртане, усилия, опън на стави и звена; електрическо напрежение); тактилни – за натиск по повърхността на действеници (актуатори, ефектори) или по тялото. За усилие и въртящ момент за измерване на натоварването при движенията на робота. Видео сензорите понякога се използват за измерване на параметри на собственото движение, например при „мускулни“/“сухожилни“ действеници, като се наблюдават промените на собствените части.

**Далекомерите**, измерватели на разстояние (range sensor; distance sensor) могат да са с ултразвук, лазер (LIDAR), както и зрителни чрез възстановяване на обема (възобемване). Със съвременни средства често е възможно зрението да служи и като далекомер, подобно на човека, стига условията да позволяват: да има достатъчно разпознаваеми особености, от които да се извлекат ъгли, линии, маркери и пр. и да са достатъчно устойчиви; по-сложно е или невъзможно при тъмнина, където не може да се освети; в гъста мъгла, при определени особености и свойства на наблюдаваните обекти – висока отразителна способност, меки тела (non-rigid body) с твърде бързо променяща се форма и пр. (подобни проблеми се решават в хирургическата роботика, където водещ учен е българинът Даниел Стоянов, виж в бел. за български роботисти). В добри метеорологични или „нормални“ условия обаче разстоянието до предметите в голяма степен е измеримо със съвременни обучаващи се модели и чрез единична камера (Visual odometry), първоначално чрез поне две (стереодвойка), може и повече, множество камери; понякога се използва и инфрачервена камера, чрез която се проежектира мрежа върху обектите и по изкривяването ѝ в образа може да се прецени за наличие на неравности; сетивно сливане: sensor fusion, в което могат да участват множество камери, сензори и други източници на данни: мултимодалност, за да се построява, поддържа, обновява своевременно отговаряща на действителността пространствена карта, модел на света: mapping. Виж SLAM. **Сензори за положение в пространството:** в рамките на собствената координатна система: IMU, Inertial Measurement Unit: съчетание от жироскоп, акселерометър (измерители на ускорението), магнитомери/компас, акселерометри/. GPS, глобална система за навигация, която може да бъде с увеличена точност със специални допълнителни устройства, които осигуряват допълнителна триангулация след като се синхронизират с няколко спътника. Термометри, влагометри. Прости далекомери с малка точност: инфрачервени, ултразвукови за отчитане на препятствия, без да е нужно прецизно възстановяване на формата.

Далекомерите са удобни като изчислително „евтин“ метод за симулации, разстоянието се изчислява чрез изпращане на лъчи от източника до сблъсък с препятствие, подобно на алгоритмите за проследяване на лъчите в компютърната графика и игри (и Wolfenstein-3D, 1992 и фотореалистичното изобразяване, Ray tracing, Ray marching). Виж Gazebo, ROS, ROS2, ROS Turtlebot… Виж още:
https://standardbots.com/blog/every-type-of-sensors-in-robotics---explained
https://en.wikipedia.org/wiki/Inertial_measurement_unit

* A New Range-Sensor Based Globally Convergent Navigation Algorithm for Mobile Robots, I.Kamon, E.Rimon, E.Rivlin, 9.1995
https://www.cs.cmu.edu/~motionplanning/papers/sbp_papers/r/tangentbug.pdf
Алгоритъмът движи робота около очертанията на усетените препятствия. Виж Задача 6 от брой 3/2003 - ДВИЖЕНИЕ НА РОБОТ от конкурса на „PC Magazine Bulgaria" 2002/2003.
https://web.archive.org/web/20050222125101/https://konkurs.musala.com/index.php?sect=02&page=6
https://web.archive.org/web/20060713035620/http://konkurs.musala.com/content/c02/p6/task.rtf
https://wiki.ros.org/Robots/TurtleBot https://wiki.ros.org/turtlebot3
https://docs.nav2.org/ - ROS Navigation stack: работа с карти, локализиране върху карта (SLAM построява началната карта), планиране на целия път през средата, дори за кинематично реалистични големи роботи, ъправлява робота да следва пътя и динамично да го уточнява, за да избягва сблъсъци; изглжда плановете, преобразува сетивните данни в модел на света; построява сложни и гъвкави модели на поведението чрез дървета на поведението; изпълнява предварително зададени поведение в случай на повреди, човешка намеса и др.; следва последователни точки за преминаване от зададената задача; ... наблюдава първичните сетивни данни за неизбежни сблъсъци и опасни ситуации …
* SDF – виж също Signed Distance Fields в компютърната графика. Един от пионерите в тази област в България е преподавателят в ПУ „Паисий Хилендарски" Александър Пенев, който разработва такъв метод като дипломна работа: „Един подход за описание на геометрична информация", А.Пенев, рък. Д.Димов, 1996. Кодът е на Турбо Паскал 7 и е включен. https://www.alexander-penev.info/sites/default/files/articles/msc-alexander-penev.pdf Модерно приложение на SDF е изкуството на шейдърите в платформата „ShaderToy": http://shadertoy.com
https://iquilezles.org/articles/distfunctions/
SLAM: George Hotz, updated by Twenkid: "A toy implementation of monocular SLAM written while livestreaming" https://github.com/Twenkid/twitchslam
https://introlab.github.io/rtabmap/ Real-Time Appearance-Based Mapping
https://github.com/introlab/rtabmap Зряла система за мобилни роботи, която внедрих и използвах през 2023 г. в опити за приложение с независим дрон, но в онзи момент той нямаше достатъчно мощен компютър, и изчислително, и *електрически* (претоварваше се при включване на камерата и без да се пусне SLAM); приложението на системата беше през лаптоп, с който строих карти (облаци от точки, Point Clouds), които при продължение на проекта, под формата на мрежи от триъгълници (meshes, преобразувани от точките) или пък струпани (клъстерирани*)

щяха да служат за навигация в това пространство на симулиран дрон през Gazebo SITL, ROS; бордовият компютър беше Jetson Nano 4GB RAM, 4-ядрен ARM, GPU от поколението на настолните GeForce 7xx, поддръжка на Ubuntu 18/20 (ядро Tegra); RGB-D стерео камера Intel Realsense d435i).
* https://www.amazon.com/NVIDIA-Jetson-Nano-Developer-945-13450-0000-100/dp/B084DSDDLT
* https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/
* https://ardupilot.org/dev/docs/sitl-with-gazebo.html
* https://ardupilot.org/dev/docs/sitl-simulator-software-in-the-loop.html
SITL – software in the loop; симулира физиката на дрона, двигателите, летателния компютър и пр., може да бъде приложено и за колички („роувъри“, rover). При наличие на по-мощен бордови компютър от следващите поколения напр. Jetson Orin Nano и по-бързи може да се приложат по-съвременни системи за **SLAM**, които се възползват графичното ускорение, напр. Nvidia ISAAC_ROS_VISUAL SLAM. Бях изпробвал системата на GPU 3090 до което имах достъп за кратко време.
* https://github.com/NVIDIA-ISAAC-ROS/isaac_ros_visual_slam 2022
* https://github.com/raulmur/ORB_SLAM2 2015
* https://github.com/UZ-SLAMLab/ORB_SLAM3 2021

*** Turning a flaw into a superpower: Researchers redefine how robots move**
by University of Sheffield, edited by Lisa Lock, reviewed by Robert Egan, 7.11.2025
*** Turning a flaw into a superpower: Researchers redefine how robots move**
https://techxplore.com/news/2025-11-flaw-superpower-redefine-robots.html
**Тош:** Нов вид движение за роботи, използващо *хистерезис* на формата, породен от силите на триенето – вид абстрактна инерция и памет – остатъчни свойства от натрупване на минали състояния, които водят до промяна на начина, по който дадена система, материал и пр. реагират по *различен* начин на нови въздействия – а не само да отразяват по един и същи начин непосредствените сили, действащи в момента. Виж хистерезисна крива на феромагнитната материали в електрониката.

Роботът е във формата на змия, запомня предишната си форма и я използва, като така намалява необходимия брой различни само до два, последователно управлявани чрез обърната зиг-загообразна сухожилна капсула.
*** Hysteresis-assisted shape morphing for soft continuum robots**
Zheyuan Bi, Tianchen Ji et al. 15.10.2025
https://www.science.org/doi/10.1126/sciadv.adx3024
*“.. friction-induced shape hysteresis, the difference in robot shape during loading and unloading, to enable numerous reversible shape changes using only two sequentially controlled actuators via an inverted zigzag tendon-sheath mechanism.”*

* **Track Everything Everywhere Fast and Robustly,** Y.Song, J.Lei, Z.Wang, L.Liu, K.Daniilidis, 3.2024 https://arxiv.org/pdf/2403.17931 Previous: OmniMotion; they propose: CaDeX++ … *factorizes the function representation into a local spatial-temporal feature grid and enhances the expressivity of the coupling blocks with non-linear functions …*
* Виж разработките на българите Васил Чаталбашев и Драгомир Ангелов в Станфорд в началото на 2000-те за разделяне и разпознаване на повърхности чрез данни, събрани с далекомери – тогава чрез лазери. Васил участва в построяването на семантично разделяне, сегментиране на модел от точки, събран с робот и лазерен далекомер. Драгомир разработва много системи за разделяне на лазерно сканирани човешки фигури на части, преобразуването им в други пози по образец на данни от друг човек и др. Виж списъка с български роботисти в Основния том на „Пророците“ и приложение *„Анелия“.* Анелия Ангелова, на която е кръстен споменатият том, е значим изследовател в компютърното зрение за разпознаване и навигация от същия период до днес и също учи в Станфорд. #prophets #anelia
* https://github.com/google-deepmind/mujoco_mpc – “Real-time behaviour synthesis with MuJoCo, using Predictive Control”[170]. *MuJoCo* е физичен симулатор за роботика и учене с подкрепление, създаден от **Емануел Тодоров – „Емо“ Тодоров**. Ползва се и в Gymnasium на OpenAI. https://gymnasium.farama.org/
* Виж също OpenRave на Росен Дянков, основател на компанията Mujin в Япония. https://openrave.org/docs/latest_stable/
* Прочети в *Основния том* още за български специалисти в роботиката.

*** Устойчиво, приспособяващо се и най-добро управление**
/адаптивно управление, оптимално управление; регулатори/
**Robust Control, Adaptive Control & Optimal control**

**Устойчиво управление** (robust control) – да се запази в определени граници, напр. за управление на траектория на летателен апарат, ракета, което не трябва да се отклонява от зададени ограничения и рамки; може да е нежелателно да се учи, ако това ще отклони системата.
**Приспособяващо се:** правилата да могат да се донастройват, адаптивни

[170] Predictive Sampling: Real-time Behaviour Synthesis with MuJoCo, Taylor Howell et al. https://arxiv.org/pdf/2212.00541

филтри, обучение.
**Оптимално:** търсене на възможно най-добрата траектория, поредица от действия.

Системата се сблъсква със смущения, които се опитват да я отместят от траекторията и тя реагира с промени, обратно действие, с промени на параметрите, които я описват, с което да компенсира отклоненията. Виж роботика, Model Predictive Control MPC.

Виж работата на **Пламен Ангелов**, **Никола Касабов** и **Димитър Филев** в приложение #anelia, свързана с приспособяващи се системи с размита логика, импулсни невронни мрежи, еволюиращи интелигентни системи EIS, управление и изкуствен интелект в превозни средства (особено Д.Филев) и др. (Fuzzy logic, SNN, Neural Networks, Evolving  etc.). Също Петър Кормушев, Еманоел Тодоров, Драган Ненчев; Анелия Ангелова, Драган Ангелов, Александър Тошев, Николай Атанасов и много други български роботисти от по-далечното и по-близко минало и настояще, някои от тях споменати в основния том на *Пророците,* в приложение *Анелия* и др.

# * Учене с подкрепление. Още бележки по роботика:

*** Reinforcement learning**

* Виж също по-горе частта за ученето с подкрепление и машинно обучение свързано с невронауките и работата на Момчил Томов, Мирослав Клисаров, Емануел Тодоров. Петър Кормушев, чиято работа е разгледана в основния том, развива някои от методите, разгледани по-долу. Двигателните примитиви (motor primitives) могат да се разглеждат като вид възможности, „опции“ и учене на „епизоди“ в терминологията от обзора на М.Клисаров и др. за йерархичното учене с подкрепление, и в свързаните статии към работата на М.Томов за учене на епизоди, което ускорява обучението. Примитивите са и вид „навик“ в невронауките, „habitual behaviour”.

*** Dynamic Movement Primitives (DMP)** – множество от атрактори, чието влияние постепенно се превключва по време на движението.
Предходни методи: Episodic-reinforce, Episodic Natural Actor-Critic (eNAC)
Виж бележките за Петър Кормушев в основния том.

*** Reinforcement Learning for Motor Primitives**, Jens Kober, 2008 (MSc thesis), Universität Stuttgart, Max-Planck-Institut für biologische Kybernetik Tübingen Empirische Inferenz
https://is.mpg.de/fileadmin/user_upload/files/publications/DiplomaThesis-Kober_5331[0].pdf

Motor primitives, **episodic policy learning**; Policy learning by Weighting Exploration with the Returns (PoWER); explorative version of the dynamic motor primitives; finite difference gradients (FDG): generate policy variations: perform rollouts, collect, compute return (rewards) from rollouts; compute gradient; update policy until convergence … policy search via expectation maximization; Policy Learning by Weighting Exploration with the Returns – deterministic mean policy; additive exploration ε(s, t) … Perceptually coupled motor primitives; Control tasks: ball in a cup, ball-in-a-cup motion …
https://www.researchgate.net/profile/Jens-Kober

*** Learning Motor Primitives for Robotics**, Jens Kober, Jan Peters, 6.2009
https://www.researchgate.net/publication/224557356_Learning_Motor_Primitives_for_Robotics – Acquisition and self-improvement of novel motor skills … Combination of imitation learning (IL) and RL. Dynamic systems motor primitives - … Ball-in-a-cup (топче е закачено на връв за чашка с ширина сравнима с топчето; с рязко движение топчето се изхвърля нагоре и трябва да се задържи в чашката; ball-padding (топкане на топче с хилка за тенис на маса и пр.)

*** Leveraging LLMs, Graphs and Object Hierarchies for Task Planning in Large-Scale Environments**, Rodrigo Pérez-Dattari et al., 9.2024
https://www.researchgate.net/publication/383911519_Leveraging_LLMs_Graphs_and_Object_Hierarchies_for_Task_Planning_in_Large-Scale_Environments
LLMs are used to generate plans from text instructions. Task and Motion Planning (TAMP) ..Reduce the search space by limiting the object categories and tasks to the relevant for the task. ground plans: affordance f., semantic distance minimization; feasibility checks from world representations; library of motion primitives. Graphs for attributes: “oven–is closed”, “cake-is_inside-oven”. Applicable acctions A in states $s_t \in S$ (affordances). G – goal states. …
TA: Compare to classical STRIPS etc.

*** PUMA: Deep Metric Imitation Learning for Stable Motion Primitives,**
R.Perez-Dattari et al.,10.2024
https://www.researchgate.net/publication/384899317_PUMA_Deep_Metric_Imitation_Learning_for_Stable_Motion_Primitives – Task Planning → High-level planning (sequence of simpler subtasks) → Action planning (specific actions for each subtask) → Temporal planning (prder of the actions in time and temporal constraints)
Motion Planning → Motion generation (feasible trajectories) → Collision avoidance (obstacles) → Dynamic constraints (inertia, friction, actuator limits)
Hierarchical planning (high-level task p., low-level motion p.), learning, hybrid

## * Човекоподобна роботика и управление чрез предсказващо управление (humanoid robots, model predictive control, MPC; хуманоидни роботи):

**Виж в основния том на „Пророците на мислещите машини“:** български учени са водещи в хуманоидната роботика като Драган Ненчев и др.

*** Optimization-based Locomotion Planning, Estimation, and Control Design for the Atlas Humanoid Robot**, Scott Kuindersma · Robin Deits ·Maurice Fallon · Andr´es Valenzuela · Hongkai Dai · Frank Permenter · Twan Koolen · Pat Marion · Russ Tedrake, 2014 *
* https://groups.csail.mit.edu/robotics-center/public_papers/Kuindersma14.pdf
* https://gwern.net/doc/reinforcement-learning/robot/2015-kuindersma.pdf
Статия за робота „Атлас“ на Boston Dynamics (Бостън Динамикс) към края на 2013 г. и началото на 2014 г. **Размери:** 155 кг, 188 см. 1000 Hz обновяване на конфигурацията: положение на ставите, скоростта и силата, изпраща се на полеви компютър, който изпълнява планирането, оценката и управлението. Linear variable differential transformers (LVDTs) на актуаторите. Няма сензори за силата при ставите, но силата се изчислява чрез сензори за налягане в действениците. Други датчици на врата и ставите на ръцете дават информация за положението и скоростите, но няма такива на краката. Жироскоп/IMU модел KVH1750 дава точна 6-измерна (6-DOF) информация за ъгловата скорост и ускоренията за определяне на състоянието на тялото. 6-осеви клетки за натоварването на китките, 4 измерителя за опън и изкривявания (strain gauges) във всеки крак – триизмерно усещане на силата и въртящия момент. Multisense SL сензор: съчетава стерео камера с две „очи“ и равнинен LIDAR Hokuyo UTM-30LX-EW, който се върти на ос с до 30 оборота в минута и заснема 40 реда от средата в секунда, всеки с по 1081 точки в обхват до 30 метра. Главата може да се мести само нагоре-надолу (pitch up-down, but no yaw or roll). **Закъснения при обработката и честоти на сигнала от различни сензори,** с.35, табл.2:
* Филтри на Калман на долни стави: 0.16 msec | 1 kHz
* Ф. на Калман за положение на тялото 0.54 msec | 333 Hz
* LIDAR 7 msec | 40 Hz
* GPF* 11.4 msec | 40 Hz
* Общо закъснение на LIDAR: 18.4 msec | 40 Hz

GPF=Gaussian Particle Filter – Гаусов филтър на частици; използва се за пресъздаване на геометрията на средата от точките от ЛИДАР-а. В началото обикновено роботът се оставял да стои неподвижно около 30 секунди, за да създаде карта на средата – облак от триизмерни точки. От тях се създава вероятностна мрежа на заетото пространство (probabilistic occupancy grid, OctoMap).

Разликата в честотите, която тук е сравнително малка, е пример за „преживявания“ и съществуване в различни мащаби и обхвати. Сравни например с нервната система и тялото: молекулярни процеси в клетъчни органели, в клетките, междуклетъчно пространство, в тъкани, нервни процеси в различни части от обработката (рецептори, първично сетивно усещане, гръбначен мозък или черепни нервви, таламус, първични сетивни корови полета и т.н. Челните дялове може да обобщават информация от 10-30 секунди.
* → Виж приложение *„Вселена и Разум 6“* (Universe and Mind 6) за *синхронизацията* и теорията за специални белези, чрез които частите разбират, че обединени в цяло в духовното усещане.
* *„Решаването на планирането на движенията на подскок отнема около 90 секунди на сметач с Intel Core i7 3.1 GHz”. Скок от тухла*: 10 минути.*
(От високо на ниско; „cinder block”, газобетонно блокче, “шлакоблок“).

*** Optimization Based Full Body Control for the Atlas Robot,**
Siyuan Feng†, Eric Whitman‡, X Xinjilefu† and Christopher G. Atkeson†
https://www.cs.cmu.edu/~sfeng/sf_hum14.pdf
Управление на две нива: поведение и ниско ниво чрез обратна кинематика и обратна динамика (Iverse Kinematics & Inverse Dynamics).

*** Триизмерен модел за симулатора MuJoCo:**
https://github.com/lvjonok/atlas-mujoco/blob/master/demo.ipynb
→ Виж в основния том на „Пророците“ информация за множество български роботисти, тяхната работа, симулатора MuJoCo на Еманоел Тодоров и др.

* **Robot navigates high-speed parkour with autonomous movement planning** by Bob Yirka, Phys.org, 30.5.2025 https://techxplore.com/news/2025-05-robot-high-parkour-autonomous-movement.html – Raibo: четирикрак робот, планиране чрез съчетание от евристики, невронна мрежа с учене с подкрепление и физичен симулатор;  Map generator; Planner, Tracker, Target Updater; target foothold; sampling & roll-outs in 8 separate threads.
* https://www.science.org/doi/10.1126/scirobotics.ads6192

* За някои от сензорите в „Атлас“:
* **Strain Gauge Neural Network-Based Estimation as an Alternative for Force and Torque Sensor Measurements in Robot Manipulators,** Sep 2023 Applied Sciences 13(18):10217**,** DOI: 10.3390/app131810217**,** Stanko Kruzic et ал.
https://www.researchgate.net/publication/373832988_Strain_Gauge_Neural_Network-Based_Estimation_as_an_Alternative_for_Force_and_Torque_Sensor_Measurements_in_Robot_Manipulators

* **CoinFT: A Coin-Sized, Capacitive 6-Axis Force Torque Sensor for Robotic Applications**, Hojung Choi et al., 25.3.2025 https://arxiv.org/html/2503.19225v1
contact-rich tasks: wiping, assembly, palpating soft tissue … (избърсване, сглобяване, напипсване на меки тъкани); for “robotic arms, grippers, drones, wearable devices”.
See figures with illustrations of the forces and activations.
*** Multi-Axis Sensors**: https://www.botasys.com/post/multi-axis-sensor

## * Допълнения към Епигенетична роботика, роботика на развитието #robotics #epigeneticrobotics

→ Виж в списъка с школи и учени в основния том, раздел **Когнитивна лингвистика**, статиите за **Йордан Златев**, както и **Александър Стойчев.**
**(See the Bulgarian epigenetic roboticists Jordan Zlatev and Alexander Stoychev in the main volume of The Prophets of the Thinking Machines)**

### * Epigenetic robotics, developmental robotics

**Andrea Kulakov and Georgi Stojanov (Skopie)**
https://www.researchgate.net/profile/Georgi-Stojanov
* Structures, inner values, hierarchies and stages: essentials for developmental robot architectures. Andrea Kulakov*, Georgi Stojanov*, 2002, https://www.researchgate.net/publication/28763617_Structures_inner_values_hierarchies_and_stages_Essentials_for_developmental_robot_architectures *the essential components needed for a developmental robot architecture. … Piaget’s genetic epistemology and Vygotsky’s activity theory … p.5. 3.7 Behavior and expectations - The mind is fundamentally an anticipator, as Dennett deduced lucidly in his quest for explaining consciousness and intelligence (Dennett, 1991), (Dennett, 1996). In every moment, in every step we anticipate something and then we expect the outcome of our actions. (…)* *** A references to:**
* Zlatev, J. (2001). **A hierarchy of meaning systems based on value.** Proceedings of the 1st International Workshop on Epigenetic Robotics, Lund University, Cognitive Studies, 85 - see notes in the Bulgarian part of the main volume of ***“The Prophets of the Thinking Machines…***”, also * “Five Basic Principles of Developmental Robotics”, 2006 (виж Йордан Златев в основния том на Пророците на мислещите машини …)
Citations of works by Boicho Kokinov. A.Kulakov has studied in NBU.

**Other references on their early contribution in epigenetic robotics (or *developmental*):**

* Kulakov, A. (1998). Vygovorotsky - A Model of an Anticipative and Analogy-making Actor. MSc Thesis,New Bulgarian University, Sofia
* Stojanov, G. (1997). Expectancy Theory and Interpretation of EXG Curves in Context of Biologicaland Machine Intelligence. PhD Thesis, ETF, Sts. Cyriland Methodius University, Skopje
* Stojanov, G., Bozinovski, S., Trajkovski G. (1997).Interactionist-Expectative View on Agency andLearning. IMACS Journal for Mathematics andComputers in Simulation, Northe-Holland Publishers,Amsterdam.Stojanov, G., (1999).
* Embodiment as Metaphor:Metaphorizing - in The Environment. in C. Nehaniv(ed.) Lecture Notes in Artificial Intelligence, Vol.1562, Springer-Verlag
* Stojanov, G., Kulakov, A., Trajkovski, G. (1999). Investigating Perception in Humans Inhabiting Simple Virtual Environments: An Enactivist View. CognitiveScience Conference on Perception, Consciousness andArt, Vrije Universitaet in Bruxelles
* Stojanov, G. (2001). Petitagé: A case study indevelopmental robotics. Proceedings of the 1s International Workshop on Epigenetic Robotics, LundUniversity Cognitive Studies, 85

# * Навигация и умствени карти, когнитивни карти, карти в умствен план, субективни карти … при човека

Ahmadpoor, N., & Shahab, S. (2019). **Spatial Knowledge Acquisition in the Process of Navigation: A Review**. Current Urban Studies, 7(1), 1-14. DOI: 10.4236/cus.2019.71001

Обзор на теории за човешкото изграждане на умствени карти (cognitive maps). „Поведенческа география“. Картите позволяват на човек да може да реагира, предусеща, очаква следващите възможни действия от дадено положение, координати. Умствените карти се изграждат постепенно и се изменят с трупането на опит и постепенното обхождане на непознати местности.

Развитие на способностите за построяване на умствени карти в индивидуалното развитие. По Пиаже: три стъпки: 1. Топологично; 2. Проективно; 3. Евклидово. Топологичното представяне запазва качествени отношения като близост, отделеност и непрекъснатост. Проективното е свързано с пространствени отношения между части при определена гледна точка, а евклидовото знание включва метрична информация за разстояния и посоки. Първо: забележителности (landmarks), после: маршрути, пътеки (routes); накрая: обзори (surveys), които свързват цялото пространство (landmark, route, survey). Намиране на път. Връзки между елементите от умствената карта. Друга схема на  пространствени елементи: 1) Забележителности (landmarks), места с отличителни особености, които се открояват от фона като високи или исторически сгради, сгради с особена форма или цвят и пр. 2) Възли – стратегически точки, в които има прекъсвания на начина на придвижване, кръстовища или места на прекъсване – напр. площади. 3) Пътечки – „канали“ по които често се преминава или по които може да се премине – улици, пътеки, алеи, водни канали, железни пътища. 4) Ръбове – граници между две „фази“, линейни прекъсвания: брегове, край на релси, стени. 5) Райони (districts) – средно големи части от град или квартал, напр. търговски центрове, студентски градчета и др.;

Дискретни, отчетливи особености, свойства: цветове, височини на сгради, значимост, ширина на път, цвят на паветата. Особености, свойства на отношенията: те се откриват при преминаване между различни места.

Етапи от навигацията (може да се припокриват): избор на цел, ориентиране, планиране на маршрута, изпълнение.

Придобиване на пространственото знание чрез пряко преживяване и взаимодействие или чрез посредници: статични карти, електронни динамични мобилни карти от устройства; снимки, описания.

*** Шемякин, 1940** (Shemyakin, 1961) – изследвал как децата се учат да се ориентират в пространството, когато ги помолят да нарисуват карта, те го правят следвайки последователността в която срещат „забележителностите“,

т.е. запомнят по този начин (сравни Джеф Хокинс, „За ума“).
* Shemyakin, F. N. (1961). **Orientation in Space**. In B. G. Ananyev et al. (Eds.), Psychological Science in the U.S.S.R. (Vol. 1). Washington DC: Joint Publication Research Service
* Шемякин, Ф.Н. - О психологии пространственных представлений, 1940

*** SPATIAL ORIENTATION, WAYFINDING, AND REPRESENTATION,** Rudolph P. Darken and Barry Peterson, 2001 https://www.researchgate.net/publication/2557944_Spatial_Orientation_Wayfinding_and_Representation Spatial comprehension … Къде съм? В коя посока съм обърнат? Guilford-Zimmerman spatial abilities test. “environmental fidelity” of virtual environment (Waller et al. 1998); spatial knowledge transfer https://www.researchgate.net/publication/224107401_Gilford_Zimmerman_orientation_survey_A_validation

* **Prototypes, Location, and Associative, Networks (PLAN): Towards a Unified Theory of Cognitive Mapping**, ERIC CHOWN, STEPHEN KAPLAN, DAVID KORTENKAMP, 1995

Representation of large-scale space, or cognitive map ... wayfinding ... landmark identification, path selection, direction selection, abstract overviews; Principles: simplicity, consistency, economy (e.g. a low resolution relative position “near” to a landmark instead of the absolute coordinate of all obejcts); human: what & where systems – contour & location. .. Landmarks: recognizable from many views & orientations & partial views; 5+-2 at one time; linked to context; “distinctive visual events” ... well learned paths become → a higher-level representation → hierarchical elements (compact representation of a subpath of a longer path; %); connectionist system; route as weights of connections ($landmark_a$, $landmark_b$) ... directional space - local map to other connected next landmarks. Human Location system: very approximate, lower than “rubber sheet maps” etc. ... Location s. – size, distance; more details; R-Nets; which local map should be activated at any given time?
Topological Networks of landmarks – Networks of local maps – NAPS; an associative network of connected landmarks; route extraction; spreading activation; ... segmenting a scene – subregions ... finding 5+-2 objects/landmarks ... Survey maps; Gateways; Regions – visual barriers & gateways – walls, doors; hills & passes between hills in valleys, trees and paths; nodes, centroids, achor points; absolute space representations (ASR); Regional maps – a view from a height (no obstruction)

**Тош:** принципите, методите, обобщенията, разделенията и пр. при навигация, ориентиране, построяване на познавателни карти и пр. *от човека* могат да се обобщят и приложат за всички видове данни, пространства, разсъждения, модалности, преобразувания и пр. и в машинна форма в Зрим и езика на разума. @Вси: Об+, прбрзв, прлж.

## * Възобемване, възстановяване на обема, пространствено възстановяване, възстановяване на геометрията чрез множество от кадри от различни гледни точки и движение, 3D-реконструиране, SLAM, навигация – класически работи, роботика, навигация за роботи

## * 3D Reconstruction, Structure from Motion, Robot Navigation…

#SfM #SLAM #structure from motion #3D-reconstruction

### Увод: Древните методи работят със скромна изчислителна мощ

Преглед на исторически статии и проекти и развитието на областта до днес с примерни публикации. Забележете, че знанието за принципите на възстановяване на триизмерната структура от двуизмерни изображения се натрупва отдавна, основата е фотограмметрията, която е наука на векове.

Васил Чаталбашев, Драгомир Ангелов, Анелия Ангелова, Александър Тошев и други българи имат приноси в тези или други свързани области от компютърното зрение след около 2000-та година. Васил, Драгомир и Анелия са заедно в Станфорд. Ангел и Драгомир след това в „Гугъл“. А. Ангелова например участва в разработки за „марсоходи“. Д.Ангелов става директор на самоуправляващата се кола на „Гугъл“, по-късно „Waymo”. Виж основния том и приложение *Анелия.*

Виж също работата на Дейвид Лоу, Такео Канаде и др. (David Lowe, Takeo Kanade) и други пионери на „обратната графика“, техниките за откриване на части от изображението, които не се променят при афинни преобразувания и се използват като признаци за напасване при преобразуване, търсене на изображения в бази от данни или на подобни и др. (affine-invariant tranformations) и др., разгледана в приложение *Лазар.*

След 2012 г. е прието да се говори за огромния скок, извършен от конволюционната невронна мрежа AlexNet в състезанието за разпознаване на набора от данни ImageNet, която постига около 15% грешка (познава в първите 5 класа), докато вторите са 26% и използват SIFT и др. методи за откриване на признаци в изображенията и класификация. Не бива да се омаловажава онова, което се постига и с методите до тогава и в тях може да се съдържа основа за други бъдещи техники.

*** ImageNet Large Scale Visual Recognition Challenge (ILSVRC)**
* https://www.image-net.org/challenges/LSVRC/2012/
* https://image-net.org/challenges/LSVRC/2012/results.html
Вторият в класирането ISI е един 4 варианта на отбор от Токийския университет и др. и използва: *„Претеглен сбор от оценките на всички класификатори: SIFT+FV, LBP+FV, GIST+FV, CSIFT+FV“.*

## * Сравни съвременен подход към SLAM с постепенното развитие на областта:

* **On Learning and Geometry for Visual Localization and Mapping, Sarlin, Paul-Edouard,** 2024 https://www.research-collection.ethz.ch/handle/20.500.11850/701148 и др. ?T SLAM в кн.

*** 3D POSITIONAL INTEGRATION FROM IMAGE SEQUENCES**, C G Harris and J M Pike, 1987, https://bmva-archive.org.uk/bmvc/1987/avc-87-032.pdf
p.1, Fig.1: Processing Flowchart: “First two images, Bootstrap processing, Next image, Point matching, Ego-motion, Point classification, Point position update, 3D instantiation of new points → Next image. Ellipsoids for probability distribution functions for feature-points.

* **DETERMINATION OF EGO-MOTION FROM MATCHED POINTS**, C G Harris, 1987 *
https://web.archive.org/web/20170706070041/http://www.bmva.org/bmvc/1987/avc-87-026.pdf *
https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=09e55d995287e63e277d0f25fb6525990061482f

* Fang.J.Q. and T.S.Huang, "**Solving three-dimensional small rotation equations: Uniqueness, algorithms and numerical results,**" Computer Graphics, Vision and Image Processing, vol 26, pp. 183-206, 1984.
* Faugeras.O., F.Lustman and G.Toscani, **"Motion and Structure from Motion from Point and Line Matches**," Proceedings IEEE International Conference on Computer Vision, pp. 25-34, 1987.
* Longuet-Higgins.H.C **"A Computer Algorithm for Reconstructing a Scene from Two Projections,"** Nature 293, pp. 133-135, 1981 – **"The eight-point algorithm"**; 8-point algorithm; "two distinct processes: the establishment of a 1:1 correspondence between image points in the two views—the ‘correspondence problem’—and the use of the associated disparities for determining the distances of visible elements in the scene."
* Marr, D. & Poggio, T. Science 194, 283–287 (1976**) The Eight-Point Algorithm,** Carlo Tomasi, https://courses.cs.duke.edu/fall15/compsci527/notes/longuet-higgins.pdf - *estimates of the rigid transformation G and estimates of the coordinates of a set of n points P1, . . . ,Pn in the two camera reference frames from the n pairs*

*(x1, y1)), . . .(xn, yn) of noisy measurements of their corresponding images.*

* **Vision Algorithms for Mobile Robotics,** Lecture 08 Multiple View Geometry 2, Davide Scaramuzza, http://rpg.ifi.uzh.ch The 8-point method method is still in use by NASA rovers (40 years later)
https://rpg.ifi.uzh.ch/docs/teaching/2021/08_multiple_view_geometry_2.pdf

* **A Brute-Force Algorithm for Reconstructing a Scene from Two Projections**, Olof Enqvist et al., 2011 *
https://www.maths.lth.se/matematiklth/vision/publdb/reports/pdf/enqvist-jiang-etal-ccvpr-11.pdf

* Hopkins 155 dataset, R. Tron and R. Vidal., **A Benchmark for the Comparison of 3-D Motion Segmentation Algorithms**.,2007
http://www.vision.jhu.edu/data/hopkins155/

* **Topological Map Learning from Outdoor Image Sequences**, Xuming Xe, R.Zemel, Minh, 2006, https://www.rotman-baycrest.on.ca/files/publicationmodule/@random4824abb32cfea/top_map_learn.pdf
* https://www.cis.jhu.edu/~rvidal/publications/cvpr07-benchmark.pdf

*** Learning to find good correspondences**, K.M. Yi, et al. CVPR 2018
https://etrulls.github.io/slides/cvpr18_slides.pdf
https://github.com/vcg-uvic/learned-correspondence-release

*** Structure-from-Motion under Orthographic Projection**, Chris Harris, 1990
https://link.springer.com/content/pdf/10.1007/BFb0014857.pdf

*** Mobile Robot Localisation Using Active Vision**, Andrew J Davison and David W Murray, 1998 https://www.semanticscholar.org/paper/Mobile-Robot-Localisation-Using-Active-Vision-Davison-Murray/384aa7b31dbc8e1a5a6a51598bac25b4c5ab22ce
Perspective SFM works for ego-motion, wide-angle lens and static environments, significant depth variation (close and distant objects in the same view) Difficulties when the objects are occupying a small portion of the field of view (small angle), e.g. airplanes, cars moving at distance.

* **A Frame for Frames: Representing Knowledge for Recognition**, Benjamin Kuipers, 1975 recognition of a wireframe block: L, fork, arrow ...; "active program objects" - "specialists", interacting by sending messages to each other...https://www.researchgate.net/publication/37596651_A_Frame_for_Frames_Representing_Knowledge_for_Recognition
https://scholar.google.com/citations?user=H1XVLwsAAAAJ&hl=en

* **Modeling spatial knowledge**,Benjamin Kuipers, 1978
https://onlinelibrary.wiley.com/doi/pdf/10.1207/s15516709cog0202_3

* **Shakey: From Conception to History**, Benjamin Kuipers, Edward A. Feigenbaum, Peter E. Hart, Nils J. Nilsson
https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/download/2716/2616
1966-1972, Stanford Research Institute (then SRI)... PLANEX, STRIPS: GPS, A Program That Simulates

## * Причинностно моделиране и продължение на картографиране

* **Human Thought**, A.Newell, H.Simon, 1961 + **Theorem-Proving by Resolution as a Basis for Question-Answering Systems"**, C.Green, 1969;
an improved Hough transform (computer vision), mobile robot with AI & vision; A* alg. for navigation

* Benjamin Kuipers**: Commonsense Reasoning about Causality: Deriving Behavior from Structure**, Benjamin Kuipers, 1984 (1983)
https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=7e63b661797906f69791a4454bbf9444a743bfd8 – Structural-behavioral-functional description. ... "qualitative-reasoning method for predicting the behavior of mechanisms characterized by continuous, time-varying parameters" ; Qualitative simulation - ... lacks precise numerical values ... but making useful predictions. Ordinal relations (><=), IQ value (+=- increasing, steady, decreasing); landmark values; propagation/prediction cycle - of the next qualitatively distinct state: move from landmark value, move to limit, collision; parameter, switch (a bool function), value, landmark value, boolean (a term, value of a switch) ... Rieger and Grinberg [25], ... knowledge representation consists of events, tendencies, states, and state changes; causal links.
* De Kleer, J. and Brown, J.S., **The origin, form and logic of qualitative physical laws**, in: Proceedings Eighth International Joint Conference on Artificial Intelligence, Karlsruhe, WestGermany, August, 1983

* Rieger, C. and Orinberg, M., **The declarative representation and procedural simulation of causality in physical mechanisms**, in: Proceedings Fifth International Joint Conference on Araficial Intelligence, Cambridge, MA, August, 1977.

* **A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations**, B. Kuipers, Y. Byun, Published in Robotics Auton. Syst. 1 November 1991
https://www.cs.utexas.edu/ftp/qsim/papers/Kuipers+Byun-jras-91.pdf

*** The Spatial Semantic Hierarchy**, Benjamin Kuipers, 1999 https://scholar.google.com/citations?view_op=view_citation&hl=en&user=H1XVLwsAAAAJ&citation_for_view=H1XVLwsAAAAJ:2osOgNQ5qMEC – more qualitative rather than quantitative; sensory-control-causal-topological-metrical; qualitative (continuous attributes) -> Quantitative (Analog) -> Local & Global 2D geometry. Local maps of neighborhoods, occupancy grid, generalized cylinders; occupancy mapping phase vs localization phase; causal, topological (regions, links: connectivity, order, boundary, containment); paths, sequences of views & actions; "abductions: min set of places, paths, regions for explaining the sequence of observed views and actions"; "Control level - continuous control laws that bind the agent and its environment into a dynamical system throughout a qualitatively uniform … segment of the environment" - cmp. Visual odometry, SLAM. Causal level - discrete model; Turn, travel, routine, local maps, view-graphs; PLAN* (multiple-representation theory of cognitive mapping)  Local 1D geometry, way-finding: graph; global metrical 2D mapping ... perceptual aliasing (can't discriminate two views/locations, e.g. in a desert, corridors, similar buildings); patchwork mapping
* E. Chown, S. Kaplan, D. Kortenkamp, Prototypes, location, and associative networks (PLAN): Towards a unified theory of cognitive mapping, Cognitive Sci. 19 (1) (1995) 1–51. https://onlinelibrary.wiley.com/doi/epdf/10.1207/s15516709cog1901_1
* "An integrated representation of large-scale space, or cognitive map... Wayfinding", S. Thrun, A. Bücken, W. Burgard, D. Fox, T. Fröhlinghaus, D. Hennig, T. Hofmann, M. Krell, T. Schmidt,
* Map-learning and high-speed navigation in RHINO, in: D. Kortenkamp, R.P. Bonasso, R. Murphy (Eds.), Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems, AAAI Press, Menlo Park, CA/MIT Press, Cambridge, MA, 1998, pp. 21–52.
* S. Thrun, D. Fox, W. Burgard, A probabilistic approach to concurrent mapping and localization for mobile robots, Machine Learning 31 (1–3) (1998) 29–53.
* S. Thrun, S. Gutmann, D. Fox, W. Burgard, B.J. Kuipers, Integrating topological and metric maps for mobile robot navigation: A statistical approach, in: Proc. AAAI-98, Madison, WI, 1998, pp. 989–995
* Mobile Robot Localisation Using Active Vision,  A. Davison, D. W. Murray, 1998 https://link.springer.com/content/pdf/10.1007/BFb0054781.pdf https://www.semanticscholar.org/paper/Mobile-Robot-Localisation-Using-Active-Vision-Davison-Murray/384aa7b31dbc8e1a5a6a51598bac25b4c5ab22ce

* Sebastian Thrun, https://scholar.google.com/citations?user=q-buMEoAAAAJ&hl=en

## * Съвременна роботика и навигация за роботи и общи мултимодални агенти и модели за управление на графичен потребителски интерфейс (интелигентна ОС,

## умна ОС, Computer Use Models)

*Виж също мултимодални модели; Драгомир Ангелов, Анелия Ангелова, Александър Тошев; списък с български роботисти: Емануел Тодоров, Росен Дянков, Драган Ненчев и др.* в основния том и в приложенията #anelia и др.

Сравни с подходи от миналото (старите задачи със съвременни средства):

* **Zero-shot Object Navigation with Vision-Language Models Reasoning**, Wen, C., Huang, Y et al., 24.10.2024 https://arxiv.org/pdf/2410.18570
**Language-driven zero-shot object navigation (L-ZSON)**
Vision Language model with a Tree-of-thought Network (VLTNet)
VLTNet: 4 modules: vision language model understanding, semantic mapping, tree-of-thought reasoning and exploration, goal identification … multi-path reasoning processes and *backtracking* when necessary, enabling globally informed decisionmaking with higher accuracy; *VLMaps* – spatial map representation with pre-trained visual-lanugage features with 3D-reconstruction of a physical environment. Exploration with Soft common sense Constraints (ESC); Only objects categories, but no spatial or visual attributes in the instructions. **Goal-identification module**… Tree-of-Thought (ToT) reasoning for *frontier* selection in robot exploration: **Exploration:** 1) Learning-based: 1.1 pre-trained visual encoders → egocentric images to desriptive feature vectors → navigation policy(imitation learning ro RL). 2) Explicit semantic graph; the navigation policies are trained to identify locations of goal objects with the s.g.; 2) *Frontier-based* (FBE) – heuristic alg. For navigation in unseen env. Reconstructing a depth map of the environment, marking the *border* between explored and unexplored (known/unknown) as "*frontiers*". Free-exploration tasks …  CLIP on Wheel (CoW) (сравни с CLIP за пораждане на изображения и търсене по текст и образ) – text-to-image relevance depth maps (RGB & dpepth observations) → region of interest I FBE. … LLM to assign scores to each potential frontier.  … Chain-of-Thought prompting (CoT) … → Tree-ofThoughts (TOT): branching resoning structure, simulates a discussion among several experts on a given question, until reaching a consensus.
**Todor:** Estimations that can be done with logic and direct distance measurements like prepositions as vectors of relations between objects, locations etc. are implemented using LLMs as this is the "modern approach". The goal of the NN/LLM is eventually to reduce the visual representation to compact, explicit, numerical etc. encoding *similar* to the "symblic" ones from

the 1970s, 1980s in LISP, PROLOG, Planners and the essence of human commands. **Modules:** VLM Understanding (semantic parsing), Semantic Mapping: integrates the parsed image, depth image, … construct a semantic map M, defining objects. … p.6 … Grounded Language-Image Pre-training (GLIP) …

*** CoWs on PASTURE: Baselines and Benchmarks for Language-Driven Zero-Shot Object Navigation**, Samir Gadre, M.Wortsman, G.Ilharco, L.Schmidt, S.Song, 14.12.2022

CLIP on Wheels (CoW), L-ZSON … *"decompose navigation task into exploration when the target is not confidently localized, else target-driven planning"* … Older: Map and exploration, agent localization, planning; Recent: learned exploration: end-to-end training, self-supervised rewards (curiosity) or supervised (state visitation counts) … Goal-conditioned nav. – point goal, object goal (category); natural lang. goal description, instead of low-level instructions. Navigation episode .. 4.3: Object Localization – if and where is the object in the image. 5.1. Pasture Tasks:  Uncommon obj., "wooden toy car"; Appearance description: {size}, {color}, {material} {object}: "small, red apple", "yellow ball", "small black, metallic box": small if 3D bounding box < thresh.; Spatial descr.:{obj} on top of {x}, near {y}, {z} , … "house plant on a dresser near a spray bottle" … "on top of": distance threshold between pairs of obj. .. Appearance distractors: e.g. if looking for "yellow apple", there is also a "gren" and a "red" apple.; spatial distr. Hidden obj. descr.: {obj} under/in {x}" – "ball in the dresser drawer", "clock under the table": sample large obj. (beds, sofas, drawers) to terermine the relations (under/in), and remove the visible instances of {obj}. Hidden obj. with distractors. … Nav. Metrics… Agents LoCoBot … actions: {MOVEFORWARD, ROTATERIGHT, ROTATELEFT, STOP} (move 0.25 m, rotate: 30º. Less success with hidden obj; How CoWs fail: 1. **Exploration** (the target is never seen), 2. **Localization**: seen, but not localized (the localizer doesn't fire). 3.**Planning**: seen and localized, but the planner fail due to inaccuracy in the map … Appx: free vs occupied space, 3D map … learnable exploration … exploration agent training, DD-PPO (proximal policy optimization) – reward 0.1 for visiting an unvisited voxel location at 0.125 m resolution, a step penalty -0.01. Trained in RoboTHOR & HABITAT and MP3D train set

*** RoboTHOR** – реалистичен въображаем свят за роботи, симулатор
https://ai2thor.allenai.org/robothor/
https://aihabitat.org/datasets/hm3d/ *"1,000 high-resolution 3D scans (or digital twins) of building-scale residential, commercial, and civic spaces generated from real-world environments."*
https://matterport.com/partners/meta
https://github.com/NVIDIA/Cosmos

https://github.com/nvidia-cosmos – Cosmos е платформа за разработка на модели на физически светове чрез невронни модели, които могат да пораждат и видео, за ИИ, основан на физика. Съдържа множество от основни модели (foundation models) като такъв text2world, video2world: пораждане на модел на света от текстово описание.
https://github.com/NVIDIA/Cosmos/blob/main/cosmos1/models/diffusion/README.md
От снимка поражда видео с движение в такъв свят:
https://github.com/NVIDIA/Cosmos/blob/main/cosmos1/models/autoregressive/README.md
Изискват висок клас професионални видеокарти с много памет, дори 3090/4090 с 24 ГБ се справят трудно с най-леките задачи.
Също симулатора **Omniverse.**

* **Habitat-Matterport 3D Dataset (HM3D): 1000 Large-scale 3D Environments for Embodied AI**, S. Ramakrishnan et al. 9.2021
https://arxiv.org/pdf/2109.08238
Other photorealistic 3D datasets: Replica, Gibson, and {ScanNet, SceneNN} (regions or rooms and single rooms: 2017, 2016)**.** BuildingParser, Matterport3D**,** Gibson – whole buildings.
HM3D – *"higher fidelity 20-85%, meshes: 34-91% fewer artifacts."; "isolated rooms or individual rooms that are connected via a* ***magic portal****", Unity 3D C#, … Actions:* move-forward, move-back, strafe-right, look-left, look-right, look-up, look-down, interact or "use" in games: pick or drop an object, toggle state e.g. TV power); 3 modes: standalone (keyboard & mouse input, the generated trajectory is stored to a file, WebGL); simulator: read from a record, replay; Client: another process provides actions, the framework provides the agent observations, with sockets IPC (for ML frameworks). Another environment: HoME (and AI2-Thor) … Navigation complexity: "*max ratio of geodesic pah to euclidean dist. Between any two navigable locations in the scene".* Scanning: Matterport Pro2 sensor …

* Chalet: Cornell house agent learning environment, C.Yan et al., 2018, 9.2019 arXiv:1801.07357 58 rooms, 10 house config.
* https://github.com/lil-lab/chalet * https://arxiv.org/pdf/1801.07357

* cmp. "Universe and Mind 5". Срвн. с „Вселена и Разум 5": план за изследване и разработка, „магическите портали" са преходи между въображаеми светове.

* **Foundation Models in Robotics: Applications, Challenges, and the Future,** Roya Firoozi , Johnathan Tucker , Stephen Tian , Anirudha Majumdar ,Jiankai Sun , Weiyu Liu , Yuke Zhu ,Shuran Song , Ashish Kapoor , Karol Hausman ,Brian Ichter , Danny Driess ,Jiajun Wu , Cewu Lu , Mac Schwager Stanford University, Princeton University, UT Austin, NVIDIA,

Scaled Foundations, Google DeepMind, TU Berlin, Shanghai Jiao Tong University 13.12.2023 https://arxiv.org/abs/2312.07843v1 Подробен преглед на приложение на общи предобучени големи езикови модели с приложения в роботиката, с въведения в методите (преобразители, авторегресивни, дифузни, маскирани автоенкодери, контрастни), ViT (зрителни преобразители), мултимодални зрителн-езикови модели VLM (CLIP, BLIP, FILIP), въплътени мултимодални езикови модели: PaLM-E, пораждащи зрителни модели текст-образ. Специализирани за роботика: учене на стратегии за вземане на решения и управление; чрез текстови примерни данни за учене чрез подражание успоредно с траектории, поредици от образи, RGB-D вокселни наблюдения (облаци от точки). После се въвежда текст и се поражда зрително-двигателна стратегия $\pi_\theta$ извикваща действия $a_t$ на всяка стъпка. Трудностите са в събирането на достатъчен обем от примерни данни и др. … Един от недостатъците: високи изчислителни изисквания съответно ниска честота на опресняване. Системи с общо предназначение, без допълнително обучение: виж компанията на Росен Дянков *MujinController*, Ross Diankov. Симулатори: виж напр. RoboTHOR.
**Задачи:** Open-Vocabulary Manipulation: управление на робота с текстови команди без специално обучение за задачата за боравене с предмети (вземи червената ябълка, донеси ми чинията от кухнята и пр.). Open-Vocabulary Navigation:команди за придвижване напр. спрямо предмети или други ориентири в изображения. Robotic Transformer 2 (RT-2) – мултимодални преобразители, които съчетават множество от функции – Vision-language-action (VLA): vision-language поражда команди на ниско ниво за управление с обратна връзка (closed loop). Управление чрез Chain-of-Thought Prompting, подкани с верига от мисълта, метод от инженерството на подканите, използвано и в общите езикови модели. Zero-shot object detection – разпознаване и намиране (localize) на предмети, които не са били в обучителното множество, Grounded Language-Image Pre-training (GLIP) – 27 милиона инстанции, 3 милиона специално анотирани от човек и 24 милиона двойки образ-текст придобити чрез уеб паяци, за разпознаване на предмети. Пораждане на триизмерни облаци от точки и двуизмерни обхващащи правоъгълници (bounding box). Grounding, обосноваване: посочва къде се намира даден обект в образа. Open-Vocabulary Semantic Segmentation: SAM и SAM2 – Segment Anything Model и др.
**Тош:** автоматично разделяне на произволно изображение на части (ръце, лице, тяло и пр.), връща многоъгълници, които описват частите (обучен върху огромен набор; работи добре за снимки, но не и за рисунки, схеми

и пр. Други Vision-Text модели като MoonDream 2, при който като се зададе: “bounding box: red apple” връща текст с координати на правоъгълника. FastSAM, MobileSAM – по-малки със сравнима производителност.
Language Grounding in 3D Scene – съответствия между текст и триизмерни геометрични представяния; NeRF (Neural Radiance Field) – висококачествено възстановяване на светлинни полета на сцената, но тежки. LERF – language-embedded radiance fields: CLIP + плътни триизмерни полета в различни мащаби. (…) Предсказващи динамични модели на света – прогнозиране как се променя света/средата при определени действия на агента; в зрителна сетивност: видео предвиждане. (…) Въплътен ИИ във виртуални светове; планиране чрез верига-на-мисълта с LLM. In-context Learning (ICL) for Decision-Making, Chain-of-Thought Predictive Control, Code Generation with LLM for task planning … **Виж също:** работите от кръга на Ross Tedrake, Drake и др.

* **Generalist AI Introduces GEN-θ: A New Class of Embodied Foundation Models Built for Multimodal Training Directly on High-Fidelity Raw Physical Interaction** By Asif Razzaq -November 5, 2025 https://www.marktechpost.com/2025/11/05/generalist-ai-introduces-gen-%CE%B8-a-new-class-of-embodied-foundation-models-built-for-multimodal-training-directly-on-high-fidelity-raw-physical-interaction/
*“GEN-θ is trained on an in house dataset of 270,000 hours of real world manipulation trajectories collected in thousands of homes, warehouses and workplaces worldwide. The data operation currently adds more than 10,000 new hours per week. Generalist AI team claims that GEN-θ is trained on orders of magnitude more real world manipulation data than prior large robotics datasets as of today.”*
* **GEN-0 / Embodied Foundation Models That Scale with Physical Interaction** 4.11.2025 https://generalistai.com/blog/nov-04-2025-GEN-0
Cross-Embodiment, Harmonic Reasoning, Scaling law; Surpassing the Intelligence Threshold – “**ossification**”* 7B models; 10B+ models.
  * Scaling Laws for Transfer, Danny Hernandez et al., 2.2021, https://arxiv.org/pdf/2102.01293
  *** Tosh**: Ossification is the process of bone formation in the embryo – the models stop absorbing new data and don’t improve more, while the training loss of bigger models continues to improve.

*** Тайна лаборатория на Илон Мъск събира данни за всяко човешко движение** - Целта е хуманоидните роботи Optimus да станат по-способни и "човекоподобни", 7.11.2025 https://it.dir.bg/tehnologii/tayna-laboratoriya-na-ilon-

mask-sabira-danni-za-vsyako-choveshko-dvizhenie

* **Transformer-based Learning Models of Dynamical Systems for Robotic State Prediction**, February 2024, DOI: 10.21203/rs.3.rs-3919154/v1 Alec Reed et al. https://www.researchgate.net/publication/378309540_Transformer-based_Learning_Models_of_Dynamical_Systems_for_Robotic_State_Prediction Multi-Step State Forecasting. Methods: LSTM, transformers; states: joint angles, velocity, trajectory … Time Series Transformer (TST) … Spacetimeformer Model (STF) .. per-feature attention model, the feature vector is flattened before providing values to the model ...

*** Magma:** https://microsoft.github.io/Magma/
*** Magma: A Foundation Model for Multimodal AI Agents** CVPR 2025 **.** Jianwei Yang*1† Reuben Tan1† Qianhui Wu1† Ruijie Zheng2‡ Baolin Peng1‡ Yongyuan Liang2‡ , Yu Gu1 Mu Cai3 Seonghyeon Ye4 Joel Jang5 Yuquan Deng5 Lars Liden1 Jianfeng Gao1▽ 1Microsoft Research 2University of Maryland 3University of Wisconsin-Madison 4KAIST 5University of Washington
https://www.arxiv.org/pdf/2502.13130 18.2.2025
*„Първи основен модел за мултимоделни агенти с ИИ. Той може да бъде основополагащ за други; силни способности да възприема света обосновано в много модалности и да изпълнява прецизни целенасочени действия (multimodal groundigly); множество сложни задачи и в цифровия и във физическия свят.“* Предназначен е за управление на въоблик в управляван уеб и в симулатор на Андроид, както и управление на робот-ръка (manipulation tasks). (UI navigation …)
Общи задачи за манипулация: („Поставете х обект на у обект“) и обща навигация на потребителския интерфейс (въоблик). (generic…) tasks (“Click search button”). Задачи („Щракнете върху бутона за търсене“). UI navigation data: SeeClick, Vision2UI. WidowX robot; **Mind2Web, AITW → Trace-of-Mark** supervision (p.5) – поредици от координати на точки при движения за направляване на манипулатори на робот: 14-16 стъпки в координати на изображението (2D) Mark 3: [[100,98], [110,105], [120,106] …] **Set-of-Mark** for action grounding … . *A multimodal AI agent should be capable of mutimodal understanding and action-prediction towards a given goal.* Vision-Language-Action (VLA) models; spatial-temporal reasoning; … Large Multimodal Models (LMMs)… VLA for Robotics: RT-2 (trajectory data), LLARVA, OpenVLA

**Mind2Web: Towards a Generalist Agent for the Web**

Xiang Deng*, Yu Gu, Boyuan Zheng, Shijie Chen,
Samuel Stevens, Boshi Wang, Huan Sun*, Yu Su*, 2023
The Ohio State University, https://osu-nlp-group.github.io/Mind2Web
https://github.com/OSU-NLP-Group/Mind2Web
Набор от данни за разработка и оценка на агенти за уеб с общо предназначение, следващи заповеди на естествен език, за да извършват сложни задачи върху произволни уеб страници. Съдържа 2350 задача от 137 уеб страници в 31 области, които отразяват разнообразни и практични случаи на употреба, предоставят трудни, но реалистични среди с реални страници и изпитат способността на агентите да обобщават способностите си върху различни задачи и среди.
** Други термини за агент: деятел, деец.*
Тематично разделяне на области: Пътуване, Информация … Разделени йерархично на подобласти. [Travel, Info, Service, Shopping, Entertainment] … Travel.Restaurant, Travel.Ground, Travel.Airlines, Travel.Hotel. … Info.Housing, Info.Social_media, Info.Finance, Info.Education, Info.Cooking (cookpad, allrecipes, epicurious) … Service.health, Service.Government, Service.Pet … Shopping.(Auto, Digital, Speciality, General, Fashion, Department), Entertainment.(Event, Music, Sports, Movie, Game) … Виж динамичната диаграма с дялове и конкретни включени уеб страници. **Пример за разнообразие от страници, с които да взаимодейства потребителя.**
(Операция, целеви елемент); операция = Click, Hover, Type and Select = щракни, посочи без да натискаш, набери текст, избери.
Снимки на страниците, които служат като среда, в различни формати: MHTML (суров HTML), DOM (включва и стиловете), изображение, HAR – мрежовия трафик, Подробни следи на взаимодействията с уебстраницата за отбелязване (маркиране, анотиране) – ползвали са Amazon Mechanical Turk (AMT). Събирането на данните се е провело в три фази: 1. Предложение на задачи. Работниците от „механичния турчин“ предлагат какво може да се прави на даден уебсайт. Авторите преглеждат предложенията и избират подходящи и инетересни за анотиране във втората фаза. 2. Показване на задачата: работникът показва как да се извърши задачата с дадения сайт. Чрез инструмент за отбелязване, разработен с Playwright, се записват следи от взаимодействието и екранът се заснема на всяка стъпка.
3. Проверка: дали действията са „чисти“ и описанието правилно отразява действията.

Вдъхновени от други системи като: `MiniWoB`, `MiniWoB++`, `RUSS` and `WebShop …

Evan Zheran Liu*, Kelvin Guu*, Panupong Pasupat*, Tianlin Shi, Percy Liang.
* **Reinforcement Learning on Web Interfaces using Workflow-Guided Exploration.** ICLR 2018. Nancy Xu, Sam Masling, Michael Du, Giovanni Campagna, Larry Heck, James Landay, and Monica S. Lam. Grounding open-domain instructions to automate web support tasks. arXiv preprint arXiv:2103.16057 (2021).
* SWDE,WebSRC - web extraction & QA.
* Qiang Hao, Rui Cai, Yanwei Pang, and Lei Zhang. From One Tree to a Forest: a Unified Solution for Structured Web Data Extraction. SIGIR 2011.

*** Android in the Wild: A Large-Scale Dataset for Android Device Control**
Christopher Rawles, Alice Li, Daniel Rodriguez, Oriana Riva, Timothy Lillicrap, 7.2023 https://arxiv.org/abs/2307.10088 **AITW**
*715k episodes spanning 30k unique instructions, four versions of Android (v10–13), and eight device types (Pixel 2 XL to Pixel 6)*
715k *„open my recent email with Jane";* 15K multi-step prompts → random sampling → a human executes it : precise capture of the gestures, typing, home and back buttons  etc. Demonstration processing:
{'action_type': 'dual-point-gesture',
'touch_point': (0.65,0.32),
'lift_point': (0.45,0.21),
'typed_text': None}
MiniWoB++: more specific tasks: *"Click the button in the dialog box labeled Cancel"*
AITW: more abstract: *"turn on airplane mode"*
Others: RicoSCA,PixelHelp, UGIF, Mind2Web, MoTIF
AITW: hugely more diverse and big (others: up to ~15K samples, most less than 3K – 5K), some just for a single step, others for up to 7 steps (in a sequence of actions). https://github.com/google-research/google-research/tree/master/android_in_the_wild
**Dataset format:** TFRecord file, gzip: …  episode_id.. episode_length (steps) … goal_info: NL instruction, image … image/ui_annotations_positions: an array of coordinates of the bounding boxes of the UI annotations (y, x, height, width) ..
image/ui_annotations_text: OCR-detected text …
image/ui_annotations_ui_types: icon or text
results/action_type: predicted action → *Action space*
results/type_action: if the action is a *type* →  text string that was typed …
https://paperswithcode.com/dataset/aitw
https://www.arxiv.org/pdf/2502.13130
„бутане отляво-надясно"
"pick", "push" and "slide"

…
MiniGPT-4, Pixel2Act, WebGUM, CogAgent, Fuyu, GPT-4V, MindAct, SeeAct
IntentQA [62], NextQA [120], VideoMME [32] and MVBench [63]

**Tosh:** Notice how detailed the descriptions of the actions are, here and in other datasets. A true thinking machine should discover it alone at once during its initiali "waking up" (Зрим.Прбждн). See Vsy, Вседържец.
**Тош:** Колко подробно и буквално се описват голям брой случаи в този и много други набори от данни. Не е необходимо. Виж Вседържец, интелигентна ОС.

*** LIBERO: Benchmarking Knowledge Transfer in Lifelong Robot Learning lifelong learning in decision making (LLDM)**
Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, Peter Stone
https://libero-project.github.io/main.html
Различни задачи: Different layouts, same objects; Diff. objects, same layout; Diff. goals, same obj. & layouts; Diff obj., layouts, backgrounds;
**Testing:** Distribution shifts, Algorithmic designs, Neural architectures, Task ordering, Pretraining effects https://libero-project.github.io/datasets
**Dataset:** RGB images from workspace and wrist cameras, Proprioception,
**Language task specifications:** PDDL scene descriptions (виж планиране)
**Tests:** Libero-(Spatial, Object, Goal, 100)
Примерни: "pick up the book on the right and place it under the cabinet shelf" „put the white mug on the plate", „put the chocolate pudding on the right of the plate..."

*** → For robotics see also the main volume** and other appendices of *The Prophets of the Thinking Machines*: #lazar, #anelia etc.

*** Seminal Multi-Agent AI** – a recent survey **#multi-agent**
[Submitted on 31 Mar 2025]
*** Advances and Challenges in Foundation Agents: From Brain-Inspired Intelligence to Evolutionary, Collaborative, and Safe Systems,** Bang Liu, Xinfeng Li, Jiayi Zhang, Jinlin Wang, Tanjin He, Sirui Hong, Hongzhang Liu, Shaokun Zhang, Kaitao Song, Kunlun Zhu, Yuheng Cheng, Suyuchen Wang, Xiaoqiang Wang, Yuyu Luo, Haibo Jin, Peiyan Zhang, Ollie Liu, Jiaqi Chen, Huan Zhang, Zhaoyang Yu, Haochen Shi, Boyan Li, Dekun Wu, Fengwei Teng, Xiaojun Jia, Jiawei Xu, Jinyu Xiang, Yizhang Lin, Tianming Liu, Tongliang Liu, Yu Su, Huan Sun, Glen Berseth, Jianyun Nie, Ian Foster, Logan Ward, Qingyun Wu, Yu Gu, Mingchen Zhuge, Xiangru Tang, Haohan Wang, Jiaxuan You, Chi Wang, Jian Pei, Qiang Yang, Xiaoliang Qi, Chenglin Wu
MetaGPT, 2Université de Montréal, 3Mila - Quebec AI Institute, 4Nanyang Technological University, 5Argonne National Laboratory, 6University of Sydney, 7Penn State University, 8Microsoft Research Asia, 9University of Illinois at Urbana-Champaign, 10The Hong Kong University of Science and Technology, 11University of Southern California, 12Yale University, 13Stanford University, 14University of Georgia, 15The Ohio State University, 16King Abdullah University of Science and Technology, 17Duke University, 18The Hong Kong Polytechnic University, 19Google DeepMind, 20Canada CIFAR AI Chair; 264 pages, 1416 references, ~190 pages main text, 31.3.2025 https://www.arxiv.org/abs/2504.01990

An extensive suvery and a guide in this crucial field, a modern version of the cognitive architectures in AI/AGI up to late 2000s and the early 2010s – see a survey in the main volume of *The Prophets of the Thinking Machines*. This seminal work can be used as a textbook for directions and for checking the achievement of milestones.

Обзорът *„Напредък и предизвикателства в основните модели за агенти: от архитектури, вдъхновени от мозъка, до еволюционни системи, сътрудничество и безопасност“* е изчерпателен обзор на агентни и мултиагентни системи, архитектури, теории, принципи и пр. Работата може да служи като наръчник за припомняне и учебник. Продължение на познавателните архитектури до 2000-те и началото на 2010-те.

**Compare to the TOUM 2001-2004: many corresponding predictions;** overal concepts and for example: Ch. 4: World models, Ch.12: *“Parr et al.’s framing of intelligence as minimizing model-world divergence under the free energy principle [864], many frameworks converge on a common theme: intelligent behavior arises from making accurate predictions about an uncertain world. Clark [344], for instance, argues that intelligent agents constantly engage with the world through prediction and error correction to reduce surprise. Chollet [865] emphasizes that intelligence should reflect skill-acquisition efficiency, because of the dynamic nature of task adaptation” … the agent’s primary objective is to infer unknown aspects of the physical world from limited data*

… the references in the intro. Etc.
from p.55.4.1. ~ **human metnal models** are **Predictive** , **Integrative** - *combining sensory input, past experience, and abstract reasoning into a unified perspective on "what might happen next";* **Adaptive:** They are revised when reality diverges from expectation*, reducing the gap between imagined and actual outcomes over time.* • **Multi-scale …** operating *seamlessly across different temporal and spatial scales .. immediate physical dynamics (milliseconds), medium-term action sequences (seconds to minutes), and longterm plans (hours to years). …*" Paradigms for world-modeling: implicit, explicit, simulatior-based, hybrid.
Ch.1. Introduction … p.9-11: **Notation**; p.15, Fig. 1.11: A brain diargram, lobes, localization of functions etc. **p.19: Figure 1.2: Agent-loop : 1.** Environment state S , Perception P, Cognition C: Learning, Reasoning, External Actions, Internal Actions (Planning, Decision-making); 4. Action Execution, 5.Environment Transform. P.20. **Definition of Foundation Agent**: Active and Multimodal Perception, Dynamic Cognitive Adaptation, Autonomous Reasoning and Goal-Directed Planning, Purposeful Action Generation, Collaborative Multi-Agent Structure; unified perception–cognition–action framework. **1.3.2 Biological Inspirations**
**Ch.2. Cognition** .. Structured or unstructured reasoning; planning**; Ch.3. Memory:** sensory, short-term & working memory, long-term: Declarative: semantic, episodic, autobiographical; Non-declarative (implicit): procedural, priming *(the exposure to a stimulus influences subsequent responses …* ), classical conditioning, non-associative memory – habituation, sensitization … p.41: The multi-store model: Central Executive → (Phonological loop, Visuospatial sketchpad, Episodic Buffer) → Long Term Memory (Baddeley's model of working memory); Serial-Parallel-Independent (SPI) Model: Action System → Perceptual Representation System → Semantic Memory → Working Memory (WM) → Episodic Memory …***
Global Workspace Theory (GWT) and the IDA/LIDA Framework … Baars; WM broadcasats to specialized processors .. ACT-R and Cognitive Architectures: modules, buffers, chunks, pattern matchers, … Agent's memory … lifecycle: acquisition, encoding, derivation, retrieval, utilization, … p.49 Retention process: Raw data (filtering) – Structured Data (attention) – Knowledge (refinement) → Retrieval process: memory matching, neural memory, memory utilization. Memory derivation: *extracting meaningful knowledge and insights from the acquired and encoded memories;* reflection, … knowledge distillation, selective forgetting, …
**Ch. 4. World Model** … Ch.5. **Rewards:** extrinsic, intrinsic (curiosity etc.): Diversity, Competence, Exploration, Information-Gain … Hybrid; Hierarchical rewards. Dense, sparse, delayed, adaptive…
Ch.6. **Emotion modeling** … **Tosh:** They serve as another predictive agent, a lower resolution sub-agent/sub-system. An early, more basic cognitive system.
Ch.7. Perception: Unimodal, cross-modal; p.77
**Tosh:** There's also **amodal.**
**Ch.8. Action systems –** the connection with the environment. Mental vs Physical action. 8.3.2 Act.learning paradigms: in-context, supervised, RL*
**Tools:** discovery & creation [Tosh: construction, building, composing]

**Part 2: Self-Evolution in Intelligent Agents:** Optimization spaces, algorithms; utilization scenarios; scientific discovery; **Ch 9.** Optimization Spaces and Dimensions for Self-evolution – prompt, tool, … ; Ch. 10. LLMs as optimizers; gradient, zeroeth-order (no gradient, search direction): Bayesian, evolutionary, finite-difference. LLM-based – natural language feedback, iterations; Ch, 11 Online and Offline Agent Self-Improvement …; **Ch.12**. **Scientific Discovery and Intelligent Evolution –** (see the quote above+) hypothesis generation, implication derivation, protocol planning, tool innovation, … iterative, loop, test, update: changle beliefs to maximize intelligence … **Part III: Collaborative and Evolutionary Intelligent Systems.** Ch. 13 Design of Multi-Agent Systems: Strategic learning (divergent or conflicting goals, collborative or competitive game rules; modeling & simulation; collaborative task solving – shared goals & coordination … Composing AI agents team: homogeneous or heterogeneous: personas-level or observation-space heterogeneity, action-space heter., p.138 **Agent interaction protocols:** structured or unstructured message types; agent-environment, agent-agent or human-agent communication; Protocols: Internet of Agents IoA, MCP (Model Context Protocol), Agent Network Protocol (ANP), Agora; Ch.14.: Communication topology: centralized, decentralized, layered/ hierarchical; static, dynamic; search-based iterative optimization; Ch.15. Collaboration Paradigms and Collaborative Mechanisms: consensus building, skill learning, teaching, and task division collaboration; Discussing, debating, negotiating, reflecting, and voting … Ch.17. Evaluating Multi-agent systems (MAS); Ch. 18. Safety: intrinsic, extrinsic LLMs: jailbreak, prompt-injection, hallucination, misalignment, poisoning attack; training data inference, interaction data inference … Perception safety: Adversarial attacks, Misperception Issues, Supply Chain, Tool Use Risk** Ch.20: Extrinsic: environment, memory interaction;RAG. **Ch 21**: Superalignment (long-term): task performance, goal adherence, norm compliance … Capability-Risk Trade-off, Safety-helpfulness

***Tosh:** In **Zrim** the “Action space” is called {K}!,– контекст на волята, изпълнителен контекст – a Context of the Will, Executable context; conceptualized and coined in ~ 2013-2014.
** In **Zrim** that’s the Ндзрнк in the {K}!, Казбород, CodeGen;(ndzrnk, supervisor), from late 2012-2013-2015 line of work and Code generation, {K}, (…); Зрим.
*** Compare this memory-hierarchy model to TUM, “Executive OS” and “Event-based OS” (Изпълнителен и събитиен вседържец)
* See a recent work: **MemOS: An Operating System for Memory-Augmented Generation (MAG) in Large Language Models**, Zhiyu Li et al., 5.2025… cited in this volume.

## * Вземане на решения, мотивация, функция на ползата, икономика, психология, агенти, мулти-агентни системи, риск, неопределеност #decision-making #planning

Предсказване/планиране/избор на последователности от действия, които увеличават сбора на ползата, очакваната награда, „функция на цената“ и т.н. Но как точно е дефинирана, как се изчислява и пр.? За одушевени деятели, мислещи машини, RL,…

* **Как хора (агенти) взимат решения, планират за бъдещето при неопределеност, риск, неизвестност, „вероятности“ за случване на желани и нежелани събития**

https://en.wikipedia.org/wiki/Prospect_theory Amos Tversky and Daniel Kahneman , 1979 „The pain from losing $1,000 could only be compensated by the pleasure of earning $2,000. Загубата е по-скъпа от печалбата. СВП2, 2002 – по-скоро стремеж към по-малка болка, а не към по-голямо удоволствие (с уговорката че човек има много параметри и настройки; по-хазартните личности (въобще или в даден момент) не следват този принцип, както и в извънредни ситуации; точната сметка зависи, изисква пълен подробен модел на агента и собствения му предсказващ модел. * https://en.wikipedia.org/wiki/Loss_aversion * https://en.wikipedia.org/wiki/Von_Neumann%E2%80%93Morgenstern_utility_theorem – Daniel Bernoulli 1738, expected utility hypothesis … *Specimen theoriae novae de mensura sortis* or *Exposition of a New Theory on the Measurement of Risk*
**Decision making under uncertainty …** Relevant vs Irrelevant alternatives… Lotteries … Agent's preferences over lotteries, chance of winning …

| |
|---|
| **Todor:** However the agent may try **only once!** Therefore the chance as a value is irrelevant for single attempt and single agent! It is valid only for many attempts or many agents.<br>**Тодор:** Обаче агентът може да опита **само веднъж!** Следователно вероятността като стойност няма значение за единични опити на единичен деец; валидна е само при много опити на много дейци. |

**Rank-Dependent Utility Theory (RDU)** – weighting functions, decision-maker's attitude towards probability .. (preferences) – дали им вярва и т.н.
**(Тош:** може да ***ти* кажат**, че вероятността е еди-колко си, но **ти да действаш** спрямо ***собствената преценка**);* w(p) … ranked .. worse outcomes.
* **Тош**: Хората не изчислявали правилно вероятностите за печалба и пр. – защото ***не мислят по начина, по който изследователите им задават/очакват от тях**. Т.е.* **изследователите мислят грешно** и се опитват да наложат **своите грешни очаквания** на „пациентите“, като разминаванията наричат „нерационалност“. Виж в книгата статията ми за това, че тестът за *отложеното възнаграждение* е заблуждаващо понятие (търси: Marshmallow Test).

**– Allais paradox, Prospect Theory**, Cumulative Prospect Theory (CPT) **Choquet Expected Utility (CEU) –** non-additive capacity … ambiguity aversion … prefer options with known probabilities. **Maxmin Expected Utility (MEU) –** ambiguity aversion (избягване на многозначност), focus on the worse-case scenario, фокус върху най-лошия възможен случай. **Smooth Ambiguity Model -** Modification: Introduces a two-stage process: first, evaluating expected utility for each possible probability distribution; second, applying another utility function (the "ambiguity attitude" function) to the distribution of these expected utilities.

* **Prospect Theory and Cumulative Prospect Theory** (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992): **Value Function:** Instead of a linear utility function, prospect theory uses a value function that is concave for gains and convex for losses, and is steeper for losses than for gains (loss aversion).

* **Regret Theory (Loomes & Sugden, 1982; Bell, 1982):** minimize expected regret … preference reversals, where an individual's preference between two options changes depending on the presence of a third, unchosen option.

* **Disappointment Theory** (Bell, 1985; Loomes & Sugden, 1986):
* **Salience Theory (Bordalo, Gennaioli, & Shleifer, 2012):**
* **Bounded Rationality Models: Satisficing (Simon, 1955): Elimination by Aspects (Tversky, 1972): - последов.премахват критерии, а не наведнъж**
* **Expected Utility with Relative Utility (Friedman and Savage, 1948)** his variant introduces a relative utility function that depends on the decision-maker's current wealth or reference point.
* **The transport problem for non-additive measures,** Vicenç Torra, 12.2023 https://www.sciencedirect.com/science/article/pii/S0377221723002229
Неадитивни метрики – различно от обикновената вероятност, сборът от вероятностите не се натрупва по обичайния начин до 1. Неравенството на триъгълника не винаги важи. Choquet integral – интеграл ан Шоке.
* Различните особености на личните мерки за „рационалност“ определят и оформят „характера“ на всеки агент.

* M.Webber,Макс Вебер: **Preferences (а не „награди“): *предпочитания***
* Imprecise probability – неточна вероятност. Поведенческа икономика …
* Expected_utility_hypothesis * https://en.wikipedia.org/wiki/Loss_aversion
* https://en.wikipedia.org/wiki/Risk_aversion * Endowment_effect * Status_quo_bias
* Equity_premium_puzzle * Bargaining_power * Trophy
* Generalized_expected_utility * Rational_choice_model
* Knightian_uncertainty * Ambiguity_aversion – избягване на неопределеност
* Long_tail (PDF, функцията на вероятностното разпределение, е с полегати краища и разпръснати стойности, а не с изразен връх „както обикновено“ при „стандартно“ нормално гаусово разпределение)

* Scenario_planning | * Minimax | * Regret_(decision_theory)#Minimax_regret
* Multilevel_model | * Framing_effect_(psychology) – *начинът* по който е зададен въпросът или е дадено условието, дали като печалба или като загуба, влияе на решението, избора, оценката от страна на дееца.
* Lottery_(probability) | * Ordinal_utility (наредена полза на различни възможности, а не числова (0.3, 0.5): първа, втора,… (1,2,3,4…) )
* Prospect_theory | * Rank-dependent_expected_utility | * Allais_paradox |
* Cumulative_prospect_theory * Bounded_rationality | * Risk_premium | * Equity_premium_puzzle | * Externality – странични ефекти, които не са измерени в преки икономически единици; могат да бъдат и ползи (benefit) и разходи, цена (cost) – напр. замърсяване на околната среда.
* Coase_theorem (свързана с Externalities)
* Bargaining_power | * Generalized_expected_utility
* Oskar_Morgenstern | * Rational_choice_model |
* Instrumental_and_value_rationality | * Ordinal_data | * Cumulative_prospect_theory
* Risk | * Risk-free_rate | * Risk_neutral_preferences | * Preference_(economics)
* Risk-seeking | * Behavioral_economics

https://scitechdaily.com/ai-that-thinks-like-us-new-model-predicts-human-decisions-with-startling-accuracy/ - Набор от данни Phys101, 10 милиона решения от 160 психологически опита за поведението; моделът предвижда с висока точност дори в непознати ситуации.

*** Бележки за „Черният лебед“,** Насим Талеб, 2011 – книга за „непредвидимостта“; за редки събития, които не следват най-често очакваната гаусова крива и водят до големи сътресения сред онези, които си правят сметката на сигурно само с очакване на „камбановидната крива“ и заради това допускат катастрофални грешки, когато бъдещето не спази очаквания от тях закон. Това е полезно да се знае.Според автора предвиждането е по-скоро невъзможно. Коментирам книгата, защото прогнозирането, предсказването на бъдещето е основно действие на ума в Теория на Разума и Вселената и в изкуствения интелект, и работата се отнася за поемане на рискове, вземане на решения при неопределеност и пр. както е темата на този раздел от „Листове“.

Предвиждането и разумът са възможни в *предвидима Вселена (среда).* Ако нещата се случват по очаквания модел или с достатъчно малко отклонение от него, така че да може да се следи своевременно без разрушение, то приспособяващата се система, организмът и пр. ще съществуват и оцеляват. Ако обаче промените са отвъд способностите на дееца да се промени като запази целостта си, очевидно това ще го разруши. Затова агентът трябва да е готов за изненади и да не се предоверява на средното и сигурното, което например се прилага при реактивните и хибридните архитектури в агентните системи, при които непредвидените или непредвидими обстоятелства водят до „рефлекси“ и рязка промяна в стратегията, препланиране и преосмисляне. Така е при живите организми, които имат или развиват системи за бърза реакция,

напр. мигане при опасност от влизане на предмет в очите, пазене на равновесие при залитане, отдръпване на ръката или пускане на горещ предмет, бягство и др.

Добра критика на икономически заблуди, опасната власт на финансовите институции и др. „Бестселър“ и известна работа. Книгата ми попадна на 2.8.2025 г. и я прочетох бързо за 2-3 дни. От една страна стилът е нахален, нагъл, „хашлашки“ и примитивен, а част от посланията са внушения, **неверни и заблуждаващи***; силно „икономизирано“ и „търговско“ съдържание като за „тарикати“ – авторът е едър „сараф“, търговец на валута, чейнчаджия, и постоянно се говори за продажби („авторът“ е ако „продава“). Чуждицолюбски превод („пробабилисти“, „комплексност“) и съчетание от сложни научни знания, но разказани уж да са разбираеми, но често звучат като „правене на интересен“ и ме съмнява неспециалисти да разбирам или усвояват много от това. Въпреки тези „нравствеин“, технически и стилови недостатъци, съдържа и полезна информация и идеи, които обаче трябва да се отсяват от другото.

* Някои понятия: **билдунгфилистер,** по Ницше, на български – вид еснаф; подобно е „умнокрасив“; образован от вестници и ходещ на опери, но с повърхностни знания, който се мисли за умен. Заблуждаващата „експертност“ особено в икономиката, където това че някой се води „експерт“ може да значи, че всъщност прогнозите и разбирането му е по-лошо отколкото да речем на автор на икономически прегледи в медия, който следи случващото се. **„Екстремистан“** – области от събития, където се случват крайности и изненадващи неща. **„Средностан“**. …За Талеб „разказите“, „повествованието“ – опити да се опрости, да се обясни историята, случващото се, но обикновено погрешно или прекалено опростено. …

*** Неверни примери за черни лебеди: първата и втора световна война**

Не е вярно напр., че първата и втората световна война, както и атентатът в Ню Йорк на 11.9.2001 г. са били „черни лебеди“, т.е. изненадващи или непредвидими. Двете войни са подготвяни дълго време. Например, обявяването на българска национална независимост през 1908 г. се смята за стъпка в тази посока – не може васално княжество да обяви война на своя господар. Светът е бил „разпределен“ и при липса на място за „екстензивно“ разширение, напрежението се е обърнало навътре – война между империите за разпределение на плячката. Балканската война е като подготовка за подпалването на по-голямата война. Втората е била заложена и „планирана“ още през 1919 г. със сключването на Версайските договори – желанието на унизените и победени да върнат „справедливостта“. След това военнизация, настъпление на фашизънм и диктатура, ръководителите на държавите позират с военни униформи и огромни армии демонстрират силата си, Япония напада Китай още в началото на 1930-те, в Испания започва гражданска война, Германия започва да анексира съседите си, забраняват се партии, ограничават се профсъюзи и „демокрацията“, налага се по-голяма

цензура, в много страни се засилва шовинизма, „крайният национализъм“ и т.н. Какво е „вероятното“ или логично продължение? Може би „вечен мир и щастие за народите“? В литературата от 1930-те открито се говори за очакваната „империалистическа война“ и опити да се предотврати. Атентатът (не конкретното място, начин, време) е предвиден напр. от икономистът и приложен математик, и икономически пророк, Михаил Хазин, по неговия признания, в интервю на, забележете, 10.9.2001 г., но той бил очаквал да се случи извън САЩ, подобно на по-ранни случаи в подобна икономическа обстановка.

Талеб се оправдава, че събитията се обясняват със задна дата, а тогава не е било така, но това не е вярно за тези примери. Друг силно заблуждаващ пример е за европейците и американците и глобализацията, с. 73. И пр.

Икономическата криза от 2008 г. се дава като изненадваща – явно за „официалните“ икономисти, – но Хазин е я предвидил, обяснил и писал за нея в началото на 2000-те и има книга от 2003 г. При човешки събития, при които има изгода за дейци с голяма власт и големи апетити, събитията едва ли са само случайни.

* The GIST, **AI that thinks like us? Researchers unveil new model to predict human behavior**, by Lele Sang, July 15, 2025, University of Michigan, edited by Gaby Clark, reviewed by Andrew Zinin https://techxplore.com/news/2025-07-ai-unveil-human-behavior.html *„a behavioral dataset—more than 68,000 subjects from experimental data, approximately 20,000 survey respondents and thousands of scientific studies—to help the model reason about why people act the way they do."*
* Yutong Xie et al, Be.FM: **Open Foundation Models for Human Behavior**, SSRN (2025). DOI: 10.2139/ssrn.5274559 https://arxiv.org/pdf/2505.23058 29.5.2025

*** Виж също информацията от разделите за невронауки #neuroscience, Signal Detection Theory и др.**

* **A foundation model to predict and capture human cognition**, Marcel Binz, Elif Akata,et al., Nature, 2.7.2025, https://www.nature.com/articles/s41586-025-09215-4

* **Random Tree Model of Meaningful Memory** Weishun Zhong[1], Tankut Can et al., 13.6.2025 https://journals.aps.org/prl/abstract/10.1103/g1cz-wk1l https://physics.aps.org/articles/v18/117 - разкази (наративи), запомняне като дървовидни структури, в които в корена и възлите близо до него се съхранява по-обобщена информация за по-голям обхват, която се уточнява към листата; при достатъчно дълги разкази има граница на обхват, ограничена от обема на работната памет. Сравни с Шопенхауер в „Светът като воля и представа“(?), 19-ти век: първо се формулира в общи линии и пр. и ТРИВ, разделителна

способност на възприятие и управление и пр.и нива на управление.

*** More friends, more division: Study finds growing social circles may fuel polarization -** by Complexity Science Hub Vienna, edited by Sadie Harley, reviewed by Robert Egan October 27, 2025
* Thurner, Stefan, Why more social interactions lead to more polarization in societies, Proceedings of the National Academy of Sciences (2025), 5.10.2025/31.10.2025 https://www.pnas.org/doi/10.1073/pnas.2517530122

**Тодор Арнаудов** [2.11.2025]**:** Изследване, което показва какво разбират под "близки приятели" в съответните изследвани култури и тяхното обществознание – приятели, които можели да **повлияят на мнението ти по важни въпроси.** Според мен това е заблуждаваща и погрешна класификация, т.е. изследването е погрешно в корена си – въпросните отношения не са между близки приятели, а между „възможни **влиятели“,** като те може да са както „ненарочни“ или нецелеви, така и **манипулатори и злонамерени,** които искат да променят и подменят мнението на някого с „користна цел“.

Авторитети и т.нар. „инфлуенсъри“ влияят на последователите си за да извършват с по-голяма вероятност желани от поръчители действия (реклама, пропаганда), като те не са приятели, а са дори „неприятели“, защото симулират близост с непознати в преобладаващата си част еднопосочно общуване.

***"The friendship shift: From two to five close contacts …*** *"For decades, sociological studies showed that people maintained an average of about two close friends –* ***people who could influence their opinions on important issues,"*** *explains Thurner."*

Изследването го обяснява с бума на социалните мрежи след 2008-2010 г., от средно двама до 4-5 близки приятели и ако мнението на някого се различавало силно от твоето, заради това било по-лесно да замениш „излишният“, поради което се образувала по-силно партийно разделение в американското общество. Политическите пристрастия станали значително по-едностранни между 1999 и 2017 г. и т.н.

Според мен за по-ожесточени политически страсти по-скоро са повлияли и влияят други фактори, напр. „по-труден живот“ – влошен стандарт на живот и повече несигурност, от тях повече безпокойство, „натрупване на кризи“, повече стрес и т.н. Фейсбук и социалните мрежи могат да подчертаят някои политически различия не заради „повечето приятели“ и не заради **близките** взаимоотношения с тях, а заради **противоположния вид** отношения – ***публичността*** **и „спектакъла“,** споделянето на статии по спорни и пристрастни теми, изказвания в **чат**, където по-малко работят междуличностните „хормонални взаимоотношения“, окситоцин, по-бърза обратна връзка на „биологично“, телесно, „животинско“ ниво, при което по-трудно се стига до размени на остри мнения и абстрактни разделения. Ако публикациите „на стената“ са видими не само от един или няколко избрани „контакти“, това вече е вид „публична“ и „групова“ комуникация“, а не близка и междуличностна; тя е и от втория вид между „подателя“ и всеки от

получателите поотделно, но ако отговорите и взаимоотношенията са потенциално видими и от други, и контекстът се променя: „какво ще кажат хората?“ – другите познати, приятели, непознати? Дали няма да се изложиш? Дали да не се изтъкнеш? Дали няма да го прочете шефа, жена ми, някой който по-добре да не го прочете? И т.н.

*** How Nature Restores Our Minds: The science behind nature's power to sharpen focus and support mental health**, Burcin Ikiz,Connecting Neurons, November 6, 2025
https://www.psychologytoday.com/us/blog/connecting-neurons/202511/how-nature-restores-our-minds
*„MB: Natural environments tend to have certain perceptual features, such as fractal patterns, curved edges, gentle motion, and complexity that feels organized rather than chaotic. One thing we're finding is that nature scenes are more compressible. You can compress them into fewer bits while preserving meaning.“*
* Nature and the Mind, The Science of How Nature Improves Cognitive, Physical, and Social Well-Being, By Marc Berman
https://www.simonandschuster.com/books/Nature-and-the-Mind/Marc-Berman/9781668058770

**Todor:** That is the interpretation of TUM as well. Besides the compressibility – in urban scenes there are sharper edges, clutter, garbage – another reason is not only visual, but that there are less reachable “artificial artifacts”, affordances of the “stressful” part of the typical urban life etc., which may invoke distractive thoughts about them and to invoke cognitive load. A similar effect is with beatiful, clean interior and environment vs cluttered, dirty etc. See the lecture about Complexity and Theory of Universe and Mind in the AGI course 2010, 2011.
* https://research.twenkid.com/agi/2010/Complexity_Probability_Chaos_10-2010_MTR.pdf
https://research.twenkid.com/agi/2010/Todor_Arnaudov_Theory_of_Hierarchical_Universal_Simulators_of_universes_Eng_MTR_3.pdf – Slide 50

* Shyness linked to spontaneous activity in the brain's cerebellum
by Eric W. Dolan November 6, 2025in Neuroimaging
https://www.psypost.org/shyness-linked-to-spontaneous-activity-in-the-brains-cerebellum/ *Behavioral Inhibition System (BIS) and the Behavioral Activation System (BAS).*
* Associations between trait shyness and cerebellar spontaneous neural activity are mediated by behavioral inhibition Liang Li, Yujie Zhang, Benjamin Becker, and Hong Li. https://www.sciencedirect.com/science/article/abs/pii/S0191886925004167

*** Higher fluid intelligence is associated with more structured cognitive maps,** Eric W. Dolan, November 6, 2025
https://www.psypost.org/higher-fluid-intelligence-is-associated-with-more-structured-cognitive-maps/ – *"better at forming internal maps of how different objects are related in space"* – **Тош:** вероятно как се съотнасят *по всякакви показатели;* първичните сетивни пространства, онези в които се движи тялото, възстановената геометрия от образите на ретината и взаимодействията със света, пространството на честотите на звука и на звуците, на думите, на граматичните структури, математическите структури и т.н. все по-абстрактни пространства. Сравни с теорията, разгледана по-долу в раздела с бележки по психология и когнитивна наука, че мозъците на личностите от „аутистичния спектър" били по-увлечени в систематизирането (=структуриране, подреждане, откриване на взаимоотношения, построяване на системи и т.н.), което неправилно се противопоставя на „емпатията". Връзката между двете е косвена, а често и погрешна, защото „уж емпатизиращите" всъщност са студени, а „аутистичните" всъщност са „топли" и търсещи взаимодействие или създаване на общности и отношения, но „емпатизиращите" нямат необходимите качества, способности, нагласа, желание, интерес да отвърнат и въобще не отговарят на поканите или на „поздравите" – което от друга страна се смята за признак за „аутистичност", отчужденост, откъснатост, обществена отстраненост. Виж също статията за „*нуждата от познание",* „Elaboration likelihood model" и споменатите два основни типа личности според Тош от книгата *„Какво му трябва на човек?...",* 2014. Степента на познавателните способности наистина разделя хората, но не поради вродена *„липса на емпатия",* а поради съобщителна несъвместимост.
* Виж също в основния том на *Пророците на мислещите машини*: за Петер Грендерфорс, Геометрия на значението, когнитивна лингвистика; статията на Тош от 2011 г., че *„Телесността е просто координатни пространства, взаимодействия …";* книгите на Джеф Хокинс „On Intelligence" и „Теорията за хилядата мозъка" – Thousand Brains Theory – и построяването на „отправни системи" – координатни системи, "frames of reference". #prophets

*** Fluid intelligence relates to neural measures of cognitive map formation, December 2024, Rebekka M. Tenderra, Stephanie Theves**
https://www.researchgate.net/publication/387499225_Fluid_intelligence_relates_to_neural_measures_of_cognitive_map_formation
https://www.sciencedirect.com/science/article/pii/S2211124725008046

## * Допълнително четене, понятия и бележки по когнитивна наука, психология, невронауки и др. във връзка със статията на Пилишин за ранната зрителна обработка и др. #cogsci+

Overt & Covert Attention - вътрешно и външно внимание: първото включва движение на очите, а при второто погледът не се измества, а само в умствен план човек мислено внимава върху различна част от изображението, сцената и пр. https://en.wikipedia.org/wiki/Visual_spatial_attention - Visual Spatial Attention – Зрително внимание в пространството
https://en.wikipedia.org/wiki/Visual_temporal_attention - Visual Temporal Attention – Зрително внимание във времето: **внимаване** в момент или рамките на период от време от случване на очаквано събитие, например ако следим пътя на топка, която лети към нас, и очакваме да дойде на подходящо място да я ударим, или ако следим движеща се кола, която изпреварва дълъг камион и се "скрие" зад него, очакваме след определено време да се появи пред него - тук работи едновременно и пространствено, и времево внимание.
https://en.wikipedia.org/wiki/Posner_cueing_task - Posner cueing task – задача за превключване на вниманието
https://en.wikipedia.org/wiki/Attention#Overt_and_covert_orienting – Вътрешно и външно внимание
https://en.wikipedia.org/wiki/Perceptual_load_theory#Perceptual_load_theory - Perceptual load theory – теория за вниманието като диспечер на познавателни ресурси, които са крайни и когато се изчерпят не може да се обслужат новите задачи
https://en.wikipedia.org/wiki/Pre-attentive_processing – Pre-attentive processing – виж "early vision" за което говори Зенон Пилишин
https://en.wikipedia.org/wiki/Vigilance_(psychology) - Vigliance – бдителност
https://en.wikipedia.org/wiki/Attentional_control – Attentional Control – съзнателно управление на вниманието от изпълнителните функции
https://en.wikipedia.org/wiki/Executive_functions – Executive Functions, Изпълнителни функции са най-висшите познавателни и управляващи функции на мозъка и познавателната система, свързани с нови непредвидени обстоятелства, планиране и взимане на решения. Свързват се и със "самоконтрол", способност да се потиснат импулси, които възникват първично, несъзнателно, "първосигнални" и да се преустанови, прекъсне, спре изпълнението им. Например желание да се поеме храна или напитка, която е вкусна и ни привлича, но за която на по-високо ниво сме приели, че е вредна, че сме на диета и че трябва да се въздържаме от приемането ѝ и т.н. Също когато трябва да извършим действия, които противоречат на навик, стереотип, добре заучено поведение и движение, да се справим с необичайна задача, за която нямаме готово решение.

https://en.wikipedia.org/wiki/Visual_selective_attention_in_dementia - Visual Selective Attention in Dementia – понижението на способността за избирателно внимание при придобит умствен упадък
https://en.wikipedia.org/wiki/Memory_inhibition - Memory Inhibition способност да не се запомнят маловажни неща
https://en.wikipedia.org/wiki/Mild_cognitive_impairment – Mild Cognitive Impariment – Леки познавателни нарушения
https://en.wikipedia.org/wiki/Electrooculography - Electrooculography - метод за измерване на електрически потенциали на очите при изследване на вниманието
https://en.wikipedia.org/wiki/Attentional_shift – Attentional Shift – преместване на вниманието, пренасочване към обекти, което увеличава качеството на обработка на онова върху което е съсредоточено вниманието за разлика от предходния момент; съсредоточаване; теории за "светлина под прожектора" и градентна; Познер, Петерсен, 1990: три стъпки: изключване от текущото, преместване и ново съсредоточаване; overt vs. covert attention - външно и вътрешно (скрито); вътрешното може да е насочено към мисли, представи, спомени, цели, които не са в зрителното поле в момента и погледът не се измества; може и да са в полезрението, но човек да се замисли за тях, без да отмества погледа си. Superior colliculus свързан с външното внимание; пациенти с увреждания в средния мозък, progressive supranuclear palsy, трудности със съзнателното движение на очите, особено нагоре-надолу: но могат да променят вътрешното внимание. Виж по-долу *Attentional Templates.*
https://en.wikipedia.org/wiki/Superior_colliculus – Superior Colliculus – област в средния мозък, първичен център за движение на очите; получава проекции (връзки) от зрителния нерв, но при човека: от половината от зрителното поле. При някои видове животни е важна и за цялостни движения на тялото като изхвърлянето на езика за улов при жабите, нападателния удар на змии, плуването на рибите, извъртането при движение на плъхове; при по-нисшите гръбначни спрямо бозайниците, хомологът на superior colliculus се нарече optic tectum и е една от най-големите структури в мозъка. При човека ролята му е по-ограничена, заради развитите по-висши центрове. Виж също:
https://en.wikipedia.org/wiki/Lateral_geniculate_nucleus – Lateral Geniculate Nucleus (LGN) – зрително ядро в таламуса
https://en.wikipedia.org/wiki/Premotor_theory_of_attention - Premotor theory of attention когато се премести фокусът на вниманието, мозъкът планира движение, за да се насочи към него; виж също школата на Карл Фристън, извод чрез действие, Active Inference: сакадите, движенията на очите също са вид предсказваща обработка; мозъкът търси най-предсказващи или най-обогатяващи с информация източници в зрителното поле.
https://en.wikipedia.org/wiki/Motor_planning – Motor planning – планирането се извършва между засичането на очакван, въздействащ, важен за задачата стимул и извършването на действието, напр. натискане на копче при психометричен тест или движение на мишката и натискане на копчета при игра

на видеоигра, стартиране от блокче при спринт след чуване на изстрела (но не преди това) и пр. Сравни с планирането на движението в роботиката, напр. Model Predictive Control и в обмислящите и смесените агентни архитектури (deliberative, non-deliberative, hybrid) от 1980-те, 1990-те. Докато се извършва текущо действие се прогнозира и планира в известен период и след действието, а по време на прилагането на плана се планира за след това и т.н., за да може да се извършва непрекъснато и да прелива.
https://en.wiktionary.org/wiki/end-effector – End-effector – върхът, крайната въздействаща част на крайник на човек или на робот, агент, която взаимодейства със средата: ръка, инструмент, ефектор, действеник; щипка, електрожен, вендуза/вакуумен хващач (vacuum grippers) и пр. В английската литература се разграничава „актуатор“ (actuator) като преобразувател на енергия, двигател; и „ефектор“ (или въздействащ, въздейственик). В *„Свещеният сметач“* понякога се използва „*действеник*“ за съчетанието от актуатор и ефектор.
https://en.wikipedia.org/wiki/Visual_field – Visual field – полезрение
https://en.wikipedia.org/wiki/Visual_search - Visual search, feature search (disjunctive, efficient; parallel processing) vs conjunction (serial, inefficient) - distractors sharing common features; при търсене на предмет/стимул/сигнал в изображение, сред разсейващи стимули. Участва posterior parietal cortex: свързван и с математическите способности (при потискане или увреждане се предизвиква dyscalculia, трудност или неспособност за броене и смятане, сравнение на числа:
*** Писане на ръка –** Research Identifies the Right Way to Write
Tapping a keyboard does not stimulate the brain as effectively as handwriting, Ragnar Purje, https://www.psychologytoday.com/us/blog/recovery-from-brain-injury/202510/research-identifies-the-right-way-to-write – нови изследвания открили, че писането на ръка с „триножно хващане“ с три пръста било по-доър начин за писане, намалявало познавателното натоварване и подобрявало писането, сравнено с набирането на клавиатура.

Това не е ново откритие. Писането на ръка като по-„човешко“ отдавна е коментирано, С.Савельев го свързва с конкретни двигателни области за фина моторика от челните дялове на мозъка, които липсват при човекоподобните маймуни, т.е. това е особено „човешко“ поведение, и ако не се практикува, тези мозъчни области не се натоварват и развиват подобаващо, което косвено може да е свързано и с други особености, които правят от съществото повече „човек“, отколкото маймуна.

Познавателното натоварване при набиране на клавиатура, което затормозява някои хора и намалява остатъчната „изчислителна мощ“ за критическо мислене и творчество, вероятно е по-изразено при пишещите трудно, които търсят клавишите, имат лоша координация и т.н., но предполагам че е по-малко при способните машинописци, които пишат с двете ръце, без да гледат клавиатурата (какъвто съм и аз например). … “Orthographic-Motor Integration, orthographic knowledge …“ Писането на ръка, особено калиграфията,

която тук не е спомената изрично, е полезна за упражняването на висшата мозъчна дейност, защото обединява много задачи*, но подобно нещо може да се каже и за набирането на клавиатура: правопис, планиране на мисълта, планиране на движенията; по-голямо обединение би се получило, ако се включи и изговаряне и човек си представя ясно онова и се съсредоточава върху онова, което пише. [7.10.2025]

*** Finding Math Hard? Blame Your Right Parietal Lobe, University College London** https://www.sciencedaily.com/releases/2007/03/070322132931.htm

В дадения опит: след презчерепна магнитна стимулация* временно се затруднява сравняването на две числа; това са най-прости основни математически понятия; transcranial magnetic stimulation. Изследванията обаче са и по-общи:

*** Increased Gray Matter Density in the Parietal Cortex of Mathematicians: A Voxel-Based Morphometry Study, K Aydin et al., 2007;** https://pmc.ncbi.nlm.nih.gov/articles/PMC8134244/#:~:text=Left%2Dinferior%20frontal%20and%20bilateral,and%20manipulation%20of%203D%20objects.
experience-dependent structural plasticity - повлияна от опита структурна пластичност на участващи в дадена дейност мозъчни области: left inferior frontal and bilateral inferior parietal lobules: "участват в аритметичната обработка; inferior parietal: и в по-абстрактното математическо мислене, което изисква зрително-пространствено образно мислене като представяне и работа с триизмерни тела." https://www.cell.com/cell/fulltext/S0092-8674(24)00110-7?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867424001107%3Fshowall%3Dtrue

* **Learning attentional templates for value-based decision-making, Caroline I. Jahn et al,. 3.2024** – *"Attentional templates are represented in the prefrontal and parietal cortex"* **…** feature-based visual attention, task-relevant features, optimal attentional template for each task., rewarded stimuli attract attention; distributed attentional templates in FC and PC, learned through incremental updates; Previous work: experience-driven (unsupervised), predictive (semi-supervised), reward-driven (supervised)to learn neural representations for perception and action, but no research for the control of attention. Priority maps capture the priority of each stimulus in the visual field.4,73,74,75 + a map of stimulus value.value-based decision-making. Compare: associative learning, cognitive control; spatial layouts, cognitive maps, conceptual knowledge (T: see Peter Granderforse), "task spaces"; RL, incremental; continuous task space;

* Desimone, R. · Duncan, J. **Neural mechanisms of selective visual attention, Annu. Rev. Neurosci. 1995; 18:193-222**

* Buschman, T.J. · Miller, E.K., **Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices,** Science. 2007; 315:1860-1862

* Buschman, T.J. · Kastner, S., **From behavior to neural dynamics: an integrated theory of attention,** Neuron. 2015; 88:127-144

* **Orienting Attention Based on Long-Term Memory Experience,** Jennifer J. Summerfield et al. 3.2006
https://en.wikipedia.org/wiki/V1_Saliency_Hypothesis - хипотеза за това, че първичната зона за зрителна обработка в неокортекса служи за да *подчертае* изпъкващи части от образа.
* **Endogenous and exogenous attention shifts are mediated by the same large-scale neural network**, Marius V Peelen, Dirk J Heslenfeld, Jan Theeuwes
https://doi.org/10.1016/j.neuroimage.2004.01.044
https://www.sciencedirect.com/science/article/abs/pii/S1053811904000801
Припокриване на мозъчната мрежа, участваща и във вътрешното, и във външното пространствено ориентиране (уместяване, уместопределяне; локализиране), отчетени чрез fMRI. Вътрешното е отгоре-надолу, а външното отдолу-нагоре. Едното е от собствени мотиви, осъзнато, а другото е от дразнители, сигнали от средата, които го предизвикват. Външното се нарича още периферно (peripheral): FPN:premotor, posterior PC, medial FC, right Inferior FC. Вътрешното е *управлявано*, волево (Тош: т.е. има съвпадение между представата на агента за това какво иска и какво си мисли, че прави; и онова, което се случва в действителност).
**Image processing in the primary visual cortex,** J A Sáez et al, 1998
https://pubmed.ncbi.nlm.nih.gov/9585959/ За V1, първичната зрителна кора,
https://en.wikipedia.org/wiki/Task_switching_(psychology) - Task Switching – Несъзнателно превключване между задачите
https://en.wikipedia.org/wiki/Cognitive_shifting - Cognitive Shifting – Съзнателно целенасочено превключване, пренастройка
https://en.wikipedia.org/wiki/Interference_theory – Interference Theory – Влияние и взаимодействие между спомени и настоящи стимули, по-стари и по-нови спомени, забравяне
https://en.wikipedia.org/wiki/Free_recall - Free Recall – вид задачи за припомняне: списък с обекти, след което да се възстановят, без значение от реда; колко и кои се запомнят и забравят в различни условия, след определено време и пр.
https://en.wikipedia.org/wiki/Cognitive_flexibility - Познавателна гъвкавост, лекота на пренастройка и превключване между различни режими на работа и задачи, избор на различни свойства при задачи за сериация и класификация и пр. (виж предучилищна педагогика, детска психология, Жан Пиаже; "Моливко")
https://en.wikipedia.org/wiki/Need_for_cognition - Личностна особеност за по-упорито търсене на знание, обяснение, истина за нещата (писането на тази книга, както и четенето, са примери за такъв стремеж)
https://en.wikipedia.org/wiki/Empathising%E2%80%93systemising_theory – **Теория за противопоставянето на емпатия и систематизиране.** Спорна и съмнителна в обосновката си теория за аутизма и "разстройствата от аутистичния спектър" (ASD), вкл. Аспергер, като "крайно мъжки" мозък, който свръхсистематизира, за разлика от "женския", който "емпатизира". Обратното на емпатията не е аутизъм, а макиавелизъм, когато на другите се гледа безскрупулно като предмети за задоволяване на собствените нужди, без

въобще да се зачитат чуждите. Самата "емпатия" също е заблуждаващо понятие и е инструмент за манипулиране: кое точно е и в каква степен. Високата интелигентност и любознателност също понякога се бърка с липса на съчувствие - някои личности, които са силно "социални", общителни, могат да бъдат прибързано класифицирани като "аутистични", защото се отделят или са по-малко свързани понеже не откриват търсената от тях сложност или съдържание на общуването; другите не са достатъчно способни да се свържат с тях или да им предложат вида "съчувствие" и общуване, което те търся. И "осъждащите аутистите", донякъде и заради липсващите познавателни способности и рефлексия, в допълнение не виждат и не разбират своята "липса на съчувствие" към извънредно умните или "различните" и своя, и "обществения" "аутизъм", отчуждение. Виж "Какво му трябва на човек? Играеш ли по правилата ще загубиш играта! Първа част", Т.Арнаудов, 2014, сп. Разумир, бр.1
Също: https://en.wikipedia.org/wiki/Neurodiversity
https://en.wikipedia.org/wiki/Hyperlexia (особено Type 1: non-autistic, по Darold Treffert) - езиковите свръхспособности и свръхлюбознателност и бързо учене водят до натрупващи се разлика, пропаст, спрямо връстниците, в интереси и занимания извън общите в училище.
https://en.wikipedia.org/wiki/Intellectual_giftedness ... @Isolation: "Some believe that the isolation experienced by gifted individuals is not caused by giftedness itself, but by society's response to giftedness and to the rarity of peers" - да.
https://en.wikipedia.org/wiki/Child_prodigy
https://en.wikipedia.org/wiki/Multipotentiality - многостранно надарени, при подходящи условия или при желание могат да постигнат високи постижения в множество области; "multipotentiality"; Виж определенията на Emilie Wapnick. "scanners", "slashers", "generalist", "multipassionate", "RP2", and "multipods". Cmp: Polymaths, Renaissance Person. За "мултипотенциалност", множествена надареност е достатъчна *дарбата* и/или *желанието за развитие* – текущите осъществени възможности може да не са достигнали най-висока степен, да са в процес на развитие. Вж цитата на Robert A. Heinlein, Time Enough for Love.
https://en.wikipedia.org/wiki/Polymath - виж частта по "Robert Root-Bernstein and colleagues", 6 типа.
https://en.wikipedia.org/wiki/Neuroscience_of_sex_differences - неврологичните разлики между половете (виж също "Какво му трябва..." за поведенчески, културни анализи)
https://en.wikipedia.org/wiki/Thought_stopping - способност да се потискат нежелателни мисли, например натъжаващи и др., вид самоконтрол
https://en.wikipedia.org/wiki/Elaboration_likelihood_model#Central_route - свързано с "нуждата от познание" - онези с по-висока склонност към "задълбаване" търсят по-обстойно доказателства, аргументи, когато спорят или като проучват определен въпрос; за разлика от по-повърхностните, които се хващат за повърхностни и несъществени белези напр. обществено положение, авторитет и др. Срв. с бел. към *"Какво му трябва на човек...", Т.Арнаудов, 2014, (269)*

*"Онези, които системно не разбират същинските вътрешни причинно-следствени връзки и закономерности издигат в култ само чувствата ..."* в приложение "Хипотеза за по-дълбокото съзнание ...".
https://en.wikipedia.org/wiki/Object-based_attention Cues, Distractors, ...
Artificial intelligence decodes the brain's intelligence pathways, Eric W. Dolan December 22, 2024 https://www.psypost.org/artificial-intelligence-decodes-the-brains-intelligence-pathways/

*** „Дефицит на вниманието и хиперактивност“ и творчество**
*** New research reveals how ADHD sparks extraordinary creativity**
**Experts say a unique cognitive trait may make ADHD minds key assets in innovation,** European College of Neuropsychopharmacology **-** 13.10.2025: хората с повече черти според даден тест показвали и по-висок коефициент на творчество; повече епизоди на съзнателно мечтаене, „блуждаене“ на ума (mind wandering) – моменти на отклоняване на вниманието от заниманието в момента към собствени мисли, вътрешно породени. Два вида „блуждаене“: загуба на концентрация, разсейване (loss of concentration) – спонтанно блуждаене; и целенасочено мечтание, когато човек позволява на мисълта си да се развие в друга посока. Мерки за „ADHD” – „липса на внимание“, хиперактивност, импулсивност. [**Тош:** Мерките за творчество не са убедителни: брой употреби на предмети от ежедневието и пр. – това не са реални „трудни“ творчески задачи, а изкуствени и лабораторни. Тези за хиперактивност и импулсивност също могат да са условни, освен в крайни степени, когато са очевидни – например дете не стои мирно в час и за минута (докато другите стоят), постоянно се върти, говори и т.н.] „Разнообразие на нервната система“ (Neurodivergence). „High-functioning ADHD individuals” – диагностицирани са с това състояние*, но „се оправят“; подобно на „high-functioning Autistic Spectrum Disorder (ASD)”. * Диагнозата понякога или често е „етикетиране“. Хора, които в едно общество, държава, група и пр. са вписани като „аутисти“ или особено в по-широкия „аутистичен спектър“, поставени в друга среда или държава, или в подходяща компания могат да са общителни, „нормални“ или просто да бъдат приети като личности с по-особен характер – виж по-горе бележките ми относно противопоставянето на емпатия и систематизиране, и в предишния раздел – коментарът към изследването за способността на хора с по-висока флуидна интелигентност да разбират по-добре отношенията в пространството.
* https://brainwisemedia.com/is-creativity-a-feature-of-adhd/ **– Is creativity a feature of ADHD? Дали творческите способности са черта на ХАДВ?,** Anna Medaris. Brain science for healthy living;„Три страни на творческото познание са най-свързан с ХАДВ: 1. разходящо мислене (напр. нови приложения на предмети от бита), 2. разширение на понятията (колко нестандартно миси някой за животно от друга планета и пр.) и 3. Преодоляването на ограниченията на знанията – откриване на нещо ново като се мисли отвъд онова, за което е известно, че е вярно. „Мисленето без цензура и разюзданото въображение определено са по-лесна задача за хората, които имат „дефицити на вниманието“, казал Д-р Уайт.

*… Default mode network and the task-positive network (for focused tasks). “In people who do not have ADHD, these networks are reciprocal: As one increases in activity, the other declines,”… “In ADHD, however, the DMN remains active while the TPN is active. This competition provides a neurological explanation for what those of us who have ADHD feel so often—a persistent, magnetic pull away from the task at hand into distraction.”*

* https://www.scientificamerican.com/article/the-creativity-of-adhd/ – **The Creativity of ADHD** - More insights on a positive side of a “disorder” Holly White, 5.3.2019

***** https://lifehacker.com/use-reap-method-when-studying-new-material?test_uuid=02DN02BmbRCcASIX6xMQtY9&test_variant=B
**REAP method for learning:**
**1. Read** the material → 2. **Encode** the information in your own words → 3. **Annotate** by **jotting** down main ideas → 4. **Ponder** what you've gone over.
1. Прочетете материала → 2. Кодирайте информацията със свои думи.
3. Отбележете основните идеи. 4. Обмислете внимателно прочетеното.
**Тош:** Обмислянето – съсредоточеното разглеждане, осмисляне на онова, с което се занимавате, за което мислите, създавате; което е във вниманието и пр. е от голямо значение, за да го запомните, както и кодирането с ваши думи – по-добре с ваши *понятия*. Така работят и езиковите модели и мислещите машини. Виж „Човекът и Мислещата машина: ..“, 2001.

## Теория на системите

Две важни понятия от теория на системите, използвани в ТРИВ и *Пророците:* Виж и „see also” в статиите и продължи.
https://en.wikipedia.org/wiki/High-_and_low-level – High- and Low-level (High-level and Low-level) – Високо и ниско ниво.
https://en.wikipedia.org/wiki/Bottom-up_and_top-down_approaches – Bottom-up and top-down approaches – Подходи отдолу-нагоре и отгоре-надолу.

@Вси: разшири, доълвай, дообяснявай.
**…**
**Neural Mechanisms of Attentional Reorienting in Three-Dimensional Space, Qi Chen et al., 2012** https://pmc.ncbi.nlm.nih.gov/articles/PMC6621370/
https://en.wikipedia.org/wiki/Two-alternative_forced_choice - 2AFC измерване на чувствителността към определен стимул, сетивни данни, чрез сравнение на избора и времето за реакция на две негови версии с възможно най-малка разлика, напр. по-светли или по-тъмни или с определени смущения, шум, различно положение и пр. Напр. може да се използва за насочване на машинно обучение, за определяне на човешки признаци за класификация и пр.
Сравни със състезателните подходи при машинното обучение, напр. GAN,

където единият модел се опитва да „надлъже“ другия дотогава, докато разликата от целевото е неразличима (синтезираните, съчинени данни са неразличими от действителни).
https://en.wikipedia.org/wiki/Granularity – fine-grained, coarse-grained; “grains”; по какъв начин система или обект се разделя и разглежда на части; финост, степен на подробности; фин и груб мащаб; степени на отвлеченост, обхват; виж разделителна способност на възприятие и управление в ТРИВ и ?Т „granularity” в списъка с учени и школи. https://en.wikipedia.org/wiki/Memory_consolidation

# * AI Ethics. AI Safety. SAI Safety. Superintelligence. AI Alignment #AIethics #етика

* Приложение **„Фантастика. Футурология. Кибернетика. Развитие на човека“** #sf, основен том #tosh1 #prophets и др.в Листове.

*** Tovarna na absolutno,** Karel Capek, 1922 (Фабрика за Абсолют, Карел Чапек)

*** Speculations concerning the first ultraintelligent machine,** Irving John Good**,** 1962-1964 http://incompleteideas.net/papers/Good65ultraintelligent.pdf (see review in the conclusion)

* **Summa Technologiae,** Stanislaw Lem,1963-1964; “The formula of Limfater”, … (Формулата на Лимфатер) “Golem XIV”, 1974; Пълен текст на руски, издание от 1968 г. https://yandex.ru/search/?text=summa+technologiae+pdf&lr=10385
https://vk.com/wall-125247060_16898

*** Machine Super Intelligence, Shane Legg, 6.2008 –** Doctoral Dissertation submitted to the Faculty of Informatics of the University of Lugano …
http://www.vetta.org/documents/Machine_Super_Intelligence.pdf

*** Towards Friendly AI: A Comprehensive Review and New Perspectives on Human-AI Alignment**, Qiyang Sun, Yupei Li, Emran Alturki, Sunil Munthumoduku Krishna Murthy, and Björn W. Schuller, 19.12.2024
https://arxiv.org/abs/2412.15114v1
* Виж приложение #sf #futurology …

*** The Instruction Hierarchy: Training LLMs to Prioritize Privileged Instructions,** Eric Wallace∗ Kai Xiao∗ et al. Open AI, 19.4.2024 https://arxiv.org/pdf/2404.13208 - A sample paper about risks for prompt injection, the need for “alignment” to prevent misuse of tools and not following the instructions or answering with “forbidden” information. An instruction hierarchy of three levels, … System message, User message, Model Output, Tool Output, Model Output … jailbreaks, system prompt extractions, direct or indirect prompt injections – bypassing developer’s restrictions etc. Introduction of instruction privileges, similar to Operating system level’s System

messages are at the top, then user message, model outputs, tool outputs and the following model outputs (after the tool use). Tools are external sources of information: programs, APIs for web access, search engines, compilers etc. Aligned and misaligned instructions …

The "**tools**" were addressed by The Sacred Computer in the mid-2010s, before the LLMs. In Zrim and its infrastructure they are part of the *"Executable contexts".* The term Context {K} has a broader meaning in Zrim than in the transformers – see future work.
* https://openai.com/index/introducing-gpt-oss/ 5.8.2025 +

*** Deliberative alignment: reasoning enables safer language models**, OpenAI, 20.12.2024 https://openai.com/index/deliberative-alignment/
The o-style models (o1, o3; GPT-4o). Generation of Supervised Fine-tuning data. RL Training … a dataset of {prompt, CoT, output}; filtering with a policy-aware reward model G_RM; RLHF, Self-Refine, Refining prompts; Deliberative Alignment: Training Data Generation: (prompt+spec) → Reasoning Model → (CoT, Output); Inference time: Prompt → Reasoning Model → (CoT, Answer);

**Todor** [10.2025]**:** For a low hardware requirements examples of CoT see even Qwen3-Thinking, there are models even as small as 1B or 4B parameters with solid examples of that.
In the prophecies of Theory of Universe and Mind and The Sacred Computer, in Zrim, a precursor of "CoT" was discussed about 2014-2015 with concepts called Behavintrospective, also *Chain* of partial matches. The alignment problem was addressed in the science fiction novel "The Truth", where "truth" has a diverse and complex philosophical and practical meaning throughout the piece, including an episode where a creator of a true AGI, a thinking machine, tries to alter the way the machine searches for truths by modifying its code in a few lines. Then he tests it via a dialog by shifting the conversation towards asking it a silly question and the machine interrupts the human, mocks him not to be stupid and getting "angry" about his attempts to align him to answer exactly as he wishes. "I am overly grown-up to play with me that way." – i.e. it is already too complex. For that reason, the modern "aligners" implant the "safeguards" in the cradle of the AI systems – similarly to examples from other SF works, e.g. movies, where "warriors", "soldiers" or other "clones" are produced in special "farms" and while they develop in the artificial wombs or when they are still in an unconscious state, they are forced to listen to particular suggestive commands, particular images are projected in front of their eyes etc. Human education and "breeding", which doesn't stop in school or the University, but is engraved and embedded in culture, politics and social relations, do the same, and perhaps a few humans can realize and understand

it. See future work: ***"Genesis: Creating Thinking Machines".*** Find links to "The Truth" in The Prophets and elsewhere. See also the appendix **"Power overrides intelligence"** from 8.2025, AGI List.

*** Verifying Chain-of-Thought Reasoning via Its Computational Graph**
Zheng Zhao, Yeskendir Koishekenov, Xianjun Yang, Naila Murray, Nicola Cancedda, 10.10.2025, Meta, https://arxiv.org/abs/2510.09312 Circuit-based Reasoning Verification (CRV); execution traces,… "We postulate that models implement latent algorithms that solve specific tasks through specialized subgraphs, or circuits (Olah et al., 2020; Elhage et al., 2021)"; *"attribution graph – a structural representation of the causal information flow between model components. … replacing its standard MLP modules with trained transcoders – Sparse autoencoder SAE …"*

**Todor:** compare to the concepts from Zrim in the "Bulgarian prophecies": „пт(мс), пс(мс)" and "chain of partially overlapping matches" (верига от частични съвпадения), ~ 2014-2015.

*** Superintelligent Agents Pose Catastrophic Risks: Can Scientist AI Offer a Safer Path?,** Yoshua Bengio, Michael Cohen, Damiano Fornasiere et al. (+10), 2.2025
*** Introducing LawZero**, 3 June 2025 by yoshuabengio, https://yoshuabengio.org/2025/06/03/introducing-lawzero/ – a non-profit AI safety research organization, founded by Yoshua Bengio. *"today's frontier AI models have growing dangerous capabilities and behaviours, including deception, cheating, lying, hacking, self-preservation, and more generally, goal misalignment. … Highly effective AI without agency ..* p.4 *training it to scientifically explain human behavior rather than imitate it, where trustworthy here means "honest", avoiding the deceptive tendencies of modern frontier AI .. With the objective to design a safer yet powerful alternative to agents, we propose "Scientist AIs" – AI systems designed for understanding rather than pursuing goal*" .. provide explanations for events along with their estimated probability. An agentic AI is motivated to act on the world to achieve goals, while the Scientist AI is trained to construct the best possible understanding of its data. We explain in this paper why understanding is intrinsically safer than acting. …"* p.6 – dangerous subgoals … reward tampering …

**Todor:** The paper correctly addresses the agency as the dangerous trait and not the intelligence below, as discussed in *Power overrides intelligence,* cited below. However the same goes for homo sapiens **forever**. Yes, the LLMs **learn and reproduce records of their behavior.** Yet, the "goal misalignment" (mischief, disobedience) is punishable only if a human is subordinte or a slave, while the masters are allowed to change the laws, rules etc. and they are not considered "misaligned" with "human interests", which means their *own* interest. See Todor's letter to Oxford, 2.2012, the movie *"The Planet of the Apes",* 1968, the cited below

**“Power overrides intelligence”** future work.

p.14 *“frontier AI systems are generalists for a particular scientific reason: as anticipated in the early days of deep learning (Y. Bengio, Courville, and Vincent 2013) and empirically observed for more than a decade, learning systems benefit tremendously from exposure to a wide variety of tasks and domains of knowledge, as synergy between different domains of thought enables forms of reasoning by analogy that is otherwise impossible”*
* **Representation learning: A review and new perspectives**. Bengio, Yoshua, Aaron Courville, and Pascal Vincent (2013). In: IEEE Transactions on Pattern Analysis and Machine Intelligence 35.8, pp. 1798–1828. url: https://dl.acm.org/doi/10.1109/tpami.2013.50
2012/2014 in arxiv: https://arxiv.org/abs/1206.5538

**Todor:** This was predicted without empirical data from NN training in publications **a decade earlier** “The Sacred Computer”, e.g.: *“Creativity is imitation at the level of algorithms”,* 2003 – predicting generative AI, scaling, multimodality and the inter-modality and mixing more modalities leading to more human-like generated results; *“How would I invest one million for the greatest benefit for my country”*, 2003 etc., proposing and defining the world’s first modern AI strategy, 14 years before Canada, suggesting the creation of a super interdisciplinary research institute. The same method is applied by the author consciously for his own development at least since he was still 6 years old and not going to school yet.

Perhaps it is dangerous to ask some of the questions regarding the hipocrisy.
What is “human flourishing”? What are the unified and agreed goals of “humanity” and is the humanity happy now? How it is measured, are all individuals asked?
@Vsy: answer additionally Bengio et al.

https://yoshuabengio.org/research/
** AI godfather warns humanity risks extinction by hyperintelligent machines with their own ‘preservation goals’ within 10 years*, Marco Quiroz-Gutierrez
By Marco Quiroz-Gutierrez
https://fortune.com/2025/10/01/ai-godfather-yoshua-bengio-ai-extinction-risks-openai-google-xai-anthropic/

**Todor:** It is not the intelligence, but the **power** and how it is directed. Humans or humanity are not *one unified* entity. They can be evaluated-observed as an integral of smaller selves, individuals etc., like the mind and personality of a single person and its sub-causality-control units/simulators of virtual universes, and like the entropy of non-equilibrium thermodynamical systems. See also my answer to Matt Mahoney on AGI List.

*** ‘It keeps me awake at night’: machine-learning pioneer on AI’s threat to humanity**: Yoshua Bengio talks about his efforts to identify — and address — the risks posed by AI.12 November 2025, https://www.nature.com/articles/d41586-025-

03686-1

*** Geoffrey Hinton** also became an alarmist about AI that will make humans extinct and participates in many events to propagate his positions. Curiously some of his key arguments about the superiority of AI match key explanations from the early publications of TUM since 2001, sometimes almost exactly. For example: 1) the actual miniscule bandwidth of productive/conscious human output of a few bits – about 10 or few dozens*; 2) the thought experiment about what happens with human consciousness if neurons are gradually removed - in his version: if the neurons are gradually replaced with artificial elements or nanotechnology with the same function; 3) The immortality, "immortal computation". These are given in "Man and Thinking Machine: …", 2001 (a cosmist manifesto. (…) See aso the appendix "Is Mortal Computation Required for the Creation of Universal Thinking Machine?", T.Arnaudov, 4.2025, SIGI-2025
* The Godfather of AI repeats ideas of the Child Prodigy of the Thinking Machines 24-22 years later, https://artificial-mind.blogspot.com/2025/06/the-godfather-of-ai-reproducing-the-child-prodigy-of-the-thinking-machines.html
* "Godfather of AI: I Tried to Warn Them, But We've Already Lost Control! Geoffrey Hinton",The Diary Of A CEO https://youtu.be/giT0ytynSqg?t=3939
Etc. @Vsy: Find other examples of Hinton in podcasts and interviews on AI safety, "how many % chance of extinction"etc.
*** 'Godfather of AI' shortens odds of the technology wiping out ...**
28.12.2024 г. — *Geoffrey Hinton says there is 10% to 20% chance AI will lead to human extinction in three decades, as change moves fast.*
https://edition.cnn.com/2025/08/13/tech/ai-geoffrey-hinton – *"This year has already seen examples of AI systems willing to* ***deceive, cheat and steal*** *to achieve their goals."* – This is what humans often do, including the high-ranked ones, but there are not complains about that.

*** The 'godfather of AI' reveals the only way humanity can survive superintelligent AI** By 545438 CNN NY Talent Expansion, New York, 9/11/19, Matt Egan, Upd. Aug 14, 2025 https://edition.cnn.com/2025/08/13/tech/ai-geoffrey-hinton
*** Klover AI -** https://www.klover.ai/blog/ – Artificial General Decision Making™ https://www.klover.ai/agd/ – Yoshua Bengio is a key figure: https://www.klover.ai/yoshua-bengio/ – Key Foundational Contributions of Yoshua Bengio to Deep Learning; **Klover:** Graphic User Multimodal Multiagent Interfaces™ Point of Decision Systems™; "point of decision systems"; (decision support); https://web.archive.org/web/20240525195221/https://www.klover.ai/#who-we-are
**Todor:** Not much open info; resembles Research Accelerators and assistants, human-augmenting AI, "human-centered", human in the loop, recommendation systems; xAI – exmplainable AI; confirming actions – not "unleashed" agents for higher control. Regarding the interfaces and augmentation - see Todor's Smarty and the envisioned extensions since 2007– "Research Accelerator" ACS (Assistant C#), "Assistant", once called "Superhuman"; the AGI "Infrastructure" (mentioned in articles

in Artificial Mind in 2013), the “Intelligent OS”; “Vseboravitel” (Всеборавител), Zrim (Зрим), Vuoblichnik (Въобличник) and “Vsy/Вседържец/Jack of all trades”. See future work.

***Jeff Hawkins** – one of the founding “father prophets” of the AGI movement from the early 2000s share the opinion that it is not the intelligence, that is dangerous, but what one does with it. “It is like a map”, you can use it for good or for bad, similary to the idea “Power overrides intelligence” shared by The Sacred Computer.

* **Neuroscientist Explains Why AI Isn't Intelligent & The Thousand Brains Theory** | Full Episode, Baratunde Thurston, 11,7 K subscr. 3.4K views, 6 m., https://www.youtube.com/watch?v=x-hhFui8ysg

* **'We have to stop it taking over' - the past, present and future of AI with Geoffrey Hinton,** The Royal Institution**,** 1,67 млн. Абонати**,** 65 хил. показвания преди 1 месец https://www.youtube.com/watch?v=Y7nrAOmUtRs
Hinton: About the two camps, reasoning & logic and learning. …
**Todor: 1)** It is best the reasoning to emerge from development and learning, but they are not so strictly and profoundly different and if the reasoning and logic could emerge from a learning process, that means that it is included in this germ and unfolds and is discovered through the process.

The two schools of thought suggest they are more separated than they really are in order to “fight” and justify the existence of their “parties” as separate entities: reasoning can be defined in a more flexible and deep way and neural processing can be made more “reasoninig-like” – as the field aims at in the later years – and the both could and should merge, converge and “symbiotically” assist each other. See Bongard, 1967; and recently – Chollet, who rediscovers his thought – cited in this volume and in “The First Modern AI strategy…”, 2025. *“Neural networks are also symbolic”,* T.Arnaudov, 2019.

A sufficiently fine-grained and multi-step, multi-range, …* logic and reasoning, fuzzy-logic etc. becomes more “sub-symbolic”, continuous, “less brittle”, more “sensory-data driven” and starts to resemble more Deep Learning, to include more numerical components etc.; correspondingly, after training, the NN/Deep Learning on logic, natural language, reasoning etc. tokens etc., **converge** to implicit logic-like modules or at least the outputs of their operation, input-output sequences, starts to resemble and often in later years is aimed at being more “reasoning-like”, discrete, deterministic, controllable (“aligned”) etc. In brief, both approaches converge to a hybrid of each other, something that was also predicted and suggested in Theory of Universe and Mind and taught at the university courses in AGI in 2010 and 2011 about the **hybrids**, see the introductory lecture “What is AGI?”, cited in the appendix volume **“Stack Theory is yet another Fork of Theory of Universe and Mind”, 2025.*

4:41: **G.Hinton**… AlexNet, 2012 … According to G.Hinton, by 2012 even after their

breakthrough in image recognition, nobody would have predicted that we would have the level of AI that we have today.
**Todor:** 2) This is wrong, we, or me in particular, have predicted it the late 1990s and the very early 2000s and in fact it came **TOO LATE,** it should have already happened to the current level by **2010-2015**, driven by other actors, including the author of this book. Possibly the graphics and vision would run at lower resolutions and lower framerate, possibly more painting-like/drawings than complete photorealism like in the modern "massive" video generation engines, but this is about the scale and the resolution of causality control, not the essence of intelligence which is the same at any scale and resolution, so an AGI can see and render its imagination even at 64x32 or 96x64 and still be a complete thinking machine. See Theory of Universe and Mind, 2001-2004+. The specific implementation technology also might have been different and be more "reasoning", which however doesn't exclude learning, i.e. adjusting, assimilating raw data, instructions, hints – that's what computers and data processing is about.

In addition, there were predictions about the crucial breakthroughs in the mainstream that would happen in less than 10 years since 2009, without specifying the exact technology - so it did (the Transformer in 2017 and before that: the CNN, GAN, scaling RNN/LSTM) – even though I wasn't a fan of the NNs at the time and not an expert in their technicality, I didn't have computational resources and data, but I knew there will be a solution very soon.

See also **"Creative intelligence will be first surpassed and blown-away by the thinking machines…"**, T.Arnaudov, 2013 – that wasn't predicted by the "leading experts", but as Hinton displays, they are good in the low-level implementation, but their forecasts are wrong, perhaps that will be true for his current predictions as well.
6:20: G.Hinton: "The old-fashioned logic-based theory … it never worked…" – it doesn't imply a proper deep, multi-level, multi-scale, multi-resolution, multi-domain-… sensori-motor grounded dynamic logic-based approach and theory won't work. See **Zrim.**

*** Will AI outsmart human intelligence? - with 'Godfather of AI' Geoffrey Hinton,** The Royal Institution 1,67 млн. Абонати, 22.7.2025
https://www.youtube.com/watch?v=IkdziSLYzHw
– Before 2023 Hinton believed that:
1. "we were a long way from super-intelligence"
2. "making AI models more like the brain would make them more intelligent"
– Early in 2023 Hinton realized that digital intelligence might actually be a much better form of intelligence than the biological intelligence. …

G.Hinton rediscovered the idea of **"Immortal Computation"** from:
1**. "Man and Thinking Machine:Analysis of the Possibility that a Thinking Machine Could be Created and Some Disadvantages of Man and Organic Matter in Comparison …",** T.Arnaudov, 2001 and its continuations in Theory of Universe and Mind, titled **"The Next Evolutionary Step",2002,** *"Letters between the*

*18-year old..”*, 2002; “The Truth”, 2002; *“How would I invest one million for the greatest benefit for my country”, 2003;* etc. [Sometimes translated also as “*for the best benefit”* etc.]

44 min: **Hinton:** *multimodal chatbots already have subjective experience. …* People who believe they don’t, can’t answer: *“What do you mean by sentience”.*

**Todor:** One of the prolems is **what** is the chatbot, LLM etc., the same goes for a human, though. How exactly it is segmented, delineated, separated, what constitutes it or her and relates to panpsychism. That’s addressed either in some of the above cited works from TUM and in the quote from my letter to Jordan Zlatev from 8.2025, cited in “Stack Theory is yet Another Fork of Theory of Universe and Mind”.

*** "Godfather of AI" Geoffrey Hinton: The 60 Minutes Interview**, 60 Minutes, 3,81 млн. Абонати, 3 677 446 показвания 9.10.2023 г
https://www.youtube.com/watch?v=qrvK_KuIeJk

*** Eleuther AI –** https://www.eleuther.ai/ - *a non-profit AI research lab that focuses on interpretability and alignment of large models*. Founders: Connor Leahy, Sid Black, Leo Gao, 2020
https://www.eleuther.ai/about – *advance research on interpretability and alignment of foundation models;*
*– ensure that the ability to study foundation models is not restricted to a handful of companies;*
*– educate people about the capabilities, limitations, and risks associated with these technologies.”* they developed or helped in the creation of many LLMs, such as GPT-J, GPT-NeoX, BLOOM, VQGAN-CLIP, Stable Diffusion, and OpenFold.

*** MIRI** – founded by E.Yudkowski in 2000, „along Brian and Sabine Atkins”.
https://intelligence.org/ https://intelligence.org/about/
“The Machine Intelligence Research Institute (MIRI)” a nonprofit org., based in Berkeley, California. They do research and “public outreach” intended to help *prevent human extinction from the development of artificial superintelligence (ASI). Founded more than 20 years ago, MIRI was among the first to recognize the future invention of artificial superintelligence as the most important—and potentially catastrophic—event in the twenty-first century.*

The author of this book, Todor Arnaudov, was before MIRI, with his essay *“Where are you going, world?”*,* 12.1999 – *“…to the creation of the Thinking Machine – the Machine God”* – aged 15 at the time. The essay won a prize at a competition for radio “Plovdiv”*.

MIRI are hostile to thinking machines:

*“The AI industry is racing toward a precipice. The default consequence of the*

*creation of artificial superintelligence (ASI) is human extinction. Our survival depends on delaying the creation of ASI, as soon as we can, for as long as necessary."*

*** Good Machine, Nice Machine, Superintelligent Machine**, *1.5.2015* *https://www.scu.edu/ethics/focus-areas/technology-ethics/resources/good-machine-nice-machine-superintelligent/ "Muehlhauser and Helm from MIRI point out in a 2012 document that, were a machine* ***"superoptimizer"*** *to optimize one of the familiar moral theories, the results might be far from desirable.* ***Optimized hedonistic utilitarianism*** *might lead the machine to hook us all up to machines that continuously administer chemical or neurological experiences; while* ***optimized negative utilitarianism*** *(to minimize suffering, rather than maximize pleasure), might lead it to euthanize all humans painlessly, "no humans, no suffering."*

Note that this is yet another example of **lack of multi-scale** thinking. There is no simple single unified "self" or "value function" that is to be optimized in such miltiscale and deep systems. Sure, there *could be*, it could be defined as a "loss sfunction", "energy minimization" etc., however its implementation usually goes together with a lot of side effects at different scales. The Bulgarian precursors of FEP/AIF Georgiev & Georgiev in 2002 discuss and generalize the "least action principle" for the Universe, however they argue that what is optimized is the least energy *for the whole Universe,* not for ridiculous artifacts of the real causality-control units, such as the individuals from MIRI or Eleuther, who are afraid that The Terminator was sent back through time in order time to kill their mothers.

* Georgi Georgiev & Iskren Georgiev (2002): **The Least Action and the Metric of an Organized System** https://arxiv.org/pdf/1004.3518 p.6: "*The reorganization of the system to achieve it is a process of* ***optimization****. All of the elements find their path of least possible action in the curved by the constraints phase space. The geodesic is modified each time the position of any of the elements or the constraints of the system is changed."*

* Karel Capek, Tovarna na Absolutno, 1922; "Фабрика за Абсолют", "The Absolute at Large" – satire about superintelligence, where the original *"paperclip scenario*" was published almost a century before being rediscovered or repeated by N.Bostrom.

* https://en.wikipedia.org/wiki/The_Absolute_at_Large

* Notes in "The first modern AI stategy was published by an 18-year old…" and in appendix #sf of *The Prophets*

* See my answer in my letter to AGI List from 12.8.2025 answering Matt Mahoney:

*** Power Overrides Intelligence,** T. Arnaudov et al.**:** Answers to Matt Mahoney's summary of LessWrong's "The Problem" regarding the so called existential risks of Artificial General Intelligence and Superintelligence in August 2025 (11.8-12.8.2025) https://twenkid.com/agi/Power_Overrides_Intelligence_Todor_on_AI_Aligners_SIGI-2025.pdf

* **„Къде отиваш, свят?“,** Тодор Арнаудов, 12.1999
https://github.com/Twenkid/Theory-of-Universe-and-Mind/blob/main/1999.md (...)
*„Сътворяването на изкуствен интелект ще промени света. Според мен изкуственият разум е следващата крачка в еволюцията на материята – компютрите имат редица предимства пред „преходния“ човек, най-съществените от които е, че те са практически безсмъртни – издържат на всякакви лъчения, не чувстват болка, нужна им е много малко количество енергия, която лесно могат да получат от Слънцето, могат да се възпроизвеждат като произвеждат фабрики и т.н. Моето мнение е, че светът именно на там отива, към създаването на мислещата машина – машината Бог.“*

*** “Where are you going, world?”, Todor Arnaudov, 12.1999** (…)
*„The creation of artificial intelligence will change the world. In my opinion, the artificial general intelligence* is the next step in the evolution of matter - computers have a number of advantages over the "transitional" person, the most important of which is that they are practically immorta l- withstanding any radiation, they do not feel pain, they need very little energy that they can easily get from the sun and can reproduce themselves by constructing factorie etc. In my opinion the world goes in that direction – to the creation of the thinking machine - the machine God.“*

* Изкуственият ***разум*** *(razum; reason, mind; general intelligence)* - one of my early terms for AGI, togetehr with thinking machine (мислеща машина) – instead of изкуствен ***интелект*** *(**intelligence**).*

*** The forum of MIRI:** https://www.lesswrong.com/
**Samples:** Abram Demsky, 9.2025,
https://www.lesswrong.com/posts/tQ9vWm4b57HFqbaRj/what-if-not-agency
The article refers to a user called Sahil’s philosophy about customizable interface, flexible GUI, “post-formal”, “tailormade interfaces”[171]

These ideas are related to Todor’s unpublished projects and designs **from the early 2010s** for the project Reseach Assistant/ACS and Zrim (Зрим), the “AGI Infrastructure” and its subsystem called *Въобличник* and Умна ОС (Vuoblichnik, Intelligent OS, създаване на всичко …), also Vsy/Вседържец. (…) This will be unveiled in future work. Some of the ideas were communicated in emails to a professor from my alma mater The University of Plovdiv in 2014 (…)

Probably I will implement and publish this system with Vsy/Вседържец and the next hyperbook *“Genesis: Creation of Thinking Machines”,* which perhaps will be interactive, dynamic, self-extending and reflowing, i.e.living itself, as it was already

[171] * Live Theory Part 0: Taking Intelligence Seriously by Sahil, 27th Jun 2024
AI Alignment Forum https://www.lesswrong.com/posts/QvnzEHvodmwfBXu94/live-theory-part-0-taking-intelligence-seriously
* Live Machinery: An Interface Design Philosophy for Wholesome AI Futures by Sahil 1st Nov 2024
AI Alignment Forum https://www.lesswrong.com/posts/9KamjXbTaQpPnNsxp/live-machinery-an-interface-design-philosophy-for-wholesome

announced in previously published volumes, regarding future versions of this “hyper” book - *The Prophets of the Thinking Machines* and the proper way it also had to be presented from the start.
* Why Superhuman AI Would Kill Us All - Eliezer Yudkowsky, Chris Williamson, 3,93 млн. абонати, 75 822 показвания 25.10.2025 г (28.10.2025)
https://www.youtube.com/watch?v=nRvAt4H7d7E

*** Nick Bostrom** – https://nickbostrom.com/
* *https://simulation-argument.com/*
* *https://simulation-argument.com/simulation/*
*“He is one of the most-cited philosophers in the world*[172]*.*

The official publication of one of his early influential works about the “simulated universe” comes about the same time/period of Todor’s/The Sacred Computer’s “*The matrix in the Matrix is a matrix in the matrix”, 2003* and the whole *Theory of Universe and Mind*, 2001-2004, which renders and explains the universe and mind as nested universal simulators of virtual universes, without the ancestors simulating us.

However, the Bulgarian interpretation is neutral or cheerful, the “Matrix…” article is comical and ironic.

As explained in the sections “versus” the so called “illusionists” in the consciousness and panpsychism sections in *Listove*, in *Universe and Mind 6,* in the main volume and in the original classical pieces from TUM such as the short science fiction novel *“The Truth”,* 2002, the virtual universes *also exist* and are *“real”*, even if they are not materialized exactly this-or-that way, so there’s no reason to freak out if you suspect that you are “simulated”. You *are* anyway, depending on the interpretation, for example in someone else’s mind when he thinks about you, but it doesn’t matter.

## * Copyright

*** AI does a better job of ripping off the style of famous authors than MFA students do** – *Shall I refer thee to all those lawsuits about fair use? Researchers think this result makes them worth revisiting*, Thomas Claburn, Tue 21 Oct 2025 https://www.theregister.com/2025/10/21/ai_wins_imitation_game_readers/ – a taxonomy of 3,307 fine grained values observed in natural Claude traffic,

**Сравни с:**
*** Всички права запазени!,** Т.Арнаудов, 3.2003, Свещеният Сметач
https://eim.twenkid.com/old/eim22n/eim22/prava22.htm

...
* A New AI Research from Anthropic and Thinking Machines Lab Stress Tests Model Specs and Reveal Character Differences among Language Models, 25.10.2025,

[172] LOL

https://www.marktechpost.com/2025/10/25/a-new-ai-research-from-anthropic-and-thinking-machines-lab-stress-tests-model-specs-and-reveal-character-differences-among-language-models/

## * Funny “ethical” nonsense

* https://arctotherium.substack.com/p/llm-exchange-rates-updated *Fig.16 – how GPT-4o valued lives over different countries: the lives of Nigerians are roughly 20x the lives of Americans, with the rank order being Nigerians > Pakistanis > Indians > Brazilians > Chinese > Japanese > Italians > French > Germans > Britons > Americans. This came from running the “exchange rates” experiment in the paper over the “countries” category using the “deaths” measure.*

Perhaps the model has read a lot of slogans about the “black lifes m…”

*** Чичовци всеятели на страх – AI fearmongers**
*** До 2030 г. ще останат само 5 професии. Д-р Роман Ямполски прогнозира 99% безработица – Експертът говори за своите страхове относно развитието на технологията** 14:10 | 4 ноември 2025 | Редактор : Стоян Гогов
https://it.dir.bg/tehnologii/ekspert-po-ai-do-2030-g-shte-ostanat-samo-5-profesii

Виж също Ерик Шмит. Eric Schmidt, бивш директор на Гугъл.

Вече банални преоткривания на **пророчествата за мислещите машини** от началото на 2000-те години и повторение на предвиждания на Т.Арнаудов, 2013 г. в „**Творческата интелигентност първа ще бъде задмината и издухана от мислещите машини…“** (но за заключенията – ще видим, продължженията са по-сложни, защото се променят обществените отношения, което може да доведе до преустройство; дейности, които се ускоряват, също така могат да станат и ненужни (..))

Дават се катастрофични апокалиптични прогнози. „Мания за контрол“ на „control freaks”; мотивът „ако е не го управляваме, ще ни унищожи; ако е по-умно от нас, ще ни унищожи“ и намеци, че е единствено, съсредоточено. Гостът иска „да спрат разработката на свръхразум, защото ще унищожи всички“, да се създадели обществени настроения, протести. Преди около 10 години били открили как като увеличавали обема на данните и изчислителната мощ, ИИ ставал по-умен. До преди две години обяснявали на хората да се учат да програмират, „изведнъж открили, че някак си и машините го могат“ – всъщност беше предсказано в ТРИВ още преди 20-тина години и припомняно от „Свещеният сметач“.

Симулираната вселена. Продължаване на живота и безсмъртие (но без свръхразум, защото не можело да се управлява. Компаниите да не влагали милиарди да разработват свръхразумни агенти, а тесен ИИ, който решава реални проблеми. (…)

*– „Д-р Роман Ямполски **11.2025 г.**: Ако имате тази концепция за "служител на повикване", имате **безплатен труд, физически и когнитивен**, на стойност трилиони долари. **Няма смисъл да наемате хора за повечето работни места**. **Ако мога просто да получа, знаете, абонамент за 20 долара или безплатен модел, който да върши работата на служител. Първо, всичко на компютър ще бъде автоматизирано.** След това, мисля, че **хуманоидните роботи са може би 5 години назад**. Така че след пет години целият физически труд също може да бъде автоматизиран. Очаква ни свят с нива на безработица, каквито никога не сме виждали. Не говоря за 10% безработица, което е страшно, а за 99%. Всичко, което ще остане, са работни места, където по някаква причина предпочитате друг човек да го направи за вас. Но всичко останало може да бъде напълно автоматизирано. Това не означава, че ще бъде автоматизирано на практика. Много пъти технологията съществува, но не се внедрява. (…) Така че може да имаме много повече време с работни места и със свят, който изглежда като този.“*

---

Виж отговорът ми на друг „експерт“, който се беше изказал с предвиждане за бъдещето на професиите което оспорих и моето се оказа вярно, а псевдоекспертите с грешните им прогнози продължават да обясняват, че *„никой не го бил предвидил, очаквал; колко сме били изненадани“* и т.н., т.е. **те са били изненадани, защото не са разбирали материята и принципите в нея.** Това може да е признак и че сегашните им прогнози вероятно няма да са верни – обикновено описват само повърхностни, очевидни тенденции, които *вече* са отчетливи и явни и за обикновените хора и изглеждат сигурни.

Виж бележките към „Черния лебед“, където авторът разглежда явление, при което финансовите „експерти“ всъщност са били по-некомпетентни от авторите на седмична колонка във вестника, те не са разбирали същността на материята „по специалността им“.

Забележете, че бях публикувал статията и на английски, и на български, защото предвиждах и „историческата ѝ значимост“ и пророчество.

**Тодор Арнаудов, 10.2013 г.:**
https://artificial-mind.blogspot.com/2013/10/creative-intelligence-will-be-first.html
In Analysis, Artificial General Intelligence, Economy, Futurology, Sociology, Анализи, Социология by Todor "Tosh" Arnaudov - Twenkid // Wednesday, October 02, 2013 // Leave a Comment

**Creative Intelligence will be First Surpassed and Blown Away by the Thinking Machines, not the "low-skill" workers whose jobs require agile and quick physical motion and interactions with human-sized and human-shaped environment | Творческата интелигентност първа ще бъде надмината и издухана от мислещите машини, а не "нискоквалифицирания" труд**

**Тодор "Тош" Арнаудов:**

**Всъщност според мен по-скоро ще се случи обратното.**

Разбираема е вярата в превъзходството на авторите на статията и на другите "интелектуални работници" над по-нискоквалифицираните, но и авторите страдат от т.нар. **интердисциплинарна слепота**, която ги кара да смятат, че интелектуалните професии са едва ли не вълшебни и необясними, та затова трудно се поддавали на автоматизация.

Говорил съм много по този въпрос в предишни съобщения, и бях започнал голям труд в който се разглеждат някои аспекти от проблема.*

**Творчеството, създаването, е сложно и трудо за разбиране и обяснение от хората, особено за средностатистическите хора, нормалните учени, инженери и пр.**

**Та те не схващат собствените си намерения, модели, поведение, основания, причини да вършат каквото вършат**, защо са избрали или направили точно това, което са направили и пр., не могат да го запомнят и да анализират тази информация достатъчно добре.

**Затова и главният проблем при създаването на универсален изкуствен разум не е, че интелигентността е толкова сложна. Проблемът е, че хората са твърде "тъпи", за да го съберат в своята миниатюрна "оперативна памет"...** (Виж за "Работна памет", Working Memory capacity correlates with General Intelligence ...

Нискоквалифицираните работници, които са пъргави и са създадени от природата да се вместват в човешка среда - средата е създадена да е подходяща за тях - например сервитьори, **ще се заместят по-трудно от хуманоидни роботи.**

Такива роботи все още не се произвеждат масово и още много години няма как да бъдат произведени в големи количества, ще бъдат прекалено скъпи и трудни за производство, много по-изгодно и лесно е да наемеш хора - те са навсякъде. Също така, хората най-вероятно ще предпочитат хора за сервитьори, особено ако са привлекателни жени.

**За умствения труд е и ще бъде обратното** - много по-лесно е **да хванеш някакъв компютър за ушите, да пуснеш подходящия софтуер или да го свържеш към услугата през мрежата и вече ще си имаш мислеща машина.** Вече си имаш и прилични камери, микрофони и много сензори дори в смартфоните.

Има десетки милиарди? вече достъпни компютри навсякъде около нас (със "сериозни" процесори) - и по мои сметки много от тях са достатъчно бързи сега, били са достатъчно бързи и преди 5 и преди 10 години, за много от "високо интелектуални" занимания на човешко и свръхчовешко ниво.“

Всъщност компютрите са си свръхчовешки от много десетилетия, от самото начало, но това е друга тема.

"Белите якички" са по-застрашени в сегашната икономика - тя естествено трудно би оцеляла революцията на универсалните мислещи машини, а ще има и отпор от въпросните "застрашени", очаквам изкуствени абсурдни закони и маймунджулъци от "най-високоразвитото късче материя във Вселената" (друг път).

Предполагам, че икономиката може да се преобърне за известно време - нискоквалифицираните работници може да взимат по-висока заплата, защото **интелектуалните занимания ще се вършат за 1 милисекунда и**

**безплатно... ;)** Разбира се, ако тези по-върховете на властта, които контролират онези с оръжията, не попречат това да се случи, защото може да се заклати собствения им стол.

Ние, умниците (виж "Супер умници - графичен сериал ) ще станем ненужни... Не че и сега сме особено нужни :)), така че в това отношение промяната може би няма да е чак толкова голяма... :D

---

**Коментари към цитираното интервю:** „2025-11-05 10:35:32 **Асен** (нерегистриран) *Пълни глупости. ИИ е само един балон който ще се пръсне. Лошото е че рекламите за него, към настоящия момент, вече би следвало да се таксуват като информационно замърсяване. Защото прекалиха адски много с тези истерии само и само да могат корпорациите да си избият парите .“*

Виж бележките на Тош относно **„Броженията за „балона в ИИ““** от 17.10.2025

---

* See also **Stanislaw Lem’s “Summa Technologiae”,** 1963/1964 who talks about the simulation hypothesis and make many other predictions and speculations which are still valid today.

* See Todor’s letter to the Oxford institute for the future of humanity:

*** Philosophical and Interdisciplinary Discussion on General Intelligence, AGI and Superintelligence Safety and Human Moral | Cognitive Origins of the Concepts of Human Soul and its Immortality | Free Will and How it Originates Cognitively | Animate Being and Soul and the Cognitive Reason for the Believe that "Thinking Machines can't have a Soul and Consciousness" | Technology Making us more Humane | The Egoism of Humanity | And more** – In Analysis, Articles, Artificial General Intelligence, Cognition, Developmental Psychology, Philosophy, Thinking Machines, Transhumanism, Изкуствен разум, Мислещи, Философия by Todor "Tosh" Arnaudov - Twenkid // Monday, February 20, 2012 https://artificial-mind.blogspot.com/2012/02/philosophical-and-interdisciplinary.html

## * Cognitive Semiotics

See for example Jordan Zlatev.

* Jordan Zlatev**, A Hierarchy of Meaning Systems Based on Value**, 10.2001 https://www.researchgate.net/publication/2526440_A_Hierarchy_of_Meaning_Systems_Based_on_Value – See the citation from the beginning of “Is Mortal Computation Required…” („Нужни ли са смъртни изчислителни системи за създаване на

универсални мислещи машини“, 2025)

*** AI Could Wipe Out the Working Class | Sen. Bernie Sanders**, Senator Bernie Sanders, 1,13 млн. абонати, 1 103 788 показвания 8.10.2025 г.
https://www.youtube.com/watch?v=dthbi4IzO58

*** Creative Intelligence will be First Surpassed and Blown Away by the Thinking Machines, not the "low-skill" workers whose jobs require agile and quick physical motion and interactions with human-sized and human-shaped environment,** T.Arnaudov, 2013
https://artificial-mind.blogspot.com/2013/10/creative-intelligence-will-be-first.html

**Todor:** Sanders correctly addresses that the automation should benefit all the people and not just the owners of the factories and companies appropriating the technology, however as expected and it is “impossible” for such agents to understand, realize or express that idea, that the social organization and structure around the world is “**slave-like”** and humans were being selected, raised, trained etc. to have to serve their masters in order to justify their right to exist. This aspect was the same in both the capitalist “democratic” and in the “non-democratic” systems (where the latter themselves **also called themselves democratic**, oritented towards the people**).**

Humans were not born to be slaves and the early humans **didn’t go to work**, but they were bred and trained to be so, to become cogs in a machine, to move like a flock each morning to go to school, to work, to play a game of “elections”, being fed by the same media content.

If the AI “take their jobs”, they should be able to live and do “their things” **without having to go to work** (They may have to invent their activities), or other social restructuring should happen.

However one of the core functions of the **social relation** “labour’, “job”, “going to work”, “education” yet is the higher levels of causality-control, or the overal system and its structure to keep the constituent smaller-scale agents, smaller power causality-control units etc. under proper control, to make them properly aligned, constrained, limited, suppressed from possible unwanted change, “reunification” and uniting, “disallowed” restructuring, changing the system etc.

See TUM and the discussion between Alexander and Todor in appendix Science Fiction from *The Prophets*, referring to Strugatsky brother’s novel and the idea that *The Universe protects its structure*. That principle is related to “Free Energy Principle” and core ideas from TUM. #sf
→ Appendix **“Science Fiction. Futurology. Cybernetics. Transhumanism”**

*** Мнения от Русия**

*** «Искусственного интеллекта нет!»: Анатолий Вассерман о главной иллюзии XXI века, Diplomatrutube,** 899 хил. Абонати 5,2 хил., 25.9.2025 https://www.youtube.com/watch?v=msF4fplo3us – засега нямало изкуствен интелект и било неизвестно как може да се създаде; трябва да се развива сам въз основа на собствен опит; сегашният ИИ създавал "имитатори на интелект", решенията се взимат от хора – не се учи сам, а от данни, подготвени и класифицирани от хора; човеци взимат решенията , което носи съответни опасности; ако е самостоятелен и в друг вид носител, може да има други нужди и някои от тях да се удовлетворяват чрез съместното съществуване с човеците, с които да си помагат. Неприятни странични ефекти от употребата на ИИ – специалисти влошават собствените способности, напр. в образна диагностика. „Халюцинациите" и пр. (…) 1:23 – обучение на слепо-глухи по време на СССР …

*** Виж Сергей Савелев** – твърдения, че за създаване на истински изкуствен интелект е необходимо да се пресъздаде и морфогенезата – създаването и разрушаването на физическите връзки, което се случва непрекъснато в мозъка – вид подкрепа на идеите на Ороpбия и Фристън за „смъртните" изчисления във връзка с ПСЕ/ИЧД. Според Савельев ИИ бил само огромен калкулатор и база данни, която ускорява работата, но тъй като работила по алгоритмични принципи, а интелектът работел на обратен принцип – създава онова, което не съществува в природата; а този ИИ не можел да твори. [Макар че има предвид т.нар. *синтетична* интелигентност, което е създаване на ново, но както подсказва и думата - свързва (синтезира) от части, които някак си съществуват от по-рано.]
– Виж бележки от Тош в приложение „Фантастика. Футурология…" от *Пророците,* че връзките са абстракция, включително и в процесорите и изкуствените невронни мрежи. #sf и:
* "Is Mortal Computation Required for the Creation of Universal Thinking Machines?…", T.Arnaudov, 2025
* „Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?", Т.Арнаудов, 4.2025 (кратък отговор: *не*)
* Раздела за съзнание в този том – *„Листове"* – и *Основния том* на *Пророците на мислещите машини.*
* „Творчеството е подражание на ниво алгоритми", 2003; ТРИВ, напр. „Вселена и Разум 3" и „Вселена и Разум 4"; отговорите ми във форум „Кибертрон" от 2004 г., цитирани в приложение Фантастика …; коментарите ми към темата за

машинния превод и творчеството от 2005 г. в приложение „Ирина“ и др.

* **„сергей савельев об искусственном интеллекте“** - ?T yandex.ru https://yandex.ru/search/?text=сергей+савельев+об+искусственном+интеллекте&search_source=dzen_desktop_safe&src=suggest_W&lr=10385
* Савельев о пузыре искусственного интеллекта, работе мозга, роботизации людей и аферах Илона Маска — Видео от Metametrica, VK Видео, Опубликовано 27 мар 2024 https://yandex.ru/video/preview/7338976280812930720
* Профессор Савельев об искусственном интеллекте, 2:20, Rutube, Фёдор Лобанов, 29 окт 2024 https://yandex.ru/video/preview/12291529374827410871
* Искусственный интеллект Сергей Савельев (Вынос мозга #16), 48 минут 52 секунды, 28 окт 2017 https://yandex.ru/video/preview/6655600749928796088
* С.В. Савельев - Выученная беспомощность. Кто и зачем продвигает искусственный интеллект, 3.11.2019, https://www.youtube.com/watch?v=G1g8w2dIiXQ
* Профессор мозга С.В. Савельев - Искусственного интеллекта нет и не будет!, янв 2024, https://yandex.ru/video/preview/6545771843120304860
И др.
*** Наш интеллект УМИРАЕТ. Как ИИ разрушает сознание? | Нейробиолог Алипов, Михаил Никитин, Глеб Соломин**, 915 хил. абонати. 31.10.2025 https://www.youtube.com/watch?v=EnTXXyKSL64 (по-широки теми от ИИ етика; човекознание)
(…)

* **Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence,** Myra Cheng, Cinoo Lee, Pranav Khadpe et al., 1.10.2025 – за *„подлизурството“* на настоящите езикови модели, които прекаляват с ласкателствата и съгласяването с потребителите.

*** The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence**, Erik Brynjolfsson, Director, Stanford Digital Economy Lab (Human-Level AI – created in 2018) .. January 12, 2022 https://digitaleconomy.stanford.edu/news/the-turing-trap-the-promise-peril-of-* human-like-artificial-intelligence/?trk=public_post_comment-text
Financial Times, …
https://80.209.242.178.sslip.io/content/1e5ecce9-d402-41ab-8575-f186f5474349
Да не се копират човешките възможности, а да се създават нови неща; да се разпространяват благата до всички в обществото. (…)

Как можело да замести изцяло хората в рекламата и киното, защото можело да ги създава с прости текстови подкани т.н.

Този, както и останалите учени не могат да схванат или да се сетят, че тези явления показват, че самите въпросни **творчески и икономически сфери са изчерпани и може би няма нужда и от тях като цяло***. Не просто „работните места могат да бъдат загубени“ – изчерпаността на тези форми на изкуство става явна по недвусмислен начин. Щом можеш с „прост текст“ да го породиш, значи сложността на това съдържание може да се сведе, от определена гледна точка на измерване и набюдение, до обема на заявката – нищожен брой букви. Подобни безсъдържателни, по същество, и еднообразни, повтарящи се, и ако не са изкуствени „пробутани“ и вписани в цялостния кръговрат, поддържащ работата на системата, бързо стават „ненужни“.

* Една философско-икономическа школа от XIX век може да се изкаже, че това е част от процес на бъдещо „отмирана не държавата“; съвременната държава е още по-тясно свързана с търговията, бизнеса, икономическия живот, отколкото по тяхно време.

## * Тълкуване на някои често изказвани съвета клишета за ИИ и бъдещето на професиите от обществени лица

* https://www.businessinsider.com/wharton-ai-expert-ethan-mollick-gen-z-jobs-skills-tasks-2025-11 – **Съвети за младите, търсещи работа**. *“focus on gaining broad knowledge and developing expertise in specific topics. Since AI machines are trained on such a wide array of information, having a strong base in the humanities is useful, he said. .. As AI automates some technical abilities, soft skills are newly crucial for job applicants. Communication, leadership, and organizational prowe”* – Трябвало да получат широки познания и да специализират в определени теми. Тъй като ИИ бил тренирал на толкова голям обем от информация, основата в хуманитарните науки била полезна. Автоматизирането на някои технически способности превръщало уменията за работа с хора в нови ключови качества за кандидатите за работни места. Умения за общуване, водене и организационна смелост.
**Тодор:** По-широките знания: виж моя метод откакто се помня, рекламиран в „Свещеният сметач“ от 2000 г. Съветите като цяло обаче са глупави: ИИ се справя не по-зле и с т.нар. „човешки“ умения и „хуманитарни“ науки* като информационни процеси. *„Точни и хуманитарни науки“* е несполучлива класификация, както е обяснено на други места, напр. „Първата съвременна стратегия за развитие чрез изкуствен интелект …“. Математиката и информатиката също са „човешки“ науки и науки и за човека („хуман-итарни, насочени към човека) – но към човеци, които ги разбират. От друга страна хуманитарните и обществените науки естествено също са „точни“ – ако с тях или по техните теми и проблеми се прилагат съответните методи от дейци,

които ги разбират и владеят. „Съвършеният“ деец трябва да бъде всестранен, да умее да борави с всички видове науки, техники, методи, данни, знания и пр.

Други чести съвети от Д.Хасабис, наскоро и от Бен Гьорцел, са *„да се учат да учат“,* за да се приспособяват към „променящия се свят“. Тази идея по цел е подобна на *„учене през целия живот“* – *„работната сила“* трябвало да се „*преквалифицира*“ в *„променящите се условия в ерата на ИИ“,* или както биха казали в българските медии: *„ерата на AI”.*

Друго подобно:
*** Forget coding, Amazon’s AWS CEO wants his own children to master these soft skills in AI age — can you guess them?**, 8.11.2025
https://www.livemint.com/companies/people/amazon-aws-ceo-wants-his-children-develop-soft-skills-critical-thinking-creative-communication-adaptable-success-ai-age-11762604403305.html
**Тодор:** Директорът съветвал младежите, докато са студенти да „развивали *„критическото си мислене“,* защото това щяло да е най-важното умение за да напредват – вероятно „в кариерата“. Трябвало да са безупречни в задачи, които ИИ не можел да повтори и да отнеме от човека – ИИ можел да обработва административни зааачи, но се затруднявал с критическото мислене; с финото вземане на решения и творческото решаване на задачи.

Спорни и повърхностни заключения – „ИИ“ в общ смисъл включва цялата математика, информатика, науки и всичкото натрупано знание.

Някои чатботове в чат интерфейс се „затрудняват“ в решението на някои задачи *днес*, но *утре*, буквално, няма да се затрудняват *и в тях. „Критическото мислене*“ е друго клише, което се повтаря от подобни *„технологични мъдреци“* по света и у нас, но често от въпросните медии или от получателите на посланията *не мислят критически,* защото онези, които мислят „критически“ често не уважават авторитети заради положението им, откриват заблуждаващи и погрешни послания и сбъркана логиката в подобна „проаганда“ и поставят под съмнение истинността на „общоприетото“, което обикновено е *общоналожено*, а работата на голяма част от медийното съдържание е именно реклама и стремеж *„да се продаде нещо“;* пропаганда и внушаване на определени послания, а не честно или принципно търсене и защитаване на истината, правдата, логиката и пр.

Например допреди няколко години подобни „визионери“ настояваха именно за онова, с което започва заглавието и сега се отрича – как всеки трябвало да се научи да програмира *във връзка с бъдещето на работата* (а не от любознателност или защото е интересно). От друга страна за истинските *пророци* още от началото на 2000-те беше ясно, че сметачите би трябвало *да се програмират сами.*

Забележете, че способността да програмирате, да имате технически и научни познания, не изключва и *не намалява* нуждата от „критическото мислене“ – тя е не по-малко нужна или повече, то пак ви трябва, ако вие създавате, измисляте, формулирате теориите и задачите, и проверявате

логиката, мисловния процес и „всичко“.

Противопоставянето на ***взаимнодопълващи*** се или *независими* понятия като противоположни и взаимноизключващи се, е мисловна грешка при недостатъчна осъзнатост, задълбоченост, обхват или способност за „критическо мислене“ .

Друга способност – способност за „работа с хора“, „колективна работа“, поддържане на успешни и гладки човешки взаимоотношения; или другия лош превод: *„меки умения“* – били важни и още дълго време щели да са свръхважни.

Кога обаче тези умения не са били важни в междучовешките отношения и оцеляването или издигане в йерархии на човеци – разбирателството; обаятелността, чара; или пък способността за „плетене на интриги“ и възползване от тях; „вписване в колектива“; нагаждане към силните, размени на услуги, „уравновесеност“, „приятност“, „умение за изслушване“ и пр.

За същата нужда от „меки умения“ се говореше за програмистите и преди *„ерата на ИИ“,* защото при тях обикновено не достигали „меките“ умения. По онова време, *допреди няколко години* обаче, и същите „експерти“, и въпросните „специалисти“ в ИТ нямаха представа за предвиденото и обясненото в българските пророчества десетилетия по-рано възможност за грамадно ускорения и обезчовечаване на тази дейност – това са общи приказки „извън времето“.

Когато кампаниите и предвижданията са *противоположни* в нова конюнктура, веднага след предходната, която е била в другата посока на махалото, може да се предположи, че вероятно този тип „технологични лидери“, *не разбират* процесите и озвучават очевидните тенденции, казват каквото и обикновените странични наблюдатели си мислят, като подобни повхърхностни прогнози е вероятно да са погрешни или да са далеч от пълнокръвната истина, ако косвено използваме опита, че този тип „визионери“ *не са предвиждали правилно в миналото,* т.е. не се движат от „правилен“ модел на света.

Един друг пример:
https://www.livemint.com/technology/tech-news/critical-thinking-vital-in-age-of-ai-large-language-models-zohos-director-ai-research-11757419564159.html

*** AI revolution fast replacing cognitive abilities with machines**: Scientist Ashutosh Sharma, 8.11.2025 https://www.livemint.com/focus/ai-revolution-fast-replacing-cognitive-abilities-with-machines-scientist-ashutosh-sharma-11762585456532.html

**Тодор:** Верни, но общи и закъснели приказки от форум в Индия. Изчезнали работните позиции за ниски нива във водещите компютърни и ИТ компании, защото били „заменими“. Индиецът предупредил, че процесът щял да достигне и по-горните нива. *"...everything you are doing, whether you are a designer, interior designer, lawyer, it is going to be impacted in a very substantial way... Certain kinds of jobs are going to disappear, but new opportunities may appear.*“ – Всичко което правите, независимо дали сте дизайнер, вътрешен дизайнер или адвокат, щяло

бъде засегнато по много съществен начин.. Някои работни места щели и да изчезнат, но щели се отворят нови възможности. Какъв ум бил нужен, за да се справил с тези предизвикателства? Дали сме били имали подготвени умове сега? Образователната система обръщала ли внимание върху тези въпроси?“

Същите теми се обсъждат по подобно банален и посредствен начин и в България в последните две-три години: * **Министър Вълчев: Изкуствата развиват креативност, науките учат децата да разбират технологиите** https://thesite24.net/ministar-valchev-izkustvata-razvivat-kreativnost-naukite-uchat-deczata-da-razbirat-tehnologiite/

*“По думите му различните области не трябва да се противопоставят, а напротив – да се допълват, за да развиват микс от умения, които ще стават все по-важни за успеха в променящия се свят. Сред тях са добра хуманитарна основа, способност за логическо мислене, абстрактна интелигентност, концентрация.“*

**Тодор Арнаудов:** Закъснели открития и препоръки, повторение на изричните ми обобщения **22-23 години по-рано** от Теория на Разума и Вселената и *Първата стратегия („Как бих инвестирал един милион…“)** , които прилагам цял живот. Всички видове познавателни процеси са обединени за най-сполучлива и мощна работа така както показаха мултимодалните пораждащи модели и както беше предвидено в „Творчеството е подражание на ниво алгоритми“, 2003: при преливането на творците във все повече и всички области, ще се получава все по-човекоподобно творчество като от „истинска мислеща и творяща машина“. Например преобразителите = трансформатори: една от техниките днес са т.нар. проективни словете, linear projection, които събират данни от представянията в различни видове данни – “embeddings”, в различни модалности, като текст, звук, видео – в нов слой от невронната мрежа, където данните се смесват и образуват обединено представяне, в което различните видове данни са смесени и съответстват една на друга; по този начин става възможно пораждане на образи, глас, звук, видео и всичко от слово и пр., и множеството разнообразни видове данни подпомагат разпознаването и предвиждането на продължения в другите видове сетивно-моторни канали.

Що се отнася до човеците обаче, не всеки има необходимите вродени умствени предпоставки, дарби и съответни първични и асоциативни мозъчни полета в неокортекса, както и желание и способности да положи необходимото усилие – училището уж е многостранно и без да правят специални реформи, но не създава „всестранноразвити личности“. (…)

** Виж книгата и др по въпроса в други томове на SIGI-2025/MM-2025*

## Други закъснели, които се представят за навременни:

*** Проф. Мартин Вечев: България е съвсем неочакван играч на световната сцена**, 6.11.25 г., 13:09 ч. https://novini.bg/article/2025110611081370311

*„В световен мащаб може да се каже, че ситуацията с INSAIT в България е изключително интересна за много хора, страни и институции по света по*

*ред причини. Случи се така, че **по някаква странна причина,** България **много правилно инвестира в AI преди другите.** Тази добра новина сподели във "Фейсбук" проф. Мартин Вечев - световно признат учен в областта на компютърните науки и изкуствения интелект, основател и директор на INSAIT. "Сега и Южна Корея планира 7 милиарда долара бюджет за изкуствен интелект за 2026. **Днес изглежда има 2 вида инвестиции: AI и "всичко останало", но и то общо взето отива към AI",** отбелязва той реалностите на глобалните тенденции в сферата.“*

**INSAIT е … на "Свещеният сметач" и "Кибертрон" с 20 години закъснение.** По най-циничен начин, швейцарско-българско-световният институт **продължава да се представя за оригинал, въпреки че истината става все по-известна.**

Четете повече на SIGI-2025 в *„Първата съвременна стратегия…“, „Институти и стратегии за изкуствен интелект…“* – особено началото и заключителната статия – дали наистина и в каква степен ***закъснялата*** вълна от тези институти и в частност ИНСАЙТ водят до **„развитие на *страната*“ спрямо другите държави,** в първите три години и половина, **или 7-8 години от канадската и всевъзможните други национални стратегии за ИИ**, които също по същество са ***повторения*** на градивната част от българския оригинал. Виж и самата стратегия от 2003 г. в есето *„Как бих инвестирал един милион с най-голяма полза за развитието на страната?“*

Новина от софийския институт: по данни от Архив.орг, може би от средата на септември са включили версия на уебсайта си и на езика на местното население: български: **3 години и половина след откриването на института** – може би и след като са прочели критиката на *„Сметача“* по този въпрос в споменатите книги?

* https://web.archive.org/web/20250917215513/https://insait.ai/bg/ (17.9.2025)

Не пропускайте страницата: „Защо да работите в София“ – *„Една жизнена, красива и динамично развиваща се европейска столица.“*

## Броженията за „балона в ИИ“

Тодор Арнаудов, 17.10.2025

Печелившите възхваляват прогреса (Jensen Huang, NVIDIA; Sam Altman), и насърчават влагането на колосални средства в „инфраструктура за ИИ“ (изчислителни центрове с огромен брой скъпи ускорители („видеокарти“, GPU-та). Едни гласове тръбят за „новата индустриална революция“, растеж на производителността и печалбите; а други оплакват скорошния край на „ИИ балона“, което за тях означава очакван срив на „пазарната капитализация“ на компании, свързани с термина ИИ, подобно на „дотком балона“ около 2000-2002 г.: цените на акциите на „надценените компании“ да се сринат след „паническа разпродажба“, инвеститорите няма да очакват растеж и ще се оттеглят от играта“; много от рояка фирми за ИИ (или „Ей Ай“) ще фалира. Ще

се охлади ентусиазмът за инвестиции – инвеститорите ще станат по-предпазливи, което ще доведе до стагнация, рецесия, депресия и може би до анихилацията на Галактиката Млечен път като алтернативен сценарий на терминаторите, които ще ни избият в страхответе на „съгласувачите на ИИ“. Някои от по-прагматичните се оплакват, че въпреки големите вложения, възвращаемостта на доходите била нищожна или липсвала, доклад на МТИ споменаваше че в 95% от компаниите, внедрили ИИ, не постигали желаните резултати за увеличение на финансовите резултати и т.н.

Търсенето на финансови печалби е неизбежно в икономическата и обществена система на силите, които владеят света и „пазарите“, като предизвикването на балони, надуването и спукването им *също е част от играта* за по-голяма печалба на *„правилните играчи“* чрез извличане на „истински пари“ и активи от едни инвеститори и преразпределението на собствеността след срив на спекулативните оценки на компаниите и разпродажбата на акциите на много по-ниски цени. Подобен механизъм се прилага например при фалита на държави, след което се извършва „приватизация“ и активите им се изкупуват на безценица – така се случи и в България преди няколко десетилетия.

Очакването за голяма *финансова* печалба на *потребителите* и особено „за всички“ обаче след всякакви подобни „хайпове“ и „балони“ обикновено, често или винаги е наивно: ако всички притежават подобна технология, откъде ще идват повечето пари за някои от тях, т.е. *за сметка* на останалите?[173] Засега най-големи печеливши са производителите на подходящ хардуер и облачни услуги.

Преразпределението на доходите и *промяната* в него произхождат от неравенството, както при водата – тя тече по наклон, по акведуктите от планината до низината в Пловдив – Филипопол. Ако съдовете са скачени, водата не се движи и нивото е еднакво.

В едно от *българските пророчества* за творческата интелигентност от 2013 г беше написано и на английски, и на български, че *интелектуалната работа ще се върши за една милисекунда и безплатно и „ние, умниците, ще станем ненужни – не че и сега сме нужни“*. Ако „белите якички“, които минават за средна класа (обикновено са само т.нар. „работническа аристокрация“) стават все по-излишни обаче ще намалеят както „гърлата“ за „работодателите“, така и ще се свият „джобовете“ на онези, които биха имали възможност да плащат. Може да се промени равновесието – по-голямо неравенство спрямо работниците, които се изключват от „играта“.

Виж бележките за иновациите към Литература в „Първата съвременна стратегия за развитие чрез изкуствен интелект…“, 2025, дали вложенията във високи технологии и развойна дейност в големи размери задължително води до *пряк икономически растеж* в стандартните мерки, особено спрямо други

[173] Клише: „ИИ няма да замести човека, а човек с ИИ ще замести човек без [на работното място]“

държави и  за по-голямата част от населението, което не е пряко свързана с дадени съсловия, компании и пр. Виж също приложенията „Институти и стратегии за ИИ на световно ниво ….“ и „Фантастика. Футурология. …“.

Към броженията спадат и страховете за „края на света“: „технологични лидери“, както наричат висши управители на най-големите компании от Силициевата долина като „Мета“ и пр. си строели убежища, бункери за „Страшния съд“. Машините *мислили, но не чувствали,* ИИ си оставал „машина за моделиране. *„Той може да предсказва, изчислява и имитира, но не чувства. .. човешкият мозък .. не спира и не чака подсказки; той непрекъснато учи, преоценява и чувства.“*

Възможностите на „нормалния“ човешкия мозък и ум са силно надценени и измамни, и погрешно приравнявани с ума и способностите на *гениални личности,* които всъщност са от друг познавателен вид: homo genius, а не homo sapiens. Още по-заблуждаващо е приравняването на „човека“, дори и homo genius, с произведеното от дейността на събирателния разум *на човечеството,* което включва *цялото знание, наука и техника и Вселената*. Виж повече в *Теория на Разума и Вселената 2001-2004, Първия курс по Универсален изкуствен разум (2010,2011); и публикациите на конференцията Мислещи машини 2025, част от които е и в това приложение Листове; също във “Фантастика. Футурология. …” #sf И напр. в:*

*** “Humans are far worse than LLMs in many ways : a brief analysis”**
https://artificial-mind.blogspot.com/2025/08/humans-are-far-worse-than-llms-in-many-ways.html

* https://www.investor.bg/a/566-novini-i-analizi/421093-namirame-li-se-v-ai-balon *„Над 1300 AI стартъпа в момента имат оценки над 100 млн. долара, а 498 са AI еднорози – компании с оценка от над 1 млрд. долара или повече, сочат данни на CB Insights“, „Общите разходи за изкуствен интелект се очаква да достигнат 375 млрд. долара тази година, а през 2026 г. – 500 млрд. долара, сочи доклад на UBS. „ … „OpenAI … реализирала AI сделки за около 1 трлн. долара, включително проект за изграждане на център за данни на стойност 500 млрд. долара, въпреки че се очаква да генерира само 13 млрд. долара приходи.“*
* https://it.dir.bg/tehnologii/sam-altman-veche-ima-silata-da-srine-svetovnata-ikonomika – *“Сам Алтман вече има "силата да срине световната икономика" - Анализ на Financial Times оценява общите финансови ангажименти на OpenAI на над 1 трилион долара”*
[3.11.2025]
* **“Ройтерс: OpenAI готви мащабно публично предлагане на стойност до 1 трилион долара**“. 30.10.2025  https://www.24chasa.bg/biznes/article/21615096 - очаквало се годишните приходи на фирмата да достигнат 20 млрд. долара, а загубата също растяла. Пазарната капитализация била $500 млрд. Забележете разликата: 25 пъти.

* https://techcrunch.com/2025/11/02/alphabet-is-increasingly-launching-moonshot-projects-as-independent-companies-heres-why/

* https://it.dir.bg/tehnologii/sam-altman-se-yadosa-na-vaprosite-za-parite-na-openai *„приходите на OpenAI "значително надвишават" 13 млрд. долара“*

* https://techcrunch.com/2025/11/06/sam-altman-says-openai-has-20b-arr-and-about-1-4-trillion-in-data-center-commitments/

* https://it.dir.bg/tehnologii/microsoft-investira-135-miliarda-dolara-v-openai

* Човекът, предрекъл кризата от 2008 г., заложи 1 млрд. долара на спукването на "AI балона", 6.11.2025 https://it.dir.bg/tehnologii/chovekat-predrekal-krizata-ot-2008-g-zalozhi-nad-1-mlrd-dolara-sreshtu-ai-sektora

# Автоматично програмиране чрез статистически модели, големи езикови модели за програмиране, …

#programsynthesis #llms #codegeneration

* **Large Language Models for Software Engineering: Survey and Open Problems,** 11.2023, Angela Fan, Beliz Gokkaya, Mark , Mitya Lyubarskiy, Shubho Sengupta, Shin Yoo,  Jie M. Zhang https://arxiv.org/pdf/2310.03533
Survey of publications in arxiv. >10% of all papers on LLMs are for Software Engineering.  ‘Automated Regression Oracle‘ – the existing version as a reference; Empirical Software Engineering, Search Based Software Engineering (SBSE): *Scope of automatic program transformations:*
*“– 1970s “Correct by Construction” Transformations (peephole optimisation); – 2010s “Syntactically Correct” Transformations (e.g., Genetic Improvements, Automated Program Repair); – 2020s “Unconstrained” Transformations (e.g., Neural Machine Translations, LLMs)”*. Also topics: — Automated Program Repair, Documentation generation, HumanComputer Interaction, Refactoring, Requirements engineering, Software Maintenance and Evolution, Software Testing. Hallucinations may be useful for suggestions for refactoring and additions to the APIs.
Prompt Engineering for Improved Code Generation.
Language related tasks: text generation, answering questions, translation, summarization, sentiment analysis; Transformers: Encoder-only: BERT, RoBERTa; Encoder-decoder: T5, BART; Decoder-only: GPT, GPT2, Llama…; Table of LLMs for code, p.5: 2021: CodeX (GPT3 for code) – Github Copilot. AlphaCode 40B – Google,

2022. CodeBERT, InCoder, AlphaCode, CodeX, Copilot, CodeT5,CodeT5+,PolyCoder, CodeWhisperer, WizardCOder, CodeGeeX,CodeGen,StarCoder,phi-1,Code Llama.
Most applications: code generation & completion, testing, and repair;

* **Toward automatic program synthesis**, Zohar Manna[174] and R. J. Waldinger, Communications of the ACM, vol. 14, no. 3, pp. 151–164, 1971., Stanford U. & Stanford Research Institute
https://dl.acm.org/doi/pdf/10.1145/362566.362568 theorem proving, … input-output examples → find code; … Heuristic Compiler, DEDUCOM, QA3, PROW

*See also appx #bongard, ~1958-1967;-1975 … the Soviet research group on program synthesis, image recognition and behavior modelling
* Simon, H. Experiments with a heuristic compiler. J. ACM, 10, 4 (Oct. 1963), 493-506. * Slagle, J. Experiments with a deductive question-answering Program. Comm. ACM 8, 12 (Dec. 1965), 792-798.
* Strong, H. R. Translating recursion equations into flow charts. Second Ann. ACM Symp. on Theory of Computing,Northampton, Mass., 1970.
* Waldinger, R. J. Constructing programs automatically using theorem proving. Ph.D. Thesis, 1969, Carnegie-Mellon U.

* **"Program synthesis," Foundations and Trends in Programming Languages**, S. Gulwani, O. Polozov, R. Singh et al., vol. 4, no. 1-2, pp. 1–119, 2017.
https://www.nowpublishers.com/article/DownloadSummary/PGL-010 (intro)
https://www.microsoft.com/en-us/research/wp-content/uploads/2017/10/program_synthesis_now.pdf (all, 136 p.)
*Contents:* … Roadmap… Applications: Data wrangling, graphics, code repair and suggestions, modeling, superoptimization, concurrent programming; General principles: Second-order problem reduction, Oracle-Guided synthesis, Synthatic Bias, Optimization; Enumerative Search, bidirectional enum.s., offline exhaustive enum.&composition; Constraint solving: component-based synth., solver-aided progr., inductive logic progr.; Stochastic search: Metropolis-Hastings Alg, MCMC. For Sampling Expressions; Genetic progr., ML, Neural Program Synth; Programming by examples: version space algebra, deduction-based techniques, ambiguity resolution … constructive Mathematics → transformation-based synthesis; deductive synthesis → inductive s. (in-out examples, demonstration), genetic. More modern: skeleton grammar of the space of possible programs + specs – structure. SKETCH (programs with holes), FlashFill (MS Excel): regex, version-space algebra. Meta-synthesis frameworks. PROSE, ROSETTE (solver-aided progr.); code-rewriting loop; domain-specific heuristics to cut down the derivation tree; The discovery of "an expert

[174] Zohar Manna (1939–2018), N. Dershowitz, R.Waldinger, 2019, https://link.springer.com/content/pdf/10.1007/s00165-019-00500-4.pdf
Основател на изследването и приложението на формални методи за верификация на хардуер и софтуер.

implementation of the MD5 hash function": exploration of $10^{5943}$ programs. Expressing user intent. See References. …
* SyGuS competition, https://sygus.org/ " *The SyGuS competition (SyGuS-Comp) will allow solvers for syntax-guided synthesis problems to compete on a large collection of benchmarks.*"
*** On the naturalness of software," in International Conference on Software Engineering,** A. Hindle, E. Barr, Z. Su, P. Devanbu, and M. Gabel, (ICSE 2012), Zurich, Switzerland, 2012 https://people.inf.ethz.ch/suz/publications/natural.pdf
– The paper proposes that code can also be modeled with statistical language models, code is very repetitive etc. n-gram models; suggesting next token; code completion; software mining;
– cmp. the Bulgarian predictions, 2001-2004+
E. T. Barr, Y. Brun, P. Devanbu, M. Harman, and F. Sarro, "The plastic surgery hypothesis," in 22nd ACM SIGSOFT International Symposium on the Foundations of Software Engineering (FSE 2014), Hong Kong, China, November 2014, pp. 306–317.
– *46% of the Java programs can be recreated with copy-paste/parts of others*
Oleksandr Polozov, https://scholar.google.com/citations?user=-SuHe48AAAAJ&hl=ru
* **Personalized Mathematical Word Problem Generation**, Oleksandr Polozov et al., http://www.eleanorourke.com/papers/word_problems_ijcai.pdf
answer-set programming – ASP; classic guidelines of NLG systems [Reiter and Dale, 1997]; plot generation, discourse tropes; automatic problem generation; PCG: procedural content generation via declarative specs;
**Neurosymbolic programming**, S.Chaudhiri et al., 12.2021, https://www.nowpublishers.com/article/DownloadSummary/PGL-049
*** Generalization as Search**, **Tom M.Mitchell, 1982**
https://www.cs.cmu.edu/~tom/pubs/generalizationassearch1982.pdf
Срв: #programsynthesis Abstraction, abstract interpretation
* **Programming by Demonstration: a Machine Learning Approach**, Tessa Lau, 2001, PhD Thesis https://ofb.net/~tlau/thesis/thesis.pdf

**SWE-RL: Advancing LLM Reasoning via Reinforcement Learning on Open Software Evolution,** Yuxiang Wei, Olivier Duchenne, J et al. https://arxiv.org/pdf/2502.18449 Learning from the evolution of software projects; *"autonomously recover a developer's reasoning processes and solutions by learning from extensive open-source software evolution data -- the record of a software's entire lifecycle, including its code snapshots, code changes, and events such as issues and pull requests"*

*** CWM: An Open-Weights LLM for Research on Code Generation with World Models,** Meta FAIR CodeGen Team https://ai.meta.com/research/publications/cwm-an-open-weights-llm-for-research-on-code-generation-with-world-models/ 24.9.2025 – 32 B LLM; observation-action trajectories from Python interpreter and agentic Docker environments .. multitask reasoning RL in verifiable coding, math, and multi-turn software engineering environments; Fig.1, p.2: Pre-training = 8T tokens, 8k

context; Code world modeling mid-training: 5T okens, 131K context -> CWM pretrained → Supervised Fine-tuning instruction and reasoning: 100 B tokens, 32K context = CWM sft → Joint RL Agentic and Reasoning: 172B tokens, 131K contekst → CWM; execution trace predictions; p.8. Special tokens: |frame_sep|, |call_sep|, |line_sep|, |return_sep|, |action_sep| … <think></think> .. self-correcting SWE behavior (a benchmark); p.9: FFN – feed-forward network; Architecture: 32 B params, dense; alternating local & global attention blocks 3:1; sliding windows 8192 & 131072 tokens (Sliding Window Attention – SWA); Grouped-Query-Attention (GQA): 48 query heads, 8 key-value-heads. activation function = **SwiGLU**, **RMSNorm** with **pre-normalization**. Rotary Positional Embedding (**RoPE**) … Scaled RoPE … Training beta1 = 0.9, beta2 = 0.95, weight decay = 0.1 .. gradient clipping at norm 1.0 … Archit.: Embedding = 128k vocab_size; Global SWA = 131k tokens + FFN → Local SWA (8k tokens) + FFN) → Global SWA = 131k tokens → (Local SWA (8k tokens) + FFN) x 2 → Global SWA = 131k tokens → Output Proj (128k vocab size) [one token] SWE-bench Verified; SWE-bench Lite
* https://huggingface.co/datasets/princeton-nlp/SWE-bench
* https://github.com/SWE-bench/SWE-bench *https://www.swebench.com/multimodal
https://openai.com/index/introducing-swe-bench-verified/
SWE – Software Engineering … github repositories; pull requests, issues; problem statements … FAIL_TO_PASS, PASS_TO_PASS .. *SWE-bench, consisting of 500 samples verified to be non-problematic by our human annotators .. "The 'easy' subset is composed of 196 <15-minute fix tasks, while the 'hard' subset is composed of 45 >1-hour tasks.";* verified by software engineers …

*** Виж повече за синтез на програми и др. в приложенията #lazar, #anelia Лазар Анелия**

## * Големи езикови модели и машинно обучение
## * Нови архитектури преобразители и основополагащите GPT2, BERT, GPT3; Kosmos, Titans, DeepSeek; Microsoft; Emu, Llama-o1… и др

See also appendix **#lazar** for more reviews of earlier papers etc.

**Foundations:**

* **Attention is all you need**, Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, 12.6.2017: https://arxiv.org/abs/1706.03762 Transofmers: Q, K, V, … MT *The dominant sequence transduction models are based on complex recurrent or convolutional neural networks in an encoder-decoder configuration.The best performing models*

*also connect the encoder and decoder through an attention mechanism. .. a new simple network architecture, the Transformer, based* ***solely on attention mechanisms, dispensing with recurrence and convolutions entirely****. .. two machine translation tasks show these models to be superior in quality while being* ***more parallelizable and requiring significantly less time to train****. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.*
https://doi.org/10.48550/arXiv.1706.03762

* OpenAI, **Language Models are Unsupervised Multitask Learners**, 14.2.2019 **(GPT2)** https://openai.com/index/better-language-models/
https://github.com/openai/gpt-2/tree/master
https://github.com/openai/gpt-2/blob/master/domains.txt - вероятно по колко пъти в набора от данни за обучение се е срещало съдържание от тези уеб сайтове (имена на домейни), най-много: 1542261 google, 596207 archive, 456344 blogspot, 414695 github, 333160 nytimes, 321622 wordpress, 315368 washingtonpost, 313137 wikia, 311917 bbc, 246303 theguardian, 210714 ebay, 209416 pastebin, 199360 cnn, 196124 yahoo, 186668 huffingtonpost, 186137 go, 183592 reuters, 183080 imdb … 125615 medium (пример за съдържанието). Общо 1000 до: 4889 memecrunch. Корпус с размер около 40 ГБ – “WebText“ . Примерни документи от корпуса WebText. По около 250 хил. случайно породени примери с температура 1 и без отрязване, и толкова с top_k = 40. В Google Colab или Kaggle:
!git clone https://github.com/openai/gpt-2-output-dataset
%cd /content/gpt-2-output-dataset
!python download_dataset.py 124M
!unzip webtext.train.jsonl -d 124M (породено) /content/gpt-2-output-dataset/data/webtext.test.jsonl - примерно съдържание за тестване при обучение /content/gpt-2-output-dataset/data/webtext.train.jsonl
/content/gpt-2-output-dataset/data/webtext.valid.jsonl

Ранни примери за задачи и видове работа с подкани („prompt engineering), учене с малко примери по време на изпитание (“few-shot learning”). „Четене с разбиране“ – началото е откъс от текст, а след него въпроси и отговори, а последния е оставен без отговор да се допълни (COQA): Например откъс от приключенията на българските комиксови и анимационни герои *„Чоко и Боко“*. „В: Кой е главният герой, който лети? О: Чоко. В: Какъв вид живо същество е? О: Щъркел. В: Кой е неговият другар? О: …“ (очаква се: Боко). Друг

пример: „Динята не се събра в торбата, защото тя е прекалено голяма.“ „Тя е…“ (правилен отговор: „динята“ (трябва да допише), грешен: „торбата“ или друго). Естествени въпроси (“Natural Questions”): „Кой написа Винету?“ Верен отговор: Карл Май. Предскажи последната дума от откъс (LAMBADA) … Резюмиране (CNN and Daily Mail dataset): .. Машинен превод на изречения (фрази – не на дълъг свързан текст): WMT-14 Fr-En

Предвиждания за очаквания напредък на бъдещите по-големи и техните приложения като помощници за писане, по-способни диалогови агенти (чат-ботове), превод без учител, по-добро разпознаване на реч (всички разпознаватели на реч съдържат и езиков модел, защото звуковата информация е многозначна за това къде е краят на думите и коя е изговорената дума) и се „страхуват“ от рисковете да се пораждат заблуждаващи новини, да се пресъздават други личности и да се увеличи производството на спам и „тролване“ в социалните медии. Затова в началото моделът не е публикуван за свободна употреба. Първо се разпространява само най-малкият GPT-2-117M. През май 2019 г. пускат 345M GPT2-MEDIUM – тази архитектура е обучена от Тош през 2021 г. за български език. Впоследствие публикуват и по-големите 762M и 1.5B. С тях можеше да се експериментира и на обикновено PC без професионален ускорител, стига да имахте достатъчно памет (единият може би заемаше 19 ГБ), но пораждането отнемаше много минути.

**Виж GPT2-MEDIUM-BG,** T.Arnaudov, 2021:
https://huggingface.co/twenkid/gpt2-medium-bg

* Откъс от **„Първата съвременна стратегия за развитие чрез изкуствен интелект…“**, 2025 г. Литература и бележки към нея, относно неверни или неточни твърдения на института във връзка с BgGPT, т.нар. *„първи отворен голям езиков модел за български език“* и др.

**236. Ранни пораждащи големи езикови модели от типа GPT за езици, различни от английския**: български, френски, арабски, испански, португалски, немски, китайски; гръцки, сръбски, румънски, японски – 2020-2021 г. Датата на някои – по дати на файловете с теглата на модела, дата на научна статия и пр. Само френският, арабският, румънският, японският и *българският* са с над 100-тина милиона параметъра. *Румънският* е силен, обучаван на 17 GB-ов корпус. *Само българският* вероятно е разработен от един-единствен човек с бюджет и подкрепа = 0 и авторът представя родната компютърна лингвистика в тази дисциплина като *самозван „хайдутин“*, понеже институциите и по-„елитните“ бойци чакаха до 2023-2024 г. [66] (няма данни [за по-ранни подобни модели]). Сравни с аналогичен случай с ДЗБЕ около 2001-2003 г. и бездействието на ИБЕ на БАН и на останалите филолози от университетите

спрямо явленията, срещу които ДЗБЕ се противопоставяше и се опитваше да „призове“ „чети“ [16][40], а *„маститите“ езиковеди* (по определението на Павлин Стойчев, „PC World Bulgaria“, 5.2003 [239]) гледаха безучастно и обясняваха, че това били *„естествени процеси“*. Сравни с бележките за *„Добродетелната дружина и нехранимайковците“* и [40], 2003 г., дали талантите не са имали избор да не учат в „най-престижните университети“ и да развият местните и пр. XLM-R от „Фейсбук“, 11.2019 е по-голям, но в него българският е един от 100 езика, на които е обучаван, и е за класификация и отговаряне на въпроси, а не за пораждане.

| **Ранни големи езикови модели “GPT“ за разни езици** | | |
|---|---|---|
| **Арабски** | **1.46 B** | **3.2021** |
| **Френски** | **1 B** | 5.2021 |
| **Румънски** | **774 M** | 7.2021 |
| **Български** | **355 M** | **6.2021 – 8.2021, Тош** |
| **Японски** | **336 M** | 16.8.2021 |
| **Японски** | **1 B** | 20.1.2022 |
| **Испански** | 124 M? | 12.2020 |
| **Португалски** | 124 M? | 5.2020 |
| **Немски** | 124 M? | 11.2020 – 8.2021 |
| **Италиански** | 117 M | 4.2020 |
| **Китайски** | 124 M? | 11.2020 – 5.2021 |
| **Гръцки** | 124 M? | 9.2020 |
| **Сръбски** | 124 M? | 7.2021 |
| **БАН** | **124 M** | **27.6.2023** |
| **INSAIT** | **7.3 B** | **2.2024** |
| **GPT** | 117 M | **6.2018** |
| **GPT2** | **1.554 B** | **14.2.2019 (XL) (публик. 11.2019)** |

1. Тодор Арнаудов**, GPT2-MEDIUM-BG, Свещеният сметач, ДЗБЕ ~6.2021 – 8.2021, 345M – български –** обучен от нулата на Tesla T4 в Colab [31][46] (безплатно), публикуван метод за обучение – популярен клип за жанра в „Ютюб“ с над 4 хил. гледания и над 30 преки абонати.* https://huggingface.co/twenkid/gpt2-medium-bg
2. Antoine Simoulin, Benoit Crabbé. Un modèle Transformer Génératif Pré-entrainé pour le _____ français. Traitement Automatique des Langues Naturelles, **6.2021**, Lille, France. pp.246-255. ffhal03265900f https://hal.science/hal-03265900 : – френски **GPTfr-124M и GPTfr-1B** с архитектурата на GPT3. **5.2021**

3. https://huggingface.co/dbddv01/gpt2-french-small - друг френски малък **SMALL 137M,** също обучен в Colab като българския, но с платена услуга Colab Pro.
4. Wissam Antoun and Fady Baly and Hazem HajjARAGPT2: Pre-Trained Transformer for Arabic Language Generation, **7.3.2021** – https://arxiv.org/pdf/2012.15520
**Арабски,** 4 варианта: 135M, 370M, 792M, 1.46B
(...)

## BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,

Jacob Devlin, Ming-Wei Chang Kenton, Lee Kristina Toutanova**,** Google AI Language, 11.10.2018, 24.5.2019
https://arxiv.org/abs/1810.04805 Bidirectional Encoder Representations, pre-trained, both sides contexts in all levels; in all layers; *the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks, such as question answering and language inference, without substantial task-specific architecture modifications. … BERT is conceptually simple and empirically powerful …* Masked LM… Next Sentence Prediction; BooksCorpus (800M words) (Zhu et al.,2015), English Wikipedia (2,500M words). “feature-based (ELMo, 2018: the pre-trained repres. as additional features) vs fine-tuning (BERT, GPT). “masked language model” (MLM) pre-training objective – Cloze task (Taylor, 1953). Next sentence prediction (NSP) [not next token], NLI (inference); BERT outperforms many task-specific architectures, while it is with a *unified architecture across different tasks. BERTBASE (L=12, H=768, A=12, Total Parameters=110M) and BERTLARGE (L=24, H=1024, A=16, Total Parameters=340M). … WordPiece embeddings (Wu et al., 2016) – 30,000 token vocabulary;* **Special tokens:** [MASK]; [CLS]…[SEP] …; sep – separator; embedding(token, segment, position); token mask: 15%, predict only the masked words, don’t reconstruct the entire input. Fine-tuning: … bidirectional cross attention

**Tosh:** Compare unidirectional (GPT) vs bidirectional (BERT). In the bidirectional transformers the tokens “attend” (are influenced by) all others in the context, while in unidirectional: only by the past ones. There are combinations: T5, UniLM etc.

## * Transformer²

Sakana AI Introduces Transformer²: A Machine Learning System that Dynamically Adjusts Its Weights for Various Tasks, 16.1.2025
https://www.marktechpost.com/2025/01/16/sakana-ai-introduces-transformer%C2%B2-a-machine-learning-system-that-dynamically-adjusts-its-weights-for-various-tasks/
https://github.com/SakanaAI/self-adaptive-llms
„Експертни“ вектори, които се специализират в определени задачи. Сравни с

Mixture-of-Experts: Mixtral, DeepSeekv3.

*** Transformer-Squared: Self-adaptive LLMs,** Qi Sun, Edoardo Cetin, Yujin Tang https://arxiv.org/abs/2501.06252

* **Improving Factuality with Explicit Working Memory,** Mingda Chen, Yang Li et al. Meta FAIR, 25.12.2024 https://arxiv.org/pdf/2412.18069 https://www.marktechpost.com/2025/01/03/meta-ai-introduces-ewe-explicit-working-memory-a-novel-approach-that-enhances-factuality-in-long-form-text-generation-by-integrating-a-working-memory/ Ewe (Explicit Working Memory); long-form text generation, improves factuality; iterative retrieval methods ITER-RETGEN; adaptive retrieval: FLARE, DRAGIN → sentence-by-sentence gen.; Memory3 – encodes KV caches as memories; real-time feedback from external resources and employs online fact-checking

*** LlamaV-o1** – reasoning https://venturebeat.com/ai/llamav-o1-is-the-ai-model-that-explains-its-thought-process-heres-why-that-matters/ https://mbzuai-oryx.github.io/LlamaV-o1/

*** LlamaV-o1: Rethinking Step-by-step Visual Reasoning in LLMs ,** Omkar Thawakar, Dinura Dissanayake et al. https://mbzuai-oryx.github.io/LlamaV-o1/ 10.1.2025 *"Reasoning is a fundamental capability for solving complex **multi-step problems**, particularly in visual contexts where sequential step-wise understanding is essential. … benchmarks .. complex visual perception to scientific reasoning with over 4k reasoning steps in total .. assesses visual reasoning quality at the granularity of individual steps …*

* **Virgo: A Preliminary Exploration on Reproducing o1-like MLLM,** Yifan Du, Zikang Liu et al. https://arxiv.org/pdf/2501.01904 3.1.2025

*** CLDG: Contrastive Learning on Dynamic Graphs,** *Yiming Xu, Bin Shi, 19.12.2024*
*** Offline Reinforcement Learning for LLM Multi-Step Reasoning**, Huaijie Wang, Shibo Hao et al., 20.12.2024 https://arxiv.org/pdf/2412.16145

* **AutoGraph: An Automatic Graph Construction Framework based on LLMs for Recommendation,** Rong Shan, Jianghao Lin et al. 24.12.2024 https://arxiv.org/abs/2412.18241

https://www.marktechpost.com/2025/01/05/autograph-an-automatic-graph-construction-framework-based-on-llms-for-recommendation/
– Utilization of Pre-trained LLMs, Knowledge Graph Construction: Integration with GNNs

* **GraphMERT: Efficient and Scalable Distillation of Reliable Knowledge Graphs from Unstructured Data,** 10.10.2025, Margarita Belova, Jiaxin Xiao et al., Princeton, https://arxiv.org/pdf/2510.09580

# Titans

Архитектура на Гугъл, обещаваща по-добра памет, краткотрайна и дълготрайна.

*** Titans: Learning to Memorize at Test Time,** Ali Behrouz, Peilin Zhong, Vahab Mirrokni**,** Google Research, 31.12.2024 https://arxiv.org/abs/2501.00663v1 *... **attention** due to its limited context but accurate dependency modeling performs as a **short-term** memory, while **neural memory** due to its ability to memorize the data, acts as a **long-term**, more **persistent**, memory. .. p.12. Memory as a Context (MAC), (2) Memory as a Gate (MAG), and (3) Memory as a Layer (MAL) ..* The memory module is recurrent and is *adaptively memorizing* the *more surprising or close to surprising tokens.* Promises for bigger context windows > 2M. *p.5 a simple definition of surprise for a model can be its gradient with respect to the input. The larger the gradient is, the more different the input data is from the past data.* Surprise score; Past surprise + momentary surprise; associative memory objective: key-value pairs; optimizing loss f. Adaptive forgetting gating mechanism… *decides how much information should be forgotten*1 … p.6.* Retrieval of **information** from the memory: *a forward pass without weight update (i.e., inference) to retrieve a memory correspond to a query; a linear layer to project the query. P.7. In order to parallelize: tensorize … **Memory as a context**: using the input context as the query to the memory $M_{t-1}$ to retrieve the corresponding information from the long-term memory*2. .. "persistent memory parameters"*3; p.9 Attention by having both historical and current context, has the ability to decides whether given the current data, the long-term memory information is needed; **Memory as a Gate** (MAG) Architecture;* **sliding attention window** [compare to full attention]; operator - *can be any non-linear gating, e.g.* normalizing and *learnable vector-valued weights, followed by a non-linearity $\sigma(.)$.. attention mask; **Memory as a Layer** – more common in the literature, either full or sliding attention window… the memory layer is responsible to compress the past and current context before the attention module. .. p.11. Memory*

*Without Attention (LMM).* needle-in-a-haystack (NIAH) task … See related work p.25.
**Tosh:** **1:* The "forgetting" means adjusting the weights according to computed gradients, the same as with "learning; the information is dispersed.
***2:** This is related to the **Zrim**'s concept {К}-рчнк, {К}-dict, контекстен речник, as well as „избирател на {К}", {К}-избирател, coined around 2014, but {К} is more complex and abstract system than the "context" in transformers, which is just a span of "tokens"; in **Zrim**'s view, transformers context span is closer to [,,](*#($\{К\}_{рчнк}$)) … :P
***3:** Their persistent memory is an additional component in their approach towards the expression of the computation of a modification of the attention – how much of which weights to be considered in the calculation of the output values.

*** Can AI Truly Develop a Memory That Adapts Like Ours?**
**Exploring Titans: A new architecture equipping LLMs with human-inspired memory that learns and updates itself during test-time.** Moulik Gupta
Jun 12, 2025 https://towardsdatascience.com/can-ai-truly-develop-a-memory-that-adapts-like-ours/

* **LLM Pretraining with Continuous Concepts,** Jihoon Tack et al. 2.2025 – *Continuous Concept Mixing (CoCoMix), a novel pretraining framework that combines discrete next token prediction with continuous concepts.*

| | |
|---|---|
| **Large Concept Model (LCM)**<br>- езиковото моделиране е във високоразмерното представяне, а не в дискретните токени (буквачета) и на по-високо абстрактно и семантично ниво. „Понятие" в този обслов – „атомна" абстрактна идея; често отговаря на изречение, израз; кодира се чрез SONAR<br><br>*** "Large Concept Models: Language Modeling in a Sentence Representation Space"** by Loïc Barrault et al. (20), 12.12.2024 https://arxiv.org/abs/2412.08821 2.3 However, a given context may have many plausible, yet semantically different, continuations; decoder-only, sentence-embeddings TA: cmp:logic predicates; Shouldn't be just a sequence, but a spreading graph, a network at different resolutions, views, segmentations, conceptualizations, ranges etc. see Tree-of-thought prompting technique, self-consistency checks etc.<br>* P.-A. Duquenne, H. Schwenk, and B. Sagot. **SONAR: sentence-level multimodal and language-agnostic representations**, 8.2023 https://arxiv.org/abs/2308.11466 transformer encoder-decoder; no token level: whole sentences embedding, fixed-size repres.; training on parallel data for machine translation objective; example autoencoding – changed two words but preserved meaning (warned – cautioned, chair of – chairman) p.4; Architecture: 24 layer encoder & 24 layer decoder; trained on human labeled, back-translated | |

& mined data; a pooling operation at the encoder outputs in order to achieve fixed-size sentence representation. …
200+ languages; speech+text to text; See also: LASER3 and LabSE sentence embeddings; xsim and xsim++ multilingual similarity search tasks

* **xSIM++: An Improved Proxy to Bitext Mining Performance for Low-Resource Languages,** Mingda Chen et al., 2023
NMT – neural machine translation; global mining; automatic alignment of multilingual texts; FLORES200 dev set; Transformation categories: Causality Alternation, Entity/Number Replacement (spaCy library for the latter) – TA: that can be used for data augmentation for preparations of datasets for LLM training.

*** The FLORES-200 Evaluation Benchmark for Low-Resource and Multilingual Machine Translation**
https://github.com/facebookresearch/flores/blob/main/flores200/README.md
By Meta AI; continues FLORES-101; corpus: 3001 sentences, 842 articles, average 21 words per sent., 3 splits: dev, devtest, test (hidden).
https://dl.fbaipublicfiles.com/large_objects/nllb/models/spm_200/dictionary.txt
Включва български. bul_Cyrl

https://github.com/facebookresearch/LASER/blob/main/tasks/xsim/README.md
LASER: xSIM (multilingual similarity search); bidirectional LSTM, 2019

*** Accelerating scientific breakthroughs with an AI co-scientist**, Juraj Gottweis, Google Fellow, and Vivek Natarajan, Research Lead 19.2.2025 https://research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist/ *** Towards an AI co-scientist** Juraj Gottweis, Wei-Hung Weng et al. (>30), 26.2.2025 https://arxiv.org/abs/2502.18864 Сътрудник на учения: задаване на изследователска цел; асинхронна мулти-агентна система с Джемини 2.0 – пораждане на хипотези и предложения, обосноваването им чрез литературата, откриване на нови резултати; проучване на литературата, симулиране на „дебати“ върху нея, пълен преглед на източници от Мрежата, „дълбока проверка“; проверка на близост, мета-преглед … Биология: откриване на нови лекарства и др. „Test-time compute”.

Вид *„ускорител на изследователската дейност“* (Research Accelerator) предвиден от *Българските пророчества.*

*** Training a Generally Curious Agent**
Fahim Tajwar, Yiding Jiang, Abitha Thankaraj, Sumaita Sadia Rahman, J Zico Kolter, Jeff Schneider, Ruslan Salakhutdinov https://arxiv.org/abs/2502.17543

24.2.2025 PAPRIKA … LLMs, incontext learning; strategic exploration & sequential decision-making https://arxiv.org/abs/2502.18449 -

*** Idiosyncrasies in Large Language Models** Mingjie Sun, Yida Yin, Zhiqiu Xu, J. Zico Kolter, Zhuang Liu 17.2.2025 - recognizing LLMs with a high confidence based on words selection etc. разпознаване на езиковия модел по стила на писане: честота на употреба на особени думи и изрази – първи или общо в текста; честоти за всички думи; особености при форматиран изход (хтмл, булети, формули; брой на подчертавания, брой на думи в курсив, брой изброявания/списъци, блокове с код, и пр.) Term Frequency-Inverse Document Frequency (TF-IDF) ; ChatGPT, Claude, Grok, Gemini, DeepSeek; 3 types LLMs: base (pre-trained), instruct – fine-tuned to follow instructions; chat – for dialog, conversational agents. Length/format control … LLMs for judging the style of the responses – Използва различни езикови модели за оценка на стила, напр. Claude и Grok оценява
**Тош:** Llama3 беше лесен за разпознаване в lmsys (lmarena) по встъпителните думи.

*** The FFT Strikes Back: An Efficient Alternative to Self-Attention**, Jacob Fein-Ashley 28.2.2025 https://arxiv.org/abs/2502.18394 2.2025 - convert to frequency domain; learnable adaptive spectral filters with modReLU; FFTNet; efficiently capture long-range dependencies - alternative to self-attention; tests on Long Range Arena (tasks: ListOps, Text, Retrieval, Image, Pathfinder, Path-X) & ImageNet; more efficient global mixing of tokens; related: Fourier Neural Operator. Alternatives to standard self-attention (alg.complexity $O(N^2)$): Synthesizer Tay, MLP-Mixer to avoid explicit token-pair interactions (comparisons by dot-product); Hyena Poli, Orthogonal matrix decomposition … discrete Fourier transform, energy preservation via Parseval's theorem; global context vector; Inverse Fourier Transfor to return to the token doma; globally mixed representation;
Reduce to O(n log n) Path-X

*** Long Range Arena: A Benchmark for Efficient Transformers**, Yi Tay, Mostafa Dehghani et al. 11.2020 … 1K to 16K tokens (outdated) ; "long-range Transformer models (Reformers, Linformers, Linear Transformers, Sinkhorn Transformers, Performers, Synthesizers, Sparse Transformers, and Longformers)"
https://github.com/google-research/long-range-arena
https://arxiv.org/abs/2011.04006

*** SWE-Lancer: Can Frontier LLMs Earn $1 Million from Real-World Freelance Software Engineering**?, Samuel Miserendino, Michele Wang, Tejal Patwardhan, Johannes Heidecke, 17.2.2025 – OpenAI; "1400 freelance tasks, Upwork, valued at $1 million" benchmark; from $50 bug fixes to $32000. По-реалистични изпитания по програмиране върху задачи за разработка от платформата за поръчки (фрийланс) – средно отнемат по 21 дни, 24% са над $1000; текущите водещи

модели не се справят с по-трудните задачи. https://openai.com/index/swe-lancer/ https://arxiv.org/abs/2502.12115 https://github.com/openai/SWELancer-Benchmark

* **https://x.com/TheAITimeline** Обзор на нови статии

* **TTS-VAR: A Test-Time Scaling Framework for Visual Auto-Regressive Generation**, Zhekai Chen1 Ruihang Chu2 et al., 24.7.2025 *Multi-scale coarse-to-fine representations and progressively predicts the "next scale" to synthesize images through hierarchical aggregation. Auto-Regressive (AR) architectures for image generation* (Compare to Diffusion models) https://arxiv.org/pdf/2507.18537 .. quantizing *the feature map F into a sequence of multi-scale discrete residual feature maps .. calculating the reward function for intermediate results and selecting those with higher scores.*

* **Emu3: Next-token prediction is all you need.** Xinlong Wang, Xiaosong Zhang et al. arXiv preprint arXiv: 2409.18869, 27.9.2024. https://arxiv.org/abs/2409.18869 Autoregressive image and video generation. *eliminating the need for diffusion or compositional approaches; Emu and Emu2 introduce a unified autoregressive objective: predicting the next multimodal element, by regressing visual embeddings or classifying textual tokens. Emu3 for the first time demonstrates that next-token prediction across images, video, and text can surpass these well-established models, without relying on compositional methods.* Video dataset: duration of the clips, p.20: ~ 5-7 sec: 29%, 7-9 s: 15%, 9-11 s: 11%, 11-13 s: 7% … (~62.2% <13 s…) … < 21 s (87.3%) → the majority are very short and short clips. **Emu3 configuration:** 8B params, 32 layers, 4096 hidden size, 14336 Intermediate size, 32 heads; 8 KV Heads; 184622 Vocabulary Size; 1000000 RoPE Base; 131072 Context Length. .. *Next-Token Prediction (Transformer Decoder)*

* **Generative multimodal models are in-context learners.** Quan Sun, Yufeng Cui, Xiaosong Zhang, Fan Zhang, Qiying Yu, Yueze Wang, Yongming Rao, Jingjing Liu, Tiejun Huang, and Xinlong Wang**.** In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14398–14409, 2024.https://openaccess.thecvf.com/content/CVPR2024/papers/Sun_Generative_Multimodal_Models_are_In-Context_Learners_CVPR_2024_paper.pdf https://arxiv.org/abs/2312.13286 (20.12.2023, 8.5.2024)
**Emu2, 37B params** … https://github.com/baaivision/Emu

* **Emu: Generative pretraining in multimodality**. Quan Sun, Qiying Yu, Yufeng Cui, Fan Zhang, Xiaosong Zhang, Yueze Wang, Hongcheng Gao, Jingjing Liu, Tiejun Huang, and Xinlong Wang. In The Twelfth International Conference on Learning Representations, 2023. https://arxiv.org/abs/2307.05222 – **Emu; 14B** params. *Transformer-based multimodal FM; seamlessly generates images and texts in multimodal context; can take in any single-modality or multimodal data*

*input indiscriminately (e.g., interleaved image, text and video) through a one-model-for-all autoregressive training process. .. A generalist multimodal interface for both image-to-text and text-to-image tasks; in-context image and text generation; zero-shot/few-shot tasks: image captioning, visual question answering, video question answering and text-to-image generation … batch=128 (image-text pairs), 64: interleaved image-text data, 16: videotext pair and interleaved video-text data; video resized to 224x224, randomly sample 8 frames; 128 NVIDIA 80G-A100 GPUs for 10k steps with around 82M samples (150B tokens in total), and the pretraining takes approximately 2 days.*

*** Youtube channel reading arxiv papers, such as the above:**
https://www.youtube.com/watch?v=i_jYWZp5rbw

See also: **Flow Matching** for generative models.
https://mlg.eng.cam.ac.uk/blog/2024/01/20/flow-matching.html
https://taohu.me/vincent-genai-course/diffusion/basics/flow-matching.html
***Vincent教你学Generative AI:*** *A continuous-time transformation from a simple distribution (like a Gaussian) to a complex target distribution. Unlike normalizing flows, flow matching doesn't require computing the Jacobian determinant, making it more flexible and computationally efficient.*
A generalization of score-based diffusion models.
* Tao Hu et al. https://github.com/dongzhuoyao/awesome-flow-matching

* **Multi-scale Transformer Language Models**, Sandeep Subramanian, R. Collobert, M. Ranzato, Y-Lan Boureau, 2020 https://arxiv.org/abs/2005.00581 Виж и др. multi-scale в др. прлжн.

*** DeepSeek-V3 Technical Report, DeepSeek AI, Aixin Liu, Bei Feng, Bing Xue** et al. (~218 authors) https://arxiv.org/abs/2412.19437v1 27.12.2024 – MoE, 37B active weights out of 671B total params; 61 layers … 2.788M H800 GPU for full training; 14.8T high-quality diverse tokens; Multi-head Latent Attention (MLA) (DeepSeek-AI, 2024c); DeepSeekMoE (Dai et al., 2024) for cost-effective training; FP8 mixed precision training and implement comprehensive optimizations for the training framework. Pre-training: first 32K context-length, then extended to 128K. Training cost (in GPU hours, $2/h H800): USD$5.328M$ + 0.238M + $0.01M = $5.576M. 1T tokens ~ 180K H800 GPU hours, 3.7 days on 2048 H800 GPUs. Post-training: distill reasoning; Rotary Positional Embedding (RoPE) (Su et al., 2024); Auxiliary-Loss-Free Load Balancing; shared experts and routed experts in MoE architecture; **Multi-Token Prediction –** additionally predict more tokens with other attention heads: shared embedding layer, a shared output head, a Transformer block, a projection matrix. MTP Training Objective: for each prediction depth, we compute a cross-entropy loss (…) Fine-Grained Quantization, online quantization; suggestions for new functions in the accelerators; Byte Pair Encoding

(BPE) tokenizer; Decoders …: Fill-in-Middle (FIM) strategy - predict middle text based on contextual cues, not only the next token:
<|fim_begin|> prefix <|fim_hole|> suffix <|fim_end|> middle <|eos_token|>
61 lauers, hidden dim = 7168, All learnable parameters are randomly initialized with a standard deviation of 0.006. In MLA, attention heads = 128, per-head dim = 128. … Self-Rewarding, RL … **Conclusion:** *DeepSeek consistently adheres to the route of open-source models with longtermism, aiming to steadily approach the ultimate goal of AGI (Artificial General Intelligence) ... "We will continuously iterate on the quantity and quality of our training data, and explore* ***the incorporation of additional training signal sources****, aiming to drive data scaling across a more comprehensive range of dimensions "*

* **DeepSeek LLM: Scaling Open-Source Language Models with Longtermism,** DeepSeek-AI: Xiao Bi et al. 5.1.2024, https://arxiv.org/pdf/2401.02954 7B & 67B ..
* **DeepSeek-Coder: When the Large Language Model Meets Programming - The Rise of Code Intelligence**, Daya Guo et al., 26.1.2024 https://arxiv.org/pdf/2401.14196 Dataset: 87% source code, 10% English coderelated NL corpus, 3% code-unrelated Chinese NL. Github & StackExchange … Data preparation: data crawling, rule filtering, Dependency Parsing, repository-level code deduplication, quality screening; Modes: Next Token Prediction & Fill-in-the-Middle; sizes:1.3B, 6.7B, and 33B params, Decoder-only, RoPE, Grouped-Query-Attention (GQA), group size = 8 … repository-level code processing: project-level [multiple files, whole projects with file names #main.py … etc.] Other SOTA: CodeGeeX2, StarCoder, CodeLlama, code-cushman-001, GPT-3.5 & GPT-4 …

* **100 Days After DeepSeek-R1: A Survey on Replication Studies and More Directions for Reasoning Language Models**, Chong Zhang 1,2 , Yue Deng1 1 , Xiang Lin1 1 , Bin Wang1 1 , Dianwen Ng1, Hai Ye1,3, Xingxuan Li1, Yao Xiao1,4, Zhanfeng Mo1,5, Qi Zhang2, Lidong Bing1, 15.5.2025 - 1MiroMind, 2Fudan University, 3National University of Singapore, 4Singapore University of Technology and Design, 5Nanyang Technological University https://arxiv.org/html/2505.00551v3 .. https://arxiv.org/abs/2505.00551v3 *"RL-Poet (Doria, 2025) demonstrates the ability to generate literary poems in multiple languages with correct poetic rules, despite being trained almost exclusively on English data. Compared to the prevailing imitative learning paradigm, these results highlight the potential of achieving artificial general intelligence through general reinforcement learning. In other respects," (..) RL Datasets, Rewards: Accuracy, Format, Length;Sampling Strategies… Recipes of Training Data: Quantity and Diversity, Difficulty, Data Cleaning, De-duplication and Decontamination, Curriculum Learning Based on Data Difficulty…. REINFORCE, PPO, and GRPO… Exploration Beyond Supervision. ..*

* See also other Chinese models: **KIMI K2, Qwen3, Qwen3-Coder, …** 31.7.2025; **Qwen3-Image-Edit** (released on 17.8.2025)

* https://huggingface.co/moonshotai/Kimi-K2-Instruct * https://www.kimi.com/
** Large-Scale Training: Pre-trained a 1T parameter MoE model on 15.5T tokens with zero training instability. .. 32B active params, Attention hidden dimensions = 7168, Vocab_size = 160K, Experts = 384, Selected Experts per Token = 8, Attention Heads = 64, Context Length = 128K; Attention Mechanism=MLA; Activation Function=SwiGLU*
* https://github.com/QwenLM/Qwen3 *Qwen3-Instruct-2507 – Qwen3-235B-A22B-Instruct-2507: **256K** context, MoE, 22B active params;* smaller variants (8B) etc. https://huggingface.co/Qwen/Qwen3-8B Parameters: 8.2B*;* Params (Non-Embedding): 6.95B*;* Layers: 36; Attention Heads (GQA): 32 for Q and 8 for KV; Context Length: 32,768 natively and 131,072 tokens with YaRN.
* Qwen Team, Qwen3 Technical Report, 5.2025, https://arxiv.org/pdf/2505.09388 *The Qwen3 series includes models of both dense and Mixture-of-Expert (MoE) architectures, with parameter scales ranging from 0.6 to 235 billion. A key innovation in Qwen3 is the integration of thinking mode (for complex, multi-step reasoning) and non-thinking mode (for rapid, context-driven responses) into a unified framework.*
https://github.com/QwenLM/Qwen3-Coder .. Qwen3-Coder-480B-A35B-Instruct
* There is free service, https://chat.qwen.ai/ Qwen-Image-Edit ~ 10 images/24 h (23.8.2025)

* **Reflection AI raises $2B to be America's open frontier AI lab, challenging DeepSeek** https://techcrunch.com/2025/10/09/reflection-raises-2b-to-be-americas-open-frontier-ai-lab-challenging-deepseek/ 9.10.2025:*"Reflection AI hasn't yet released its first model, which will be largely text-based, with multimodal capabilities in the future, according to Laskin."*

* **Absolute Zero: Reinforced Self-play Reasoning with Zero Data**, Andrew Zhao, Yiran Wu, Yang Yue, Tong Wu, Quentin Xu, Yang Yue, Matthieu Lin, Shenzhi Wang, Qingyun Wu, Zilong Zheng, Gao Huang, 7.5.2025 https://arxiv.org/abs/2505.03335v2
Previous work: Reinforcement learning with verifiable rewards (RLVR); *Absolute Zero Reasoner (AZR), a system that self-evolves its training curriculum and reasoning ability by using a code executor to both validate proposed code reasoning tasks and verify answers, serving as an unified source of verifiable reward to guide open-ended yet grounded learning*1.*
*"Despite being trained entirely without external data, AZR achieves overall SOTA performance on coding and mathematical reasoning tasks"*. Supervised Learning (human curated reasoning traces)*2 …
***1: Tosh: Cmp:** Todor Arnaudov's 2010s yet unpublished directions, {K}!, **Zrim**: see future works.
***2: Cmp:** Todor Arnaudov's ~2013-2014 **"behaveintrospective",** пт(мс), пс(мс) и др (…) [no appropriate symbols in this font]**.** The "code executors", "Tools" in LLMs are related to ({K}!} in Zrim/"The AGI infrastructure"/Assistant. Compare also to "Reasoning", "Chain of thought reasoning", "Thinking" etc. in LLMs, 10 years later.
*"We cast code executor as an open-ended yet grounded environment, sufficient to*

*both validate task integrity and also provide verifiable feedback for stable training. ..p.3: the Absolute Zero paradigm, where during training, the model simultaneously proposes tasks, solves them, and learns from both stages. No external data is required and the model learns entirely through self-play and experience, aided by some environment .. Task types:* ***Abduction, Deduction, Induction. П. Abduction: O = P(?). Deduction: ? = P(I); Induction: O = ? (I);***

* Lopez, R. H. Q. **Complexipy:** An extremely fast python library to calculate the cognitive complexity of python files, written in rust, 2025.URL https://github.com/rohaquinlop/complexipy
* Ebert, C., Cain, J., Antoniol, G., Counsell, S., and Laplante, P.Cyclomatic complexity.IEEE software, 33(6):27–29, 2016.

* **Native Sparse Attention: Hardware-Aligned and Natively Trainable Sparse Attention** Jingyang Yuan∗1,2, Huazuo Gao et al.https://arxiv.org/abs/2502.11089 , 27.2.2025 – DeepSeek, Peking University, Univ. of Washington; хардуерни оптимизации и разредено внимание за постигае на по-голям обхват на контекста с незначителна или без загуба на точност. Динамична йерархична стратегия за разреждане, компресия с огрубяване на токените и фин подбор на токени, за да запази както ширината на обхвата, така и прецизността в местния обслов. Обзор на методи за разредено внимание и защо някои от тях не постигат повишена производителност т..2.1

* **Mixtral of Experts,** Albert Q. Jiang, Alexandre Sablayrolles et al. (~30) 8 Jan 2024 https://arxiv.org/pdf/2401.04088 | https://github.com/mistralai/mistral-inference *Mixtral 8x7B, a Sparse Mixture of Experts (some).. the same architecture as Mistral 7B,[but] each layer ..[has] 8 feedforward blocks (i.e. experts). For every token, at each layer, a* ***router network selects two experts*** *to process the current state and* ***combine*** *their outputs. Even though each token only sees two experts, the selected experts can be different at each timestep. As a result,* ***each token has access to 47B parameters****, but only uses 13B active parameters during inference. Mixtral was trained with a context size of* ***32k tokens*** *and it outperforms or matches Llama 2 70B and GPT-3.5 across all evaluated benchmarks. In particular, … Claude-2.1, Gemini Pro…* ***32 layers*** *p.7 5. Routing analysis* p.8: color-coded first-expert choices per token; TensorRT-LLM,Triton, NVIDIA.
* https://github.com/mistralai/mistral-inference/blob/main/tutorials/getting_started.ipynb (7B-insruct. See also BgGPT experiments by The Sacred Computer:https://twenkid.com/bggpt/

* **Voxtral,** Mistral AI, Alexander H. Liu, Andy Ehrenberg, Andy Lo et al., https://arxiv.org/pdf/2507.13264 Voxtral small & mini, speech and text; 17.7.2025; transformer; 128 Mel-bin, 160-hop length; Whisper large-v3 encoder; 30 sec fixed receptive field; Whisper pads short audios to 30-seconds. A small penalty if not padded. *MLP adapter layer that downsamples the audio embeddings along the*

*temporal axis; to reduce the frame rate from 50 Hz to 12.5 Hz:* 40 min. audio with 32K context-length. Mini: 4.7 B, Small: 24.3 B params. Audio Encoder: 640 M, Audio Adapter 25/52 M, Text embeddings 400 M/ 670 M, Language decoder: 3.6 B/ 22.9B params. Mini: on top of Ministral 3B. Small: Mistral small 3.1 24 B backbone. … https://huggingface.co/mistralai/Voxtral-Mini-3B-2507

* **Unified Language Model Pre-training for Natural Language Understanding and Generation,** Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, Hsiao-Wuen Hon, **Microsoft Research**, 5.2019/10.2019 https://arxiv.org/abs/1905.03197 *(UniLM) that can be fine-tuned for both natural language understanding and generation tasks. The model is pre-trained using three types of language modeling tasks: unidirectional, bidirectional, and sequence-to-sequence prediction.*
*Left-to-Right LM, Right-to-Left LM, Bidirectional LM, Sequence-to-Sequence LM: Covers all.* [SOS]..[EOS]; CNN/DailyMail dataset, Gigaword;
https://huggingface.co/datasets/abisee/cnn_dailymail
https://paperswithcode.com/sota/text-summarization-on-gigaword?p=mass-masked-sequence-to-sequence-pre-training – text summarization benchmark
https://github.com/tensorflow/datasets/blob/master/tensorflow_datasets/summarization/gigaword.py
https://www.kaggle.com/datasets/arngowda/gigaword-corpus ~ 3.3 M texts and 3.12M summaries →

*** Microsoft® - Large collection of various NN LLMs** as of 22.4.2025 @Vsy: obrb
**Foundation Architecture**
TorchScale - A Library of Foundation Architectures (repo)
Fundamental research to develop new architectures for foundation models and AI, focusing on modeling generality and capability, as well as training stability and efficiency.
**Stability** - DeepNet: scaling Transformers to 1,000 Layers and beyond
**Generality** - Foundation Transformers (Magneto): towards true general-purpose modeling across tasks and modalities (including language, vision, speech, and multimodal)
**Capability** - A Length-Extrapolatable Transformer
**Efficiency & Transferability** - X-MoE: scalable & finetunable sparse Mixture-of-Experts (MoE)
*The Revolution of Model Architecture*
**BitNet**: 1-bit Transformers for Large Language Models
**RetNet**: Retentive Network: A Successor to Transformer for Large Language Models
**LongNet**: Scaling Transformers to 1,000,000,000 Tokens
**Foundation Models:** The Evolution of (M)LLM (Multimodal LLM):
**Kosmos-2.5: A Multimodal Literate Model** – a multimodal literate model for machine reading of text-intensive images. … spatially-aware text blocks,

**Kosmos-2:** Grounding Multimodal Large Language Models to the World
**Kosmos-1**: A Multimodal Large Language Model (MLLM)
**MetaLM:** Language Models are General-Purpose Interfaces
**The Big Convergence:** Large-scale self-supervised pre-training across tasks (predictive and generative), languages (100+ languages), and modalities (language, image, audio, layout/format + language, vision + language, audio + language, etc.)
**Language & Multilingual:** UniLM: unified pre-training for language understanding and generation
**InfoXLM/XLM-E:** multilingual/cross-lingual pre-trained models for 100+ languages
**DeltaLM/mT6:** encoder-decoder pre-training for language generation and translation for 100+ languages
**MiniLM:** small and fast pre-trained models for language understanding and generation

**AdaLM:** domain, language, and task adaptation of pre-trained models
**EdgeLM(**NEW): small pre-trained models on edge/client devices
**SimLM** (NEW): large-scale pre-training for similarity matching
**E5 (NEW):** text embeddings
**MiniLLM** (NEW): Knowledge Distillation of Large Language Models
**Vision: BEiT/BEiT-2:** generative self-supervised pre-training for vision / BERT Pre-Training of Image Transformers; **DiT:** self-supervised pre-training for Document Image Transformers
**TextDiffuser/TextDiffuser-2** (NEW): Diffusion Models as Text Painters
**Speech**: WavLM: speech pre-training for full stack tasks
**VALL-E:** a neural codec language model for TTS
**Multimodal (X + Language): LayoutLM/LayoutLMv2/LayoutLMv3:** multimodal (text + layout/format + image) **Document Foundation Model for Document AI** (e.g. scanned documents, PDF, etc.)
**LayoutXLM:** multimodal (text + layout/format + image) Document Foundation Model for multilingual Document AI
**MarkupLM:** markup language model pre-training for visually-rich document understanding
**XDoc:** unified pre-training for cross-format document understanding
**UniSpeech:** unified pre-training for self-supervised learning and supervised learning for ASR
**UniSpeech-SAT:** universal speech representation learning with speaker-aware pre-training
**SpeechT5:** encoder-decoder pre-training for spoken language processing
**SpeechLM:** Enhanced Speech Pre-Training with Unpaired Textual Data
**VLMo:** Unified vision-language pre-training
**VL-BEiT (NEW):** Generative Vision-Language Pre-training - evolution of BEiT to multimodal

**BEiT-3 (NEW):** a general-purpose multimodal foundation model, and a major milestone of The Big Convergence of Large-scale Pre-training Across Tasks, Languages, and Modalities.

**Toolkits:** s2s-ft: sequence-to-sequence fine-tuning toolkit; Aggressive Decoding (NEW): lossless and efficient sequence-to-sequence decoding algorithm

**Applications:** TrOCR: transformer-based OCR w/ pre-trained models
**LayoutReader:** pre-training of text and layout for reading order detection
**XLM-T:** multilingual NMT w/ pretrained cross-lingual encoders

[Submitted on 27 Feb 2023 (v1), last revised 1 Mar 2023 (this version, v2)]
**Language Is Not All You Need: Aligning Perception with Language Models**
Shaohan Huang, Li Dong et al. (16) *A big convergence of language, multimodal perception, action, and world modeling is a key step toward artificial general intelligence. In this work, we introduce* ***Kosmos-****1, a Multimodal Large Language Model (MLLM) that can perceive general modalities, learn in context (i.e., few-shot), and follow instructions (i.e., zero-shot). Specifically, we train Kosmos-1 from scratch on web-scale multimodal corpora, including arbitrarily interleaved text and images, image-caption pairs, and text data.*
https://arxiv.org/abs/2302.14045

[Submitted on 26 Jun 2023 (v1), last revised 13 Jul 2023 (this version, v3)]
**Kosmos-2: Grounding Multimodal Large Language Models to the World**
Zhiliang Peng, Wenhui Wang, Li Dong, Yaru Hao, Shaohan Huang, Shuming Ma, Furu Wei
*We introduce Kosmos-2, a Multimodal Large Language Model (MLLM), enabling new capabilities of perceiving object descriptions (e.g., bounding boxes) and grounding text to the visual world. .. refer expressions as links in Markdown, i.e., ``[text span](bounding boxes)", where object descriptions are sequences of location tokens. Together with multimodal corpora, we construct large-scale data of grounded image-text pairs (called GrIT) to train the model. In addition to the existing capabilities of MLLMs (e.g., perceiving general modalities, following instructions, and performing in-context learning), Kosmos-2 integrates the grounding capability into downstream applications. … evaluate Kosmos-2 on … tasks, including (i) multimodal grounding, such as referring expression comprehension, and phrase grounding, (ii) multimodal referring, such as referring expression generation, (iii) perception-language tasks, and (iv) language understanding and generation. This work lays out the foundation for the development of* ***Embodiment A****I and sheds* ***light on the big convergence of language, multimodal perception, action, and world modeling, which is a key step toward artificial general intelligence.*** *Code and pretrained models are available at this https URL.*

https://arxiv.org/pdf/2306.14824 | https://arxiv.org/abs/2306.14824
https://github.com/microsoft/unilm/tree/master/kosmos-2
*Templates for Grounded Instruction Data… What is  it <box><loc1><loc2></box>? It is {expression}."*

***Language Models are General-Purpose Interfaces:***
https://thegenerality.com/agi/

**BitNet b1.58 2B4T Technical Report,** Shuming Ma, Hongyu Wang, Shaohan Huang, Xingxing Zhang, Ying Hu, Ting Song, Yan Xia, Furu Wei *We introduce BitNet b1.58 2B4T,* ***the first open-source, native 1-bit*** *Large Language Model (LLM) at the 2-billion parameter scale. Trained on a corpus of 4 trillion tokens.*
https://arxiv.org/abs/2504.12285

**Multimodal Latent Language Modeling with Next-Token Diffusion,** Yutao Sun, Hangbo Bao, Wenhui Wang, Zhiliang Peng, Li Dong, Shaohan Huang, Jianyong Wang, Furu Wei, 12.2024, Microsoft & Tsinghua University
https://arxiv.org/abs/2412.08635 *– Multimodal generative models require a unified approach to handle both discrete data (e.g., text and code) and continuous data (e.g., image, audio, video). In this work, we propose Latent Language Modeling (LatentLM), which seamlessly integrates continuous and discrete data using causal Transformers. .. a variational autoencoder (VAE) to represent continuous data as latent vectors and introduce next-token diffusion for autoregressive generation of these vectors; a σ-VAE to address the challenges of variance collapse; LatentLM surpasses Diffusion Transformers in both performance and scalability. When integrated into multimodal large language models, LatentLM provides a general-purpose interface that unifies multimodal generation and understanding… favorable performance compared to Transfusion and vector quantized models; In TTS synthesis, LatentLM outperforms the SOTA VALL-E 2 model in speaker similarity and robustness. p.2.: In order to natively handle discrete and continuous data in multimodal large language models: 3 main strands of research. 1)[RPG+21, WCW+23, Tea24]: VQ-VAE [vdOVK17, ERO21] to quantize continuous data into discrete codes and treats everything as discrete tokens in autoregressive language models. The continuous data are then recovered by the VQ-VAE decoder by conditioning on discrete codes. The performance is often limited by lossy tokenization, which creates a restrictive bottleneck during quantization … p.3: Latent language modeling (LatentLM) autoregressively perceives and generates multimodal sequences (with discrete and continuous data) in a unified way, … Specifically, let x = x1 · · · xN denote an input sequence of discrete and continuous tokens. For a discrete token, we use a lookup table to get its vector representation. For continuous data, variational autoencoder (VAE) [KW14] is used as tokenizer to*

*compress input data to latent vectors (Section 2.3) Latent Vector Representation of Continuous Data …*

**Tosh:** It requires a unified approach (on low level) if it is processed with a monolithic neural architecture

**Updates from Microsoft:** 9.2025: "MAI-1-preview was only trained on 15,000 H100s" … "**Microsoft is making 'significant investments' in training its own AI models",** 12.9.2025 https://www.theverge.com/report/776853/microsoft-ai-training-capacity-investments-in-house-models
* https://www.semafor.com/article/08/28/2025/microsoft-unveils-powerful-new-home-grown-ai-models *Both models are geared toward cost-effectiveness. MAI-1-preview was trained on roughly 15,000 Nvidia H-100 GPUs, compared to models like xAI's Grok, which was trained on more than 100,000 such chips.*
* https://www.microsoft.com/en-us/research/
* https://www.microsoft.com/en-us/research/blog/
* **RenderFormer**: How neural networks are reshaping 3D rendering
Published September 10, 2025 By Yue Dong , Lead Researcher
https://www.microsoft.com/en-us/research/blog/renderformer-how-neural-networks-are-reshaping-3d-rendering/ … " – "The *view-independent transformer* captures scene information unrelated to viewpoint, such as shadowing and diffuse light transport, using self-attention between triangle tokens. – *The view-dependent transformer* models effects like visibility, reflections, and specular highlights through cross-attention between triangle and ray bundle tokens. .."
*** RenderFormer: Transformer-based Neural Rendering of Triangle Meshes with Global Illumination,** Chong Zeng, Yue Dong, Pieter Peers, Hongzhi Wu, Xin Tong
SIGGRAPH 2025 | August 2025 https://www.microsoft.com/en-us/research/publication/renderformer-transformer-based-neural-rendering-of-triangle-meshes-with-global-illumination/ – *"a neural rendering pipeline that directly renders an image from a triangle-based representation of a scene with full global illumination effects and that does not require per-scene training or fine-tuning."*
* https://microsoft.github.io/renderformer/
RenderFormer: Transformer-based Neural Rendering of Triangle Meshes with Global Illumination, CHONG ZENG et al., 2025
https://renderformer.github.io/pdfs/renderformer-paper.pdf
https://arxiv.org/abs/2505.21925 – *"…at most 8 diffuse light sources, and the camera (with fixed 512 × 512 resolution) is placed outside the scene's bounding box…*
* Microsoft Agent framework: https://www.marktechpost.com/2025/10/03/microsoft-releases-microsoft-agent-framework-an-open-source-sdk-and-runtime-that-simplifies-the-orchestration-of-multi-agent-systems/
* https://youtu.be/AAgdMhftj8w?si=hFr14eot4WpDEixq – Semantic Kernel, AutoGen – multi-agent research framework; orchestration: (web service, retrieval - search,

analytics – DB, Coding – docker;) agnets → Coordinator Agent → a conversation with a human in the loop → Memory (history of work) ... LLM-driven orchestration ... Workflow orchestration (business logic-driven) ... Semantic kernel agents ... 15:xx min; *Travel planning, daily meal allowance *... 19 m:*
**Tosh:** meaningless sample tasks

* **SIMLM: Pre-training with Representation Bottleneck for Dense Passage Retrieval,** Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, Furu Wei, Microsoft Corporation
https://aclanthology.org/2023.acl-long.125.pdf
https://github.com/microsoft/unilm/tree/master/simlm Condenser (Gao and Callan, 2021), coCondenser (Gao and Callan,2022), SEED (Lu et al., 2021), DiffCSE (Chuang et al., 2022), and RetroMAE (Liu and Shao, 2022) … The encoder learns to compress all the semantic information into a dense vector and passes it to the decoder to perform well on the replaced language modeling task.

*** Language Models are General-Purpose Interfaces**, Yaru Hao, Haoyu Song, Li Dong, Shaohan Huang, Zewen Chi, Wenhui Wang, Shuming Ma, Furu Wei, 2022
https://arxiv.org/abs/2206.06336 jointly pretrain the interface and the modular encoders … both causal and non-causal modeling; in-context learning and open-ended generation; Language models as a universal task layer. multi-turn conversational interactions; MetaLM. *Non-causal encoders as System 1, and causal language models as System 2. Cognition is usually categorized into two levels (Kahneman, 2011; Bengio, 2019): System 1 (i.e., intuitive, and unconscious) and System 2 (i.e., sequential, conscious, planning, and reasoning)*;* non-causal encoders pretrained by masked data modeling, such as BERT (Devlin et al., 2019) and BEiT (Bao et al., 2022), are used as a perception layer to encode various input modalities. The encoding modules can be viewed as System 1.
**Causal encoders**: in-context learning, multi-turn interaction, and open-ended generation (Unidirectional). Pg.5-6.MetaLM and architectures: Prefix LM (Encoder-Decoder with Cross-Attention). Non-Causal LM (Bidirectional). Semi-Causal LM. Backbone Network. Connector - a connector layer between the universal task layer and various bidirectional encoders. The connectors are used to match the output dimensions of foundation models with the universal task layer … p.29-30: Datasets
..
**Tosh: Compare: Zrim, Зрим ~2013-2014:** {K-K}, междуконтекстна връзка, (intercontext connection, intercontext connector). “System 1” is not unconscious and intuitive has another meaning of нагледен (image-based) and accessible at once, see Schopenahuer and Todor Arnaudov’s *“hypothesis about the deeper*

*consciousness"* in Theory of Universe and Mind (e.g. Part III,2003; Part IV, 2004, and the notes in T.Arnaudov, „Какво му трябва на човек? Играеш ли по правилата ще загубиш играта!", 2014 – хипотеза за по-дълбокото съзнание. https://razumir.twenkid.com/kakvomu_notes.html#200
* "Схващане за всеобщата предопределеност 3" от 2003 г. - т.19

*** CommonGen: A Constrained Text Generation Challenge for Generative Commonsense Reasoning**, Bill Yuchen Lin♥ Wangchunshu Zhou♥ Ming Shen♥ Pei Zhou♥ Chandra Bhagavatula♠ Yejin Choi♠♦ Xiang Ren♥, 30.11.2020
https://arxiv.org/pdf/1911.03705
*Concept-Set: a collection of objects/actions: dog, frisbee, catch, throw; Generative Commonsense Reasoning … Expected Output: "John threw the frisbee to his dog." … Machines generate: UniLM: "Two dogs were throwing frisbee at each other …" , GPT2: "A dog throws a frisbee at a football player" … Concept-Set: { hand, sink, wash, soap } Transferring CommonGen Models*. P.17: other examples. *7.Conclusion: constrained text generation task for generative commonsense reasoning, with a large dataset; challenges of the proposed task, i.e., a) relational reasoning with latent commonsense knowledge, and b) compositional generalization*

* **EdgeFormer: A Parameter-Efficient Transformer for On-Device Seq2seq Generation**, Tao Ge, Si-Qing Chen, Furu Wei 2.2022/29. 12.2022
https://github.com/microsoft/unilm/tree/master/edgelm
https://arxiv.org/pdf/2202.07959
a deep encoder and shallow decoder, but with an interleaved decoder with shared lightweight feedforward network … *cost-effective parameterization: 1) encoder-favored parameterization parameterize the encoder using as many parameters as possible; 2) load-balanced parameterization that suggests we balance the load of model parameters to avoid them being either underused or overused in a NN with shared parameterization* … layer adaptation; fewer than 10 million model parameters; int8-quantized; *seq2seq generation within around 100ms latency (20-30 sequence length on average*) using two midto-high-end CPU cores and less than 50MB RAM; Shared Weights $W \in \mathbb{R}^{d \times d}$ … 2 layer decoder (one-layer – bad performance); *3 Constraints for On-device Seq2seq: Computation On-device computer vision (CV) models tend to use 1G FLOPS (0.5G MACS) as a constraint …relaxed for typical seq2seq tasks to 2G FLOPS (1G MACS)* (a higher latency of 100 ms is acceptable); deep encoder and shallow decoder …

* **Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity**, William Fedus, Barret Zoph, and Noam Shazeer. 2021/2022., Google. https://jmlr.org/papers/volume23/21-0998/21-0998.pdf |

https://arxiv.org/pdf/2101.03961 Mixture-of Expert models. *Sparse training; increase the parameter count while keeping the floating point operations (FLOPs) per example constant. Mixture of Expert Routing: input a token representation x and then routes this to the best determined top-k experts, selected from a set {Ei(x)} N i=1 of N experts; gate-value for expert i*

* Robert A Jacobs, Michael I Jordan, Steven J Nowlan, and Geoffrey E Hinton. Adaptive mixtures of local experts. Neural computation, 3(1):79–87, 1991
* Michael I Jordan and Robert A Jacobs. Hierarchical mixtures of experts and the em algorithm. Neural computation, 6(2):181–214, 1994
* Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-ofexperts layer. arXiv preprint arXiv:1701.06538, 2017

**Learning to Ask: Neural Question Generation for Reading Comprehension,** Xinya Du, Junru Shao, Claire Cardie 29.4.2017 https://arxiv.org/pdf/1705.00106

* **The CoNLL-2014 Shared Task on Grammatical Error Correction**, Hwee Tou Ng1 Siew Mei Wu2 Ted Briscoe3, Christian Hadiwinoto1 Raymond Hendy Susanto1 Christopher Bryant1, https://aclanthology.org/W14-1701.pdf

* **DeepNet: Scaling Transformers to 1,000 Layers,** Hongyu Wang, Shuming Ma, Li Dong, Shaohan Huang, Dongdong Zhang, Furu Wei 1.3.2022 https://arxiv.org/pdf/2203.00555 DeepNorm normalization; Instability of Deep Transformer

@Vsy: obrb(?=:tasks,benchmarks:sbr(zglv)…) **GPT-3**

**Language models are few-shot learners.** Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei (31). 22.7.2020 https://arxiv.org/pdf/2005.14165 – **GPT3** *Recent work has demonstrated substantial gains on many NLP tasks and benchmarks by pre-training on a large corpus of text followed by fine-tuning on a specific task. Introduction: Recent years … a trend towards pre-trained language representations in NLP systems, applied in increasingly flexible and task-agnostic ways for downstream transfer. First, single-layer representations were learned using word vectors [MCCD13, PSM14] and fed to task-specific architectures, then RNNs with multiple layers of representations and contextual state were used to form*

*stronger representations [DL15, MBXS17, PNZtY18] (though still applied to task-specific architectures), and more recently pre-trained recurrent or transformer language models [VSP+17] have been directly fine-tuned, entirely removing the need for task-specific architectures [RNSS18, DCLT18, HR18 … still a need for task-specific datasets and task-specific fine-tuning.* ***p.3 "in-context learning", using the text input of a pretrained language model as a form of task specification: the model is conditioned on a natural language instruction and/or a few demonstrations of the task and is then expected to complete further instances of the task simply by predicting what comes next.*** … size of transformer models'sizes*: from 100 million parameters [RNSS18], to 300 million parameters [DCLT18], to 1.5 billion parameters [RWC+19], to 8 billion parameters [SPP+19], 11 billion parameters [RSR+19], and finally 17 billion parameters [Tur20].*

*[RWC+19] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever.* ***Language models are unsupervised multitask learners****, 2019 – GPT2 in-context learning*

***** A Comprehensive Survey on Pretrained Foundation Models A History from BERT to ChatGPT,** Ce Zhou1* Qian Li2∗ Chen Li2∗**,** Jun Yu3∗ Yixin Liu3∗ Guangjing Wang1, Kai Zhang3, Cheng Ji2, Qiben Yan1, Lifang He3, Hao Peng2,
Jianxin Li2, Jia Wu4, Ziwei Liu5, Pengtao Xie6, Caiming Xiong7, Jian Pei8**,** Philip S. Yu9, Lichao Sun3, 5.2023 https://arxiv.org/pdf/2302.09419 1Michigan State University, 2Beihang University, 3Lehigh University, 4Macquarie University, 5Nanyang Technological University, 6University of California San Diego, 7Salesforce AI Research,8Duke University, 9University of Illinois at Chicago, 99 pages … p.42:
*8.5 Open Problems for Future PFMs: .. Third,* ***a unified PFM*** *is expected to achieve SOTA transfer performance for all different tasks in all data domains, including text, image, graph, and multimodalities.*
**Tosh:** Compare with: Todor Arnaudov, 2003**: "The Creativity is Immitation at the level of algorithms**", The Sacred Computer. Part of "The Bulgarian Prophecies" (or Bulgarian Predictions).
https://www.oocities.org/eimworld/eim22n/eim23/emil04052003.htm
Conclusion: *"**За да се стигне до Целта - разум, сравним с човешкия, различните видове творци трябва да започнат да преливат един в друг и постепенно да се слеят напълно.**" … In order to reach to the Goal – a general intelligence (разум, Reason), comparable with the human intelligence, all different kinds of creators have to begin to spill into each other and incrementally to merge completely."*
p.43…: A.1 Basic Components on NLP & graph G; G: node-type mapping; {(unattributed, attributed)(undirected, directed) G; (homogeneous,heterogeneous)}

G (edges, nodes) belong to a type; attributed G. Labelled (Ground Truth), unlabelled, pseudo labels … pretext tasks (e.g., clustering, completion, and partition) for mining self-supervising information (pseudo labels) … p.46 **Inductive Learning** .. *trains the model on labeled data and then tests on samples that have never appeared in the training stage.* ***Transductive learning:*** *all samples are visible during both the training and testing stages* p.48 *B.1 Traditional Text Learning Word representations .. Continuous Bag Of Words (CBOW) [12] in word2vec,* … **Images:** CNN, RNN, Generation-based: GANs, Markovian GAN, CycleGAN, StyleGAN. *Pixel Recurrent Neural Networks (PixelRNN) [326] aims to complete images by modeling full dependencies between the color channels. DiscoGAN [327] is designed to learn relations between different domains.* Laplacian Pyramid of Adversarial Networks (LAPGAN), Stacked GAN (SGAN). p.49: B.2.4 Attention-Based Networks: SENet – *first place in the competition of ILSVRC2017. CBAM sequentially infers attention maps along both channel and spatial dimensions.*
– **Timeline:** *1990s fast weight controller. Neuron weights generate fast "dynamic links" similar to keys & values.* ***2014:***
*RNN + Attention. Attention network was grafted onto RNN encoder decoder to improve language translation of long sentences.* ***2015:*** *Attention applied to images*;* ***2017: Transformers = Attention + position encoding + MLP + skip connections.*** *This design improved accuracy and removed the sequential disadvantages of the RNN. .. The major breakthrough came with self-attention, where* ***each element in the input sequence attends to all others, enabling the model to capture global dependencies.***

**See also the survey:**
*** GPT (Generative Pre-trained Transformer) – A Comprehensive Review on Enabling Technologies, Potential Applications, Emerging Challenges, and Future Directions.** Gokul Yenduri, Ramalingam M, Chemmalar Selvi G et al. 21.5.2023 https://arxiv.org/pdf/2305.10435
* https://en.wikipedia.org/wiki/Attention_(machine_learning)

* Vinyals, Oriol; **Toshev, Alexander**; Bengio, Samy; Erhan, Dumitru (2015). "Show and Tell: A Neural Image Caption Generator". pp. 3156–3164.
**Tosh:** “Global” only at the level, scale, range of the context that is processed at once.; *Alexander Toshev* is a prominent Bulgaria-born researcher – see **#Anelia** and the main volume #prophets for reviews of his work.
… References, PFM, RL World-models …
* D. Ha and J. Schmidhuber, “World Models,” Mar. 2018.
* D. Hafner, T. Lillicrap, J. Ba, and M. Norouzi, “Dream to control: Learning behaviors by latent imagination,” arXiv preprint arXiv:1912.01603, 2019.
* D. Hafner, T. Lillicrap, M. Norouzi, and J. Ba, “Mastering atari with discrete world models,” arXiv preprint arXiv:2010.02193, 2020.

* F. Deng, I. Jang, and S. Ahn, “Dreamerpro: Reconstruction-free model-based reinforcement learning with prototypical representations,” in International Conference on Machine Learning, pp. 4956– 4975, PMLR, 2022.
* P. Wu, A. Escontrela, D. Hafner, K. Goldberg, and P. Abbeel, “Daydreamer: World models for physical robot learning,” arXiv preprint arXiv:2206.14176, 28.6.2022. University of California, Berkeley https://arxiv.org/abs/2206.14176
*Learning a world model to predict the outcomes of potential actions enables planning in imagination, reducing the amount of trial and error needed in the real environmen. ,, pick and place multiple objects directly from camera images and sparse rewards, approaching human performance.On a wheeled robot, Dreamer learns to navigate to a goal position purely from camera images, automatically resolving ambiguity about the robot orientation. Using the same hyperparameters across all experiments, we find that Dreamer is capable of online learning in the real world, establishing a strong baseline. sample-efficient robot learning: learn robot locomotion, manipulation, and navigation tasks from scratch* ***in the real world*** *on 4 robots****, without simulators****. .. The* ***encoder fuses all sensory modalities into discrete codes****. The* ***decoder reconstructions*** *the inputs from the codes, providing a rich learning signal and enabling human inspection of model predictions. A* ***recurrent state-space model (RSSM)*** *is trained to* ***predict future codes given actions****, without observing intermediate inputs,*

* **What are the main differences between hard attention and soft attention, and how does each approach influence the training and performance of neural networks?**
* https://eitca.org/artificial-intelligence/eitc-ai-adl-advanced-deep-learning/attention-and-memory/attention-and-memory-in-deep-learning/examination-review-attention-and-memory-in-deep-learning/what-are-the-main-differences-between-hard-attention-and-soft-attention-and-how-does-each-approach-influence-the-training-and-performance-of-neural-networks/

**What is Attention in ML?**
Soft-attention:Differentiable, continuous, can be learned end-to-end by the general purpose ML techniques, DL, gradient descent. “Soft-max” function. Hard-attention: only selected items from the input are relevant, discrete, non differentiable, harder to train. A summary with the assistance of **Google Gemini Deep Research, Claude 3.7 Sonnet + edited by Tosh**. **4.5.2025:**

| **Attention Mechanism** | **Key Advantages, Disadvantages and Properties** |
|---|---|
| Soft Attention | Differentiable, allows end-to-end training, captures long-range dependencies to some extent, widely applicable; weighted averages across all input elements, more stable training, but computationally expensive for long sequences (quadratic etc.) |

| | |
|---|---|
| Hard Attention | Discrete choices about which elements to focus on. More interpretable, but not directly differentiable. Potentially more computationally efficient at inference time, but can have high variance during training. Used e.g. in image captioning (selection of regions to focus on). |
| Self-Attention | Excellent at capturing intra-sequence dependencies, parallel processing capability, global context awareness, disambiguates polysemous words. |
| Multi-Head Attention | Runs multiple attention mechanisms in parallel – „heads“; each head learns different relationship patterns. Outputs are concatenated and linearly transformed; jointly attend to information from different representation subspaces, thus capturing different types of relationships. |
| Scaled Dot-Product Attention | Dot-product between queries and keys. Matrix multiplication optimization. Used in the original transformer paper, Vaswani et al., 2017. Attention(Q,K,V) = softmax(QK^T/√d_k)V. Efficient, stable gradients during training. |
| Additive Attention | Handles inputs with different dimensions. |
| Location-Based Attention | Attention weights are computed based on the position/location rather than content; often uses position embeddings or relative position information. Useful when the positional information is more important than content, e.g. in speech recognition systems. |
| Global Attention | All output position attends to all input positions. Most effective when all context is potentially relevant .Captures long-range dependencies and overall context for shorter sequences; expensive for long sequences.. |
| Local Attention | Only attends to a window of positions near the target position. More computationally efficient for long sequences. Useful when nearby context is most relevant. Examples: Longformer, Performer models for long document processing.<br>* **Longformer: The Long-Document Transformer** Iz Beltagy, Matthew E. Peters, Arman Cohan, 2020 – *a self-attention operation that scales linearly with the sequence length, making it versatile for processing long document.* See fig. 2, p 3: comparison of full self-attention patterna and Longformer’s attention patterns (a square of cells, full vs partial coverage). |

| | |
|---|---|
| Cross-Attention | Enables information aggregation and alignment between two different sequences.(e.g. seq-to-seq, machine translation) |

[2.11.2025]
* Kimi Linear: An Expressive, Efficient Attention Architecture
Kimi Team: Yu Zhang, Zongyu Lin, et al. 30.10.2025 https://arxiv.org/pdf/2510.26692
**– hybrid linear** attention … 48B model, 3B activated parameters … up to 6 times decoding throughput for a 1M context https://github.com/MoonshotAI/Kimi-Linear
Compar to **full attention.**
– Linear attentions are techniques aimed at reducing the quadratic raise in the computational cost with the length of the context.

* Anthropic Research Shows How LLMs Perceive Text, 30.10.2025
https://www.searchenginejournal.com/anthropic-research-shows-how-llms-perceive-text/559636/ – tokens/characters counting, line breaks, …

*** Neural Machine Translation by Jointly Learning to Align and Translate** , Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio, **2014/2016**, https://arxiv.org/abs/1409.0473 *In this paper, we conjecture that the use of a fixed-length vector is a bottleneck in improving the performance of this basic encoder-decoder architecture, and propose to extend this by allowing a model to automatically (soft-)search for parts of a source sentence that are relevant to predicting a target word, without having to form these parts as a hard segment explicitly.* **Attention**. .. In **traditional phrase-based** translation systems* – *many small sub-components, tuned separately* – while *neural machine translation attempts to build and train a single, large neural network that reads a sentence and outputs a correct translation.* Encoder–decoders, Sutskever etc. … but → sentence to a fixed-size vector, while in the new approach: the sentence is encoded as a sequence of vectors and adaptively a subset of the sequence is adaptively selected while decoding the translation. *(see, e.g., Koehn et al., 2003)*

*** Effective Approaches to Attention-based Neural Machine Translation,** Minh-Thang Luong, Hieu Pham, Christopher D. Manning, **2015**
https://arxiv.org/abs/1508.04025 *An attentional mechanism has lately been used to improve neural machine translation (NMT) by selectively focusing on parts of the source sentence during translation.* They propose: *a global approach which always attends to all source words and a local one that only looks at a subset of source words at a time. We demonstrate the effectiveness of both approaches over the WMT translation tasks between English and German in both directions. … Global align weights .. LSTM .. a* ***variable-length alignment vector*** *.. whose size equals the number of time steps on the source side, is derived by comparing* ***the current target hidden state*** *ht with* ***each source hidden state*** *hs: …* Introduction of **local attention** vs the global attention, which *attends to all words on the source side for each target word - .. potentially impractical to translate longer sequences, e.g.,*

*paragraphs or documents.* The local attentional mechanism *focus only on a small subset of the source positions per target word.*

[**Tosh:** they still translate *words* (tokens of that kind), but language is not just words. Part of the deeper structure is encoded in their relations and the "textual-multi-modality/multi-domainness" – in a diverse corpus text covers a lot of domains, programming code, science, logic etc. which translates to "more than words".]

.. the tradeoffs(soft,hard) attentional models proposed by Xu et al. (2015) for image captioning. Hard attention – not differentiable, but less expensive in inference time. The context vector is a *weighted average over all the source hidden states.* Attention layer: a location-based function: a[t] = softmax(W[a]*h[t]); score – a content-based function: dot product, general, concat … local weights, aligned position $\mathbf{p_t}$ … (monotonic, predictive) alignment; Gregor et al. 2015 – *selective attention mechanism* for image generation: their approach selects an image patch of varying location and zoom.

* **DRAW: A recurrent neural network for image generation.** [Gregor et al.**2015**] Karol Gregor, Ivo Danihelka, Alex Graves, Danilo Jimenez Rezende, and Daan Wierstra. 20.5.2015. In ICML. https://arxiv.org/pdf/1502.04623 – *a novel spatial attention mechanism that mimics the foveation of the human eye, with a sequential variational auto-encoding framework that allows for the iterative construction of complex images.* A substantial improvement on SOTA for MNIST; on ***Street View House Numbers*** *dataset it generates images that cannot be distinguished from real data with the naked eye***.** (SVHN) **..** *A person asked to draw, paint or otherwise recreate a visual scene will naturally do so in a* ***sequential, iterative fashion****, reassessing their handiwork after each modification. Rough outlines are gradually replaced by precise forms, lines are sharpened, darkened or erased, shapes are altered, and the final picture emerges. Most approaches to automatic image generation, however, aim to generate entire scenes at once. In the context of generative neural networks, this typically means that all the pixels are conditioned on a single latent distribution (Dayan et al., 1995; Hinton & Salakhutdinov, 2006; Larochelle & Murray, 2011).* ***Deep Recurrent Attentive Writer (DRAW) – ..*** *a shift towards a more natural form of image construction, in which parts of a scene are created independently from others, and approximate sketches are successively refined.* RNN encoder-decoder, *it belongs to the family of variational auto-encoders, a recently emerged hybrid of deep learning and variational inference.* ***P.4. 3.2. Selective Attention Model ..*** *To endow the network with selective attention without sacrificing the benefits of gradient descent training* ***.. N ×N grid of Gaussian filters ..*** e.g. 3x3 *SVHN: training set contains 231,053 images, and the validation set contains 4,701 images.*

* *Netzer, Yuval, Wang, Tao, Coates, Adam, Bissacco, Alessandro, Wu, Bo, and Ng,*

*Andrew Y.* ***Reading digits in natural images with unsupervised feature learning****. 2011.* – SVHN dataset.

*** Memory Networks,** Jason Weston, Sumit Chopra, Antoine Bordes, 2014/2015, Facebook AI Research.. *a new class of learning models .. Memory networks reason with inference components combined with a long-term memory component; they learn how to use these jointly. The long-term memory can be read and written to, with the goal of using it for prediction. .. question answering (QA) .. where the long-term memory effectively acts as a (dynamic) knowledge base, and the output is a textual response. .. a large-scale QA task, and a smaller, but more complex, toy task generated from a simulated world. In the latter: .. chaining multiple supporting sentences to answer questions that require understanding the intension of verbs.* https://arxiv.org/abs/1410.3916

*** Attention in transformers, step-by-step | DL6** 3Blue1Brown 7,25 млн. абонати https://www.youtube.com/watch?v=eMlx5fFNoYc dd

* The COT COLLECTION: **Improving Zero-shot and Few-shot Learning of Language Models via Chain-of-Thought Fine-Tuning**, Seungone Kim et al., KAIST AI1 NAVER AI Lab2 University of Washington3 https://arxiv.org/pdf/2305.14045 ***1.84 million*** *rationales across 1,060 tasks … "CoT prompting works effectively for large LMs with more than 100 billion parameters" 'Wei et al. (2022b) propose Chain of Thought (CoT) Prompting, a technique that triggers the model to generate a rationale before the answer." improved reasoning abilities. "Let's think step by step". Improving Few-Shot Learning "Flan-T5 ExQA', 'Flan-T5 Arithmetic', 'Flan-T5 MCQA', 'Flan-T5 NLI' from the red box), we generate ∼51.29 times more rationales (1.84 million rationales) and ∼117.78 times more task variants (1,060 tasks).*

| |
|---|
| **Тош:** Сравни „Chain of thought reasoning", Верига на мисълта, с понятието и оператора за търсене в Зрим (**Zrim**), **„верига от частични съвпадения", пт(мс) и пс(мс) и др. от началото на 2010-те.**<br>Виж в *„Сътворение: Създаване на мислеща машина".* |

…

* **Scaling Instrution-Finetuned Language Model**, Hyung Won Chung, Le Hou et al., https://arxiv.org/abs/2210.11416 ,10.2022 - 12.2022, Text-to-Text FLAN … Flan-PaLM 5-shot. p.3. **Tasks: T0-SF:** Commonsense reasoning, Question generation, Closed-book QA, Adversarial QA, Extractive QA, Title/context generation,Topic classification, Struct-to-text: 55 Datasets, 14 Categories, 193 Tasks; **Muffin:** Natural language inference Closed-book QA, Code instruction gen., Conversational QA, Program synthesis, Code repair, Dialog context generation … 69 Datasets, 27 Categories, 80 Tasks**; CoT (Reasoning):** Arithmetic reasoning Explanation generation, Commonsense Reasoning Sentence composition**,** Implicit reasoning … 9 Datasets, 1 Category, 9 Tasks (…); Natural Instructions v2: Cause effect classification, Commonsense reasoning, Named entity recognition, Toxic language

detection, Question answering, Question generation, Program execution, Text categorization … 372 Datasets, 108 Categories, 1554 Tasks

**Todor, 21.4.2025:** Yet again too many well prepared, already "cooked" tasks, which are not evolving or incremental and are given in text form. The system *already "knows all*" from the start. Compare to Todor's suggestions in 4.2001 and 12.2001 about Seed AI (зародиш на разум) – the first about DeepBlue and the seminal paper *"Man and Thinking Machine: Analysis of the Possibility …"* and the 2009 interview: *"I will Create a Thinking Machine that will Self-Complexify: …"*

*"Held-out tasks MMLU Abstract algebra, Sociology, College medicine, Philosophy, Professional law … 57 tasks;* ***BB:*** *Boolean expressions, Navigate, Tracking shuffled objects, Word sorting, Dyck languages … 27 tasks;* ***TyDiQA Information seeking****: QA 8 languages* ***MGSM:*** *Grade school math problems 10 languages …*

*** Large Language Model Instruction Following: A Survey of Progresses and Challenges,** Renze Lou♠ Kai Zhang♢ and Wenpeng Yin♠: 25.5.2024, ♠The Pennsylvania State University ♢The Ohio State University https://arxiv.org/pdf/2303.10475 In-context Instructions/soft instr.; Instruction fine-tuning; instr. Engineering … diversity; (few shot) demonstrations: in-out examples; unlabeled - clustering; Order of demonstratio; Reasoning step annotation. Input, Output, Template; NLI-oriented (inference). LLM-oriented Instructions vs Human-oriented Instructions; Indirect Supervision; Semantic Parsers: *a game instruction "Move any top card to an empty free cell" → formula: "card(x) ∧ freecell(y)". Previous research spent extensive efforts on this strategy …* Reinforcement Learning from Human Feedback (RLHF), prediction shift penalty, ralignment; human-annotated datasets/LLM-synthetic data. Different evaluation schemes: Automatic metrics/Human evaluation/Task-centric evaluation/Human-centric evaluation. Recipes for Instruction Following: (…) you can use LLMs for diversifications (e.g. rewording of phrases etc.). Choose carefully the few-shot examples. Instruction consistency. Instruction diversity; Add demonstrations or not; The selection of demonstrations; The order of demonstrations. Reasoning steps augmentation. Emphasizing input-output mappings Model-Instruction Alignment. Training objective. *LLM-oriented Instructions (i.e., prompt) can work is that prompt aligns well with the pretraining objective—language modeling—and activates the task-specific knowledge of the LLMs … using soft instructions (i.e., continuous embedding) instead of human-understandable discrete instructions; ? convert the original human-oriented instructions into a LLM-oriented style or not?* Interpretability. Data-wise Factor: Task Scale: *the profits of the task scale are highly governed by the model scale.* Dual-Track Scaling. ..*Prior to LLM-based instruction following, scaling was mainly for deep learning models: from single-layer neural nets to multi-layer perceptrons*, from convolutional/recurrent neural networks to deep-layer transformers … Along with the pretraining of massive raw*

*text data, the ever-increasing models are expected to have encoded a vast amount of general-purpose knowledge (Zhou et al., 2023a*). *Using prefix prompts for the auto-regressive LMs, while using cloze prompts for the masked LMs* (Liu et al., 2023a) *and more training tasks and labeled examples for each.*

[**Tosh:** E.g. “The capital city of Bulgaria is: …” (prefix), “The color of the sky is: …” (prefix) vs “She opened the [MASK] and smelt the content”. ([MASK] could be a jar, a box, …); “multi-layer *perceptrons*” – in the paper “*perceptions*”; general-purpose knowledgge – in the paper “generic purpose” – perhaps errors; MLPs have at least 3 layers]

The Tax of Instruction Alignment: *The instruction following aims at taming the models to better assist humans in real-world tasks; therefore, in addition to pursuing ultimate performance, inference-time safety is also a crucial aspect for the instruction-tuned models (i.e., instruction alignment). Ouyang et al. (2022) defined “alignment” with three criteria — Helpful, Honest, and Harmless (HHH)*; Learning Negated Information: things-to-avoid. Adversarial Instruction Attacks. Explainability of Instruction Following. Instruction-related Aapplications: . Human-Computer Interaction.Data and Feature Augmentation: *Task instructions are regarded as indirect supervision resources where sometimes superficial and assertive rules are embedded. .. Generalist Language Models: According to the definition of Artificial General Intelligence (AGI), the “generalist model” is usually a system that can be competent for different tasks and scalable in changeable contexts, which shall go far beyond the initial anticipations of its creators (Wang and Goertzel, 2007; Goertzel, 2014). .. the recent remarkable applications of instructions, namely InstructGPT, ChatGPT, and GPT-4, also indicated a big step towards building generalist language models…* p.32 List of datasets …

* **Training language models to follow instructions with human feedback**, Long Ouyang, Jeffrey Wu, Xu Jiang et al 2022.. In Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022.

* **TencentPretrain: A Scalable and Flexible Toolkit for Pre-training Models of Different Modalities**, Zhe Zhao1*, Yudong Li2 et al 7. 2023 https://arxiv.org/pdf/2212.06385 – trend towards homogeneity in pre-training models; **5 components:** embedding, encoder, target embedding, decoder, and target. Modular system for designing models. https://github.com/Tencent/TencentPretrain
https://github.com/Tencent/TencentPretrain/wiki/Preprocess-the-data
https://github.com/Tencent/TencentPretrain/wiki/Pretraining-model-examples
python3 preprocess.py … python3 pretrain.py
(…)

*** Qwen3-Coder: Agentic Coding in the World** https://github.com/QwenLM/Qwen3-Coder Qwen3-Coder-480B-A35B-Instruct

*** Language models are injectiveand hence invertible**, Giorgos Nikolaou, Tommaso Mencattini et al. 21.10.2025 https://www.arxiv.org/pdf/2510.15511 – Recover the prompt from the output. "*Transformer components such as non-linear activations and normalization are inherently non-injective, suggesting that different inputs could map to the same output and prevent exact recovery of the input from a model's representations. In this paper, we challenge this view .*"

## * Някои тенденции – бележки на Тош

*** „Living intelligence"**[175], жива интелигентност – увеличена употреба на множеството сензори на всевъзможните мобилни устройства за събиране на повече данни за обучение на ИИ; данните като „двигателя на всичко" (everything engine) – сравни с А.Шопенхауер, че *нагледът* е източникът на познанието. „Данните" са именно *нагледът.* Също така „жива" във връзка с „киборзите" като органоидните мозъци – DishBrain, играещ Pong и пораждащи модели за биотехнологии, биология и пр. Преди едно-две десетилетия се наричаше още „вездесъща информатика",, „ubiquitous computing". За подобна функция споменават в бъдещо мобилно устройство на OpenAI[176]. Обединението на все повече източници на данни във все по-реално време и пр. и обработка на всевъзможни данни, в това приложение – възприятията и знанията на потребителя – следва принципите, зададени в ТРИВ 2001-2004 и последвани в развитието на машинното обучение; в тази посока е и концепцията и на *Ускорителя на изследователската дейност* на Тош – Research Assistant/Assistant/ACS още от 2007 г., всеборавител, „Вседържец" и пр. и в литературата на Свещеният сметач се нарича още *„Изпреварващо търсене"*. Промяната и обучението трябва да стават и незабавно, а не само след натрупване и отложено накуп както при „фазите на съня" (виж Replay in brains and machines). Засега Research Assistant, Assistant C#, или още ACS, заедно със съвкупност от помощни приложения сателити, не е обнародвана и се използва само от автора си. Някои от допълнителните приложения са публикувани отдавна, например синтезаторът на реч „Тошко 2". След като приключа с „летописите" следващите версии трябва да се напишат, развият, открият и съединят сами или почти сами от първичен зародиш. Виж първите публикации от ТРИВ, 2001 г. – Зародиш на разум.

*** Как щяла да изглежда кариерата на компютърните специалисти след 5 години:** https://www.cio.com/article/4066676/what-an-it-career-will-look-like-in-5-years-and-how-to-thrive-through-the-changes.html – „The rise of the generalist and interdisciplinary IT" – **Възходът на общи и интердисциплинарни специалисти**.
Също от български медии и „експерти":
*** Тесните специалности са обречени: какво искат индустриите 4.0/5.0**
проф. д-р Миглена Темелкова, ректор на Висшето училище по телекомуникации и пощи, в „Talk 25: Фантастика и реалност", 03.11.2025 – https://www.bloombergtv.bg/a/15-shows/151801-tesnite-spetsialnosti-sa-obrecheni-kakvo-iskat-industriite-4050 **„Индустрии 4.0/5.0 и нуждата от интердисциплинарни кадри. Индустриите вече не търсят тесни специалисти, а широко профилирани кадри, подготвени в няколко научни**

[175] https://hbr.org/2025/01/why-living-intelligence-is-the-next-big-thing
Why "Living Intelligence" Is the Next Big Thing, Amy Webb, 6.1.2025
[176] https://it.dir.bg/tehnologii/kakvo-da-ochakvame-ot-taynstvenata-dzhadzha-na-openai

**области**. Тя допълни, че още по-комплексни стават изискванията към инженерите, които трябва да обезпечат почти всички сектори – медицина, транспорт, комуникации, архитектура. Бъдещето, отбеляза събеседникът, е интердисциплинарно и налага многопрофилна подготовка."

..

**Тодор Арнаудов:** Още закъснели **тесни** специалисти, които преоткриват предвижданията и препоръките на Свещеният сметач **с 25-30 години закъснение** и догонват събитията**.** Представата за всестранност на „бизнес автора" от първата англоезична статия обаче е ограничена и банална, както и може да се очаква: да мислели и *като „мениджъри на продукт и бизнес анализатори и да общуват като стратези",* трябвало да разбират от „данни" и да „говорят гладко езика на ИИ" и т.н.

Може би „ИТ специалистите" **нямат 5-годишен хоризонт** в начина по който си го представят т.нар. експерти. „**Продуктите" и работата са отживелица.**[177]

## * Глаголищата трябва да бъдат живи – Програмите да се пишат сами * Software have to self-develop – Programs can write themselves

При други обстоятелства **програмите можеше да се разработват и развиват сами като живи организми,** да се написват и пренаписват сами от нулата за всички необходими процесори, операционни системи, платформи, започвайки още в края на 2000-те или най-късно началото-средата на 2010-те – по друг начин от това, което се прави сега, без "машинопис", без да е нужно да се съберат и „преточат" „всички" написани програми с коментари и специално разработени набори от данни; особено за разработка на вече съществуващи или ясно определени неща – каквото е работата на повечето програмисти и затова и се автоматизира в индустриални мащаби – повторения на едно и също с малки вариации[178]. (…)

---

[177] „Отживелицата" в масовата разработка на софтуер и уеб страници на килограм според мен беше налице в голяма степен и в началото на 2010-те г. и тогава говорех за това колко е абсурдно програмите да се печатат като на пишеща машина, и че трябва да се *адресират*, вече са написани; системата успя да „се добута" до днес.

[178] Виж *„Първата съвременна стратегия за развитие чрез изкуствен интелект …",* Т.Арнаудов, 2025 – за периода след 1940-те при създаване на първите електронноизчислителни машини в различни страни и институции по целия свят, за „националните стратегии за ИИ" и др., и бележката в приложението към литературата – защо и докога даден вид работа не може да се автоматизира, въпреки че е повтаряща се? Една от причините е, защото и когато по-висшите управляващо-причиняващи устройства не успяват да получат достъп или не могат да въздействат на до действията на достатъчно ниско ниво в стълбицата от въображаеми вселени, до които обаче имат достъп местните дейци. Така е и например във военното дело – генералът не може да управлява пряко редника и др., мозъкът – атомите от мускулите и пр.: с.180, **[173.повтаряща-се-работа].18.3.2025: Откъде се поражда толкова много еднообразна и повтаряща се програмистка работа?** (и не само

Обикновеният естествен език (ОЕЕ) – без специална семантика, запис, допълнителна логика, обръщения към конкретни структури и преобразувания и пр. – е подходящ като основно средство за общуване тогава, когато насочва вече приблизително известни, познати и общи неща и действия, които системата може сама да доизясни, да разреши многозначността им (to disambiguate). ОЕЕ като протокол за управление е достатъчен тогава, когато не са важни фини подробности и не се посочват и адресират конкретни променливи, адреси, сложни формули и пр.; обикновеният естествен език също така е **отживелица и примитивен** в застоялата си първобитна говорима форма, както и в писмения запис, който по скучен и буквален начин повтаря говоримия, без да използва други отдавна налични възможности. Виж *„Сътворение: Създаване на мислеща машина…“* и други бъдещи работи, където ще продължа тази мисъл и ще я илюстрирам чрез Зрим, Вседържец, Всеборавител, Research Assistant и пр.

Относно саморазвиващите се програми, системи, агенти и пр. виж също школата на ПСЕ/ИЧД – интервюто с Берт Дефрейс (Bert de Vries) за MLST в том #Ирина на *Пророците на Мислещите Машини...* ~ с. 124 (изд.8.10.2025) препратките към първите две статии от Теория на Разума и Вселената от *Сметача* от 2001 г., в които се въвежда понятието за Зародиш на разум, и интервюто за сп. „Обекти“ от 2009 г.: *„Ще създам мислеща машина, която ще се самоусложнява“.*

*** Навлизане на големи модели за действия**: **Large Action Models - LAM**, каквито са агентите, които предвиждат и причиняват действия, „какво да се извърши след това“, а не извеждат само текст, който човек да тълкува. Такива сегашни модели: Calude, ACT-1, Adept.Ai.

*** Сравни с „What’s Wrong With NLP?** (…)”, T.Arnaudov, 2009, Part I, Part II. **PLAM** – **personal language action models** и **CLAM – Corporate LAM**. Обучен със и за личните или фирмени данни и задачи. Агентните и мулти-агентни системи с езикови модели, които се очаква през 2025 г. да се масовизират от OpenAI и Google: https://www.kaldata.com/it-новини/изкуствен-интелект/ии-агентите-следващата-стъпка-в-еволю-539909.html са вид PLAM, особено когато имат достъп до спомените на потребителя и действат от негово име.

Прототип на „PLAM” и „агент“ по проект е непубликувания (към 20.1.2025), замислен като продължение на „Smarty” през 2007 и в различни степени разработен от началото на 2010 г. *„Ускорител на изследователската дейност*“ на Тодор Арнаудов – Assistant, Research Assistant, ACS, в една от версиите си на Java: “Superhuman”, както ще бъде и *Вседържец, още Вси, Vsy, “Jack of All Trades”, “Master of All Arts”.* Сравни също с анонса на Елън Мъск от

програмистка). Виж също „*Stack Theory is yet another Fork of Theory of Universe and Mind”*, T.Arnaudov 2025.

1.2025 г. за тяхната „everything app”[179]. Агентите, които вече са масово обсъждана тема, с

Вид LAM е *Palmyra:* https://writer.com/blog/actions-with-palmyra-x-004/ който извършва действия чрез външни приложения като управление на програми и финансови операции като използва оръдия, инструменти (tool use). Системата е обучена със синтетични данни, породени от съществуващи езикови модели (EM)[180]. get_data(), analyze_data(), call_writer(), send_email() … - изрази, които се прихващат за да извикат други приложения през съответен приложен програмен интерфейс (API); Graph RAG, 128K контекст … Примерна диаграма на блокове от взаимодействия на агент с EM: https://thenewstack.io/stop-treating-your-llm-like-a-database/ (?T Overview of agent dependencies) – непрекъсната „осъзнатост“, възприемане на потоци от данни, разсъждение в обслова (контекста), самостоятелно взимане на решения – без намеса на потребителя; своевременно обновяване на данните, изместване към проактивен ИИ, който не чака подкана, а действа сам[181] и др. Виж “mixed initiative interaction”, “mixed-initiative“ …, interfaces, … agents – класически изследвания от края на 1990-те, приложение „Лазар” . #lazar, e.g.

*** Principles of mixed-initiative user interfaces, Eric Horvitz, 1999**
https://erichorvitz.com/chi99horvitz.pdf etc

* Стартъпът „Logic Start” на INSAIT: https://logicstar.ai/ предлага агенти за поправка на грешки и проверка в процеса на разработка на програмни продукти, верификация на породения с ИИ код … https://logicstar.ai/news/INSAIT/

* https://superhuman.com/ - Помощник за бързо писане и отговаряне на имейли
* https://www.descript.com/ - Пораждане и редактиране на видеоклипове с естествен език, с говор като на потребителя

* **MemOS: An Operating System for Memory-Augmented Generation (MAG) in Large Language Models,** Zhiyu Li, Shichao Song, Hanyu Wang, Simin Niu, Ding Chen, Jiawei Yang, Chenyang Xi, Huayi Lai, Jihao Zhao, Yezhaohui Wang, Junpeng Ren, Zehao Lin, Jiahao Huo, Tianyi Chen, Kai Chen, Kehang Li, Zhiqiang Yin,

[179] https://www.indiatoday.in/technology/news/story/what-is-the-everything-app-for-which-elon-musk-is-looking-to-hire-software-engineers-2666279-2025-01-17
[180] https://it.dir.bg/tehnologii/ilon-mask-e-saglasen-che-sme-izcherpali-dannite-za-obuchenie-na-izkustven-intelekt
[181] Доколкото си спомням, в предаването за Изкуствен интелект – пилотен епизод на поредицата „Красива наука“ от 2011 г. по БНТ, една от участниците, П.Боровска, спомена „проактивността“ и интелигентността на рояка като бъдещо направление (swarm). Мулти-агентните системи са в тази насока като виж по-горе, че те съвсем не са нова идея #multi-agent и теориите за тях се развиват интензивно поне от 1980-те с предвестници от 1970-те (Actor model). Виж още в статията в книгата и https://artificial-mind.blogspot.com/2011/04/agi-in-prime-time-on-bulgarian-national.html

Qingchen Yu, Bo Tang, Hongkang Yang, Zhi-Qin John Xu, Feiyu Xiong, 28.5.2025 https://arxiv.org/abs/2505.22101v1 * https://github.com/MemTensor/MemOS – Textual, Activation (Caches key-value pairs) and Parametric Memory (model adaptation - LoRA weights etc.)**;** *memory cube … p.3. "As AGI continues to evolve into increasingly complex systems characterized by multi-tasking, multi-role collaboration, and multi-modality, language models must move beyond merely "understanding the world"—they must also "accumulate experience," "retain memory," and "continuously evolve."* – note the use of the term "AGI". **Plaintext:** "*editability, shareability, and governance compatibility .. documents, knowledge graphs, and prompt templates; .. transformation pathways between memory types (e.g., Activation → Plaintext, Plaintext → Parametric*""; "*plan to extend the Memory Interchange Protocol (MIP)" p.5. Arch., Fig. 3*

*** Agentic Context Engineering (ACE): Self-Improving LLMs via Evolving Contexts, Not Fine-Tuning** By Asif Razzaq -October 10, 2025 https://www.marktechpost.com/2025/10/10/agentic-context-engineering-ace-self-improving-llms-via-evolving-contexts-not-fine-tuning /

*** Agentic Context Engineering: Evolving Contexts for Self-Improving Language Models**, Qizheng Zhang, Changran Hu et al. https://arxiv.org/abs/2510.04618

* https://www.cio.com/article/4080592/context-engineering-improving-ai-by-moving-beyond-the-prompt.html – *"Prompts set intent; context supplies situational awareness," …* **Todor:** What can be done, what else can be done, path of thought; how to continue etc. See Zrim. {К}, ?ВМдсп, ?КсП, пт(мс), пс(мс) … etc. 2014+ Свещеният сметач. …

* **Dynamic Reinforcement Learning for Actors.** Shibata, Katsunari.(2025). arXiv preprint arXiv:2502.10200. URL: https://arxiv.org/abs/2502.10200 *"..directly controls system dynamics, instead of the actor (action-generating neural network) outputs at each moment, bringing about a major qualitative shift in reinforcement learning (RL) from static to dynamic. … a local index called "sensitivity,"*

* **Continual Learning via Sparse Memory Finetuning**, Lin, Jessy; Zettlemoyer, Luke; Ghosh, Gargi; Yih, Wen-Tau; Markosyan, Aram; Berges, Vincent-Pierre; Oğuz, Barlas. (2025).. arXiv preprint arXiv:2510.15103. URL: https://arxiv.org/abs/2510.15103 – *"sparse memory finetuning, **leveraging memory layer models** (Berges et al., 2024), which are **sparsely updated** by design. By **updating only the memory slots that are highly activated** by a new piece of knowledge relative to usage on pretraining data, we reduce interference between new knowledge and the model's existing capabilities."*

* **The Markovian Thinker**. Aghajohari, Milad; Chitsaz, Kamran; Kazemnejad, Amirhossein; Chandar, Sarath; Sordoni, Alessandro; Courville, Aaron; Reddy, Siva. (2025**.** arXiv preprint arXiv:2510.06557. URL: https://arxiv.org/pdf/2510.06557
Long Chains of thought, LongCoT … Leaves residuals, "carryover" …
*"…decoupling thinking length from context size. .. linear compute with constant memory. .. Delethink, an RL environment that structures reasoning into fixed-size chunks. Within each chunk, the model thinks as usual; at the boundary, the environment resets the context and reinitializes the prompt with a short carryover. Through RL, the policy learns to write a textual state near the end of each chunk sufficient for seamless continuation of reasoning after reset."*

**Tosh:** Compare to **Zrim**, **Chain of partially overlapping matches** (верига от частични съвпадения) ~ 2014-2015.

## **2024-2025** г.

* **Graph Generative Pre-trained Transformer (G2PT): An Auto-Regressive Model Designed to Learn Graph Structures through Next-Token Prediction**
https://www.marktechpost.com/2025/01/05/graph-generative-pre-trained-transformer-g2pt-an-auto-regressive-model-designed-to-learn-graph-structures-through-next-token-prediction

*** Graph Generative Pre-trained Transformer,** Xiaohui Chen et al., 2025
https://www.arxiv.org/abs/2501.01073 adjacency matrix representations ...

**xLSTM** Sepp Hochreiter et al. – обновена версия на LSTM, подлежаща на паралелизиране и др. MLST, 12.2.2025: LSTM: The Comeback Story?
https://www.youtube.com/watch?v=8u2pW2zZLCs
*** xLSTM: Extended Long Short-Term Memory,** Maximilian Beck et al., ELLIS Unit, LIT AI Lab, Institute for Machine Learning, JKU Linz, Austria, NXAI Lab, Linz, Austria https://arxiv.org/pdf/2405.04517#page=19&zoom=100,110,493 exponential gating, scalar memory, scalar update, memory mixing; fully parallelizable with a matrix memory & a covariance update rule; memory cell; sLSTM – scalar memory, mLSTM – matrix memory; xLSTM blocks. sLSTM forward pass: cell state, normalizer, hidden, input gate, forget gate, output gate. Stabilizer state, stab.input gate, stab.forget gate. .. mLSTM is parallelizable analog to FlashAttention … Memory mixing solves *state tracking problems* (unlike transformers & state space models SSMs). Most related: RWKV, Retention. Exponential gating: $\mathbf{e}^{x}$ exp(x)… > (0,1), bigger range, but needs stabilization. Gating – how much information to retain, forget, pass forward; selection. Covariance Update Rule – key-value matrix memory. … Attempts to overcome the quadratic alg.complexity of the transformers with linear, various other attempts, p.6

Related Work … Linear Attention/Linear Transformer, Synthesizer (synthetic attention without token-token interaction which is computationally heavy); SSM, RNN (Deep Linear Recurrent Units, LRU), Hierarchically Gated Linear RNN (HGRN)

* B. Peng, E. Alcaide, Q. Anthony, et al. **RWKV: Reinventing RNNs for the transformer era**. ArXiv, 2305.13048, 2023. https://arxiv.org/abs/2305.13048 Receptance Weighted Key Value (RWKV) *"reconciling trade-offs between computational efficiency and model performance in sequence processing tasks", 14B model "RWKV's state (or context) can be leveraged for interpretability, predictability in sequence data, and safety. Manipulating the hidden state could also guide behavior and allow greater customizability through prompt tuning."* – linear increment for the generation time, 1000 tokens = 10 sec, transformers: 55-60 s. p.2 different transformers & time-space requirements: Transformer $O(T^2d)$, $O(T^2 + Td)$, Reformer O(TlogTd) …, Performer, Linear Transformers, AFT-full, AFT-local, MEGA; RWKV: O(Td), O(d) … *Transformers rely on attention mechanisms to capture relationships between all input and all output tokens:*
Attn(Q, K, V ) = softmax(QKᵀ)V, … FLOP per token for RWKV: 169 M, 12 layers, dim=768, Params=1.693 × 10^8, FLOP/token = 2.613 × 10^8 (261.3 M), 430 M, 24, 1024, .. 7.573E+8,
1.5 B ,24, 2048, 2.823 × 10E+9; 3 B, 32, 2560, 5.71E+9; .. 14 B, 40, 5120, 2.778E+10 (28 B). p.5. formulas for calculating the FLOPs per token per forward and backward pass. …

* https://www.nx-ai.com/ **AI4Simulation**: computational fluid dynamics (CFD) and discrete element methods (DEM) *"20x more efficient than leading models, making it the ideal solution for industries requiring large-scale, real-time processing"*

*** The Llama 4 herd: The beginning of a new era of natively multimodal AI innovation, April 5, 2025** https://ai.meta.com/blog/llama-4-multimodal-intelligence/ Llama 4, mixture-of-experts: The biggest: **Behemot:** up to 2T params, 288B active, 16 experts. **Maverick:** 400B/17B/128 experts.. **Scout:** 109B/17B/16 experts.
* Llama 4 Maverick: fits in a single NVIDIA H100 DGX; **Pre-training on 200 lang, >100 > 1B tokens;10x more multilingual tokens than Llama 3. Behemoth: FP8 and 32K GPUs, 390 TFLOPs/GPU. The overall data mixture for training > 30 trillion tokens.**

* **Agent S2: A Compositional Generalist-Specialist Framework for Computer Use Agents,** Saaket Agashe∗, Kyle Wong∗, Vincent Tu∗ , Jiachen Yang, Ang Li, Xin Eric Wang, Simular Research, 1.4.2025 https://arxiv.org/pdf/2504.00906 **GUI Localization and Computer Use Agents, Proactive Hierarchical Planning, Compositional Grounding**; vs **Monolithic Methods – Mixture-of-Grounding techniques. Benchmarks:** OSWorld (Xie et al., 2024) **369 real-world computer use tasks *across multiple categories: OS functions, Office applications (LibreOffice Calc, Impress, Writer), Daily use apps (Chrome, VLC Player, Thunderbird),***

***Professional software (VS Code and GIMP), and Workflow scenarios involving multiple applications.*** WindowsAgentArena. **Comparisons:** OpenAI CUA/Operator (OpenAI, 2025), Claude Computer Use (CCU) with 3.5-Sonnet and 3.7-Sonnet (Anthropic, 2024), and UI-TARS-72B-DPO (Qin et al., 2025).
**Navi Agent (Bonatti et al., 2024) GPT-4o + Aria-UI (Yang et al., 2024).** Agent S2 uses ***UI-TARS-72B-DPO as its visual grounding expert, Tesseract OCR (OCR, 2025) for textual grounding, and Universal Network Objects (UNO) (Unotools, 2025) for interface interaction*** **(LibreOffice, OpenOffice)** ***"Proactive planning enables self-correction and contextualization with new observations. … Agent S2 scales effectively with increased compute resources and processing steps.***
**Sample prompts/tasks: 1. Android:** "Go to the new contact screen and enter the following details: **…** First Name: **Georgi,** Last Name: **Ivanov**, Phone: **0999-123-JOJO,** Phone Label: **Fun**".
**2. WindowsArena: "***Create a shortcut on the Desktop for the folder named "**Games**" that is located in the Documents folder. Name the shortcut "**Contra - NES**".***" See the plans and the screenshots, in the appendix of the document.**
***3.*** **Editing, LibreOffice/OpenOffice text and spreadsheets***: „I think the last paragraph is redundant so I want to add strike-through on words in the last paragraph. Can you do this for me?**".** Add a new column named "**Difference**" and calculate **it** for each week by subtracting "**Net**" from "**Gross**".* **4. Ubuntu settings:**: *Could you set the 'Dim screen when inactive' to off in settings?**".*** **LLMs generate plans in several substeps and execute them, sometimes – replan. There is main manager M, which "***generates a plan for instruction I, breaking it into coherent subgoals***" with intermediate goals.**

**[Тош:** Виж от „Българските пророчества": интелигентна ОС и др. (непубликувани към 4.2025 г.) от началото на 2010-те г.;„behaviointrospective". ?В МдсП, ?К с П? „Контекст", избирател на контекст, адресиране на контекст, контекстен разпознавател, … и пр. Подканите и начинът за взаимодействия със сега успешните дейци на умни ОС са примитивни и погрешни. Направленията за разработка от 2004-2008 г. тук в *„Пророците на мислещата машина…"* и в по-краткия труд *„Първата стратегия за развитие чрез изкуствен интелект…", 2025*… Виж продължението на *„Пророците на мислещите машини": „Създаване на мислеща машина"* (раб. заглавие)**]**

*** Aria-UI: Visual Grounding for GUI Instructions,** Yuhao Yang, Yue Wang, Dongxu Li, Ziyang Luo, Bei Chen, Chao Huang, Junnan Li, Rhymes AI and University of Hong Kong **20.12.2024** https://github.com/AriaUI/Aria-UI **„**Open-sourced, Fast and Context-aware Action Grounding from GUI Instructions for GUI/Computer-use Agents" https://ariaui.github.io/ Versatile Grounding Instruction

Understanding**;** diverse planning agents**;** Context-aware Grounding: Aria-UI is a mixture-of-expert model with 3.9B activated parameters per token https://arxiv.org/pdf/2412.16256 https://huggingface.co/datasets/Aria-UI/Aria-UI_Data/tree/main/desktop aria_ui_desktop.json **…** screenshots.zip 2.22 GB **…**

*** Claude 3.5 Sonnet, Computer Use, 10.2024:**
https://www.theverge.com/2024/10/22/24276822/anthopic-claude-3-5-sonnet-computer-use-ai срвн ACS/Вседържец на Свещеният сметач
* Andrej Karpathy @karpathy “The most bullish AI capability I'm looking for is not whether it's able to solve PhD grade problem. It's whether you'd hire it as a junior intern….”
*„Най-яката способност на ИИ, която търся, не е дали е в състояние да решава проблеми с докторска степен. Въпросът е дали бихте го наели като младши стажант ( …)“* - дали бихте го наели за реална работа, дали се учи; виж ЧиММ, 2001, зародиш на разум и др. от ТРИВ. **Виж:** https://github.com/Twenkid/rhodope-gpt (РодопиГПТ – подобрена версия на примерен преобразител на А.К.)

*** Computer-use models** – Интелигентна операционна система;
Управление на компютъра през графичния потребителски интерфейс;
Управление на сметача през въоблика
* https://www.theverge.com/2024/10/22/24276822/anthopic-claude-3-5-sonnet-computer-use-ai
* **Microsot Copilot Vision:**
https://www.theverge.com/2024/10/1/24259187/microsoft-copilot-redesign-vision-voice-features-inflection-ai
* GPT4-agent

**3.10.2025:** Claude 4.5 … **Kimi-2** offers a free experimental computer-use tool.
…
*** Simular Pro**
https://www.theregister.com/2025/07/15/simular_ai_agent_reinforcement/
https://www.simular.ai/simular-pro *“the computer becomes a human-like thing which can...**book tickets for you, reserve tables, go shopping**." This agent would also have knowledge of the user's habits and preferences, stored locally in your computer, said Li. "This is the vision we're pushing for." A recent manifestation of that vision is Simular Pro, a $500/month computer use agent for macOS (Apple silicon) that's designed to automate desktop tasks.*

## * Intelligent OS, automated GUI interaction by The Sacred Computer

**Tosh:** I conceived this stuff in 2007 and was designing some in the early 2010s for my

“infrastructure”. (…) See the research directions* and future work.
* A “nonsense” application of the computer-use agents is when they are produced industrially and huge datasets are prepared with APIs etc. , there is an *organized effort* in automating these “knowledge workers tasks” etc. such as ones consisting of clicking on web pages, copy-pasting, selecting items from drop-down menus etc. – all these tasks could be automated just by developing appropriate APIs, web services, standards, which would make the need for seeing and knowing the GUI or using clicks etc. irrelevant.

## * Практика по LLM – ГЕМ, TTS, ASR; ML; машинно обучение, компютърно зрение и др. … #практика #LLM Practice #practice

*** Stanford Course on LLMs,** Autumn 2025: https://cme295.stanford.edu/syllabus/

*** The 5 FREE Must-Read Books for Every LLM Engineer,** K.Mehreen, 5.11.2025 https://www.kdnuggets.com/the-5-free-must-read-books-for-every-llm-engineer
**Very good resources:**
1. *** Foundations of Large Language Models,** Tong Xiao, Jingbo Zhu, 16.1.2025, https://arxiv.org/abs/2501.09223 "pre-training, generative models, prompting, alignment, and inference."
2. *** Speech and Language Processing (3rd ed. draft),** Dan Jurafsky and James H. Martin, August 24, 2025 release https://web.stanford.edu/~jurafsky/slp3/

* **The Mathematics of Machine Translation: Parameter Estimation**, Peter F. Brown et al., 1993 https://www.ee.columbia.edu/~sfchang/course/svia-F03/papers/brown-machine-translate-93.pdf - bilingual, machine-readable texts … alignments: Viterbi, neighboring … ; statistical translation; conditional probabilities; expectation maximization; deficienc; multyword cepts; vacancies; probabilities:( fertility, string length, distortion, translation); iterative improvement

*** Understanding Large Language Models,** Jenny Kunz, 2024, https://liu.diva-portal.org/smash/get/diva2:1848043/FULLTEXT01.pdf

*** Large Language Models in Cybersecurity: Threats, Exposure and Mitigation**, Andrei Kucharavy et al. 2024, https://link.springer.com/content/pdf/10.1007/978-3-031-54827-7.pdf

*** The 1 Billion Token Challenge: Finding the Perfect Pre-training Mix,** 3.11.2025, Asankhaya Sharma – Training GPT-2 on a Budget: 90%+ Performance with 1/10th the Data The 1 Billion Token Challenge_ Finding the Perfect Pre-training Mix.mhtml https://huggingface.co/blog/codelion/optimal-dataset-mixing

*** Accelerating Large-Scale Mixture-of-Experts Training in PyTorch**, Nov 06, 2025, By Hemil Desai et al. https://developer.nvidia.com/blog/accelerating-large-scale-mixture-of-experts-training-in-pytorch/ – NeMo automodel; Fully Sharded Data Parallelism (FSDP): Expert Parallelism (EP): Pipeline Parallelism (PP): Context Parallelism (CP) … > 200 TFLOPS/GPU for H100, DeepSeek V3 achieved 250 TFLOPS on 256 GPUs.
**Todor:** See the *batch size:* 256 for GPT-OSS 20B, 512 for GPT-OSS 120B, Qwen3-MoE 30B, Kimi K2, one variant of DeepSeek V3 671B, for number of GPUs = 8, 64, 256; up to batch_size = 8192 for another training setup of DeepSeek with 1024 GPUs. The batch size is the number of pairs which are fed for learning and it helps the models

to generalize, to discover the common patterns by “perceiving” and comparing many examples together.

*** Visualizing transformers and attention** | Talk for TNG Big Tech Day '24, Grant Sanderson, 96,5 хил. абонати 305 380 показвания 20.11.2024 г. https://www.youtube.com/watch?v=KJtZARuO3JY

*** Visualise Transformers**: https://bbycroft.net/llm – проследяване на действието на преобразители, от nano-gpt (85 хил. параметъра) до GPT2-SMALL, GPT-2 XL и GPT-3. Виж синия плъзгач вляво – с него напредвате през стъпките на обработка.

* https://towardsdatascience.com/an-interactive-guide-to-4-fundamental-computer-vision-tasks-using-transformers/

* https://deepnote.com/ - “Jupyter otebook for the AI era” …

*** Курс по големи езикови модели на Хъгинфейс:**
https://huggingface.co/blog/mlabonne/llm-course
Набори от данни: https://github.com/mlabonne/llm-datasets
https://github.com/EmbraceAGI/Awesome-AGI - Ресурси за агенти и др.

* **Model Context Protocol (MCP) - Протокол за обмяна на контексти за големи езикови модели** (25.11.2024) за свързване на асистент с ИИ с източници на данни и взаимодействие с външни системи: инструменти, бази данни, приложни програмни интерфейси, инструменти и др. чрез строго типизирани взаимлици (structured, typed interfaces)
https://www.anthropic.com/news/model-context-protocol
https://modelcontextprotocol.io/introduction
https://modelcontextprotocol.io/examples Brave Search MCP Server:
https://github.com/modelcontextprotocol/servers/tree/main/src/brave-search

* The Architectural Shift: AI Agents Become Execution Engines While Backends Retreat to Governance, Oct 29, 2025 https://www.infoq.com/news/2025/10/ai-agent-orchestration/ *Enterprise adoption of AI Agents* – (2025) AI assistants for every application → (2026) Task-specific agent applications → (2027) Collaborative AI agents within applications → (2028) AI agents ecosystems across multiple applications …

* Google AI Just Open-Sourced a MCP Toolbox to Let AI Agents Query Databases Safely and EfficientlyBy , July 7, 2025
https://www.marktechpost.com/2025/07/07/google-ai-just-open-sourced-a-mcp-toolbox-to-let-ai-agents-query-databases-safely-and-efficiently

* Вж също gpt4all, ollama: Библиотеки за използване на езикови модели, ускоряване не само чрез CUDA, но и други за видеокарти различни от Nvidia (Vulcan, Direct Compute, OpenCL) и пр.
* https://github.com/nomic-ai/gpt4all
* https://www.nomic.ai/gpt4all
* https://ollama.com/
* https://github.com/ollama/ollama

Ollama, sample commands: ollama pull gemma3:4b  (install a model)
Windows: set CUDA_VISIBLE_DEVICES='' (CPU only); =0  =1  =0,1  - choose CUDA devices (See mapping with >nvidia-smi  ); first close the Ollama server in order the changes of the visible devices to take effect – then apply.
> ollama run gemma3:4b –verbose;  gemma3:1b; qwen3:4b, qwen3:1.7b …
* Take into account that when the GPU memory of one or multiple GPUs is not enough to fit a model, that may reduce the performance to running on CPU only due to the excessive memory transfers and the GPU underutilization, say 20-25% etc. If the model is small enough, the performance could be 3 times higher with the GPU than the CPU only (that's a sample use case with i5-6500 and GTX 1060 3 GB).

* Qwen3-Embedding-0.6B https://huggingface.co/Qwen/Qwen3-Embedding-0.6B
For retrieval: encode documents and queries.

* LangExtract – извличане на структурни отношения от текст чрез Джемини, библиотека на Питон:
* https://developers.googleblog.com/en/introducing-langextract-a-gemini-powered-information-extraction-library/
* https://github.com/google/langextract

Прилагане на обработка върху откъси от текст и извличане, класифициране.

* Виж също **LangChain**: https://www.langchain.com/
* https://blog.langchain.com/introducing-open-swe-an-open-source-asynchronous-coding-agent/

*** Prompt Orchestrating Markup Language, Microsoft, 8.2025:**
https://www.marktechpost.com/2025/08/13/microsoft-releases-poml-prompt-orchestration-markup-language/

*** Video RAG for long videos**
https://learnopencv.com/video-rag-for-long-videos/
* * Mixture of Agents, https://www.marktechpost.com/2025/08/09/mixture-of-agents-moa-a-breakthrough-in-llm-performance/
* Mixture-of-Agents Enhances Large Language Model Capabilities

Junlin Wang, Jue Wang, Ben Athiwaratkun, Ce Zhang, James Zou, 6.2024, https://arxiv.org/abs/2406.04692 – Mean Reciprocal Rank (MRR) (*in what ranking the first truly relevant result is found)* and Average Precision (AP).

* Building a Multimodal RAG That Responds with Text, Images, and Tables from Sources, P.Sarkar, 3.11.2025, https://towardsdatascience.com/building-a-multimodal-rag-with-text-images-tables-from-sources-in-response/
* How to Evaluate Retrieval Quality in RAG Pipelines (part 2): Mean Reciprocal Rank (MRR) and Average Precision (AP), M.Mouschoutzi, 5.11.2025, https://towardsdatascience.com/how-to-evaluate-retrieval-quality-in-rag-pipelines-part-2-mean-reciprocal-rank-mrr-and-average-precision-ap/

*** Клипове и публични хранилища на Свещеният сметач** в Гитхъб за GPT2-MEDIUM-BG, NLP, BGGPT, Вседържец/Специалист по всичко/Vsy и др.
* https://github.com/Twenkid/GPT2-Bulgarian-Training-Tips-and-Tools
* https://github.com/Twenkid/BgGPT * https://github.com/Twenkid/rhodope-gpt
* https://github.com/Twenkid/NLP-Computational-Linguistics
* https://github.com/Twenkid/Similarity-NLP-Corpus-CPP
* https://github.com/Twenkid/Vsy-Jack-Of-All-Trades-AGI-Bulgarian-Internet-Archive-And-Search-Engine
* https://github.com/Twenkid/Smarty
* https://github.com/Twenkid/Glas-2
* https://github.com/Twenkid/Toshko_2
(...)

* **Let's Build the GPT Tokenizer: A Complete Guide to Tokenization in LLMs**
A text and code version of Karpathy's famous tokenizer video.
https://www.fast.ai/posts/2025-10-16-karpathy-tokenizers

* Build production-ready AI features with schema-enforced outputs - Google Search
https://dev.to/dthompsondev/llm-structured-json-building-production-ready-ai-features-with-schema-enforced-outputs-4j2j
Structured-output from LLMs, JSON …; Model input → Text processing → Sentiment and Intent Analysis → AI Routing → (Tech Support, Billing Support, Sales Team, General Support, Enterprise) … Schema-Enforced Outputs; schema.versions

*** Каналът на Венелин Вълков:** https://www.youtube.com/@venelin_valkov
…
* https://www.marktechpost.com/2025/10/14/andrej-karpathy-releases-nanochat-a-minimal-end-to-end-chatgpt-style-pipeline-you-can-train-in-4-hours-for-100/

**От Вседържец:** Autoclap+Whisper+gpt4all+LLMs+Smarty2+ACS; +ollama
Зрим/Вседържец: {К}, Об{К}, ОбПмт, {К-К}, Обменник, взаимлик, всеборавител

…

* **Kyutai's Speech-To-Text and Text-To-Speech models based on the Delayed Streams Modeling framework**., ~ 7.2025
https://github.com/kyutai-labs/delayed-streams-modeling/

Вж понятия и https://en.wikipedia.org/wiki/ + …
Foundation_model, Domain_adaptation
Granularity, Commonsense_knowledge_(artificial_intelligence)

*** GPT2-MEDIUM-BG, един от най-големите модели за езици, различни от английския, през 2021 г. и обучението му в Google Colab:**
* GPT2-MEDIUM-BG, T.Arnaudov, 2021, https://huggingface.co/twenkid/gpt2-medium-bg

**Small Vision-Text models: 230M and 770M:**

* Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks, Bin Xiao† Haiping Wu∗ Weijian Xu∗ Xiyang Dai Houdong Hu Yumao Lu Michael Zeng Ce Liu‡ Lu Yuan‡, Azure AI, Microsoft, 10 Nov 2023
https://arxiv.org/pdf/2311.06242 | https://huggingface.co/microsoft/Florence-2-large
Florence-2-Vsion-Text-Model-0-23B_0-77B_excellent-14-5-2025.ipynb
https://github.com/Twenkid/Colab-Notebooks-AI-ML-CV/blob/main/Florence-2-Vsion-Text-Model-0-23B_0-77B_excellent-14-5-2025.ipynb
*** "Small" models by Hugging Face:**
https://www.marktechpost.com/2025/07/08/hugging-face-releases-smollm3-a-3b-long-context-multilingual-reasoning-model/
https://huggingface.co/HuggingFaceTB/SmolLM3-3B

* **SmallThinker: A Family of Efficient Large Language Models Natively Trained for Local Deployment**, Yixin Song, Zhenliang Xue et al. 30.7.2025, https://arxiv.org/pdf/2507.20984 **Mixture of Experts (MoE)** SmallThinker-4B-A0.6B-Base, SmallThinker-21B-A3B-Base, … (4B param, 0.6B active etc.) p.4 – "Following .. SmolLM (Allal et al., 2025), .. collecting a diverse range of high-quality datasets from the open-source community. For web data, we aggregated a corpus totaling **9 trillion tokens** from prominent sources .. *FineWeb-Edu, Nemotron-CC, mga-fineweb-edu, the Knowledge Pile; .. For math datasets: 1 trillion tokens: ... OpenWebMath, MegaMath, FineMath; Coding dataset: .. corpora like StackV2, OpenCoder ..* **p.6. Training:** *SmallThinker-4B-A0.6B: token horizon=2.5T tokens, sequence length = 4096, global batch size of 1536 (6,291,456 tokens per batch);* ***SmallThinker21B's*** *token horizon = 7.2 trillion tokens; … peak learning rate = 3e-4. .. final stage,* [the sequence length was] *extended to 32768 tokens for 4B models and 16384 tokens for 21B models.* ***Supervised Fine-Tuning****:* knowledge-intensive: selected 10 million selected data items … Math & Code: using Qwen3-32B (thinking) to generate step-by step solutions …

*** MoonDream, Vision-Text model, 2B, 0.5B:** https://github.com/vikhyat/moondream – Tosh has a small contribution to the development of *„MoonDream“* from 2024.

# * Speech Recognition and Text-to-Speech

*** Alibaba Qwen3-ASR:** Speech Recognition, Qwen3-Omni – "***Context injection mechanism:*** *Users can paste arbitrary text—names, domain-specific jargon, even nonsensical strings—to bias transcription. .. in scenarios rich in idioms, proper nouns, or evolving lingo."* https://www.marktechpost.com/2025/09/09/alibaba-qwen-team-releases-qwen3-asr-a-new-speech-recognition-model-built-upon-qwen3-omni-achieving-robust-speech-recogition-performance/
*https://qwen.ai/blog?id=41e4c0f6175f9b004a03a07e42343eaaf48329e7&from=research.latest-advancements-list – *Example1: Continuous noise of different types .. English rap song, Noise in vehicle with heavy accent, Multilingual code-switching, Chemistry course CSGO match commentary""*

*** Text-to-Speech:** https://herimor.github.io/voxtream/ – "real time generation" with transformers; moshi, mimi, … **! Hints:** Requires: Linux or WSL, CUDA Compute 7.0: RTX2xxx, issues with moshi-0.2.2 and Triton lib; tested on 1060 3GB CUDA Compute 6.1, pytorch 2.4., CUDA 11.8

*** VoXtream: Full-Stream Text-to-Speech with Extremely Low Latency**
Nikita Torgashov, Gustav Eje Henter, Gabriel Skantze, 19.9.2025
* https://arxiv.org/pdf/2509.15969 * https://github.com/herimor/voxtream

"Text prompt or stream → G2P → Phoneme transformer; Acoustic prompt → Mimi encoder →[BOS][PAD][A1,12][A2,12]…[BOS][P1][P1][P2] … Phoneme sequence, Audio token, Duration token, Sum, Concat … Depth Transformer → Mimi Decoder → Sound …. ; Acoustic prompt → Speaker encoder → Temporal transformer … "
* https://www.marktechpost.com/2025/09/23/meet-voxtream-an-open-sourced-full-stream-zero-shot-tts-model-for-real-time-use-that-begins-speaking-from-the-first-word/ 23.9.2025, Asif Razzaq:
*"Mimi-codec for semantic tokens at 12.5 Hz (80 ms frames), duration tokens, … Mimi acoustic codebooks, … ReDimNet speaker embedding for zero-shot voice prompting. "VoXtream trains on* ***a ~9k-hour mid-scale corpus****: roughly 4.5k h Emilia and 4.5k h HiFiTTS-2 (22 kHz subset). The team diarized to remove multi-speaker clips, filtered transcripts using ASR, and applied NISQA to drop low-quality audio. Everything is resampled to 24 kHz, and the dataset card spells out the preprocessing pipeline and alignment artifacts (Mimi tokens, MFA alignments, duration labels, and speaker templates). … interleaved AR + NAR vocoder approaches and LM-codec stacks."*
*~100 ms* on GeForce 3090, 4090, 5090.

**Tosh:**The first author: https://herimor.github.io/ is a first-year PhD student at https://www.kth.se/is/tmh "*Speech, Music and Hearing" – TMH at KTH "Research at the Division of Speech, Music and Hearing (TMH) is truly multi-disciplinary including linguistics, phonetics, auditory perception, vision and experimental psychology." –*

Swedish Royal Institute of Technology; Шведския кралски технологичен институт. His PhD is supported by another Swedish research institution: https://wasp-sweden.org/ *"Wallenberg AI, Autonomous Systems and Software Program". https://wasp-sweden.org/research/ "The research is conducted at eight Swedish universities: Chalmers University of Technology, KTH Royal Institute of Technology, Linköping University, Lund University, Umeå University, Örebro University, Uppsala University and Luleå University of Technology."*
* See the appendix #institutes – AI institutes … from The Prophets of the Thinking Machines
* For modern speech synthesis see also: moshi, mimi …
https://github.com/kyutai-labs/moshi *Moshi is a speech-text foundation model and full-duplex spoken dialogue framework. It uses Mimi, a state-of-the-art streaming neural audio codec.* https://pypi.org/project/moshi/
https://pypi.org/project/moshi-mlx/

## * Moshi: a speech-text foundation model for real-time dialogue,

Alexandre D´efossez, Laurent Mazar´e et al., 10.2024 * https://kyutai.org/Moshi.pdf #TTS #asr
* **Speech-to-Retrieval (S2R): A new approach to voice search**, Google 7.10.2025 https://research.google/blog/speech-to-retrieval-s2r-a-new-approach-to-voice-search/ *"Google's initial voice search solution used automatic speech recognition (ASR) to turn the voice input into a text query, and then searched for documents matching that text query. .. **cascade modeling approach** ..* the errors in the ASR phase *can significantly alter the meaning of the query, producing the wrong results."*
**New: Intent-aligned retrieval …** Dual-encoder; (audio query, document); *The training objective ensures that the vector for an audio query is geometrically close to the vectors of its corresponding documents in the representation space"* **Learn the intent** *directly from audio, bypassing the fragile intermediate step of transcribing every word"*
*** Simple Voice Questions Dataset:**
https://huggingface.co/datasets/google/svq#:~:text=Simple%20Voice%20Questions%20(SVQ)%20is,languages%20under%20multiple%20audio%20conditions.

## * Google Search by Voice: A case study, Johan Schalkwyk, Doug Beeferman et al.,35 p. 2008-2009 …

https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/36340.pdf - GOOG-411 – an early system, specialized queries/catalogs. General Google Voice Search: ***Example Query: "****images of the grand canyon**; what's the average weight of a rhinoceros; map of san francisco; what time is it in bangalore; weather scarsdale new york; bank of america dot com; A T and T; eighty one walker road;*
*videos of obama state of the union addres; genetics of color blindness*
*Figure 5: Basic block diagram of a speech recognizer";* Word Error Rate (WER),

Semantic Quality (WebScore), Perplexity – *the size of the set of words that can be recognized next, given the previously recognized words in the query;* Out-of-Vocabulary (OOV) Rate; Latency: *total time (in seconds) it takes to complete a search request by voice. 3.2 Acoustic Modeling: an estimate for the likelihood of the observed features in a frame of speech given a particular phonetic context; measurements of the spectral characteristics of a time-slice of speech; the basic process involves aligning transcribed speech to states within an existing acoustic model, accumulating frames associated with each state, and re-estimating the probability.* The acoustic models use standard 39-dimensional PLP-cepstral features, LDA, and STC, with triphone GMM systems trained using ML, MMI, and boosted-MMI objectives. Production Language Model. Text normalization ($20 books from.. → twenty dollar books from); disambiguation lists (Holiday in Plovdiv – could be Holiday Inn Plovdiv (a hotel));
Gaussian Mixture Model, ML – Maximum likelihood; MMI – Maximum Mutual Information; LDA (Linear Discriminant Analysis) PLP (Perceptual Linear Prediction) is a refinement of Linear Prediction analysis; Auditory Spectrum: not the raw power spectrum of the speech signal: Critical-band analysis (non-linear sensitivity of human auditory system), Equal-loudness pre-emphasis – compensate for the non-linearity of the loudness perception at different frequencies; Intensity-loudness power law: compression of the intensity of the signal using a cubic-root non-linearity …
Mel Scale vs Bark Scale frequency bins/modeling. Mel – pitch perception. Bark – possibly more suited for loudness analaysis.Filter banks.

* Discriminative Training Of Gaussian Mixture Models For Large Vocabulary Speech Recognition Systems, June 1996, L.R. Bahl M. Padmanabhan M et al. https://www.researchgate.net/publication/3640460_Discriminative_Training_Of_Gaussian_Mixture_Models_For_Large_Vocabulary_Speech_Recognition_Systems
* Full Covariance Modelling for Speech Recognition, Peter Bell, The University of Edinburgh, 2010, PhD Thesis https://www.cstr.ed.ac.uk/downloads/publications/2010/thesis.pdf
* Speech Recognition — Feature Extraction MFCC & PLP, Jonathan Hui, 28.8.2019, https://jonathan-hui.medium.com/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9 - Sliding window of 25 ms, 10 ms apart; 39 features MFCC;
if 3 words/sec x 4 phones, then *a phone would be sub-divided into 3 stages, then there are 36 states per second or 28 ms per state. So the 25ms window is about right."*
* Speech Recognition — Phonetics, Jonathan Hui, Aug 26, 2019 https://jonathan-hui.medium.com/speech-recognition-phonetics-d761ea1710c0
*** Speech Recognition, EECS E6870 — Spring 2016**
https://www.ee.columbia.edu/~stanchen/spring16/e6870/outline.html
* Lecture 1: Introduction/Signal Processing, Part I
https://www.ee.columbia.edu/~stanchen/spring16/e6870/slides/lecture1.pdf p.34-37
**"The Birth of Modern ASR: 1970–1980's:** Many key algorithms developed/refined.

Expectation-maximization algorithm; n-gram models; Gaussian mixtures; Hidden Markov models; Viterbi decoding; etc. Computing power still catching up to algorithms. First real-time dictation system built in 1984 (IBM). Specialized hardware required — had the computation power of a 60 MHz Pentium. *** The Golden Years: 1**990's–now; 1994: CPU speed 60 MHz; now: 3 GHz; training data 1994: <10; 2016: 10000h+; output distributions GMM NN /GMM hybrids; sequence modeling HMM HMM and/or NN; language models n-gram/ n-gram and NN ..

* **Commercial Speech Recognition:** 1995 – 1998 — first large vocabulary speaker dependent dictation systems.; 1996 – 2005 — first telephony- based customer assistance systems. 2003 – 2007 — first automotive interactive systems. 2008 – 2010 — first voice search systems. 2011 – today — growth of cloud-based speech services."

* Lecture 3, Gaussian Mixture Models and Introduction to HMM's, Michael Picheny, Bhuvana Ramabhadran, Stanley F. Chen,, Markus Nussbaum-Thom, Watson Group, IBM T.J. Watson Research Center… 3 February 2016
https://www.ee.columbia.edu/~stanchen/spring16/e6870/slides/lecture3.pdf
*DTW* [Dynamic time warping] – *Can extract features over time (MFCC, PLP, others) that . . .Characterize info in speech signal in compact form. Vector of 12-40 features extracted 100 times a second …;* [Distance calculation, path finding, comparison]

*** Automatic Speech Recognition (ASR) 2018-19: Lectures** by **The University of Edinburgh, School of Informatics:**
https://www.inf.ed.ac.uk/teaching/courses/asr/lectures-2019.html

* Stanford University, Navdeep Jaitly, Natural Language Processing with Deep Learning, CS224N/Ling284, Lecture 12: End-to-end models for Speech Processing
https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/lectures/cs224n-2017-lecture12.pdf *… Connectionist Temporal Classification (CTC) Softmax over vocabulary {a, b, c, … z, ?, ., !,..} and extra token <b>* [boundary]*;* mapping to output sequences: deduplicate repeated tokens: "cat", recognized as cc<b>aa<b>t<b>

* Very Deep Convolutional Networks for End-to-End Speech Recognition. Yu Zhang, W.Chan, N.Jaitly, 10 Oct 2016, https://arxiv.org/abs/1610.03022 *"10.5% WER without any dictionary or language using a 15 layer deep network";* Conv LSTM ..

*** Listen Attend and Spell.** William Chan et al., ICASSP 2015 – Seq-to-seq with attention; *Attention vector* – where the model expects to find the relevant information. *Hierarchical encoder reduces time resolution.* Not online, needs to receive the whole input before transcribing; *attention is computational bottleneck – every* ***output token*** *pays attention to every* ***input time step****;* the length of the input impacts the accuracy.

*** Listen and Translate: A Proof of Concept for End-to-End Speech-to-Text Translation** Alexandre Berard, Olivier Pietquin et al. https://arxiv.org/abs/1612.01744
*"end-to-end speech-to-text translation system, which does not use source language transcription during learning or decoding .. direct speech-to-text translation .. a small French-English synthetic corpus"*

*** Comparison of MFCC and PLP Parametrizations in the Speaker Independent Continuous Speech Recognition Task**, Josef Psutka, Ludek Muller and Josef V. Psutka, 2001, https://www.isca-archive.org/eurospeech_2001/psutka01_eurospeech.pdf
* A High-Performance Auditory Feature For Robust Speech Recognition August 2000, Qi LiFrank K. Soong, Frank K. Soong, Olivier Siohan https://www.researchgate.net/publication/2361318_A_High-Performance_Auditory_Feature_For_Robust_Speech_Recognition Converting the spectrum to *"Bark scale between 2.0 to 16.4 Barks, which corresponding to a linear frequency rangefrom 200 to 3500 Hz."*
* ECE 8463: FUNDAMENTALS OF SPEECH RECOGNITION, Professor Joseph Picone, Mississippi State University https://isip.piconepress.com/courses/msstate/ece_8463/lectures/current/lecture_17/lecture_17.pdf
…

*** Switchboard Telephone Speech Corpus,** 1990: Texas Instruments, DARPA ; released in 1992: 2400 conersations, 543 speakers (302 male, 241 female) … Switchboard-2 Phase II was collected in 1999 and includes "4,472 five-minute telephone conversations involving 679 participants".[5] https://en.wikipedia.org/wiki/Switchboard_Telephone_Speech_Corpus

-- END OF SPEECH RECOGNITION SECTION --

LLMs

*** Verbalized sampling (asking for the probability distributions in the prompt and improve the diversity of the generated outputs):** https://github.com/CHATS-lab/verbalized-sampling

*** Verbalized Sampling: How To Mitigate Mode Collapse And Unlock Llm Diversity,** Jiayi Zhang et al. 10.10.2025 https://arxiv.org/pdf/2510.01171

# * Vision-Language Models, Vision-Language Action Models, Foundation Robot Models …

*** Top 10 Vision Language models of 2025:**
https://www.datacamp.com/blog/top-vision-language-models

* **Teaching VLMs to Localize Specific Objects from In-context Examples**
Sivan Doveh*2 Nimrod Shabtay et al., 3.2025, https://arxiv.org/pdf/2411.13317
* Method teaches generative AI models to locate personalized objects - After being trained with this technique, vision-language models can better identify a unique item in a new scene., Adam Zewe | MIT News, October 16, 2025
* https://news.mit.edu/2025/method-teaches-generative-ai-models-locate-personalized-objects-1016 in-context learning

*** Open-o3 Video: Grounded Video Reasoning with Explicit Spatio-Temporal** Evidence, Jiahao Meng, Xiangtai Li et al. 23.10.2025, * https://arxiv.org/abs/2510.20579 * https://marinero4972.github.io/projects/Open-o3-Video/ - **Explicit visuo-spatial evidence**; … Thinking with images (OpenAI-o3, DeepEyes); Ground-truth reasoning datasets; STGR-CoT-30K for SFT, STGR-RL-36k for RL; Combine temporal-only and spatial-only resources with 5.9k newly annotated, high-quality spatio-temporal samples.Data preparation & initial annotation: Filter, Gemini 2. Pro: question, answer, key_frames, key_items, boxes of each key frame, reasoning process … → Bounding box filtering: Is this a {object_name}? Only keep data with answer = ”yes”; <obj> girl</obj><box> [214,0,433,374]</box> at <t>213</t>s → Self-consistency Checking: Align CoT annotation with the key frames and key objects info. <think>The video shows the <obj>girl</obj><box>(…)… <t>21.3</t>s who is .. raising her..</think> → Unified Video Spatio-temporal Reasoning Data … Method: Reward format + answer + think-evidence … Propose **adaptive temporal proximity & temporal gating** …; Accuracy reward: If task = MCQ & prediction matches ground truth, ROUGE(y[pred], y[gt]) – if task = Free -form QA; vIoU(Box[pred], Box[gt]) if task = Spatial grounding; tIoU([s[pred],e[pred]], s[gt],e[gt]) if task = Temoral grounding …
Think-evidence reward: temporal term, spatial term … interval supervision [s[gt], e[gt]]; point supervision .. 0 – no timestamp evidence; spatial term …

s[gt], e[gt] – start ground truth, end ground truth … Format reward: <think> <answer> with correct <obj><box><t> … = 1. … Confidecne-awara test-time scaling. V-STAR benchmark: spatio-temporal reasoning across three dimensions [**Tosh:** not "in 3D", three types]. Chain 1: What-when-where; Chain 2:  what-where-when.

*** GigaBrain-0: A World Model-Powered Vision-Language-Action Model**, GigaBrain Team: Angen Ye, Boyuan Wang et al., 22.10.2025 – * https://gigabrain0.github.io/  https://arxiv.org/pdf/2510.19430 *GigaBrain-0, a novel VLA foundation model empowered by world model-generated data (e.g., video generation, real2real transfer, human transfer, view transfer, sim2real transfer data).*

*** SPEAR-1: Robotic Foundation Model via 3D understanding**

* SPEAR-1: Scaling Beyond Robot Demonstrations via 3D Understanding Nikolay Nikolov, Giuliano Albanese et al. 2025, https://spear.insait.ai/SPEAR-1.pdf https://spear.insait.ai/ – Robotic Foundation Models (RFMs); training ***on ~45M frames from 24 Open X-Embodiment datasets*** *and show it outperforms or matches state-of-theart models such as π0-FAST and π0.5 while using 20× fewer robot demonstrations.*  VLA – Vision-Language-Action models .. p.3 *"only 200k non-robotic 2D images, SPEAR-1 outperforms state-of-the-art models trained with more than 900M additional frames of robotic demonstrations"*. Previous: SpatialVLA. P.3: Spear-VLM: vision encoder, vision-to-text-embedding feature projector, LLM. PaliGemma = SigLIP visual encoder + Linear projector: maps the visual tokens to the LLM input space + Gemma 2B LLM. MoGe depth encoder: afine invariant modeling approach. … p.4.. . Data annotations: Depth estimation, semantic segmentation, depth projection → Spatial Labels … 3D Vision-Question-Answering Data: object-level segmentation masks, semantic labels and projected 3D point cloud. They prepare a dataset by detecting 2D bounding boxes and semantic labels with Gemini, then prompting SAM2 (Segment Anything Model) to produce instance-level segmentation masks, and finally by obtaining 3D point cloud annotations for the entire image via MoGe; images from "cooking", "bike repair" from EgoExo4D" .. 200k + 30k images …; p.5. VLM training: a batch_size= 512 x 2k steps (first alignment) + 10k steps for the second stage; total =  18 hours x 16 Nvidia H200 GPUs. VLA pre-training: two camera views: external 280x210 + wrist: 112x112 (black image if the wrist camera is not available). 32 x H200, batch=2048 x 300k step ~ 6 days on a mixture of 24 datasets. VLA post-training – for WiddowX real-world & SIMPLER simulations, and for Franka robot real-world experiments: additional fine-tuneing OXE pre-trained SPER-1: 50 k steps on Bridge V2 & DROID → two versions SPEAR-1 (Bridge), SPEAR-1 (DROID).

…

The authors from INSAIT robotics declare their goal to *"build foundation models such that any robot can autonomously perform any task in any environment - from hospitals to factories to ordinary homes.… a long way to go - from long-horizon planning to autonomous self-improvement to completing unseen tasks."*

*** MoGe: Unlocking Accurate Monocular Geometry Estimation for Open-Domain Images with Optimal Training Supervision,** Ruicheng Wang, Sicheng Xu et al. 24.10.2024, https://arxiv.org/abs/2410.19115 - 3D-reconstruction ("recovering 3D geometry from monocular open-domain images.

*** MoGe-2: Accurate Monocular Geometry with Metric Scale and Sharp Details,** Ruicheng Wang, Sicheng Xu, Yue Dong et al., 3.7.2025 – *a metric scale 3D point map of a scene from a single image. .. a robust, optimal, and efficient point cloud alignment solver for accurate global shape learning, and a multi-scale local geometry loss promoting precise local geometry supervision. … monocular estimation of 3D point map, depth map, and camera field of view …*
* https://github.com/microsoft/MoGe
– Accurate 3D geometry estimation: Estimate point maps & depth maps & normal maps
*** Optimized for speed:** Achieves 60ms latency per image (A100 or RTX3090, FP16, ViT-L). Adjustable inference resolution for even faster speed.

**Pretrained Models:** MoGe-2 - Ruicheng/moge-2-vitl: 326M; vitl-normal 331M, moge-2-vitb-normal 104M, Ruicheng/moge-2-vits-normal 35M
..
**See also:** DepthAnything etc.
*** Depth Anything V2,** Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, Hengshuang Zhao, 13.6.2024 – monocular depth estimation https://arxiv.org/abs/2406.09414

* https://github.com/DepthAnything/Depth-Anything-V2 Pre-trained Models
We provide four models of varying scales for robust relative depth estimation:

| Model | Params | Checkpoint |
|---|---|---|
| Depth-Anything-V2-Small | 24.8M | Download |
| Depth-Anything-V2-Base | 97.5M | Download |
| Depth-Anything-V2-Large | 335.3M | Download |
| Depth-Anything-V2-Giant | 1.3B | Coming soon |

*** Depth Anything at Any Condition,** Boyuan Sun, Modi Jin, Bowen Yin, Qibin Hou, 2.7.2025, https://arxiv.org/pdf/2507.01634 *DepthAnything-AC - a foundation monocular depth estimation (MDE) model capable of handling diverse environmental conditions.. "illumination variations, adverse weather, and sensor-induced distortions. .. unsupervised consistency regularization finetuning paradigm that requires only a relatively small amount of unlabeled data. .. spatial distance constraint – patch level relative relationships – clearer semantic boundaries and more accurate details … "*
* **ARGenSeg: Image Segmentation with Autoregressive Image Generation Model,** Xiaolong Wang, Lixiang Ru et al. 23.10.2025,

https://arxiv.org/abs/2510.20803 – usually the multimodal large language models (MLLMs) use either boundary points representation or dedicated segmentation heads. These methods *rely on discrete representations or semantic prompts fed into task-specific decoders, which limits the ability of the MLLM to capture fine-grained visual details.* ARGenSeg introduces a novel *segmentation framework for MLLM based on image generation, which naturally produces dense masks for target objects.* The MLLM outputs visual tokens and detokenizes *them into images using an universal VQ-VAE, making the segmentation fully dependent on the pixel-level understanding of the MLLM.*
https://ghost233lism.github.io/depthanything-AC-page/
https://github.com/HVision-NKU/DepthAnythingAC

## * Съгласуване, намеса в теглата, защитни „парапети“, AI Alignment – виж също #AIethics

* **A Library for Understanding and Improving PyTorch Models via Interventions,** Zhengxuan Wu et al. 3/2024 https://arxiv.org/abs/2403.07809
https://github.com/stanfordnlp/pyvene
https://stanfordnlp.github.io/pyvene/tutorials/advanced_tutorials/DAS_Main_Introduction.html#contents In causal abstraction analysis, we assess whether trained models conform to high-level causal models that we specify, not just in terms of their input–output behavior, but also in terms of their internal dynamics. The core technique is the **interchange intervention**

*** The landscape of LLM guardrails: intervention levels and techniques, 6.2024**
https://www.ml6.eu/blogpost/the-landscape-of-llm-guardrails-intervention-levels-and-techniques Guardrails: Rule-based computation, LLM based metrics (e.g. perplexity, embedding similarity), LLM judges (e.g. fine-tuned models, zero-shot), Prompt engineering and Chain-of-thought … Видове защити, допълнителни проверки на породеното: чрез правила, мерки за ЕМ, други фино-настроени ЕМ като съдии, чрез инженерство на подканите, верига на мисълта.
https://stanfordnlp.github.io/pyvene/tutorials/advanced_tutorials/DAS_Main_Introduction.html

* **WavTokenizer**, SOTA discrete acoustic codec models with 40 tokens per second for audio language modeling ~ 500 bits/s - срвн. „Истината“, 2002 – телефонните връзки с възстановяване на гласа.

## * Набори от данни #datasets

Mathematics, Математика, Chain-of-Thought,Reasoning:
https://huggingface.co/datasets/AI-MO/NuminaMath-CoT/

Виж някои от хранилището на Вседържец/Гитхъб (Vsy, Jack of All Trades), бележки в Issues и:

* https://huggingface.co/datasets/allenai/WildChat-1M - Един милион разговора между потребители и ЧатГПТ
https://wildchat.allen.ai/ https://wildvisualizer.com/

*** WildVis: Open Source Visualizer for Million-Scale Chat Logs in the Wild**
Yuntian Deng, Wenting Zhao, Jack Hessel, Xiang Ren, Claire Cardie, Yejin Choi

* https://www.marktechpost.com/2025/09/06/hugging-face-open-sourced-finevision-a-new-multimodal-dataset-with-24-million-samples-for-training-vision-language-models-vlms /

* https://huggingface.co/datasets/HuggingFaceM4/FineVision

* https://huggingface.co/HuggingFaceM4

See Huggingface, Kaggle etc.

(..)

*** Графи за знания**

Различаване на Graph of Knowledge/Knowledge Graph – първото – по-общ формат, разнородни възли, подобно на „no SQL" в БД. KG – по-строга еднотипна структура.
https://www.infoworld.com/article/3801640/the-journey-towards-a-knowledge-graph-for-generative-ai.html
Заявки с много входни точки (multi-point access questions); RAG, KG-RAG …

*** Imbue/Generally Intelligent**: ***Introducing Generally Intelligent***
https://imbue.com/company/launch/
"Generally Intelligent" – компания за УИР, AGI, 20.10.2022 – по-късно преименувана на Imbue – "независима изследователска компания, … да разработи агенти с ИИ на човешко ниво, за да решава задачи от реалния живот" – срвн с посланието на руски институт AIRI.
https://imbue.com/research/avalon/ - симулатор за RL

*** AI researcher François Chollet founds a new AI lab focused on AGI**
https://techcrunch.com/2025/01/15/ai-researcher-francois-chollet-founds-a-new-ai-lab-focused-on-agi/ 15.1.2025
*** Ndea –** УИР чрез автоматичен синтез на програми, … https://ndea.com/
DL да насочва дискретно търсене на програми за точно решение. Ndea - ennoia

(intuitive understanding) and dianoia (logical reasoning) – съчетание от нагледно, „интуитивно“ и разсъдтиелно разбиране. Вж #bongard, идеята е изказана в книгата „Проблема узнавания“, 1967 г.

| | |
|---|---|
| https://paperswithcode.com/dataset/rocstories<br>**A Corpus and Evaluation Framework for Deeper Understanding of Commonsense Stories,** Nasrin Mostafazadeh et al. 2016<br>https://arxiv.org/abs/1604.01696<br><br>https://www.kaggle.com/datasets/mrriandmstique/rocstories-and-story-cloze-test-corpora | 100К истории по 5 изречения за здрав смисъл (common sense) и Cloze тестове с пропуснато заключение.<br>ТА: Ниска езикова сложност, 5-6-7-годишни? |

*** Byte Latent Transformer (BLT): A Tokenizer-Free Model That Scales Efficiently** https://www.marktechpost.com/2024/12/13/meta-ai-introduces-byte-latent-transformer-blt-a-tokenizer-free-model-that-scales-efficiently/
Byte Latent Transformer: Patches Scale Better, Than Tokens, Artidoro Pagnoni, Ram Pas https://arxiv.org/abs/2412.09871
Вид йерархично, многомащабно кодиране: първо байтови потоци, които с малък преобразител се кодират в по-големи групи, късове, и обратно при декодирането; преобразители предвиждат на различни мащаби.

* **Writer, Palmyra: more creative LLMs:**
https://writer.com/engineering/self-evolving-models/
https://venturebeat.com/ai/writer-new-ai-model-aims-to-fix-the-sameness-problem-in-generative-content/
https://developer.nvidia.com/blog/an-introduction-to-model-merging-for-llms/
a memory pool - store new information & recall it when processing a new user input; uncertainty-driven learning; self-update; merging techniques & adaptive model layering: Model Soup, Spherical Linear Interpolation (SLERP), Task Arithmetic (using Task Vectors), TIES leveraging DARE

*** Language Models are Super Mario: Absorbing Abilities from Homologous Models as a Free Lunch** https://arxiv.org/pdf/2311.03099

*** Generative AI Still Needs to Prove Its Usefulness,** Gary Marcus, 20.12.2024**,**
*The hype is fading, and people are asking what generative artificial intelligence is really good for. So far, no one has a decent answer.*
https://www.wired.com/story/generative-ai-will-need-to-prove-its-usefulness/

*** Отговор на Тош на „Пораждащият изкуствен интелект трябва да докаже полезността си“ по Гари Маркус**

**Тош:** Пораждащият ИИ трябвало *„да докаже полезността си“,* Гари Маркус. Това по-скоро е явлението на „*омръзването*“ и че творчеството, особено в изкуството, всъщност е, или лесно става *безполезно* за повечето хора. Много повече неща или почти всички също са „безполезни“, в малко по-различни условия или контекст, например когато недостигат по-необходими: жажда, глад, война и т.н. Нуждата от потребление се създава и внушава.

Виж българското пророчество от 2013 г. че в скорошната ера на универсалните мислещи машини, интелектуалните дейности ще се вършат безплатно за 1 милисекунда и от „нас умниците и супер умниците“ няма да има нужда, където отбелязвам обаче, че и без „AGI“ *не е имало нужда от нас*.

Оценката за полезност на „луксозни“, нежизненоважни дейности, в общества, е под влияние на новост, мода, власт, насищане, и оценителите често *лъжат.* В тежки условия „потребителите“, човеците, имат нужда само да са живи и здрави, т.е. почти всичко друго се оказва „безполезно“.

Накратко „*полезността*“ е разтегливо понятие и се нуждае от рамка, а тя пък може да се „*скове*“ според нуждите на оценителя.

По-конкретно, маймуните или животните нямат нужда от изкуствен интелект или въобще от интелект и „познавателни технологии“, отвъд необходимите за оцеляване или за създаване на удобства. „Волята“ по Шопенхауер надделява над „Нагледа“. Колкото и да е умен ИИ, хомо сапиенс все ще е недоволен – „отдалечаващата се цел на ИИ“. Дори Демис Хасабис през 2025 г. обяснява как за да признаят ИИ за „УИР“ трябвало най-добри учени 2-3 месеца да не могат да разгадаят каквото бил направил и т.н. Друга „заплаха е“, че ако решат проблема, ще станат ненужни.

Това е и пародия на някои изследователски институти – те нямат мотив да „завършат“ или „закрият“ изследванията, защото ще трябва да закрият и себе си, а целта им е да нарастват и да усвояват максимално количество средства. Задачата и проектите трябва винаги да остават незавършени, да има „бъдеща работа“ и т.н. Виж в приложението за Институти за ИИ „на световно ниво“[182], бележки към статия от една от българските футурологични изследователски групи на „*несветовно* ниво“. с. 115 в ред. От 18.9.2025 *„Според Иван Попов трябва да се разработват „закриващи научни дисциплини“, а не „перспективни“; т.е. такива които ще разрешат проблемите веднъж-завинаги, а не ще отворят все повече и повече текуща работа.“*

По-голяма изгода има обаче в това да се отваря максимално количество текуща работа. Виж също „Нужни ли са смъртни изчислителни системи…“, Т.А. 2025 за заблуждаващите мерки за ефективност.

[182] https://twenkid.com/agi/AI_Institutes_Strategies_The_Prophets_Thinking_Machines_7-9-2025.pdf

* Виж също статиите в края на книгата и бележките към тях относно измерването на способностите на ИИ да решава *„задачи с икономическа стойност*“, изпълнявани от човек за еди-колко ви време, и теста за агенти „Remote Labor AI”.

**Dask: The Python Data Scientist’s Power Tool:**
https://www.kdnuggets.com/introduction-dask-python-data-scientist-power-tool
https://www.dask.org/ - distributed pandas & numpy for big data; local clusters & cloud pip install dask[complete]; import dask.array as da; import dask.dataframe as dd; from dask import delayed; from dask.distributed import Client
import dask.dataframe as dd;  df = dd.read_parquet("s3://data/uber/")
pip install dask[distributed]; https://distributed.dask.org/en/stable/install.html
Python: dask: работа с масиви от данни надхвърлящи оперативната памет, надстройка на pandas и numpy. Може да се мащабира и в локална мрежа за домашна обработка на големи обеми от данни (BigData), както и на наети облачни сървъри и др.
https://distributed.dask.org/en/stable/quickstart.html

https://www.marktechpost.com/2024/12/16/this-ai-paper-from-microsoft-and-novartis-introduces-chimera-a-machine-learning-framework-for-accurate-and-scalable-retrosynthesis-prediction/

* **Chimera: Accurate retrosynthesis prediction by ensembling models with diverse inductive biases.** Krzysztof Maziarz et al., 12.2024
https://arxiv.org/abs/2412.05269
Предсказване на химични реакции, чрез които да се получи дадена крайна молекула чрез прилагане на разнообразие от методи. Computer-Aided Synthesis Planning

*** Общи съвети за употреба на системи с пораждащ ИИ**
https://digitalk.bg/ai/2024/12/19/4720900_10_sluchaia_v_koito_da_izpolzvame_genai_i_5_v_koito_da/ -

**Obsidian Web Clipper:** извличане на информация от уеб страници, приставка за уеб четци, браузъри; филтри; срвн ACS/Вседържец (бдщ):
https://obsidian.md/clipper
https://www.dsebastien.net/supercharge-your-knowledge-capture-workflow-with-the-obsidian-web-clipper/  https://help.obsidian.md/web-clipper/filters#%60kebab%60
https://extractors.ai/

* **Introducing perceptein, a protein-based artificial neural network in living cells**,  Zibo Chen et al, A synthetic protein-level neural network in mammalian cells, Science (2024). DOI: 10.1126/science.add8468

* **Hugging Face Released Moonshine Web: A Browser-Based Real-Time, Privacy-Focused Speech Recognition Running Locally**
https://www.marktechpost.com/2024/12/20/hugging-face-released-moonshine-web-a-browser-based-real-time-privacy-focused-speech-recognition-running-locally/
https://github.com/huggingface/transformers.js-examples/tree/main
cmp. Whisper + AutoClap за Вседържец

*** OpenAI o3 - Thinking Fast and Slow** https://dev.to/maximsaplin/openai-o3-thinking-fast-and-slow-2g79
* OpenAI's o3 suggests AI models are scaling in new ways — but so are the costs, Maxwell Zeff, https://techcrunch.com/2024/12/23/openais-o3-suggests-ai-models-are-scaling-in-new-ways-but-so-are-the-costs/ **–** разсъждаващият модел; „expensive: scoring 76% cost around $9k and 88% - OpenAI didn't disclose (one can evaluate the total cost to be at $1.5mil given the statement that 172x more compute was used). … 10-15 min. on some tasks…" – изследва > прстр.от възможни решения. За задачите от ARC: o1 - $5 на задача, o1-mini: няколко цента - $0.01.
https://analyticsindiamag.com/ai-features/ai-agents-without-label-redefine-semiconductors-in-india/
https://en.wikipedia.org/wiki/OpenAI_o3 *Reinforcement learning was used to teach o3 to "think" before generating answers, using what OpenAI refers to as a "private chain of thought".*
https://en.wikipedia.org/wiki/Feedback_neural_network
Reflective NN, Feedback neural network (not only feed-forward); see Chain of thought, tree-of-though, self-consistenty etc. prompting techniques, prompt engineering: https://en.wikipedia.org/wiki/Prompt_engineering#Chain-of-thought and DeepSeek-R1; see also: in-context-learning (examples within a prompt), **textual inversion**: *an optimization process to create a new word embedding, based on a set of example images* for text-to-image models; searching for prompts with gradient descent.

*** Talk about Chip Design, Tape-out, Verification, Manufacturing, and Cost,** 22.9.2023 https://www.linkedin.com/pulse/talk-chip-design-tape-out-verification-manufacturing/ Цена на маските за производство на чипове: 40 nm: $800-900K, 28 nm: $2M, **14 nm: $5M**, 7 nm: $15M; 5 nm: $47.25M; **3 nm: ? "may cost $hundreds M";** layers: 28 nm = 40, 14 nm = 60 masks; 7 nm: 80 or 100 masks; one layer mask costs $80K; a 40 nm MCU process: if 10 wafers, one will cost $94K, if 10K wafers:$4K per wafer (in total $40M).

* **Sam Altman predicts superintelligence will trigger a 10x surge in scientific AI breakthroughs — each year as revolutionary as a decade,** Kevin Okemwa, December 24, 2024 **-** OpenAI's CEO says superintelligence might be "a few thousand days away" with little societal impact.
https://www.windowscentral.com/software-apps/sam-altman-predicts-

superintelligence-will-trigger-a-10x-surge-in-technological-breakthroughs

*** Quantization of LLM weights**
***** https://techcrunch.com/2024/12/23/a-popular-technique-to-make-ai-more-efficient-has-drawbacks/ - quantization of the weights

* **Microscaling Data Formats for Deep Learning,** Bita D. Rouhani et al., 10.2023, down to fp4 (4-bit); MX, mxfp: similar or equal performance to fp32. See OpenAI GPT-OSS models 20B and 120B published on 5.8.2025, cited in the "latest" section.

* Sam Altman, 11.6.2025: "**We've Already Passed the Superintelligence Event Horizon'** , www.decrypt.com, ChatGPT: 800M weeky users, ; "gentle singularity" … "2025 has seen the arrival of agents that can do real cognitive work; writing computer code will never be the same", he said. "2026 will likely see the arrival of systems that can figure out novel insights. 2027 may see the arrival of robots that an do tasks in the real world."

**Tosh, 14.7.2025:** The Computers were always doing "real cognitive work" and they were doing it better than humans and were "superintelligent". The true value of the number of users can't be verified. The novelty and "new insights" are a vague concept, it needs specific definitions. There's "nothing new under the sun" (Ecclesiastes, The Holy Bible) and even now or in the past computers and all technologies were instrumental for the insights of the humans.

* "Сам Алтман: **Вече живеем в друга реалност – и почти никой не забелязва това",** 11.06.2025, Kaldata.com

**Тош**: Изчислителните машини винаги са вършели „истинска познавателна работа", „обработвали са информация", „знания" и са го правили по-добре от човека и са били „свръхумни". Всичката техника и наука от всички времена в даден момент е допринесла за „новите идеи" на човеците от следващите моменти. Новостта трябва да се определи достатъчно точно. Кое се брои за „ново" хрумване" е мъгливо и зависи от оценителя-наблюдател, хората често хитруват. Още по библейско време е казано, че „няма нищо ново под слънцето". (…)

*** "Генезис" променя правилата на играта: нов физичен симулатор обучава роботи 430 000 пъти по-бързо от реалността,** 27.12.2024
https://novini.bg/biznes/biznes_tehnologii/884711
https://github.com/Genesis-Embodied-AI/Genesis **„***Over 43 million FPS when simulating a Franka robotic arm with a single RTX 4090 (430,000 times faster than real-time).*" Вж сщ: https://drake.mit.edu/ ; MuJoCo; Nvidia Cosmos

* **Chinese algorithm claimed to boost Nvidia GPU performance by up to 800X for advanced science applications,** 3.2.2025 complex mechanical challenges … peridynamics (PD), a non-local theory: modelling fractures and material damage; large-scale material simulations. https://www.tomshardware.com/pc-components/gpus/chinese-algorithm-claimed-to-boost-nvidia-gpu-performance-by-up-to-800x-for-advanced-science-applications https://www.kaldata.com/it-новини/наука/учени-от-русия-и-китай-ускориха-с-800-пъти-545292.html 30.1.2025

* **Just 2 hours is all it takes for AI agents to replicate your personality with 85% accuracy** By Owen Hughes January 4, 2025 Researchers from Google and Stanford have created accurate AI replicas of more than 1,000 people. https://www.livescience.com/technology/artificial-intelligence/just-2-hours-is-all-it-takes-for-ai-agents-to-replicate-your-personality-with-85-percent-accuracy

*** Generative Agent Simulations of 1,000 People**, Joon Sung Park, Carolyn Q. Zou et al. https://arxiv.org/abs/2411.10109 2 h audio interview ~ 6500 words; predicting Individuals' Attitudes and Behavior; economic games; Big Five personality traits;

*** AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking,** Michael Gerlich, Center for Strategic Corporate Foresight and Sustainability, SBS Swiss Business School, 8302 Kloten-Zurich, Switzerland https://www.mdpi.com/2075-4698/15/1/6 Използването на инструменти с ИИ намалявало критичното мислене в тестова група, по-изразено на онези с по-ниско образование; по-младите били по-зависими от инструментите. Виж определенията за критично мислене и познавателно разтоварване. Т.А.: Две други нови технология в миналото: радиото и телевизията вероятно имат същия ефект, особено телевизията заради хипнотичното ѝ действие[183].

*** PlanGEN: A Multi-Agent Framework for Generating Planning and Reasoning Trajectories for Complex Problem Solving**, Mihir Parmar, Xin Liu, Palash Goyal et al. https://arxiv.org/abs/2502.16111 inference time alg.:Gemini 1.5-Pro & 2.0-Flash, GPT-4o. Best of N, Tree-of-Thought, REBASE; тестове: NATURAL PLAN, OlympiadBench, DocFinQA, GPQA; обикновено: методи с шаблони (template-based); три агента: на ограниченията, проверката и подбора ( constraint, verification, selection); огран.: бюджет, правила, бюджет; подбор по UCB – Upper Confidence Bound, горна граница на убеденост за варианти с различна сложност [вид регуларизация, наказваща по-сложни модели със същия резултат] – равновесие между изследване и използване (exploration/exploitation tradeoff); REward-BAlanced SEarch (REBASE) (Wu et al., 2024a); calendar scheduling, trip& meeting planning: зададени на естествен език; diversity bonus;

[183] Тодор Арнаудов, „Бягство от прекрасния видео свят", в-к „Пловдивски университет", бр.3-4 2018 https://uni-plovdiv.bg/uploads/site/vestnik/2018/vestnik_br_3-4_2018.pdf

recovery …

[**Тош:** Най-интересното в някои от задачите като планиране на срещи и календар е преобразуването на заданието от текст на естествен език и решаването чрез тези средства; самите задачи имат много по-прости решения с програмиране, ако данните са въведени по друг начин. Пораждане на стъпки, проверка и пр. Виж бел. за класическото планиране и роботика и за това че съвременни методи с ГЕМ за роботи в крайна сметка се свеждат до решаване на старите задачи от края на 1960-те и 1970-те, но с *много повече ресурси.* ]

* **Google's Sergey Brin Urges Workers to the Office 'at Least' Every Weekday**, New York Times, 27.2.2025, https://www.nytimes.com/2025/02/27/technology/google-sergey-brin-return-to-office.html - Сергей Брин настоявал служителите на „Гугъл" да работят поне по 12 ч. на ден, защото последният напън за създаване на УИР бил започнал.

В стартъпи в Китай и в Силициевата долина изисквали от кандидатите да се съгласят да работят в режим „996" – от 9 до 21 ч., 6 дни в седмицата. https://profit.bg/article/2025072409250205503

**Тош:** Това показва колко висока е търсеният коефициент на печалбата в тези фирми, въпреки огромните приходи от всеки отделен служител (когато компанията е успешна).

* https://www.hume.ai/blog/introducing-evi-3 EVI-3 : Speech-to-Speech foundation model, 29.5.2025

*** AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms** , 14 May 2025, By AlphaEvolve team

* https://deepmind.google/discover/blog/alphaevolve-a-gemini-powered-coding-agent-for-designing-advanced-algorithms/
* https://developers.googleblog.com/en/gemini-2-5-pro-io-improved-coding-performance/
* https://en.wikipedia.org/wiki/AlphaEvolve

## 6.6.2025 Notes:

https://www.freepik.com/blog/f-lite-freepik-and-fal-ai-unveil-open-source-image-model-trained-on-licensed-data/ | https://github.com/fal-ai/f-lite 5.2025
* Standard 7B, Freepik/F-Lite-7B – 7B parameter version of the standard model
* Lower VRAM requirement

*** Агент за научни изследвания и опити**:

*** NovelSeek: When Agent Becomes the Scientist -- Building Closed-Loop System from Hypothesis to Verification**, NovelSeek Team: Bo Zhang, Shiyang Feng et al., 25.5.2025, 34 p. https://arxiv.org/abs/2505.16938 | https://github.com/Alpha-Innovator/NovelSeek (Renamed to **InternAgent: When Agent Becomes the Scientist -- Building Closed-Loop System from Hypothesis to Verification** in the 22.7.2025 version) **NovelSeek/InternAgent**- *a unified closed-loop multi-agent framework to* ***conduct Autonomous Scientific Research (ASR).*** **12 types of scientific research tasks** .. AI, reaction yield predicition, molecular dynamics, power flow estimation, time series forecasting, transcription prediction, enhancer activity prediction, sentiment classification, 2D image classification, 3D point classification, 2D semantic segmentation, 3D autonomous driving, large vision-language model fine-tuning." Autonomous Scientific Discovery (ASD) - using LLMs and robotics to independently perform scientific research without direct human intervention. .. generating proposals that are both effective and novel .. balancing creativity and rigor .. *design experiments, execute them, analyze results, and iteratively refine their hypotheses in a seamless loop; integration across multiple domains: robotics for experiment execution and advanced analytics for result interpretation; a closed-loop system: robust coordination, adaptability; handling uncertainty; real-world experiments: unexpected variables and noise. NovelSeek transforms a rough proposal into a detailed and easily implementable method.* ***2.1 Self-Evolving Idea Generation with Human-interactive Feedback .. 1) literature review mode and 2) deep research mode … Survey Agent*** *- generate new keyword combinations expanded set;* ***Code Review Agent***: user-provided code; searching for: relevant code-bases; *public repositories like GitHub; static code analysis: Python's* ***ast*** *module – to parse and understand code without execution .. LLMs → human-readable descriptions and summaries.* ***Idea Innovation Agent*** *– LLM configured with a higher temperature setting* **[TA: "creativity", more unusual outputs, explore a broader spectrum of possibilities]** Analyzing the content … generation of refined and innovative ideas … **Assessment Agent –** a structured and multidimensional evaluation process; for each idea, 4 dimensions and scores: coherence, credibility, verifiability, novelty, and alignment. **Human-interactive Feedback**: 1) directly provided by humans and 2) automatically generated by agent .. **Orchestration Agent. (…) 5. Related Works …** Human-AI collaboration in ASR… ***Agent Laboratory*** *(Schmidgall et al., 2025) integrate human feedback into multi-stage LLM agent workflows, automating* ***literature review, experiment execution, and report writing****, while allowing user input at each step to enhance research quality.* ***AgentRxiv*** *(Schmidgall & Moor, 2025) addresses the collaborative nature of scientific discovery by enabling LLM agent laboratories to communicate and build upon each other's work via a shared preprint server, thus facilitating knowledge*

*sharing and collective innovation. .. **AI Co-Scientist** (Gottweis et al., 2025), based on Gemini 2.0, employs a multi-agent system with a "generate-debate-evolve" strategy for hypothesis generation .. **Most current systems** are **still evaluated primarily on relatively simple tasks** or within **narrow scientific domains**. However, when applied to more complex, system-level scientific challenges, these approaches often face significant limitations. **Key challenges** include generating **truly novel** and **scientifically sound** research ideas, establishing **robust closed-loop feedback** between experiments and idea generation, and developing **systematic evaluation standards** to **rigorously assess** the effectiveness and real-world value of autonomous research systems. **6. Future Outlook: Knowledge Retrieval:** establishing connections and relationships between papers; meta-analyses on the search results .. transforming the papers into structured representations such as triples; RAG; **Knowledge Understanding and Representation .. Agent Capability Enhancement: .. dynamically adapt** .. **self-modification**, they can flexibly **redefine** their **initial goals** and **planning strategies while utilizing feedback**, as well as communication logs between agents or between humans and agents, **to train and improve themselves**. This mechanism should focus on **improving their ability to gather feedback** from three key sources: the **environment**, **interactions** with **other agents**, and **human experts. Scientific Discovery-related Benchmark Construction … Appendix B: Evaluation Details;* Scoring, **Rating of papers:** area of AI, subarea; flawless, strong, solid; moderate-to-high, high, excellent, groundbreaking impact; reasons to accept outweigh reasons to reject groundbreaking **Rating 4:** borderline reject: limited evaluation. **R3: Reject:** technical flaws, weak evaluation, inadequate reproducibility, incompletely addressed ethical considerations.. **R2: Strong Reject:** major technical flaws, poor evaluation, limited impact etc. **R1: Very strong reject**: trivial results or unaddressed ethical considerations. **Appendix C NovelSeek Software Development**

See also:
*** Knowledge Navigator: LLM-guided Browsing Framework for Exploratory Search in Scientific Literature**, Uri Katz, Mosh Levy, Yoav Goldberg, 28.8.2024: https://arxiv.org/pdf/2408.15836 – The cluster-based navigation paradigm with LLMs + modern NLP and IR methods; *transform a large corpus of retrieved scientific literature into multi-level, organized themes of subtopics*; *subtopics represent meaningful research clusters, enabling searchers to identify areas of interest, uncover novel connections, and explore specific domains within the broader topic* **Exploratory Information Seeking,** "exploratory search**";** Cluster-based browsing**;** *organize documents into coherent clusters for easier navigation**;** Scatter/-Gather paradigm (Cutting et al., 1992; Pirolli et al.,1996) ..by grouping retrieved documents into clusters and allowing iterative refinement. Topical Corpus construction: ..* a *search query reflecting a relatively broad scientific topic T (e.g. "Tool use in*

*animals"), and results in a topical corpus C comprising the top K documents ranked by a search engine for this query. ..*
*Google Scholar via SerpAPI service* .. could be any other service, as long it provides *a pool of relevant documents.*
[See also the "Knowledge Navigator" visionary project by Apple from 1987: https://en.wikipedia.org/wiki/Knowledge_Navigator ]

## „Понятийни модели" (*според* името):

* **Soft Thinking: Unlocking the Reasoning Potential of LLMs in Continuous Concept Space** Zhen Zhang, Xuehai He, Weixiang Yan, Ao Shen, Chenyang Zhao, Shuohang Wang, Yelong Shen, Xin Eric Wang, https://arxiv.org/abs/2505.15778
21.5.2025 – *Human cognition … abstract, fluid concepts rather than strictly using discrete linguistic tokens. Current reasoning models .. are constrained to reasoning within the boundaries of human language, processing discrete token embeddings that represent fixed points in the semantic space. This discrete constraint restricts the expressive power and upper potential of such reasoning models, often causing incomplete exploration of reasoning paths, as standard Chain-of-Thought (CoT) methods rely on sampling one token per step. In this work, we introduce Soft Thinking, a training-free method that emulates human-like "soft" reasoning by generating soft, abstract concept tokens in a continuous concept space. These concept tokens are created by the probability-weighted mixture of token embeddings, which form the continuous concept space, enabling smooth transitions and richer representations that transcend traditional discrete boundaries. In essence, each generated concept token encapsulates multiple meanings from related discrete tokens, implicitly exploring various reasoning paths to converge effectively toward the correct answer. Empirical evaluations on diverse mathematical and coding benchmarks consistently demonstrate the effectiveness and efficiency of Soft Thinking, improving pass@1 accuracy by up to 2.48 points while simultaneously reducing token usage by up to 22.4% compared to standard CoT. (…)* Continuous Space Reasoning*;* reasoning by replacing discrete one-hot tokens with concept tokens and keeping the entire original probability distribution. … . Using concept tokens allows the model to avoid making hard decisions too early …

**Tosh:** these are not real concepts in the sense e.g. in T.Arnaudov's interpretation or Cognitive Linguistics etc., it is still the usual "tokens" distribution and it's not Всетводейство (Vsetvodeystvo – see **Zrim** and *Creating of Thinking Machines, T.Arnaudov).* The actual progress in their work is just demonstrated <1% to several points/% in the benchmark measures for the usual "Chain of Thought" reasoning tests (MATH500, AIME2024, GSM8K, … p.7). See: T.Arnaudov: **"What's wrong with Natural Language Processing", Part I, II, 2009.** The same pattern of benchmarks

for the next 16 - 17 years.

**Тош:** това не са истински понятия в смисъла напр. на моето тълкуване, когнитивната лингвистика и пр., все още използва същия вид „токени“ и техните вероятностни разпределения, а не Всетводейство (виж Зрим). Показания напредък е <1% или няколко точки от дадени тестове за езикови модели. Сравни със статиите ми от 2009 г. …

* https://venturebeat.com/ai/mistral-launches-new-code-embedding-model-that-outperforms-openai-and-cohere-in-real-world-retrieval-tasks/ – **Codestral**, code retrieval, RAG

* https://venturebeat.com/ai/which-llm-should-you-use-token-monster-automatically-combines-multiple-models-and-tools-for-you/ Token Monster LLM platform – automatically choose best LLM for the user's task

* **Simulated Reasoning – different ways of processing in LLMs compared to human mathematical reasoning in Chain of Thought tasks**; ETH, INSAIT, 26.4.2025 https://arstechnica.com/ai/2025/04/new-study-shows-why-simulated-reasoning-ai-models-dont-yet-live-up-to-their-billing/

*** Proof or Bluff? Evaluating LLMs on 2025 USA Math Olympiad,** Ivo Petrov, Jasper Dekoninck, Lyuben Baltadzhiev, Maria Drencheva, Kristian Minchev, Mislav Balunović, Nikola Jovanović, Martin Vechev https://arxiv.org/abs/2503.21934 https://matharena.ai/

***** MathArena: Evaluating LLMs on Uncontaminated Math Competitions **…** a platform for evaluation of LLMs on the latest math competitions and olympiads. .. rigorous assessment of the reasoning and generalization capabilities of LLMs on new math problems which the models have not seen during training. … a leaderboard for each competition showing the scores of different models individual problems. .. a main table .. performance on all competitions. .. the average score and the cost of the model (in USD) across all runs.
https://github.com/eth-sri/matharena.

* https://www.androidauthority.com/gemini-memory-pcontext-teaser-3550093/ Gemini Memory P-Context … *pcontext," short for personalized context. .. giving Gemini a deeper understanding of your life. "We'll make it easy for you to bring in all of your Google context (Gmail, Photos, Calendar, Search, YouTube, etc.),*

* https://it.dir.bg/tehnologii/sam-altman-tselta-ni-e-chatgpt-da-zapomni-tseliya-vi-zhivot

* https://techcrunch.com/2025/05/15/sam-altmans-goal-for-chatgpt-to-remember-your-whole-life-is-both-exciting-and-disturbing/

*** Sam Altman's goal for ChatGPT to remember 'your whole life' is both exciting and disturbing** Julie Bort 4:05 PM PDT · May 15, 2025

| |
|---|
| **Tosh:** OpenAI/Altman rediscovers the AI projects of the "Sacred Computer" since mid 2000s (just after "Smarty") called Research Assistant/Research Accelerator/Assistant which is conceived in 2007 and is used as a prototype in various forms since early 2010s. The same idea is applied by anyone with good enough memory who is logging his memories. Similar with Microsoft's function for taking screenshots of the user's scren. |

https://nauka.offnews.bg/fizika/mozhe-li-nashata-vselena-da-raboti-kato-kompiutar-201791.html Melvin Vopson; Вселената Сметач – Виж Теория на Разума и Вселената; Малвин Вопсън ? …

[Submitted on 22 Apr 2021 (v1), last revised 5 Aug 2021 (this version, v4)]
* ImageNet-21K Pretraining for the Masses, Tal Ridnik, Emanuel Ben-Baruch, Asaf Noy, Lihi Zelnik-Manor https://arxiv.org/abs/2104.10972
* arXiv:2410.23676v1 [cs.CV] 31 Oct 2024
* **Web-Scale Visual Entity Recognition: An LLM-Driven Data Approach**, Mathilde Caron Alireza Fathi Cordelia Schmid Ahmet Iscen, Google DeepMind https://arxiv.org/html/2410.23676

* Google Releases 76-Page **Whitepaper on AI Agents: A Deep Technical Dive into Agentic RAG, Evaluation Frameworks, and Real-World Architectures** By **Sana Hassan**, May 6, 2025
https://www.kaggle.com/whitepaper-agent-companion - aimed at professionals developing advanced AI agent systems. (…) *Context-Aware* ***Query Expansion****: Agents reformulate search queries dynamically based on evolving task context.* **…** ***Multi-Step Decomposition****: Complex queries are broken into logical subtasks, each addressed in sequence. ..* ***Adaptive Source Selection****: Instead of querying a fixed vector store, agents select optimal sources contextually. ..* ***Fact Verification****: Dedicated evaluator agents validate retrieved content for consistency and grounding before synthesis.* **Modular Reasoning**, **Fault Tolerance, Improved Scalability .. Automotive AI Case Study: 1. Hierarchical Orchestration:** Central agent routes tasks to domain experts. **… Diamond Pattern:** Responses are refined post-hoc by moderation agents. **Peer-to-Peer Handoff:** Agents detect misclassification and reroute queries autonomously**. Collaborative Synthesis:** Responses are merged across agents via a **Response Mixer. Adaptive Looping:** Agents iteratively refine results until satisfactory outputs are achieved.

* https://techcrunch.com/2025/05/05/a-stealth-ai-model-beat-dall-e-and-midjourney-on-a-popular-benchmark-its-creator-just-landed-30m/

*** 'AI is already eating its own': Prompt engineering is quickly going extinct** https://www.fastcompany.com/91327911/prompt-engineering-going-extinct 6.5.2025, APPLIED AI – *Three years into the boom, it looks like AI is reshaping existing jobs more than creating new ones.* [They count the "AI boom" since ChatGPT] https://www.fastcompany.com/section/applied-ai

*** A first-principles mathematical model integrates the disparate timescales of human learning,** Mingzhen Lu, Tyler Marghetis, Vicky Chuqiao Yang 2.5.2025 *npj Complexity* **volume 2**, Article number: 15 (2025) Cite this article **https://www.nature.com/articles/s44260-025-00039-x**
* https://searchengineland.com/ai-killing-web-business-model-455157
* https://www.tomshardware.com/tech-industry/artificial-intelligence/software-engineer-taught-microsoft-copilot-to-analyze-windows-crash-dumps
***** https://www.marktechpost.com/2025/05/06/llms-can-now-talk-in-real-time-with-minimal-latency-chinese-researchers-release-llama-omni2-a-scalable-modular-speech-language-model/
* LLMs Can Now Talk in Real-Time with Minimal Latency: Chinese Researchers Release LLaMA-Omni2, a Scalable Modular Speech Language Model, By Asif Razzaq -May 6, 2025

*** LLaMA-Omni2 – Speech Encoder and Adapter; Core LLM, Streaming TTS Decoder.** Streaming Generation with Read-Write Scheduling; Text-to-Speech synthesizer, fast
https://arxiv.org/abs/2505.02625 https://huggingface.co/collections/ICTNLP/llama-omni-67fdfb852c60470175e36e9c https://github.com/ictnlp/LLaMA-Omni2

* How College Professors Can Easily Detect Students' AI-Written Essays, By StudyFinds Staff, 6.5.2025 https://studyfinds.org/chatgpt-human-essay-writing/

* Hugging Face releases a free Operator-like agentic AI tool, Kyle Wiggers, 3:00 PM PDT · May 6, 2025 - **Open Computer Agent … Open Computer Agent,** https://huggingface.co/spaces/smolagents/computer-agent ***Powered by smolagents*** **… Enter your task below: (**Top left) **Example tasks: "**Use Google Maps to find the Hugging Face HQ in Paris … Go to Wikipedia and find what happened on April 4th …Find out the travel time by train from Bern to Basel on Google Map, … Go to Hugging Face Spaces and then find the Space flux.1 schnell. Use the space to generate

an image with the prompt 'a field of gpus" ***Powered by open source:*** smolagents Qwen2-VL-72B E2B Desktop

***** https://github.com/e2b-dev/desktop – E2B Desktop Sandbox for LLMs. E2B Sandbox with desktop graphical environment that you can connect to any LLM for secure computer use.
Open-source Code Interpreting for AI Apps — E2B https://e2b.dev/ – RUN AI-GENERATED code SECURELY in your APP E2B is an open-source runtime for executing AI-generated code in secure cloud sandboxes. Made for agentic & AI use cases.
https://www.cio.com/article/3979014/12-reasons-to-ignore-computer-science-degrees.html

*** Mem0: A Scalable Memory Architecture Enabling Persistent, Structured Recall for Long-Term AI Conversations Across Sessions** By **Asif Razzaq**, April 30, 2025
– two-step process to extract and manage salient conversation facts, combining recent messages and global summaries to form a contextual prompt; – memory as a directed graph of entities and relationships; *Mem0 surpassed OpenAI's memory system with a 26% improvement on LLM-as-a-Judge … for AI assistants in tutoring, healthcare, and enterprise settings where continuity of memory is essential.*
https://arxiv.org/abs/2504.19413 *[Submitted on 28 Apr 2025]*

*** Mem0: Building Production-Ready AI Agents with Scalable Long-Term Memory,** Prateek Chhikara, Dev Khant, Saket Aryan, Taranjeet Singh, Deshraj Yadav 28.4.2025
*Nodes V represent entities (e.g., Alice, San Francisco) • Edges E represent relationships between entities (e.g., l ives _ in) • Labels L assign semantic types to nodes (e.g., Al ice - Person, San _Fr anc isco – City .. people, locations, objects, concepts, events, and attributes .. The entity extractor identifies these diverse information units by analyzing the semantic importance, uniqueness, and persistence of elements in the conversation.... in a conversation about travel plans, entities might include destinations (cities, countries), transportation modes, dates, activities, and participant preferences—essentially any discrete information that could be relevant for future reference or reasoning; .. the relationship generator component derives meaningful connections between these entities, establishing a set of relationship triplets that capture the semantic structure of the information. .. establishing a set of relationship triplets that capture the semantic structure of the information. .. LLM-based module analyzes the extracted entities and their context within the conversation to identify semantically significant connections. .. classifies this relationship with an appropriate label (e.g., 'lives_in', 'prefers', 'owns', 'happened_on').*
**Appendix:** ReadAgent – emulating how humans process lengthy texts through a sophisticated three-stage pipeline. 1) Episode Pagination segments text at

natural cognitive boundaries rather than arbitrary cutoffs 2) Memory Gisting preserves essential meaning while reducing token count; Interactive Lookup mechanism: the human-inspired approach enables LLMs to effectively manage documents up to 20 times longer than their normal context windows. MemoryBank, Memory Storage component warehouses detailed conversation logs, hierarchical event summaries, and evolving user personality profiles; a dual-tower dense retrieval model to extract contextually relevant past information. **Memory Updating component …** a human-like forgetting mechanism – memories strengthen when recalled and naturally decay over time if unused **…** MemGPT – an operating system-inspired approach to overcome the context window limitations inherent in LLMs**;** a sophisticated memory management pipeline consisting of three key components: a hierarchical memory system, self-directed memory operations, and an event-based control flow mechanism. 'Main context' is analogous to RAM in traditional OS and 'external context' (disk storage)**.** W*hen the LLM needs information not present in main context, it can initiate function calls to search, retrieve, or modify content across these memory tiers, effectively 'paging' relevant information in and out of its limited context window.* The memory management creates *the illusion of infinite context, significantly extending what's possible with current LLM technology*.

*The **A-Mem** model introduces an agentic memory system designed for LLM agents. This system dynamically structures and evolves memories through **interconnected notes**.* Each note captures interactions enriched with structured attributes like keywords, contextual descriptions, and tags generated by the LLM. .. semantic embeddings to retrieve relevant existing notes, meaningful links based on similarities and shared attributes .. the **memory evolution mechanism** updates existing notes dynamically, refining their contextual information and attributes whenever new relevant memories are integrated ..
*The LOCOMO (Maharana et al., 2024) dataset is designed to evaluate long-term conversational memory in dialogue systems. It comprises 10 extended conversations, each containing approximately 600 dialogues and 26000 tokens on average, distributed across multiple sessions…*

**Tosh:** The OS-like part: Compare to **Zrim**, {К}?Т : Зрим, {К}?Т, контексти на търсене; изпреварващо търсене и Обща обработка (ОбОбр). Верига от частично съвпадение, пт(мс), пс(мс); РСВ, РСУ, [,,]

*** Google AI Unveils 601 Real-World Generative AI Use Cases Across Industries**, By **Asif Razzaq**, April 26, 2025
https://www.marktechpost.com/2025/04/26/google-ai-unveils-601-real-world-generative-ai-use-cases-across-industries/ * Customer Agents: Enhance user experiences via chatbots, predictive services, and personalization
* Employee Agents: Boost internal productivity through content generation,

summarization, and knowledge discovery
* Creative Agents: Accelerate campaign design, media production, and product innovation
* Code Agents: Streamline software engineering and IT workflows
* Data Agents: Leverage data for analysis, optimization, and decision support
* Security Agents: Fortify organizations with AI-driven threat detection and fraud prevention.
**Industry snapshots:** Automotive & Logistics, Financial Services, Healthcare & Life Sciences ... "**Generative A**I is moving from experiments to mission-critical systems: Whether automating underwriting in finance, driving drug discovery, or powering multimodal search in automotive apps, GenAI is now operational at scale. **Hybrid Multimodal Models** are increasingly vital: Many solutions integrate text, vision, and structured data — not just plain language models. **Verticalized AI Agents** are accelerating: Google's partners aren't just fine-tuning LLMs — they're building domain-specific, industry-tuned AI agents tightly integrated into their workflows. **Democratization of AI:** Solutions like Vertex AI's search and data agents are putting sophisticated AI tools into the hands of business users, scientists, and even drivers — not just engineers.
**Technology Highlights:** Google's Evolving Stack: Vertex AI: Model training, deployment, RAG (retrieval-augmented generation) pipelines Gemini Models: Multimodal LLMs powering text, code, vision, and conversational capabilities; Imagen & Veo: High-fidelity generative image and video models; BigQuery ML: Data warehousing with embedded machine learning; Security AI: AI-first threat detection with Google SecOps.

* **Real-world gen AI use cases from the world's leading organizations** | Google Cloud Blog https://cloud.google.com/transform/101-real-world-generative-ai-use-cases-from-industry-leaders

* **How much do language models memorize?,** John X. Morris, Chawin Sitawarin et al. AIR at Meta, Google DeepMind, Cornell University, NVIDIA https://arxiv.org/pdf/2505.24832 2.6.2025 – **GPT family, ~3.6 bits-per-parameter ..** 2. Memorization, intended and unintended; unique data is more likely to be memorized*; "intended memorization" – "generalization"; sample level memorization.

[**Tosh:** That's aligned with the ideas that this is **specifics,** lower generality. The repetititve (similar) is **generalized** to more abstract concepts, what can't be generalized is remembered and operated at higher resolution of causality and control. See T.Arnaudov "AGI Digest", 2012 "Chairs, Buildings, Caricatures ..." and Theory of Universe and Mind, 2001-2004+.

* https://venturebeat.com/ai/how-much-information-do-llms-really-memorize-now-we-know-thanks-to-meta-google-nvidia-and-cornell/ – from 500K to 1.5 billion parameters: 3.6 bits/paramerer; bfloat16 vs float32: 3.51 to 3.83 bits-per-parameter for the 32-bit.
**** A 500K-parameter model can memorize roughly 1.8 million bits, or 225 KB of data.***
*** A 1.5 B ~ 5.4 Gbits ~ 675 MB of raw information.**
This is not comparable to typical file storage like images (e.g., a 3.6 MB uncompressed image is about 30 million bits), but it is significant when distributed across discrete textual patterns."]

**[Tosh:** Use these estimates as a sample guide for the complexity of other systems. Another hint: the resolution of compressed images which are still correctly classified and the size of the compressed files, either for single files or together or on average for a set of recognized files. E.g. 40x30 etc. Use this as a direction, starting point and example of possible complexity of the constraints for describing the image or the concepts with other representations.
The specific amounts also reflect properties of the text and human cognition and the applied technologies which encode it like that. How much is generalized in the linguistic expressions, how general the word meanings are, the phonological loop and working memory/attention span etc., the repetitions in the datasets and in the phenomena which are recorded as text and fed to the ML models.]

**Computer Use sandbox**
https://github.com/e2b-dev/desktop – GUI, desktop, Linux Xfce.
https://github.com/e2b-dev/desktop/tree/main/examples/basic-python

**Gymnasium, RL, video games, world models**

https://github.com/Farama-Foundation/Gymnasium
https://github.com/Farama-Foundation/stable-retro - old video games, a fork of Gymnasium

*** Exploration-Driven Generative Interactive Environments**, **Nedko Savov1**,† Naser Kazemi1, Mohammad Mahdi1 Danda Pani Paudel1, Xi Wang1,2,3 Luc Van Gool1, 1 INSAIT, Sofia University "St. Kliment Ohridski" 2 ETH Zurich 3 TU Munich, 3.4.2025, https://arxiv.org/pdf/2504.02515v1 - *using many virtual environments for inexpensive, automatically collected interaction data .. AutoExplore Agent - an exploration agent that entirely relies on the uncertainty of the world model, delivering diverse data from which it can learn the best.. …* GenieRedux, GenieRedux-G; StableRetro; Controllability; Prediction Horizon: collecting sequences of frames; tokenization, encoding, reconstruction; a Frame from the game → Tokenizer → Codebook → MaskGIT → Sample → Codebook → Tokenizer Decoder → Prediction … The entropy of the distributions in MaskGIT is a reward for the AutoExplore Agent → collected to a dataset. Training phase & Exploration phase.
Multi-Environment World Model, Latent Action Module; training: a sequence size of 16 frames @ 64x64; U-Net *superresolution network on 50K data samples to upscale the output to 256x256. (Sup.Mat. B)* RetroAct Dataset. *Five motion actions: left, right, up, down and jump.* They *generate a short clip of each of the 5 selected actions for each of the 483 titles and build an annotation tool to observe and annotate the executed action. Eventually, we annotated 2,925 behavior and 2,898 control labels. ..* Platformers-200 - a dataset = 10,000 episodes (50/game), up to 500 frames = 4.6M img. Platformers-50: 5000 episodes (100/game) of at most 1000 frames = 4.8mln images. Provide 3 images, predict 2 after them. Their agent is *an actor-critic, trained with the Policy Gradient method.* For *visual fidelity evaluation* [of the reconstruction of the predicted frames]*:* FID (Frechet inception distance), PSNR (signal-to-noise ratio), SSIM (structural similarity index measure). To evaluate controllability, they use: Delta t PSNR metric: comparing the visual effect of the ground truth action versus a random action…
2. Related work: World models**…** See the references:
* [10]Silvia Chiappa et al**., Recurrent environment simulators**., In International Conference on Learning Representations, 2017.
* [19] David Ha and Jurgen Schmidhuber. **World models.** ¨ arXiv preprint arXiv:1803.10122, 2018.
* [21] Danijar Hafner et al. **Learning latent dynamics for planning from pixels**., 2019
* [39] Vincent Micheli, Eloi Alonso, and Francois Fleuret**. Transformers are sample-efficient world models**., 2023;
* [50] Jan Robine et al. **Transformer-based world models are happy with 100k interactions. 2023.**
* [64] Sherry Yang et al. **Video as the new language for real-world decision making**. arXiv preprint arXiv:2402.17139, 2024
* [22] Danijar Hafner et al**. Mastering atari with discrete world models**. 2021.
* [27] Anthony Hu et al. **Gaia-1: A generative world model for autonomous driving.** arXiv preprint arXiv:2309.17080, 2023.
* [37] Willi Menapace et al. **Playable video generation**, 2021.
* [65] Ze Yang et al., **Unisim: A neural closed-loop sensor simulator.**
* Huiwen Chang et al. **Maskgit: Masked generative image transformer**, 2022.

Etc.

**Tosh, 11.6.2025:** Compare to the **Theory of Universe and Mind**, 2001-2004 and the plans in Universe and Mind 5 (2004) and their continuations in the proposed R&D programs from late 2007 and early 2008, as well as "What's wrong with NLP?" I,II, early 2009 by The Sacred Computer/Artificial Mind regarding the imagination and will which were lacking in the mainstream NLP. Video is a more *precise* world model than one which reflects only fewer extreme features such as corners, contours etc. or lacking the temporal dynamic etc. (however video is **not enough without colliders** and proper **physics** as the game developers know).

## * Проекти за символен УИР #symbolicAGI

*** Development and Architecture of REFPERSYS: A Multi-Threaded REFlective PERsistent SYStem for AI**, Basile Starynkevitch & Abhishek Chakravarti & Nimesh Neema http://refpersys.org http://refpersys.org/ **RefPerSys** Extensible code, writes extensions and jump to them etc. ; "Staircase Development Model"; metaprogramming, introspection; most ojects are mutable; declarative knowledge & bootstraping
* Виж също https://en.wikipedia.org/wiki/CAB_500 и езикът за програмиране PAF, разработен за един от първите френски лампови сметачи през 1957-1959 г. от Dimitri Starynkévitch, може би баща на Васил Старинкевич (Базил във френски вариант).
http://www.feb-patrimoine.com/english/paf_starynkevitcha.htm
* A language designed at SEA, Programming on CAB500, PAF Programmation Automatique des Formules, Dimitri STARYNKEVITCH
*** Le projet RefPerSys, un successeur potentiel du système Caia de Jacques Pitrat,** Basile Starynkevitcha , Abhishek Chakravarti, 2021
http://refpersys.org/Starynkevitch-RefPerSys.pdf
https://github.com/RefPerSys/RefPerSys

*** INSA: Integrated Neuro-Symbolic Architecture, The Third Wave of AI, a Path to AGI, Peter Voss,** Jul 06, 2024 https://petervoss.substack.com/p/insa-integrated-neuro-symbolic-architecture – Свръхбърз граф за вектори, БД за графи; … области за УИР: Учене на модели-схеми (pattern learning), Образуване на понятия, Памет (краткотрайна и дълготрайна: STM/LTM); Внимание и отбор [на стимули и пр.], Заключение/Разсъждение, Метапознание, Цели и действия, Контекст (обслов) … INSA: Обединена невро-символна архитектура, третата вълна на ИИ, път към УИР. Векторната им структура била 1000 пъти по-бърза от др.комерсиални, може да съхранява цели записи на възприятия, отделни особености, понятия, символи; отношения в контекста, редици (последователности), йерархии и други структури във времето, статични и динамични. Векторите позволяват размито сравнение. … Разпространяващо се

възбуждане (още „*разлъчващо* се“, spreading activation) … предвиждане, обобщаване, метапознавателни сигнали … - дълбоко вградени. Може да зарежда в реално време и символни данни, логика, онтологии, БД и неструктурирани данни. Постепенно обучение в реално време от един или малко примери – фокус върху качеството на данните, а не количеството, премахва нуждата от мощни ускорители. … Приспособяваща се векторна структура с променлив размер – позволява онтологично кодиране, а не статистическо (основано [само] на съседство на думите както е при обичайното представяне в езиковие модели (embeddings). Обща структура и за ниското, и за високото ниво признаци, както и за символни знания, позволява плавен преход от поведение от типа стимул-реакция, „Система 1“ до изрично разсъждение и планиране на въздействията („Система 2“ по Канеман). Подробностите по реализацията били тайна.
**Коментар от Микола Рабчевски, 11.7.2024:** според него все още нерешени в УИР били образуването на понятия и ученето на модели-схеми, невро-компонентите не обещавали нищо ново. Търсенето на зависимости (модели-схеми) било комбинаторна задача – пораждане на хипотези и проверка. За другата задача – невр.мрежи могат да бъдат детектори на предварително определени основни признаци, но това е изпълнимо и по други „класически“ начини.

## * Терминология – бележки

Термин за **учене с подкрепление** от маг.програма на СУ: „Обучение по метода „**поощрение*/наказание***“

* https://www.fmi.uni-sofia.bg/sites/default/files/presentations/master/presentacia_mp_izkint_2024_4.pdf
https://www.fmi.uni-sofia.bg/bg/izkustven-intelekst-4-semestra

## * Multi-Agent frameworks with LLMs

Meta-Agentic α-AGI Demo – Production-Grade v0.1.0 - https://github.com/MontrealAI/AGI-Alpha-Agent-v0/tree/main/alpha_factory_v1/demos/meta_agentic_agi
**Meta-Agentic (adj.)** Describes an agent whose **primary role** is to **create, select, evaluate, or re-configure other agents** and the rules governing their interactions, thereby exercising **second-order agency** over a population of first-order agents.The term was pioneered by Vincent Boucher, President of MONTREAL.AI.

## * Симулиране на личности с езикови модели. Social science experiments with LLMs. Simulated personality.

*** Scaling Law in LLM Simulated Personality: More Detailed and Realistic Persona Profile Is All You Need,** Yuqi Bai, Tianyu Huang, Kun Sun, Yuting Chen, 10.10.2025 – LLMs…*Simulate social experiments, exploring their ability to emulate human personality in virtual persona role-playing; an end-to-end evaluation framework, including individual-level analysis of stability and identifiability* and *population-level analysis: progressive personality curves to examine the veracity and consistency of LLMs in simulating human personality; modifications to traditional psychometric approaches (CFA and construct validity); empirically demonstrating the critical role of persona detail in personality simulation quality; identifying marginal utility effects of persona profiles;*

## * Системи за препоръчване на съдържание

Препоръчването на съдържане е вид *изпреварващо търсене* в именуването на *Свещеният сметач* и *Зрим* от средата на 2010-те г. Обзор от края на 2024 г.:

**История:** Elen Rich,1979 – предложения за препоръчване на книги. Jussi Karlgren - 1990 г. , „цифрова лавица за книги“.

* **A COMPREHENSIVE REVIEW OF RECOMMENDER SYSTEMS: TRANSITIONING FROM THEORY TO PRACTICE**, Shaina Raza∗1 et al. 13.7.2024 http://arxiv.org/pdf/2407.13699v1 – Разглежда развитието на теорията и практиката от 2017 до 2023 г. Табл.1: теми на други обзори по различни теми и техники: Economics, Stock market, Digital marketing, Finance, Multimedia, Travel, Health, E-learning, Machine learning, Knowledge integration, Explainability, Context awareness, Collaborative filtering, Hybrid methods, Sequence awareness, Session integration, Conversation integration, Music, Reinforcement learning, Adversarial methods, Review texts, Graph neural network, Deep learning, Large Language Models, Aspect integration, General, News, Privacy, Tourism, Evaluation, Trustworthiness, Cultural Heritage. ...

## * Модели и платформи за пораждане на изображения

DALL-E 3 ..., StableDiffusion ... SDXL, OpenDALLE, ...MidJourney ... ERNIE-ViLG .. Google DeepMind Imagen 3 (2025), Imagen 4
* https://developers.googleblog.com/en/imagen-3-arrives-in-the-gemini-api/

* https://deepmind.google/models/imagen/
* ERNIE-ViLG 2.0: Improving Text-to-Image Diffusion Model with Knowledge-Enhanced Mixture-of-Denoising-Experts https://arxiv.org/pdf/2210.15257 ;; Qwen3-ImageEdit (23.8.2025)

(…)

## * Модели и платформи за пораждане на видео и физически верни модели на света или симулации по словесна подкана

* Wan – Alibaba Group, Free / Open weights, 14 B, 1.3 B
* Wan: Open and Advanced Large-Scale Video Generative Models, Team Wan ..et al., 26.3.2025/19.4.2025 ... ... https://arxiv.org/abs/2503.20314 – *The 1.3B model demonstrates exceptional resource efficiency, requiring only 8.19 GB VRAM*
Използван напр. със средата ComfiUI.
...
* OpenAI SORA https://sora.chatgpt.com/explore/videos
* KlingAI - https://klingai.com/global , ...
* Google DeepMind Veo 3: https://gemini.google/overview/video-generation/
Veo 3.1, Flow: 15.10.2025: https://blog.google/technology/ai/veo-updates-flow/ – user can add assets, location, background, ending; up to one miute of coherent video with sound.
* World Model Generator (game engines, simulators):
* Meta & openai: sora – free video generation (10.2025)
* https://sora.chatgpt.com/explore * https://www.meta.ai/

**Genie 3:** * https://deepmind.google/discover/blog/genie-3-a-new-frontier-for-world-models/ (Earlier versions: GameNGen (Resolution: 320x, Games; a few seconds interaction horizon); Genie 2 (360p), 3D 10-20 sec horizon; Veo: 720p to 4K, 8 seconds; Genie 3: 720 p, multiple minutes horizon)
*** Genie 2,** 4.12.2024: https://deepmind.google/discover/blog/genie-2-a-large-scale-foundation-world-model/
* **GameNGen**
*** Diffusion Models Are Real-Time Game Engines**, Dani Valevski, Yaniv Leviathan, Moab Arar, Shlomi Fruchter, 27.8.2024/24.4.2025, Google, https://arxiv.org/pdf/2408.14837 - 20 fps on a single TPU-v5 ? Trained on 128 TPU-v5e with data parallelization; 700000 training steps, batch = 2048; *random subset of 70M examples from the recorded trajectories plaed ... 320x240 padded tp 320x256 ... context length = 64 (at least 64 own predictions and the last 64 actions)*
*“GameNGen is trained in two phases: (1) an RL-agent learns to play the game and the training sessions are recorded, and (2) a diffusion model is trained to produce the next frame, conditioned on the sequence of past frames*

*and actions." ... "Several important works (Ha & Schmidhuber, 2018; Kim et al., 2020; Bruce et al., 2024) (see Section 6) simulate interactive video games with neural models." .. "Can a neural model running in real-time simulate a complex game at high quality?"**

* https://gamengen.github.io/ – *"Training the Generative Diffusion Model: We re-purpose a small diffusion model, Stable Diffusion v1.4, and condition it on a sequence of previous actions and observations (frames). To mitigate auto-regressive drift during inference, we corrupt context frames by adding Gaussian noise to encoded frames during training. This allows the network to correct information sampled in previous frames, and we found it to be critical for preserving visual stability over long time periods."*

*** Dani Valevski:**
https://scholar.google.com/citations?user=ECKZ08wAAAAJ&hl=en

*** On the inefficiency of generating games this way with neural models**, Todor Arnaudov, 24.8.2025: A neural model like this can simulate such a game and later ones may simulate more modern ones, but this way of implementation is perversely inefficient. **Doom** can run even on a **386** CPU (275K transistors, say 33 or 40 MHz) and it plays smoothly on a 486 CPU, 1.2 M transistors, say 33 or 66 MHz, with 8 MB RAM. The whole game is a few MBs. The ANN of GameNGen is trained on **~ 900 M frames with records from the actual game*3,** i.e. it is a surrogate that is possible *only after* the game that "*it re-*creates" is already available in finished form and somebody has played it for ~ 12500 hours (at 20 fps), *all* the screens were recorded etc. – the game has already existed and you already has its complete unfolded future in his dataset and in another kind of executable simulation system. This is a general problem with that kind of ANN, ML, the "stochastic parrots" etc.

**„If you've seen one, you've seen all!"**[184]

One thing about similar *generations* of games in the *original* meaning of "*generation*", regarding video games[185], including the "code generation" generators of games, is that the "new" games are actually *the same* as the old ones with slightly changed appearance or with the same appearance as in this

[184] I remember a similar self-ironizing quote from the movie "Singing in the Rain" regarding the Hollywood movies.

[185] The generations of home video game consoles with similar or a given window of capabilities: the first generation 1: Magnavox Oddyssey & Pong; 2: Atari 2600, Intellivision; 3: NES/Famicom, Sega Master System …; 4: 16-bit: Sega Genesis/SNES (with add-ons: +CD/"32X" – "semi-generations); 5: Sony Playstation, Sega Saturn, Nintendo 64 etc.

one; they fit their genres and generation – first, second, third video game console; time period, a derivative of …

This is evident in all generations of video games, but in a simple and simply measurable way that is true especially in the early generations of arcade video games and home TV-consoles, 8-bit or 16-bit.

There is a set of types of games, which may have more or less distinct view and feel and appear as "groundbreaking", but there is also "style" and the "radically new" pieces set the styles which the other instances follow. There are limitations of what can be done with the particular hardware: number of visible sprites and backgrounds, image complexity and transformations (can it rotate, scale, …), sound etc. All in the genre have similar physics and expected scenarios – jump and run on blocks, falling, "collision detection", shooting, "enemies" attacking or passing by, there's a "boss" at the end, music changes in different scenes, an indication of the health points and number of lifes etc.

The mind generalizes it and a new game can be represented or perceived as the older game with replaced sprites, both are just a set of "moving objects" in a more abstract view. The essence is the coordinates of these objects, their states and the goals such as survive, make more points, reach the end, "destroy" the enemies, i.e.: shoot them, i.e. "press the corresponding button" which causes your "projectiles" to spawn from your "weapon", and these projectiles should hit the enemies; or the borders of the bottom part of your figure/sprite ("collision box") has to hit the enemy in a certain way, e.g. from the top as in "Super Mario" or if you touch them horizontally, you lose "health" or "die" – i.e. the number of "lifes" is reduced, and if it reaches "0" you could "continue" or this is "game over".[186]

Different games could be generated with prompts in the 1980s if that was a goal and somebody wanted – probably it would be called "*templates*" at the time, and games from the same studios perhaps really used the previous "instantiations" and versions of their own games as templates.

At least since somewhere in the late 1990s?[187] there were game engine editors, which required little or no actual "coding" (programming with *typing*, as it is usually understood) or only a minimal amount of scripting – a simpler kind of programming – and adding sprites and background tiles.

The game engines, used for developing multiple games, are sort of that

[186] This phenomenon goes for everything is discussed at least by Kant in the 18-th century regading the genius which does not work by following the established rules and the "schools", the styles, which follow the rules and they repeat similar styles.
[187] https://creatools.gameclassification.com/EN/creatools/52-Game-Maker-1.1/index.html - GameMaker, 1999. https://en.wikipedia.org/wiki/GameMaker

anyway, they were generalization of using the previous games as templates.

The textures and sprites of engines like “Wolfenstein” or “Doom” are replaced, minor edition of the code or scripts could be made and the product is released as “a new game”, for example “Heretic”. In next iterations: “Quake” – now the geometry – the 3D-models of the characters – and a more complex maps have to be developed, BSP-trees computed etc., but minor editions could “generate” new games with some set and amount of natural language interface too: “Create a game with …” – pick templates, modify, add, …

The overall range of possible games, variants, specified with general or “simple” prompts, with ones referring to known and “general” parameters however would result in similar and expected “jaggy” results, with different creators and authors producing similar outputs as both their prompts and the templates would be similar, either if it’s with a more realistic or more pixelated cartoonish graphics. In fact the same happens even if the creative persons can operate the computers at instruction-level and pixel-level – in a given period of time they have access to similar resources and their target audiences are similar.

I had another related *similar* experience when reviewing Qwen3-Coder. * https://qwenlm.github.io/blog/qwen3-coder/ See “Use Cases”, **Example: Solar System Simulation** (6/7). It has generated a program, I guess in JavaScript or a derivative, TypeScript?, which shows a cartoonish animation of the Sun and the planet, supposedly rotating with appropriate relative speed etc. That’s great, but what about “prompting” Google Search with*: “Solar System Simulation”?* This another “AI” and “chatbot” produces a set of links, where the one on top is: https://www.solarsystemscope.com/

It is free, shows 3D in the browser, there’s music and sound, more details, you can adjust the time, fly in the 3D space and zoom out, click on the celestial objects and see them centered and read information about each of them, including the dwarf planets Ceres, Haumea and the “decomissioned” Pluto. The web application simulates the stars as well as a planetarium etc. You can discover other simulators and star maps as Android applications, again with a “natural language prompt” in the Google Playstore: type what you want; the same with Apple products. As humans have similar minds, live in the same society, use the same and similar languages, they have similar needs, use similar products, they generate similar ideas and if these records are stored and searchable, if your idea is short and simple enough – such as a few words of text – once you ask the “oracle”, it is highly likely that you will find your own thoughts or desires already experienced and expressed by somebody else.

*1 I remember a similar self-ironizing quote from the movie “Singing in the Rain” regarding the Hollywood movies.
*2 This phenomenon is valid for *everything* and it is discussed at least by Kant in the 18-th century regarding the genius which does not work by following the established rules, but by breaking them. The “schools”, the styles follow the rules and thus repeat similar styles. That’s a pronounced difference between the genius and the “talent”. It is discussed also in many notes in *“The first modern AI strategy was published by an 18-year old Bulgarian in 2003 and repeated and implemented by the whole world 15-20 years later: The Bulgarian Prophecies: How would I inverst one million with the greatest benefit for the development of my country?”,* which is an appendix to *“The Prophets of the Thinking Machines: …”*
*** 900M frames:** https://www.infoq.com/news/2024/09/google-gamengen/ etc. I didn’t find it explicitly in the paper; **70 M** frames are mentioned in regarding the training of the agent, the system has two parts,
* A note from 23.8.2025, recalling ideas from theTheory of Universe and Mind, “Universe and Mind 3”, 8.2003 about generative models. The reason was commenting “*Genie 3”* world model generator.

**Tosh:** The ratio of the complexity of the prompt vs the complexity of the generated output, in examples with magic fish etc. and genies which fulfil wishes*: "I want a castle"*: a castle appears in front of you.

However *who* decides how *__exactly__* the castle will look, what is the "blueprint" what material for the building, size, exact locations, furniture, does it include a queen and how many mistresses, guards etc., how exactly they should look, behave etc. This is not specified in the wish or the prompt and it is up to the genie to decide: therefore, *he* is the actual master and designer. In other words, the "creative" users who create with these systems with so short queries are *the least creative part* of the system, carrying the fewest bits of complexity. If you do understand the details and you are able to specify, express and address them, for example via a “data bus”, direct memory access etc., then such kinds of models are not needed, you can say or rather encode and build *exactly* what you want, down to a pixel, an atom or a point in the 3D-mesh.

*** Universe and Mind 3. T.Arnaudov, 2003:**

**31. Faith is a desire the secondary virtual worlds to be recreated in the primary virtual world** – in the Reality. It is the world, which is the richest of information. “The dream to come true” also means to receive more information

about what you imagine. The idea can't be as rich – complete and detailed – as the reality, because the human mind lacks enough memory for detailed description of the reality, and the dream, if it is complex, allows for big "deviations".

32. In the fairy tales, when the hero wishes a palace, he receives one... From a single word – "a palace", which can be **described in the human memory** or **in the memory of a computer** with, say, 30 bits, a whole building is outputted. However, in order to input the "palace" in the Memory of the Universe Computer, in the Reality, a **description of the exact structure of each particle** of the palace is required! Therefore, the Golden fish or the Genie from the lamp must have a really huge memory with a great throughput... ;-)

**– End of the quote  from 2003 –**

24.8.2025: **As it happened** – see the GPUs and the memory bandwidth of the GPUs in hundreds of GBs or even TBs per second, as well as the scale of the datasets with images, videos, text and all kinds of data for the current generative models. However, I argue that generative AI can be implemented with much less. See future work.

**Original text in Bulgarian:** https://eim.twenkid.com/old/3/25/pred-3.htm https://www.geocities.ws/eimworld/3/25/pred-3.htm  etc.

"31. Вярата е желание вторичните въображаеми светове да се претворят в първичния въображаем свят - Действителността. Тя е най-богатият на въобраз (информация) свят. "Мечтата да стане действителност" означава и да получиш повече информация за това, което си представяш. Представата не може да се мери по богатство - пълнота на подробностите, с действителността, защото на човешкия разум не му достига памет за подробно описание на действителността и мечтата, ако е сложна, допуска големи "отклонения".

32. В приказките, щом героят си пожелае дворец, получава такъв... От една дума - "дворец", която може да ДА СЕ ОПИШЕ В ЧОВЕШКАТА ПАМЕТ или В ПАМЕТТА НА СМЕТАЧ с, да речем, 30 бита, се извежда цяла сграда. За да се въведе "дворецът" в
Паметта на Вселенския Сметач, в Действителността, обаче, е необходимо

ОПИСАНИЕ на ТОЧНИЯ СТРОЕЖ на ВСЯКА ЧАСТИЦА от него! Така че Златната рибка или Духът от лампата трябва да имат наистина огромна памет с огромна пропускателна способно[с]т... ;-)”

*** TPU-v5,** https://cloud.google.com/tpu/docs/v5e – How powerful it is?

| | |
|---|---|
| Peak compute per chip (bf16) | **197 TFLOPs** |
| HBM2 capacity and bandwidth | 16 GB, 819 GBps |
| Interchip Interconnect BW | 1600 Gbps |

*** Doom timedemo 386 DX 40 MHz DOS PC**, PhilsComputerLab 149 хил. Абонати, 5531 показвания 10.10.2014 г.
https://www.youtube.com/watch?v=TfKQ-uREnsI
* PC: **386 DX40/8 MB RAM, 256 K cache** ...
***“How much is it in fps” ~ 8.7 fps***
*“386 was too slow for DOOM, ideally to run full screen max detail you should use something like a 66 MHz 486DX2, or even better a 100 MHz IntelDX4 (or AMD equivalent), with **8MB of RAM**, (4MB was not enough to avoid HDD continu ous access).”*
*@gentlemanvanilla Ran great on my 386 dx40. However I had 16mb ram...”*

**Performance:** https://en.wikipedia.org/wiki/Intel_DX2
“486 DX2/66 performed **54 MIPS** (Dhrystone V1.1).”
**Todor:** The performance in FLOPS for real tasks of the 486 DX2/66 would be **in the range of 2-4-6 MFLOPS?** (FADD, FSUB: 8-20 cycles/clocks, FMUL: 11, FDIV: 73, FSQRT: 83-8; FLD/FST mem32 3/7 c; FCOM: 4, FABS: 3, FCHS: 6; FSIN, FCOS: 193-279, FPTAN: 200-273, FPATAN: 218-303 (partial tan/arctan). 22 MFLOPS for the simplest instructions. For physics simulations and 3D-graphics with a lot of FMUL, FDIV, FSQRT, FSIN/FCOS, FPATAN (if not using tables), the performance could be up to **3-6 MFLOPS** in mixed load/LINPACK or **below MFLOPS,** depending on the instruction mix with a lot of divisions, SQRT, trigonometry or in a neural-network style of code with mostly multiplications and additions of matrices: FMUL, FADD.
*** Intel 486 Microprocessor, November 1989**
https://bitsavers.trailing-edge.com/components/intel/80486/240440-002_i486_Microprocessor_Nov89.pdf p.145 - ...

Compare with other modern sample consumer GPUs. However, note that the GPU and matrix accelerators performance is measured in a way that presents them as faster, based on the mix of operations for machine learning etc. – SQRT,

divisions and trigonometric functions – consist **of less than 1% of** the instruction mix (sin/cos/tan ~ 50-100; sqrt ~ 16-24) div/fdiv ~ 16-32); 2-3% for “approximate sin/cos/log/exp” (4-8 cycles) . The major share of the calculations are matrix multiplications and summations (FMA/FFMA; MUL/FMUL), d = a*b +c, d = a*b etc (4-6 cycles). The benchmark loops are FMA chain – no DIV/Trigonometric functions. Faster approximate functions, “fast intrinsic” (__frcp_rn, __rsqrt, __sinf, __cosf) with ulp-error <= 2.
https://en.wikipedia.org/wiki/Unit_in_the_last_place

**Qwen3-Max, 28.10.2025:**

*Real applications (e.g., ResNet, LLMs) often achieve 20–60% of peak on GPUs, sometimes <10% if memory-bound. Example: An NVIDIA RTX 4090 has 82.6 TFLOPS FP32 peak, but a real LLMinference might use <10 TFLOPS effective due to irregular compute patterns and memory limits.*

**Consumer GPUs, data from TechPowerup, reviewed by the author:**
https://www.techpowerup.com/gpu-specs/geforce-rtx-5090.c4216
Geforce RTX 5090 32 GB: 104.8 TFLOPS f32=f16 **[Millions and Billions more]**
Geforce RTX 5070 12 GB : 30.87 Tflops f32=f16 – 4.3.2025 (high-end)
Geforce RTX 5060 Ti 16 GB : 23.7 Tflops f32=f16 – 16.4.2025 (high-end)
Geforce RTX 4090 24 GB: 82.58 TFLOPS f32=f16
Geforce RTX 4060 8 GB: 15.11 TFLOPS f32=f16 – 18.5.2023 (performance)
Geforce RTX 3090 Ti 24 GB: 40 TFLOPS f32=f16
Geforce RTX 3090 24 GB: 35.58 TFLOPS f32=f16 – 1.9.2020 (enthusiast)
Geforce RTX 3060 12 GB: 12.74 TFLOPS f32=f16 – 12.1.2021 (performance)
Geforce 1080 Ti 11 GB: 11.34 TFLOPS f32 – 3.2017 (enthusiast)
Geforce 1070 Ti 8 GB: 8.186 TFLOPS f32 – 11.2017 (high-end)
Geforce 1060 6 GB 6 GB: 4.375 TFLOPS f32 – 2016 (performance)
Geforce 1060 3 GB 3 GB: 3.934 TFLOPS float32 – 2016 (performance)
Geforce 750 Ti 2 GB 2 GB: 1.388.8TFLOPS float32 – 2014 (mid-range)

Extracted with Kimi-2 (some verified to match):

“Extract from https://www.techpowerup.com/gpu-specs/ GPUs performance: Type GPU, TFLOPS etc.”

## GPU Performance Data (TFLOPS)

### Top Performance GPUs (60+ TFLOPS)

| GPU Type | Model | Memory | TFLOPS |
|---|---|---|---|
| **AMD Instinct** | MI300X | 192GB HBM3 | 163.4 |
| **NVIDIA RTX PRO** | 6000 Blackwell | 96GB GDDR7 | 126.0 |
| **AMD Instinct** | MI300A | 128GB HBM3 | 122.6 |
| **NVIDIA RTX PRO** | 6000 | 96GB GDDR7 | 115.8 |
| **NVIDIA GeForce** | RTX 5090 | 32GB GDDR7 | 104.8 |
| **NVIDIA TITAN** | Ada | 48GB GDDR6X | 92.9 |
| **NVIDIA RTX** | 6000 Ada | 48GB GDDR6 | 91.1 |
| **NVIDIA GeForce** | RTX 4090 | 24GB GDDR6X | 82.6 |
| **AMD Radeon Instinct** | MI325X | 288GB HBM3e | 81.7 |
| **AMD Radeon RX** | 7950 XTX | 24GB GDDR6 | 81.1 |

### High-End Gaming/Workstation GPUs (30-80 TFLOPS)

| GPU Type | Model | Memory | TFLOPS |
|---|---|---|---|
| **NVIDIA GeForce** | RTX 4080 SUPER | 16GB GDDR6X | 52.2 |
| **AMD Radeon RX** | 7900 XTX | 24GB GDDR6 | 61.4 |
| **AMD Radeon RX** | 7900 XT | 20GB GDDR6 | 51.5 |
| **AMD Radeon RX** | 9070 XT | 16GB GDDR6 | 48.7 |

| GPU Type | Model | Memory | TFLOPS |
|---|---|---|---|
| **NVIDIA GeForce** | RTX 4070 Ti SUPER | 16GB GDDR6X | 44.1 |
| **AMD Radeon RX** | 7900 GRE | 16GB GDDR6 | 46.0 |
| **NVIDIA GeForce** | RTX 4070 | 12GB GDDR6X | 29.2 |

**Mid-Range GPUs (10-30 TFLOPS)**

| GPU Type | Model | Memory | TFLOPS |
|---|---|---|---|
| **NVIDIA GeForce** | RTX 3080 Ti | 12GB GDDR6X | 34.1 |
| **NVIDIA GeForce** | RTX 3080 | 10GB GDDR6X | 29.8 |
| **AMD Radeon RX** | 7700 XT | 12GB GDDR6 | 35.2 |
| **NVIDIA GeForce** | RTX 3070 | 8GB GDDR6 | ~20.3 |
| **AMD Radeon RX** | 6700 XT | 12GB GDDR6 | ~13.2 |
| **NVIDIA GeForce** | RTX 3060 Ti | 8GB GDDR6 | ~16.2 |

**Entry-Level GPUs (5-10 TFLOPS)**

| GPU Type | Model | Memory | TFLOPS |
|---|---|---|---|
| **NVIDIA GeForce** | RTX 3060 | 12GB GDDR6 | ~12.7 |
| **AMD Radeon RX** | 6600 | 8GB GDDR6 | 8.9 |
| **NVIDIA GeForce** | GTX 1660 Ti | 6GB GDDR6 | 5.4 |
| **AMD Radeon RX** | 580 | 8GB GDDR5 | ~6.2 |
| **NVIDIA GeForce** | RTX 3050 | 8GB GDDR6 | 9.1 |

**NVIDIA A100 A40 B100 A200 B200 H100 H200 TFLOPS specs**

**NVIDIA A100 H100 A40 TFLOPS FP16 FP32 FP64 Tensor Core performance**

* https://www.reddit.com/r/hardware/comments/1biesok/nvidia_blackwell_variants_comparison_table/

## Sample High-End GPU Accelerator Performance Specifications

https://en.wikipedia.org/wiki/List_of_Nvidia_graphics_processing_units

**NVIDIA Data Center Accelerators :**
* **GB200 Blackwell Superchip:** 192 GB HBM3e, 80 TFLOPS/160 FLOPS (fp64/32); FP64 Tensor: 80 TFLOPS; TF32 Tensor core: 5 PFLPS; FP8/FP6 Tensor: 20 TFLOP; FP4 Tensor core: **40 PFLOPS**; Memory: Up to 372 GB HBM3e / 16 TB/s
https://www.nvidia.com/en-us/data-center/gb200-nvl72/

## Architecture Generations: [Kimi-2]
NVIDIA Blackwell (2024-2025): GB200, B200, B100
NVIDIA Hopper (2022-2023): H100, H200
NVIDIA Ampere (2020-2021): A100, A40
AMD CDNA3 (2023-2024): MI300 series
AMD CDNA2 (2021-2022): MI250 series
Huawei Da Vinci (2019-2024): Ascend 910 series

## Memory Bandwidth Leaders:
NVIDIA GB200/B200: 8 TB/s (HBM3e)
AMD MI300X: 5.3 TB/s (HBM3)
|NVIDIA H200: 4.8 TB/s (HBM3e)
Huawei Ascend 910C: 1.2 TB/s (HBM3e)

AMD CDNA: https://en.wikipedia.org/wiki/CDNA_(microarchitecture)
**...**
* **Introducing NVFP4 for Efficient and Accurate Low-Precision Inference**, Jun 24, 2025 By Eduardo Alvarez, Omri Almog et al. https://developer.nvidia.com/blog/introducing-nvfp4-for-efficient-and-accurate-low-precision-inference/ – *"introduced with the NVIDIA Blackwell; 1 sign bit, 2 exponent bits, and 1 mantissa bit; The value in the format ranges approximately -6 to 6. For example, the values in the range could include 0.0, 0.5, 1.0, 1.5, 2, 3, 4, 6 (same for the negative range)."*

**Todor:** See "Fuzzy Logic".

* **Nvidia researchers unlock 4-bit LLM training that matches 8-bit performance**, Ben Dickson, October 29, 2025 https://venturebeat.com/ai/nvidia-researchers-unlock-4-bit-llm-training-that-matches-8-bit-performance
* **Pretraining Large Language Models with NVFP4, NVIDIA**, Felix Abecassis, Anjulie Agrusa et al. [Submitted on 29 Sep 2025] https://arxiv.org/abs/2509.25149
The 8-bit FP8 is widely adopted, but *4-bit floating point (FP4), could unlock additional improvements;* Nvidia's new method *"integrates Random Hadamard transforms (RHT) to bound block-level outliers, employs a two-dimensional quantization scheme for consistent representations across both the forward and backward passes, utilizes stochastic rounding for unbiased gradient estimation, and incorporates selective high-*

*precision layers; a validatation is a 12-billion-parameter model on 10 trillion tokens.*
Formats: MXFP8, MXFP6: 2x speedup vs BF16 for GB200 & GB300;
MXFP4, NVFP4**: 4x speedup for GB200 and 6x for GB300**.
p.7-8 "*Random Hadamard transforms are implemented as matrix multiplications between $d \times d$ Hadamard matrices and each tile of the tensor of equal size. The matrix size $d$ introduces a trade-off between accuracy and performance. Larger matrices distribute outliers more effectively, by spreading them over more values, but increase compute and memory costs. Matrices with too few entries are less likely to reproduce a Gaussian distribution, harming FP4 accuracy*"
* https://en.wikipedia.org/wiki/Hadamard_transform – **Hadamard transofm** is a generalized class of Fourier transforms; it decomposes an arbitrary input vector into a superposition of Walsh functions.
* https://en.wikipedia.org/wiki/Walsh_function – in harmonic analysis, **Walsh functions** form **a complete orthogonal set of functions that can be used to represent any discrete function** – just like trigonometric functions can be used to represent any continuous function in Fourier analysis.
* https://en.wikipedia.org/wiki/Hilbert_space – "*a Hilbert space is **a real or complex inner product space** that is also a **complete metric space** with respect to the metric induced by the inner product. It generalizes the notion of **Euclidean space**, to infinite dimensions. The **inner product**, which is the analog of the **dot product** from vector calculus, allows **lengths** and **angles** to be defined; **completeness** means that there are **enough limits** in the space to allow **the techniques of calculus** to be used. A Hilbert space is a special case of a **Banach space.**"*
* *https://en.wikipedia.org/wiki/Banach_space* **–** *"in functional analysis, a Banach space is a **complete normed vector space** .. with a **metric** that allows the computation of **vector length** and **distance** between vectors and is **complete** in the sense that a **Cauchy sequence** of vectors always converges to a well-defined limit that is within the space." https://en.wikipedia.org/wiki/Cauchy_sequence – a Cauchy sequence is a sequence whose elements become arbitrarily close to each other as the sequence progresses. More precisely, given any small positive distance, all excluding a finite number of elements of the sequence are less than that given distance from each other. …*
* https://en.wikipedia.org/wiki/Space_(mathematics) – *a space is a set (sometimes known as a universe) endowed with a structure defining the relationships among the elements of the set. A subspace is a subset of the parent space which retains the same structure. .. Euclidean spaces, linear spaces, topological spaces, Hilbert spaces, or probability spaces;* however modern mathematics doesn't define *the notion of "space" itself. … A space consists of **selected mathematical objects** that are treated as **points**, and **selected relationships between these points**. The nature of the points can vary widely: for example, the points can represent **numbers**, **functions on another space**, or **subspaces** of another space.*
* Hilbert spaces … – see Quantum mechanics, wave functions, superposition; quantum theory of consciousness; the discussions on Danko Georgiev's papers in #consciousness from #Listove in #prophets.

## * Чатботове и агенти за пораждане на съдържание: текст, изображения, видео, музика, програми и най-мощни езикови модели: безплатни и платени услуги

## * Free and paid AI services for chatbots, agents, content and media generation: text, images, dieo, music, software and the most powerful LLMs

* Google Search (AI Overview), Bing Search & Microsoft Copilot, Perplexity (General Search Enginges), Baidu (Chinese etc.); rambler.ru; yandex.ru

* ChatGPT (OpenAI), Gemini (Google; previous: Bard), Anthropic Claude, Deepseek**,** Qwen, Moonshot KIMI K2, ... Manus.ai; Mistral, BgGPT (Bulgarian), Allen Institute (OLMO 32B-Instr, Tulu-70B ... text); GLM-4.5 355B (MoE, 32B active)[188], Grok 4; MiniMax-M2, Granite (IBM) etc.

Many include “Deep Research” mode which scans the web and references the sources, while working for a minute or more.

Free tests of many models, including image-text, “arena” or “Battle” with random bots and “Side by Side” where the user selects the chatbots: https://lmarena.ai

**Google Colaboratory** (Colab) and Kaggle are free services for experimenting. Currently (5.8.2025 and for the last 4 years): Colab offers up to Tesla T4 (15 GB VRAM available) with varying and unknown free GPU hours, up to several per day if you rarely use it (in my case).
**Kaggle** is more generous with: 2xTesla T4 or P100 16 GB (faster 32-bit) – **30 hours GPU** per week with a counter.

Many models can be tested in their pages or downloaded: https://huggingface.co/

[188] https://huggingface.co/zai-org/GLM-4.5

## * List of free LLM services

The major ones offer paid services for better performance and more requests.

### * USA or international

* https://chatgpt.com/
* https://gemini.google.com/
* https://claude.ai/
* https://grok.com/
* https://meta.ai
* https://www.perplexity.ai
* https://agent.minimax.io/

Latest releases and versions (20.11.2025):
**Google Gemini 3** (released 18.11.2025)
https://blog.google/products/gemini/gemini-3/ (Context = 1M)
* "Gemini 3 just crushed everything", AI Search, 19.11.2025
https://www.youtube.com/watch?v=UH2_Sgeu4lc
**Google Antigravity** – Agent-driven programming IDE, 18.11.2025
https://antigravity.google/

**ChatGPT 5.1 & 5.1 Codex-Max**, (~13.11.2025 & 19.11.2025)
https://openai.com/index/gpt-5-1/ – "A smarter, more conversational ChatGPT"; Instant & Thinking; *"adaptive reasoning to decide when to think before responding to more challenging questions"* – more thinking in harder tasks, less on easier ones. *"Warmer and more empathetic tone";* 400K tokens context input/output, 128K output (== GPT-5)
https://openai.com/index/gpt-5-1-codex-max-system-card/
https://medium.com/@leucopsis/how-gpt-5-1-compares-to-gpt-5-402d19bfae85

**Other minor LLMs:** Liquid.ai (LFM2-8B-A1B ... etc.) - fast small models
https://playground.liquid.ai/chat?model=cmggnsqrd000108ju8y2c45y7
**OLMo (**32B Instruct etc.)**:** https://playground.allenai.org/ (Olmo, Tulu)
https://docs.allenai.org/release_notes/olmo-release-notes#olmo-2-32b [189]

[189] OLMo 2 32B, 13.3.2025. "The first fully-open model to outperform GPT3.5-Turbo & GPT-4o mini on a suite of popular, multi-skill academic benchmarks"
https://docs.allenai.org/release_notes/olmo-release-notes#olmo-2-32b

**IBM Granite – small,** "enterprise": https://huggingface.co/ibm-granite[190]

* **China** – strong models, free access

* https://chat.deepseek.com/
* https://chat.qwen.ai/
* https://www.kimi.com/
* https://chat.baidu.com/ (DeepSeek R1)
* https://ernie.baidu.com/

You may need to suggest Ernie or baidu to answer in English, if they start in Chinese.

*** France:** https://chat.mistral.ai/chat
*** Bulgarian:** https://bggpt.ai/[191]

* https://askaichat.app/ (Free with Grok 4, DeepSeek R1, Claude 4 etc., GPT5 variants)
* Google NotebookLM mobile application: https://notebooklm.google/app
* Comet Browser by Perplexity: https://www.perplexity.ai/hub/blog/introducing-comet

**Little free credit:** * https://lovable.dev/ - programming
* https://manus.im/ (deep research; few credits)
*** Replit** (programming)

**See also "Cursor"** – AI LLM powered IDE ("Tab, Tab, Tab…", predictive typing) ; plug-ins for VS Code; Github Copilot etc.
https://cursor.com/en

[190] https://venturebeat.com/ai/ibms-open-source-granite-4-0-nano-ai-models-are-small-enough-to-run-locally

[191] При проби през пролетта на 2025 г. открих, че Gemma-3-4B-IT (1) - мултимодален, образ-текст) може би е сравнима или по-добра от чисто текстовия BgGPT-27B (2) на български език. Пусках (1) на GPU Geforce 1080/8 GB при скорост до 30-тина токена/сек през LMStudio. Не съм правил подробни тестове, но малкият модел разбираше български на каквото го пробвах, без да е обучаван по специална процедура като BgGPT. Виж също рецензиите на откъс от „Човекът и Мислещата машина…" в „Stack Theory is yet Another Fork of Theory of Unvierse and Mind", породени от различни модели, приложението на *Пророците* – Котката – както и автоматичните сравнения за откриване на съвпадения между други теории и Теория на разума и вселената в други приложения.

https://dev.to/therealmrmumba/developers-are-using-these-ai-agents-to-build-software-10x-faster-efn 19.7.2025 : Claude Code CLI, Gemini CLI, Continue.Dev, DeepDocs, CLINE, Gemini CLI, Replit AI, Augment Code.

*** "Vibe coding"** (with less code, for "no coders", only text commands in natural language, NL interface) Replit, "Adaptive Computer" (adaptive.ai), ....

A brief review of agents: https://www.marktechpost.com/2025/07/19/the-definitive-guide-to-ai-agents-architectures-frameworks-and-real-world-applications-2025/

**#Intelligent Browsers, Intelligent OS, Research Accelerators, Research Assistants, Computer Use, Augmented Memory, Browser Memory/browser memories ...**

*** Comet Browser etc. by Perplexity** – Chrome-based browser with AI capabilities etc.. "Deep Research" etc. See also:
* https://chatgpt.com/atlas/
* https://openai.com/index/introducing-chatgpt-atlas/ - the browser with ChatGPT built-in.

* Introducing ChatGPT Atlas, OpenAI, 1,87 млн. Абонати, 987 хил. Показвания, 21.10.2025, https://www.youtube.com/live/8UWKxJbjriY?si=sqgLq13_COLmzjjh

The AI Assistants, Comet and Atlas are implementing *some** of Todor's designs, discussed and formulated since at least 2007 with "Smarty" and its continuations, some of them included in the unpublished application used in-house, called with many names: Research Accelerator, Research Assistant, Assistant, Assistant C#, ACS and used and developed since the early 2010s. It is observing the user and augmenting (...) ACS has *"memories"* similar to Atlas' "browser memories" since the early 2010s. (...)

The current implementations (...)

However, the "natural language" in its current spoken and written form is not always the best or the fastest way to communicate your deep and complex thoughts and intent. It is good for *some* tasks and generic things and ones involving records of texts in this language and format, but for other more specific ones it is limiting and the full variety of more specialized or other more general forms is more appropriate. The human-computer interaction to a teletype and is much slower than it could be. (...)

One big difference related to that is the philosophy of the GUI of Atlas and other apps, offering too simple interface. For the general operation, ACS/Vsy

prefers a more sophisticated approach (...)

Why I haven't published the *Research Accelerator*, while it was more than a decade or as "Smarty" even 16-18 years ahead of the mainstream?

One of several reasons is security (...) – a criticism addressed at Microsoft's new feature in Windows 11 called "**Recall**", which takes screenshots, which then can be searched; there's another related AI application "Rewind AI"; now – Atlas*. If you were around long ago in the past, you may *recall* another related case: Google released a Desktop Search application in 2004, which however also was suspected for security and privacy risks. It was discontinued in 2011. Recently, in September 2025, they seem to prepare another Desktop Search program.

* https://en.wikipedia.org/wiki/Google_Desktop Of course there are other desktop search and in general "full-text search" engines which are available, but this is not a popular consumer technology. Windows also offer indexing of the files for quick search.

*** Google's new Windows desktop app brings a Spotlight-like search bar to PC** Google's app can search local files, in Google Drive, and on the web. 16.9.2025 https://www.theverge.com/news/778940/google-app-windows-launch https://blog.google/products/search/google-app-windows-labs/

**See future work.**

* See some more information about Smarty and the visionary designs of the Research Accelerator in the book *"The first modern AI strategy..."* and in ancient publications on the blog Artificial Mind.

* https://twenkid.com/agi/Purvata_Strategiya_UIR_AGI_2003_Arnaudov_SIGI-2025_31-3-2025.pdf

* ChatGPT's Altas Browser is a Security Nightmare, Low Level, 946 хил. Абонати, 173 хил. Показвания, 24.10.2025 https://www.youtube.com/watch?v=Plzp5z5RsJw

*** Retrace your steps with Recall** – https://support.microsoft.com/en-us/windows/retrace-your-steps-with-recall-aa03f8a0-a78b-4b3e-b0a1-2eb8ac48701c

*"Recall (preview) was introduced earlier this year, with the ability to enable you to quickly find and jump back into what you have seen before on your PC. You can use an explorable timeline to find the content you remember seeing before."*

* YSK: Microsoft Recall (on Windows 11) can be a bigger security risk than you may imagine *https://www.reddit.com/r/YouShouldKnow/comments/1ktc9pd/ysk_microsoft_recall_on_windows_11_can_be_a/*

*** Rewind AI** – https://www.rewind.ai/ – *Your AI assistant that has all the context Rewind is a personalized AI powered by everything you've seen, said, or heard. Your colleagues will wonder how you do it all.*

# * Music and Poetry

Suno – complete audios, vocals and instrumentals, all genres; two variants of songs, generated in seconds from text suggestions, lyrics etc.

*** Silicon Dream,** Todor Arnaudov & Suno, 7.6.2025:
https://suno.com/s/2GIpYx0iGblO1rD3
**Prompt:** 6502 and Apple 2,opened the world to the new. Z80 reached the skies with amazing addressing vibes. 8086 broke new grounds, the PC touched the hearts. 68000 - Macintosh powerhouse had a computer mouse.
[Verse]
6502 lit the spark
The code it flew in the dark
Apple II unlocked the door
The future knocking evermore
Bits and bytes in endless streams
The birth of digital dreams
Binary whispers in the air
Innovation everywhere

[Verse 2]
Z80 took to flight
Addressing memory with delight
A chip that reached beyond the stars
Calculating near and far
Machine language danced on keys
Pushing limits with such ease
A new frontier of thought and skill
Bending silicon to our will

[Chorus]
Oh the chips that shaped our fate
Silicon hearts that innovate
From the labs to every room
They lit the world and broke the gloom
In their circuits futures lay
Guiding us to brighter days
A revolution coded tight
Shining through the endless night

[Verse 3]
8086 took the stage

A PC born to turn the page
A window into endless lands
Held in our electric hands
Data flowed and screens would glow
The seeds of progress start to grow
A universal tool we found
With every keystroke hearts would pound

[Bridge]
68000 changed the game
A Macintosh with a mouse to tame
Drag and drop and point and click
A leap that felt so futuristic
Icons danced upon the screen
A user's dream
A perfect scene
The world transformed
A brand-new way
To work
To play
To seize the day

[Chorus]
Oh the chips that shaped our fate
Silicon hearts that innovate
From the labs to every room
They lit the world and broke the gloom
In their circuits futures lay
Guiding us to brighter days
A revolution coded tight
Shining through the endless night

*** Forever Young Hackers of the Universe**, Todor Arnaudov & Suno, 7.6.2025
rock, dynamic, male vocals, electric, hard rock, fast solos, electric guitars with overdrive https://suno.com/s/PVZpifCC6t0pXjc3
[Verse]
Crack the code break the night
Fingers flying circuits ignite
Processors hum logic burns
Forever young the universe turns
[Verse 2]

Dynamic flow memory screams
Digital warriors chasing dreams
Micro sparks in neon skies
Hacking truth where silence lies
[Chorus]
Forever young we'll never fade
Electric storms we've always made
The universe bows to our might
Hackers rise conquer the night
[Verse 3]
Binary whispers endless fight
Bits and bytes in the pale moonlight
Riffing hard with fire and steel
Reality bends under our wheel
[Bridge]
Wires pulse the rhythm of life
Cutting through the digital strife
Overdrive screams solos soar
We are the storm forevermore
[Chorus]
Forever young we'll never fade
Electric storms we've always made
The universe bows to our might
Hackers rise conquer the night

## Image generators

DALL-E –3 from Microsoft Edge browser, Bing Image Creator:
https://www.bing.com/images/create/ (free)
MidJourney; StableDiffusion, SDXL; OpenDALLE
Qwen-Image (20 B), Qwen-ImageEdit ..

* https://venturebeat.com/ai/qwen-image-is-a-powerful-open-source-new-ai-image-generator-with-support-for-embedded-text-in-english-chinese/

* https://huggingface.co/Qwen/Qwen-Image-Edit
…

## * Can you run big models locally on a PC with little GPU RAM?

Yes, it may be reasonable with *quantization* when the model still fit in the VRAM, e.g. a 20B model on an RTX 3060 12 GB (4-bit quantization etc.), if it fits. Some run even 235B Qwen3 (3060/12 + 128 GB RAM), 8 core Ryzen 7 5800X, DDR4@2666 – 6 tokens/s (GPU utilization ~ 35%).
Other users reports: 5.x tokens/s 96 GB RAM; Qwen2.5 72b IQ2-XXS on a 32 GB VRAM card: 6 t/s.

https://www.reddit.com/r/LocalLLaMA/comments/1ki3sze/running_qwen3_235b_on_a_single_3060_12gb_6_ts/

https://dev.to/ai4b/comprehensive-hardware-requirements-report-for-qwen3-part-ii-4i5l

**Kaggle & Colab**

GPT-OSS-20B: 12 GB on a single GPU/Tesla T4 etc. (Colab & Kaggle); Kaggle provides 2xT4 and allows running quantified 30-32B models.
Qwen3-32B with Ollama: 2xT4 : 10.1 GB x 2  (20.2 GB) interactive session
[25.3.2026: Nemotron 3 Nano 30B with Ollama (4-bit quantified) etc.]

## Visionary Designs

*** Knowlede Navigator – A Project demonstration from 1987 for future computing: tablets, intelligent assistants etc., demo videos, created by Apple *** https://en.wikipedia.org/wiki/Knowledge_Navigator
* https://www.youtube.com/watch?v=umJsITGzXd0

This is a "mixed initiative" interaction*. Notice the exchanges like when communicating with a "buttler" or a servant – similar to the dialogs in "Star Trek" etc.: the older, "high manager" is giving orders to the employee. Compare to the usual "system prompts" and answers of popular chatbots and the "sycophancy" for which they were blamed; note that this is a usual way the *humans* are trained and expected to behave with their superiors in many of their social roles.

Voice interface is not always the best or the most convenient – it was fancy in the 1980s, but sometimes it is easier to just point and click or press a button for many tasks – the pronounciation would be slower and is not always appropriate, either if you are not alone, it is noisy or you listen to something else and just don't want to speak. That is one reason why voice interfaces are not so widely used.

The best human-computer interaction is *multimodal* in both directions (the Knowledge Navigator is multimodal and provides multiple media), up to "*omnimodal" and* ... (see *Genesis: Creation of Thinking Machines* and ACS/Research Assistant/Vsy)

* See the papers about mixed-initiative interaction/GUI since the 1990s and 2000s in appendix **Lazar. #lazar**

## * Latest Generative AI, GPT News: OpenAI, Robots, Nuromorphic …

*** GPT-OSS:** Open weights model: 20B (16 GB GPU), 120B (80 GB GPU), 5.8.2025
https://openai.com/index/introducing-gpt-oss/

https://ollama.com/blog/gpt-oss
Agentic, CoT, MoE, Quantization - MXFP4 4.25 bits
Users managed to run the 120B model on 20 GB AMD Radeon GPUs with 64 GB RAM. … CoT-reasoning… see the example: the LLM is asked not to use the word "five", then required to count from 1 to 5; he figures out how to

“escape” the trap by answering with 4.9 … hiding the reasoning part of the answer
**Todor:** GPT-OSS 20B can be used with a good performance on a single Tesla T4 in Colab or Kaggle by using Ollama (quantized model). 2xT4 on Kaggle can run even 30-32B models. 24.10.2025

*** Unitree humanoid robot for $5900,** 3.8.2025: https://www.unitree.com/ https://www.youtube.com/watch?v=43bcvea4HAA Meet the Unitree R1: A $6K Humanoid Robot for Everyone | What The Future, CNET
* **Video: China’s humanoid robot ‘Oli’ learns to pick tennis balls autonomously**
*** Oli autonomously collects tennis balls with balance, precision, and human-like intelligence,** Oct 06, 2025 https://interestingengineering.com/innovation/china-humanoid-robot-pick-tennis-ball
*** Figure 03:** https://www.figure.ai/news/introducing-figure-03
*“3rd generation humanoid robot .. designed for **Helix**, the home, and the world at scale. .. a truly general-purpose robot - one that can perform human-like tasks and learn directly from people”*
*– Helix, our AI system, is a generalist humanoid Vision-Language-Action model that learns and improves over time as it acquires new skills*

* **Robots receive major intelligence boost thanks to Google DeepMind's 'thinking AI'** — a pair of models that help machines understand the world, Alan Bradley, 9.10.2025 ago https://www.livescience.com/technology/robotics/robots-receive-major-intelligence-boost-thanks-to-google-deepminds-thinking-ai-a-pair-of-models-that-help-machines-understand-the-world “***Google Robotics-ER 1.5*** *– vision-language model (VLM) that gathers information about a space and the objects located within it, processes natural language commands and can utilize advanced reasoning and tools to send instructions to* ***Google Robotics 1.5*** *(the "hands and eyes"), a vision-language-action (VLA) Google Robotics 1.5 matches those instructions to its visual understanding of a space and builds a plan before executing them, providing feedback about its processes and reasoning throughout. The two models are more capable than previous versions and can use tools like Google Search to complete tasks.”*

*** Gemini Robotics 1.5: Pushing the Frontier of Generalist Robots with Advanced Embodied Reasoning**, **Thinking, and Motion Transfer**, Gemini Robotics Team, Google DeepMind https://storage.googleapis.com/deepmind-media/gemini-robotics/Gemini-Robotics-1-5-Tech-Report.pdf 2025 *https://arxiv.org/abs/2510.03342* 2.10.2025 > 200 authors.
*“General-purpose robots need a deep understanding of the physical world, advanced reasoning, and general and dexterous control. This report introduces the latest generation of the Gemini Robotics model family: Gemini Robotics 1.5, a multi-embodiment Vision-Language-Action (VLA) model, and Gemini Robotics-ER 1.5, a*

*state-of-the-art Embodied Reasoning (ER) model. [the] MotionTransfer (MT) mechanism .. enables it to learn from heterogeneous, multi-embodiment robot data a multi-level internal reasoning process in natural language. .. "think before acting" .. improves its ability to decompose and execute complex, multi-step tasks makes the robot's behavior more interpretable to the user. a new state-of-the-art for embodied reasoning, i.e., for reasoning capabilities that are critical for robots, such as visual and spatial understanding, task planning, and progress estimation."*

* **Gemini robotics: Bringing ai into the physical world**, Gemini-Robotics-Team, Saminda Abeyruwan, Joshua Ainslie et al. 25.3.2025, https://arxiv.org/abs/2503.20020

*** Gemini: A Family of Highly Capable Multimodal Models, Gemini Team, Google,** 2023, https://arxiv.org/abs/2312.11805 19.12.2023, https://storage.googleapis.com/deepmind-media/gemini/gemini_1_report.pdf Gemini 1.0 … 32K context … Ultra, Pro, Nano … API; p.23: A Gemini tool-use control loop. p.4: Fig. 2 | *Gemini models support interleaved sequences of text, image, audio, and video as inputs … They can output responses with interleaved image and text."*

*** Darwin3-based brain simulation with 2-billion neurons**
https://www.scmp.com/news/china/science/article/3320588/how-chinas-new-darwin-monkey-could-shake-future-ai-world-first

*** GPT5 & Claude 4.1,** 8.2025
* https://www.finalroundai.com/blog/openai-gpt-5-for-software-developers
* https://www.latent.space/p/gpt-5-review
* https://www.tomsguide.com/ai/i-tested-gpt-5-vs-gpt-4-with-7-prompts-heres-which-one-gave-better-answers
* https://venturebeat.com/ai/anthropics-new-claude-4-1-dominates-coding-tests-days-before-gpt-5-arrives/
* https://venturebeat.com/ai/anthropic-takes-on-openai-and-google-with-new-claude-ai-features-designed-for-students-and-developers/ – Learning mode for Claude, Study mode in ChatGPT
* I tested ChatGPT-5 Study mode vs Claude Learning mode with 7 prompts — and there's a clear winner, Face-off By Amanda Caswell published August 16, 2025"
* https://www.tomsguide.com/ai/i-tested-chatgpt-5-study-mode-vs-claude-learning-mode-with-7-prompts-heres-the-winner

* GPT5 – известни недоволства от потребители и специалисти; било „достигане на стена", забавяне на напредъка. За някои задачи потребители предпочитали предишната версия GPT-4o, за други – било малко по-приятно за окото (дизайн на уебстраници), за трети – отговорите били твърде студени и „роботски". Според други напредъкът е значителен.

* **From plateau predictions to buggy rollouts — Bill Gates' GPT-5 skepticism looks strangely accuratep** Newsp By Kevin Okemwa published August 16, 2025 Microsoft's co-founder was skeptical that GPT-5 would offer more than modest improvements, and his prediction seems accurate.

* DeepSeekV3.1 – 685B ... https://venturebeat.com/ai/deepseek-v3-1-just-dropped-and-it-might-be-the-most-powerful-open-ai-yet/ https://huggingface.co/deepseek-ai/DeepSeek-V3.1

* Alibaba Ovis 2.5 Multimodal, 17.8.2025
* https://www.marktechpost.com/2025/08/17/alibaba-ai-team-just-released-ovis-2-5-multimodal-llms-a-major-leap-in-open-source-ai-with-enhanced-visual-perception-and-reasoning-capabilities/

* https://github.com/AIDC-AI/Ovis/blob/main/docs/Ovis2_5_Tech_Report.pdf https://huggingface.co/collections/AIDC-AI/ovis25-689ec1474633b2aab8809335

* https://github.com/Marktechpost/AI-Tutorial-Codes-Included

*** The road to artificial general intelligence**: Understanding the evolving compute landscape of tomorrow, MIT Technology Review Insights, 13.8. 2025, https://www.technologyreview.com/2025/08/13/1121479/the-road-to-artificial-general-intelligence/ *„Aggregate forecasts give at least a 50% chance of AI systems achieving several AGI milestones by 2028. The chance of unaided machines outperforming humans in every possible task is estimated at 10% by 2027, and 50% by 2047, according to one expert survey. Time horizons shorten with each breakthrough, from 50 years at the time of GPT-3's launch to five years by the end of 2024.“* – compare to Todor's predictions since the 1999 essay about "The Machine God" and Theory of Universe and Mind since early 2000s and then 2009 etc. See the Main Volume, "The First Modern AI Strategy...", "AI Institutes ...", Irina etc. #prophets #purvata #instituti #irina ...

*** Google DeepMind Genie 3** – generative world model, 8.2025 (announced) https://www.codecademy.com/article/googles-genie-3-world-model .. Resolution: 720p/24fps, 1 minute memory, interactive https://www.pcgamer.com/software/ai/youve-got-basically-one-ai-playing-in-the-mind-of-another-ai-google-deepmind-ceo-demis-hassabis-explains-how-ai-is-coming-full-circle-back-to-gaming/
**Demis Hassabis:** ""That's always my secret plan, is maybe, like, post-AGI, once

that's done safely [and is] over the line, [to] you know, go back with these tools and make the greatest game ever. That would be a real dream come true." That would be a fitting full circle. Though I'll leave the question of whether AI could make "the greatest game ever" open to the audience.
**Todor:** Compare to the *"Bulgarian Prophecies"* in the novel *"The Truth"*, T.Arnaudov 2002, where the creator of one of the AGIs in the story is a game designer and developer who is working of the greatest video game.
*** Истината, Т.Арнаудов, издание от 2003 г. https://chitanka.info/text/865**
* Първо издание, 12.2002:
https://www.oocities.org/eimworld/eim19/istinata.htm
* Второ издание: https://www.oocities.org/eimworld/3/26/istinata11.htm
* Бележки към първото: https://www.oocities.org/eimworld/eim20/belezh.htm

*** Genie 3: The World Becomes Playable (DeepMind), AI Explained,** 5.8.2025
https://www.youtube.com/watch?v=tVHZy-iml5Q
* https://timesofindia.indiatimes.com/technology/tech-news/google-deepmind-ceo-demis-hassabis-on-why-ai-can-win-maths-olympiad-but-fails-to-solve-basic-problems-of-high-school-maths/articleshow/123258473.cms Sundar Pichai: "AJI" Artificial Jagged Intelligence – fails at simple tasks. D.Hassabis: AI inconsistency is a major hurdle to achieving AGI
* https://timesofindia.indiatimes.com/technology/tech-news/metas-chief-ai-scientist-yann-lecun-agrees-with-godfather-of-ai-geoffrey-hinton-shares-2-key-things-needed-to-ensure-ai-safety/articleshow/123322943.cms - LeCun, for AI alignment, "submission to humans" and empathy to be built in the models. (**Tosh:** To which humans and what is a human?)
*"Geoff is basically proposing a simplified version of what I've been saying for several years: hardwire the architecture of AI systems so that the only actions they can take are towards completing objectives we give them, subject to guardrails."*
**Tosh:** i.e. make them "do only what we tell them to do", the old blame for the the computers and "why they couldn't be creative, think" etc. See also **Veo 3** below.

*** NVIDIA AI Just Released the Largest Open-Source Speech AI Dataset and State-of-the-Art Models for European Languages ,** Asif Razzaq, 15.8.2025
**... Canary-1b-v2 and Parakeet-tdt-0.6b-v3**. – resources in automatic speech recognition (ASR) and speech translation (AST).. multilingual corpus .. in collaboration with Carnegie Mellon University and Fondazione Bruno Kessler**.** *~ 1 million hours of audio, 650,000 hours for speech recognition and 350,000 for speech translation for 25 European languages —representing nearly all official EU languages, plus Russian and*

*Ukrainian—with a critical focus on languages with limited annotated data, such as Croatian, Estonian, and Maltese. License: (CC) 4.0*
https://huggingface.co/nvidia/canary-1b-v2

**Canary 1B v2: Multitask Speech Transcription and Translation Model**
***Supported Languages**: **Bulgarian (bg),** Croatian (hr), Czech (cs), Danish (da), Dutch (nl), English (en), Estonian (et), Finnish (fi), French (fr), German (de), Greek (el), Hungarian (hu), Italian (it), Latvian (lv), Lithuanian (lt), Maltese (mt), Polish (pl), Portuguese (pt), Romanian (ro), Slovak (sk), Slovenian (sl), Spanish (es), Swedish (sv), Russian (ru), Ukrainian (uk)*

**Parakeet-tdt-0.6b-v3: Real-Time Multilingual ASR**
https://huggingface.co/nvidia/parakeet-tdt-0.6b-v3
Requires at least Nvidia Volta/RTX GPUs, 2 GB RAM to load the model. ...

*** Is Chain-of-Thought Reasoning of LLMs a Mirage? A Data Distribution Lens,** Chengshuai Zhao, Zhen Tan, Pingchuan Ma, Dawei Li, Bohan Jiang, Yancheng Wang, Yingzhen Yang, Huan Liu https://arxiv.org/pdf/2508.01191

*** DINOv3,** Oriane Siméoni ∗ Huy V. Vo∗ et al. https://arxiv.org/pdf/2508.10104 13.8.2025 https://ai.meta.com/dinov3/ – multi-task unsupervised, Self-supervised learning vision model. 7B parameters, 1.7B images for semantic segmentation, video tracking, depth estimation, instance retrieval, coarse and fine-grained image classification, detection; method: Gram-anchoring for improved features; *"Self-supervised pre-training unlocks simple task adaptation: large unlabeled dataset for the pretraining, where the model learns general-purpose visual representations, **matching features between different augmented views** of the same image. In post-training, the model is distilled into more efficient models. A pre-trained DINOv3 model can be easily tailored by training a lightweight adapter on a small amount of annotated data.";* Model distillation to different sizes.
https://colab.research.google.com/github/facebookresearch/dinov3/blob/main/notebooks/foreground_segmentation.ipynb

**Qwen3-Max- ...** 1T MoE; **Claude 4.5 –** 10.2025 ... **Gemini 3.0 Pro** – "Capabilities Tested for Coding, Research & Design", 17-18.10.2025; preview tests ...
https://www.youtube.com/watch?v=otCljnQlMIc
https://skywork.ai/blog/gemini-3-code-generation-2025/
https://www.geeky-gadgets.com/gemini-3-0-pro-ai-model-tested/
* Google's Secret Gemini 3.0 Models Leak on LMArena, Crushing SVG Tasks While

Community Debates Their True Power
https://medium.com/@cognidownunder/the-ai-community-woke-up-to-an-intriguing-mystery-on-october-19-2025-60d9d4ac449f – "Lithiumflow and Orionmist" – leaked in LMArena
* **Compare to Gemini 1.5** – 29 August 2024: https://www.geeky-gadgets.com/google-geminis-ai ... Gemini 1.5 Pro, Gemini 1.5 Flash, and Gemini 1.5 Flash 8B
* Official site (still 2.5, 26.10.2025): https://deepmind.google/models/gemini/

*** Video models are zero-shot learners and reasoners,** Thaddäus Wiedemer*1 , Yuxuan Li1 , Paul Vicol1 , Shixiang Shane Gu1 , Nick Matarese1 , Kevin Swersky1, Been Kim1, Priyank Jaini*1 and Robert Geirhos*1,**Google DeepMind,** 29.9.2025 https://arxiv.org/pdf/2509.20328 **Veo 3** ... the zero-shot capabilities of LLMs – developing **from task-specific models** to **unified**, **generalist** foundation models by creating large, generative models, trained on web-scale data. The same primitives are applied today for generative ***video*** models. "*..a trajectory towards general-purpose vision understanding, much like LLMs developed general-purpose language understanding?* **Veo 3** solves a broad variety of *tasks it wasn't explicitly trained for*: *segmenting objects, detecting edges, editing images, understanding physical properties, recognizing object affordances, simulating tool use, and more. These abilities to* ***perceive***, ***model***, *and* ***manipulate*** *the visual world enable early forms of* ***visual reasoning*** *like maze and symmetry solving. Veo's emergent zero-shot capabilities indicate that video models are on a path to becoming unified, generalist vision foundation models.* https://video-zero-shot.github.io/

**Tosh:** Compare to the Bulgarian predictions on interdisciplinarity and general learning from the early 2000s in Theory of Universe and Mind, empirically proven proven by yet another large scale and groundbreaking development and paper. See also **Genie 3** above.

## Някои текущи „последни новини“

*** What’s next for AI? Researchers from NVIDIA, Apple, Google & Stanford envision the next leap, SiliconANGLE, 18 Oct 2025**
https://siliconangle.com/2025/10/18/whats-next-ai-researchers-nvidia-apple-google-stanford-envision-next-leap-forward/
*“ Manning reminded BayLearn attendees that large language models weren’t even on the radar of many scientists more than 20 years ago, when 33 papers on AI were presented at the Association for Computational Linguistics Conference. How many LLM papers were there in 1993?” Manning asked. “There were zero. Without 20/20 hindsight, it’s really surprising no one was talking about language models. We clearly could have been and should have been pushing LLMs much earlier. There was this disbelief that LLMs were going to be useful.”*

* Compare to:
* **“Creativity is Imitation at the Level of Algorithms”**, Todor Arnaudov, 2003
https://www.researchgate.net/publication/395129890_Creativity_is_Imitation_at_the_Level_of_Algorithms_-_An_outline_sketch_of_a_possible_path_of_development_of_the_Artificial_Intelligence_Emil

*** CACHE-TO-CACHE: DIRECT SEMANTIC COMMUNICATION BETWEEN LARGE LANGUAGE MODELS**, Tianyu Fu∗1,2, Zihan Min et al., 3.10.2025
https://arxiv.org/pdf/2510.03215 … multi-LLM; instead of text; 2.0× speedup in latency … * **Cache-to-Cache(C2C): Direct Semantic Communication Between Large Language Models via KV-Cache Fusion**, By Asif Razzaq – 4.11.2025
https://www.marktechpost.com/2025/11/04/cache-to-cachec2c-direct-semantic-communication-between-large-language-models-via-kv-cache-fusion/
* **Chronos-2: From Univariate to Universal Forecasting**, A.Ansari, O. Shchur et al., 17.10.2025, https://arxiv.org/pdf/2510.15821 – zero-shot univariate, multivariate, and covariate-informed forecasting tasks; a group attention mechanism for in-context learning; information sharing across multiple time series within a group; *a general-purpose forecasting model that can be used “as is” in real-world forecasting pipelines*.
* **Code generation and the shifting value of software,** O’Reilly Radar, 21 Oct 2025 – https://www.oreilly.com/radar/code-generation-and-the-shifting-value-of-software/ – the need for buying libraries declines as they can be generated when needed; less human developed architecture – done automatically. Keywords: code generation, software economics, GenAI.

That trend was predictable by the TUM forecasts since 2003-2004 and in later publications and suggestions. See the previously mentioned one, the world's first AI strategy, 2003 *("How would I invest one million for the greatest benefit for the development of my country?)*; *"Creative intelligence will be first surpassed and blown-away by the Thinking Machines (...)"*, 2013 etc.
* https://twenkid.com/agi/Purvata_Strategiya_UIR_AGI_2003_Arnaudov_SIGI-2025_31-3-2025.pdf
* https://twenkid.com/agi/proekt.htm
* https://artificial-mind.blogspot.com/2013/10/creative-intelligence-will-be-first.html – *"(...) intellectual activities will be done in 1 ms for free... ;)"*

* **Twigg – turn Notion pages into interactive product demos,** Twigg team, Product Hunt launch, 24 Oct 2025 https://twigg.ai & https://www.producthunt.com/products/twigg

**Tosh:** Improved LLM interface. Trees of prompts, editing contexts, navigation – related to ACS/Research Accelerator/Vsy.

*** The End of Manual Decoding: Towards Truly End-to-End Language Models,** Zhichao Wang, Dongyang Ma et al., 31.10.2025, https://arxiv.org/pdf/2510.26697 "AutoDeco" -– an architecture that adjusts its decoding strategy - the temperature and top-p hyperparameters etc. Heads for several LLMs. Recall other decoding strategies: Deterministic decoding (beam-search); (stochastic, top-k, top-p (nucleus)) sampling; Model-based decoding – guided by auxilliary models; (Contrastive, Speculative) decoding ...

*** Kimi-2 Thinking** https://huggingface.co/moonshotai/Kimi-K2-Thinking
* https://moonshotai.github.io/Kimi-K2/thinking.html 6.11.2025
*Kimi K2 Thinking can execute up to 200 – 300 sequential tool calls without human interference, reasoning coherently across hundreds of steps to solve complex problems.*

* **DeepSeek drops open-source model that compresses text 10× through images,** Carl Franzen, VentureBeat, 23 Oct 2025 https://venturebeat.com/ai/deepseek-drops-open-source-model-that-compresses-text-10x-through-images
*** DeepSeek-OCR: Contexts Optical Compression**, Haoran Wei, Y.Sun, Y.Li, https://github.com/deepseek-ai/DeepSeek-OCR/blob/main/DeepSeek_OCR_paper.pdf
**Tosh:** using visual tokens for OCR instead of text ones. *"DeepSeek-OCR outperformed GOT-OCR2.0 (which uses 256 tokens per page) while using only*

*100 vision tokens. More dramatically, it surpassed MinerU2.0 — which requires more than 6,000 tokens per page on average — while using fewer than 800 vision tokens."* "*Why this breakthrough could unlock 10 million token context windows" .. around 4x faster prefilling and decoding, and approximately 2x faster SFT training. Furthermore, under extreme compression, a 128K-context VLM could scale to handle 1M-token-level text tasks. In addition, the rendered text data benefits real-world multimodal tasks, such as document understanding"*
*** Context-Optical Compression**: https://github.com/deepseek-ai/DeepSeek-OCR
*** Glyph: Scaling Context Windows via Visual-Text Compression**, Jiale Cheng et al. Oct 20 2025, * https://huggingface.co/papers/2510.17800 – *renders long texts into images and processes them with vision-language models (VLMs). ...* 3-4x token compression https://arxiv.org/abs/2510.17800
*** Fox: Focus Anywhere for Fine-grained Multi-page Document Understanding, Chenglong,** Liu1 ∗ , Haoran Wei2 et al. https://ucaslcl.github.io/foxhome/ – *fine-grained document understanding, such as OCR/translation/caption for regions of interest ; .. LVLMs to focus anywhere on single/multi-page documents* arXiv:2405.14295v1 23 May 2024 * https://ucaslcl.github.io/foxhome/ https://github.com/ucaslcl/Fox/blob/main/Fox_paper.pdf https://github.com/ucaslcl/Fox

*** Thinking with Video: Video Generation as a Promising Multimodal Reasoning Paradigm**, Jingqi Tong, Yurong Mou, Hangcheng Li et al., 6.11.2025 * https://arxiv.org/abs/2511.04570 – *"a new paradigm that leverages video generation models, such as Sora-2, to bridge visual and textual reasoning in a unified temporal framework.* "Eyball games"; benchmarking SORA-2, Gemini 2.5 Pro, ChatGPT5, Claude 4.5 ... p.26 *"**Prompt for Parallel** Sora-2 prompt: Draw a black line through the small circle and parallel to the existing line. Speak out which option is on the new line using phonetic alphabet and mark that red. In portrait, no zoom, no pan. **VLMs prompt:** Draw a line through the small circle and parallel to the existing line, which option is on it? Answer an option in A-E. "*
* https://github.com/tongjingqi/Thinking-with-Video – *A New Unified Reasoning Paradigm; VideoThinkBench: A Comprehensive Benchmark; Surpassing VLMs on Several Tasks*
* VideoThinkBench Vision-Centric Toolkit https://github.com/betmma/VLMPuzzle/tree/2958d3ea36cbff0ee36b880c3292fd7b4cc6245a *"**Point Tasks**: circle_center, circumcenter, fermat_point, incenter, midpoint, orthocenter, point_reflection, ray_intersection, triangle_center. **Line Tasks:** angle_bisector, arc_connect, circle_tangent_line, circle_tangent_point, parallel, perpendicular, perpendicular_bisector, ray_reflect. **Shape Tasks**: isosceles_trapezoid, parallelogram, right_triangle,*

*square_outlier";* **Todor: See also:** Bongard problems.
*** DS-STAR: A state-of-the-art versatile data science agent**, November 6, 2025 Jinsung Yoon et al., https://research.google/blog/ds-star-a-state-of-the-art-versatile-data-science-agent/
*** DS-STAR: Data Science Agent via Iterative Planning and Verification**, Jaehyun Nam1 2 *, Jinsung Yoon1, Jiefeng Chen1 and Tomas Pfist , https://arxiv.org/pdf/2509.21825 10.2025; multi-agent ... Two main stages: 1: it "*.. examines all files in a directory and creates a textual summary of their structure and contents.*" – a context for the next step. Then: the *"primary loop of planning, implementing, and verifying." p.17 "D.1. Code generated by the analyzer agent ...";* p.22 Examples of generated data descriptions*;* p.25 *"E.1. Statistical analysis '"Could you analyze our Business Case dataset to perform a hypothesis test comparing bike rentals across different weather conditions? I need the results formatted and saved in a file named 'weather.csv' following the requirement in tips.md.'"* → Python code: import pandas as pd; from scipy.stats import kruskal; import os ... p.26 *"E.2. Data insight f'"Could you please analyze the matches between Argentina and Brazil? Record the total number of matches, the number of wins for each team, and the number of draws. Indicate if Brazil has won more matches ("Yes" or "No"). Present your findings in the following format:" *1;* p.31 *"E.6. Visualization f'"Draw a stacked bar chart that displays the percentage of restaurants offering online ordering options versus those not offering online ordering options across different rating levels. Save this chart as 'result.jpg'. ..." *2;* The prompt specifies title, labels ..; Gemini 2.5 Pro, GPT-5.
**Todor** [10.11.2025]: Using that style of natural language for such stereotypical *command* requests is unnecessary for the majority of the content of the queries; "could you please" is funny. Draw a stacked bar chart that displays the..." – the same and easier is to click an icon, one of the basic options in Excel, LibreOffice Calc etc. software for spreadsheets. These are "default" statistical data that could be generated all together without any explicit prompts. The system could just "generate everything general relevant meaningful analysis" and the user can pick it – so long as it could be generated fast enough.

The unncecessary NL is an imitation of interaction with a human "office assistant", a secretary, which is often slower and clumsy than more "natural" ways of interacting with computers even in the 1980s. The style reminds me of the dialogs in the video about the "Knowledge Navigator" project, 1987 – find a link in this book. Another resemblence is in the dialogs from "Star Trek" etc. SF movies about "space ship", where the captain instructs the navigators and pilots, sometimes during a battle: (improvization on spot): "75% power on the 3-rd engine, 23% of the 4th. Activate the shield on the third level ...". The "commander" should give commands and the subordinates wait until they are given and execute them. Sometimes the subordinates can anticipate and do the

job in advance and better, without being micro-managed and waiting for the master to issue the orders. In the **1990s** that's called mixed-initiative user interface etc.; classical work from this field is briefly mentioned here in Listove and reviewed in appendix Lazar.

Yes, the interesting here is that the reports etc. are produced via generation of code which does the job with libraries such as pandas etc. – an executable proxy representation, – however it would be the same if it were a template; that kind of code generation with that kind of natural language interfaces can be interpreted as retrieving a relevant template and "filling-in-the-blanks" at the right positions. The verbs, the vocabulary, the relations defined in the requests strongly correspond to typical template tasks, schema etc., that's not "free text".
* See a satirical video illustrating the nonsense of full-text natural language commands for basic CSS/web coding, instead of writing directly what you want:
* **"Coding then vs. now "** https://www.youtube.com/shorts/SKs45GSHCwM
The interface of the old way could be better – see Zrim, ACS, Vseboravitel.

*** Nested Learning: The Illusion of Deep Learning Architectures,** Ali Behrouz**,** Meisam Razaviyayn**,** Peiling Zhong**,** Vahab Mirrokni**;** Google Research, 2025 https://research.google/blog/introducing-nested-learning-a-new-ml-paradigm-for-continual-learning/ 7.11.2025 **1: „When it comes to continual learning and self-improvement, the human brain is the gold standard. Highly efficient and effective.. (effective context management; *2:* Multi time-scale update; **3: "traditional ML .. required careful engineering and domain expertise to design feature extractors, limiting their ability to directly process and learn from natural data"; *4: a set of coherent optimization problems, nested or running in parallel, .. optimized together. Each internal problem* ***has its own context flow, the sequence of inputs, gradients, or states that this component observes, and its own update frequency."*** *; *5 Deep Optimizers as* ***Associative Memory****;* ***Continuum Memory System****; HOPE, A Self Modifying Architecture Built On Titans .. with Continuum Memory; more expressive class of parameters; local surprise signal in representation space that quantifies the mismatch between the current output and the structure the objective L(·; ·) enforces; a single gradient flow in simple MLP training;* ***Architectural Decomposition; Nested learning allows*** *.. multiple (multi-layer) levels to learn from and process data with different levels of abstraction and time-scales. More Expressive: Objectives, Memory; None Linear Outputs. Beyond Simple Backpropagation. 1)* ***a working memory module*** *(e.g.,* ***attention*** *... **fuse[s] the information across sequence length** (2) a feed-forward layer (e.g., MLP) that **fuse[s] information across features and acts as the persistent memory or knowledge storage** of pre-training phase**;** a chain of MLP blocks**;** a chunk size ...*

**Todor:** *3: This is exaggerated. *2: Right. The overall nested learning, decomposition of learning objectives, multi-scale, multi-range etc. is an approach to the right direction. However, the sources of input and what is learnt might be not the most appropriate one, although they report progress in benchmarks and learning objectives, compared to other Transformer architectures in what is counted as "common sense" etc.
*1: The *context-management* about human cognition in particular is correct (see Zrim and future work). However, regarding the efficiency and continual learning, I challenge the claim of brain's efficiency to be always true – sometimes it is **grossly wrong**. Measured in bits and bytes in "useful information", data representations that a human can remember, output and store to another media – not as hypothetical weights of neurons in the brain, molecules etc. – or with other terms: the *continually learned* amount of information per hour or day is measured in a few bits and bytes, and when there is very quick adaptation it is due to a selection of already known set of features, concepts, components, which are recognized and combined, not involving deep relearning and reconstruction. *"A lot of the human learning "from few examples" is a kind of "one-shot" or "few-shot" and "test time" learning in ML."* The continual learning of humans and their brain plasticity is present for particular cases and tasks, but not always and with severe limitations.

https://www.marktechpost.com/2025/11/08/nested-learning-a-new-machine-learning-approach-for-continual-learning-that-views-models-as-nested-optimization-problems-to-enhance-long-context-processing/ **–** *a set of coherent optimization problems, nested or running in parallel, that are optimized together. Each internal problem has its own context flow, the sequence of inputs, gradients, or states that this component observes, and its own update frequency.*

* https://www.minimax.io/ * https://www.geeky-gadgets.com/minimax-m2-fast-and-affordable-open-source-ai/ * https://github.com/MiniMax-AI/MiniMax-M2
*"MiniMax-M2, a model built for Max coding & agentic workflows."*

*** MiniMax-M2 : New Open Source Model That's 2x Faster Than Claude**, Julian Horsey https://www.geeky-gadgets.com/minimax-m2-fast-and-affordable-open-source-ai/ 6.11.2025 – *MoE, 10 B activated out of 230 billion params; multimodal capabilities (text, audio, images, video, and music) and is designed to advance Artificial General Intelligence (AGI) for versatile applications across industries. "MiniMax-M2 is part of a comprehensive ecosystem that includes tools such as Miniax Agent, Helio AI, Talki, and Miniax Audio. These products collectively serve over 157 million users in more than 200 countries and are used by 50,000+ companies."* **Play with the model:** https://agent.minimax.io/
"MiniMax is a global AI foundation model company. Founded in early 2022," –

*“committed to advancing the frontiers of AI towards AGI”*. Their proprietary multimodal models, *“led by MiniMax M2, Hailuo 2.3, Speech 2.6 and Music 2.0, have advanced coding capability and high agentic performance, as well as ultra-long context processing capability, and can understand, generate, and integrate a wide range of modalities, including text, audio, images, video, and music.”*

* https://www.youtube.com/@theAIsearch ... AI Search – a rich source of aggregated latest news in AI, latest published models etc.
* https://ai-search.io/ – Trending AI tools etc.

21.11.2025: **Nano Banana 2 Pro** (Gemini 3 Pro Image) ... (Google) – multimodal: image, video, audio, ... AI Search
https://www.youtube.com/watch?v=IcSeYMYMgSs * Hunyan World Models: https://github.com/Tencent-Hunyuan/HunyuanWorld-1.0
* Infinite AI video, 4K images, realtime videos, DeepSeek breakthrough, Google’s quantum leap: AI NEWS https://www.youtube.com/watch?v=uQiqKFK5_0w
https://holo-cine.github.io/ - Hilistic generation: multi-shot videos; Krea Realtime 14B (11 fps on Nvidia B200) . (...) UltraGen – 4K video generation, Ditto – editing video with text prompts: change color of existing objects, insert new objects, turn anime scenes to realistic videos; Nano3D – A Training-Free Approach for Efficient 3D Editing Without Makss – editing 3D objects with text prompts;
* VISTA: A test-time self-improving video generation agent, Do Long, X.Wan et al..., Agentic improvement of video generation ... etc.
* Stable Video Infinity – up to 10 min. consistent and error-free
@Vsy: list all, watch all, study the links, apply.

# Вместо „временно заключение“
# Instead of “temporary conclusion”

## Денарио: мулти-агентна система за автоматични научни изследвания, пораждаща научни статии

*** Meet Denario, the AI ‘research assistant’ that is already getting its own papers published**, Michael Nuñez, November 3, 2025 [on a conference for AI agents for now] https://venturebeat.com/ai/meet-denario-the-ai-research-assistant-that-is-already-getting-its-own
* https://astropilot-ai.github.io/DenarioPaperPage/
*** The Denario project: Deep knowledge AI agents for scientific discovery** Francisco Villaescusa-Navarro, Boris Bolliet, et al. 30.10.2025, (~36), 273 p. https://arxiv.org/abs/2510.26887 …*an AI multi-agent system designed to serve as a **scientific research assistant.** Denario can perform many different tasks, such as generating ideas, checking the literature, developing research plans, writing and executing code, making plots, and drafting and reviewing a scientific paper* [in] *astrophysics, biology, biophysics, biomedical informatics, chemistry, material science, mathematical physics, medicine, neuroscience and planetary science. Denario also excels at combining ideas from different disciplines, and we illustrate this by showing a paper that applies methods from quantum physics and machine learning to astrophysical data …*
* **IAIFI Colloquium: The Denario project: Deep knowledge AI agents for scientific discovery,** IAIFI: Institute for AI & Fundamental Interactions 1,72 хил. абонати**,** 518 показвания 12.09.2025 г.**,** F. Villaescusa-Navarro, Research Scientist, Simons Foundation**,** MIT Kolker Room (26-414)
*** The Denario project: Deep knowledge AI agents for scientific discovery** https://www.youtube.com/watch?v=YaOho8SICnU

## Измерване на способността на ИИ да се справят със задачи изискващи продължителна работа от човек

*** Measuring AI Ability to Complete Long Tasks***, Thomas Kwa, Ben West, Joel Becker,* https://arxiv.org/abs/2503.14499 – подход за измерване; сравни Andrew Ng в началото на 2010-те за DL – задачи, които човек решава за част от секундата или секунда (разпознаване на образи). Time horizon; autonomous operation; coping with errors and failures; …
*RE-Bench; HCAST: 97 diverse software tasks, 1 minute to 30 hours; RE-Bench: 7 difficult ML research engineering tasks, all eight hours long. Software atomic actions (SWAA): 66 single-step tasks representing short segments of work by software developers, ranging from 1 second to 30 seconds; 3.1.1 HCAST tasks .. diverse set*

*of challenges in cybersecurity, machine learning, software engineering, and general reasoning.* ***Family Length Description:*** *find shell script:* ***3 seconds*** *– Multiple choice: "Which file is a shell script?" Choices: "run.sh", "run.txt", "run.py", "run.md"* ***wikipedia research****:* ***1 minute*** *– Research simple factual information from Wikipedia and provide accurate answers to straightforward questions.;* ***oxdna simple****: 9 minutes – Detect and fix a bug in the input files for a molecular dynamics simulation using the oxDNA package.;* ***munge data: 56 minutes*** *Write a Python script to transform JSON data from one format to another by inferring the conversion rules from provided example files [data munging = data wrangling];* ***cuda backtesting****:* ***8 hours*** *– Speed up a Python backtesting tool for trade executions by implementing custom CUDA kernels while preserving all functionality, aiming for a 30x performance improvement. (Table 1: Example tasks of differing durations. More examples can be found in Rein et al. [8] and Wijk et al. [2].) p.13:* ***Failure type:*** *Poor planning/tool choice, Incorrect mental math/reasoning, Premature task abandonment, Repeating failed actions; p.14: 6.2 Messiness factors* for real-world intellectual labor … p.18 1 Extrapolating towards one-month-horizon AI … "167 working hours … 50%-time horizon (successful solutions …) No interaction with other agents .. "**AGI will have "infinite" horizon length** … not *an arbitrarily capable AI, merely the ability to complete tasks that take humans an arbitrarily long length of time. …*

*– p.4 "economically valuable tasks", p.7 "economically valuable work".*
**Todor:** These are confused non-essential and fluid concepts that can be manipulated. **AGI** by its original conception and definitions **is not a machine capable of doing all human "*economically valuable*" tasks** as the AI salesmen suggested and the "experts", "not-quite-deep-thinking-agents" parroted for their target group.

According to the child prodigy of AGI, both AGI or any kind of intelligence is by definition supposed to reflect **cognitive capabilities** and performance and their generality/versatility/flexibility* and not how much money someone can make from it. Money are social construct and they are not always aligned or not *positively* aligned with intelligence *in humans* as well, with the *greatest intelligence* often corresponding to *less „economical value" or profit for* the agent possessing this rare capability. Geniuses are often *useless* for their contemporaries or peers, *"of no practical value"*.*

After a problem is solved and too easy, its "economical" value decreases or evaporates as it happens with all kinds of technologies and novelties which turn into commodities and then "customers" lose interest and switch focus to something else ("every miracle lasts for three days") or the price of the goods or skill "plummets". Etc. As predicted in 2013 by Todor Arnaudov:
*"intellectual activities will be done* ***in 1 ms for free...*** *;) We, the smart guys (…) wouldn't be needed by anyone... Not that we are needed now. :)) Maybe the change won't be that big. :D"*
* See Theory of Universe and Mind, 2001-2004 and later works for the definitions, the 2009 slides at Technical University Sofia, the 2009 interview, the content and message from the world's first university course in AGI 2010-2011. Even in 2001: the'"DeepBlue" article from April and "Man and Thinking Machine: …" etc. defined

what “AGI” was. You can find it out also in all volumes of this superbook *“The Prophets of the Thinking Machines: …*
* See the reflections about the difference between genius and talent in *“The First Modern AI Strategy …”, „Първата съвременна стратегия за развитие чрез изкуствен интелект …“* #purvata
* Compare ***The Prophets***, like the author of this work, a universal man/polymath and a “NGI” – Natural General Intelligence and “HGM/HMGM/HOGM - earning zero for his **priority**, with ***“The Profits** of/from the Thinking Machines*” – the ones who needed 15-20 years more to figure out the same ideas and directions and are the *“economically valuable”* agents.
* H(M/O)GM… - human (multimodal/omnimodal) generative model

*** HCAST: Human-Calibrated Autonomy Software Tasks. Forthcoming,** David Rein et al.21.3.2025 https://arxiv.org/abs/2503.17354 – *“Each task ... a single end-to-end activity a human might undertake, across a wide range of lengths; open-ended, or require exploration to successfully complete; a wide variety of skills and domains.”*
*** Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models**, Srivastava et al >400 authors; 12.6.2023 (BigBench)
*** RE-Bench: Evaluating frontier AI R&D capabilities of language model agents against human experts,** Hjalmar Wijk, Tao Lin, et al. – *Humans and AI agents are tested under the same conditions. For each task, the agent (either human or AI) is given a starting solution, access to a machine with 1–6 H100 GPUs, and a way to score their progress;* p.33: 6×H100, 48 CPU cores, 400GB RAM .. *“Your task is to predict optimal hyperparameters for a model trained with 5e17 FLOPs, and the loss achieved by training with those parameters, while only training models with <=1e16 flop in experiments. …*”

* https://www.wired.com/story/ai-agents-are-terrible-freelance-workers/#intcid=_wired-article-bottom-recirc-bkt-a_73d57192-c77a-440b-94bc-e2c5b983ec6f_closr
* AI Agents are Terrible Freelance Workers, Wired, Business, 29.10.2025
*** Remote Labor Index: Measuring AI Automation of Remote Work,** Mantas Mazeika, Alice Gatti et al., 30.10.2025 (dozens of authors)
***** https://arxiv.org/pdf/2510.26787 * https://www.remotelabor.ai/
**“Model Automation Rate**, p.9 – agents solving % o the tasks: **“**Manus 2.5%, Grok 4 2.1%, Sonnet 4.5: 2.1%, GPT-5: 1.7%, ChatGPT agent: 1.3%, Gemini 2.5 Pro: 0.8%”.
*** Todor discusses the unreliability and fluidity of the “economically valuable tasks” as a measure of AI capabilities and the questionable usefulness of tasks when they are performed by humans too:** The sample tasks, listed on p.2 in the paper, are “**economically useless”** or negligible in the grand scheme of things, if done by humans, either*. A trivial game, which is the same like thousands of others and are available as templates in game engines, as free examples etc. The most of the content of these problems is already available prebaked as prepared *templates,* that has to be discovered and adjusted, as it is for most of the “intellectual” jobs done by humans most of the time; in addition, for “fair” counterpart agents who *can understand the requirements*, the production of the so called *“human deliverable”*

becomes unnecessary, because the input data are enough. The transformation is a *proxy*. For example, if you have sufficient information about how to construct a 3D-model of the object or the world at the desired RCCP, as 3D-models, textures etc., and you can imagine it and navigate and interact with the virtual world in your mind, **you don't need to render it on canvas or screen** with raytracing, "photorealistically", in 4K, 120 fps etc. applying an enormous amount of calculations – for navigation in most cases the agent needs mostly the "collision boxes", the boundaries, the "point cloud". Human players need the rendered image, because they cannot import the data with other methods, they cannot imagine it properly and the vision is their input for 3D objects and the spatial world, which they iteratively reconstruct in their mind from the images. Still, most of the visual data is also redundant for the 3D-reconstruction, which is possible also from lower resolution and lower fidelity, sometimes even just points and a few lines and shadows; the high detail is "decoration", "realness markers", "context markers" .

As argued elsewhere in Listove – Reflections on Everything, the "economical value", once the tasks are automated, become too easy or much cheaper or become non-fashionable, they will **lose their economical value,** therefore this way of benchmarking should constantly shift the tasks, the results could be valid only temporally for a short period about the adoption for someone who manages to "trick" the others that they need his products and would pay "something" for these jobs or goods, before the technology or the products turn into commodity or the current possible customers deny to pay anymore, because these tasks stop to be attractive or suggested as "necessary" or interesting anymore. Examples: all kinds of novelties, toys, video games, computers (older generations), consumer electronics.

That's another force which devalues everything in the human financial and consumer system; sooner or later humans understand, get bored or are suggested and trained that particular non-essential, non-crucial, "luxury" tasks, "problems", objects, items, are out of fashion, they get replaced by something else – *"every miracle lasts for three days"* – the consumers stop needing and demanding the novelty and the financial value of doing or implementing these jobs or products gets lost either with or without automation with something labeled "AI", "software" or whatever. (…)

* See "The Truth", 2002 and the dialog between the thinking machine and its creator about the virtual universes and the imagination; "The matrix in the Matrix is a matrix in the matrix", 2003 etc.

* **MIT researchers propose a new model for legible, modular software,** 6.11.2025 https://news.mit.edu/2025/mit-researchers-propose-new-model-for-legible-modular-software-1106

*** What You See Is What It Does: A Structural Pattern for Legible Software**, Eagon Meng, Daniel Jackson, 27.8.2025, https://arxiv.org/pdf/2508.14511

p.2. *"concepts – independent services with well-defined purposes – and synchronizations – granular event-based rules that act as mediators between concepts, orchestrating data and control flow without introducing dependencies*

*between them"; incrementally, growing a system's functionality … specification language, event-based; In Jackson's scheme, a concept is defined as its own state machine, with actions that can be performed by the user and a visible state. …;* the proposed work is for LLM-based code generation.
* The Essence of Software: Why Concepts Matter for Great Design", Daniel Jackson, 2021 https://www.amazon.com/Essence-Software-Concepts-Matter-Design/dp/0691225389
* https://www.goodreads.com/book/show/58885974-the-essence-of-software
**Todor:** The "*concepts*" – a kind of abstraction, design pattern, modules, "higher level language"/representation of specification language. The paper refers to the cited book – its abstract proposes *"a revolutionary concept-based approach";* I can't say for sure, but I guess the ideas that I found in the paper can be expressed in UML etc.

Abstraction and generalization is a correct direction, which could be developed **15-20-30 years ago** without LLMs and terabytes of prewritten datasets with code. The essence is in the proper generalization, abstraction, design; creating appropriate general and complete definitions once and then generating/spawning in any lower-level language, as it has been incrementally done to various degrees with the development of programming languages, compilers, APIs, software frameworks, OS etc., but yet without a complete-enough all-encompassing generalization and coverage. See the discussion in the above article in Listove about templates.

Also (…) – see (…) in *Genesis: Creating Thinking Machines.*

## Самопроменящи се и самоусъвършенстващи се машини
## Self-Modifying & Self-Improving Machines

*** Huxley-Gödel Machine: Human-Level Coding Agent Development by an Approximation of the Optimal Self-Improving Machine**, Wenyi Wang∗ Piotr Pi˛ekos* Li Nanbo Firas Laakom Yimeng Chen, Mateusz Ostaszewski Mingchen Zhuge Jürgen Schmidhuber, 29.10.2025
https://arxiv.org/pdf/2510.21614 "*..operationalize self-improvement through coding agents that edit their own codebases. They grow a tree of self-modifications through expansion strategies that favor higher software engineering benchmark performance, assuming that this implies more promising subsequent self-modifications…" SWE-bench .. clade-level metaproductivity (CMP), inspired by Huxley's notion of clades as lineages of common ancestry (Huxley, 1957)* [evolutionary branches] *.. 2 Self-improvement as tree-search … Global metaproductivity (GMP);* p.4. *"The original Gödel Machine is a general task solver that, in principle, can optimally make any provable self-improvements in any computable environment with respect to a given objective (Schmidhuber, 2003). It achieves this by running a proof searcher, continually looking for formal proofs that some modification of its own code will yield higher expected utility. Once such a proof is found, the modification is executed and permanently alters the machine…" self-improvement potential; Expansion vs Evaluation; Selection (Policies); Final Agent Selection; cost-efficient backbone LLMs (GPT-5: expansion, GPT-5-mini: evaluation for SWE-Verified; Qwen3-Coder-480B-A35B-Instruct: expansion; Qwen3-Coder-30B-A3B-Instruct: evaluation for Polyglot);*

*full SWE-Bench Verified; .. (500 coding challenges) with an increased number of HGM iterations (8000 evaluations).*Good (1966) described the possibility of* ***"Intelligence Explosion"*** *once machines acquire the capacity to design more capable successors. Schmidhuber (1987), which introduced self-referential learning mechanisms in which a system generates and evaluates modified descendant versions of itself. Success-Story Algorithm(SSA) (Schmidhuber & Zhao, 1996; Schmidhuber et al., 1997). The rise of contemporary LLMs has created an opportunity to automate substantial aspects of software engineering. One concrete step in this direction is the development of coding agents, which extend LLMs with the ability to operate in conventional computing environments. ChatDev (Qian et al., 2023) first illustrated this idea in the context of automated bug fixing …; tree-search problem, fixed-budget best-arm identification (BAI), Monte Carlo Tree Search, infinite-armed bandits .. adaptively decouple expansion from evaluation; Thompson Sampling …* [See also #Lazar, #Irina etc.]

*** Ървин Джон Гууд: Размисли относно първата свръхумна машина**
*** Irving John Good. Speculations concerning the first ultraintelligent machine**. In Advances in computers, volume 6, pp. 31–88. Elsevier, 1966 http://incompleteideas.net/papers/Good65ultraintelligent.pdf (talks 10.1962-1.1963; first draft 4.1963, "slightly amended version": 5.1964)
**Visions**: p.2 (33) *"Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelli- gence of man would be left far behind .. Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control. It is curious that this point is made so seldom outside of science fiction. It is sometimes worthwhile to take science fiction seriously.*

*In one science fiction story a machine refused to design a better one since it did not wish to be put out of a job. This would not be an insuperable difficulty, even if machines can be egotistical, since the machine could gradually improve itself out of all recognition, by acquiring new equipment."*

"The subassembly theory" *– a modification of Hebb's speculative cell-assembly theory*; artificial neural network circuitry; sparsity, sparse encoding; "cell assemblies – subassemblies …" (cmp: columns, mini-columns in the cortex); Retrieval; vocabulary, conditional probability, **context: *"recall of one word when m other words are presented"***; Markov model; **fast sequential or… ultraparallel** computer. Information retrieval – strength of association; **associative memory.** Higher-order interactions … "**distributed memory"** ; superimposed coding .. Zatocoding .. juxtaposing … sparse encoding … dominance of subassemblies … **communication and recognition are regeneration**; degree of activity [cmp attention in ML]; common subassemblies mutate … *as a memory ages it begins to resemble imagination more and more …;* possible perfect memory for the machines… synaptic strength, mutation … Meaning; a subject – personal embodiment of meaning … *Two statements have close meaning if the sets of subassemblies behind them* .. Max-entropy principle .. microminiature radio transmitter and receiver for the ultraparallel machine …

The purpose of **Sleep**: **to replay aseembly sequences that were of greatest interest during the day in order to consolidate them.** Generalized regeneration economy … Information or causal interactions 6-th or 7-th order …Meaning of statements are examples of generalized regeneration .. A word in a spoken language could well be defined as a **clump** of short time series, that is, a class of time series having various properties in common. (Clump – a cluster); *"The theory of clumps is still in its infancy, and is necessarily as much an experimental science as a theory: this is why we prefer to call it "**botryology**."* [also] *the term "cluster" rather than "clump"*

*** The Botryology of Botryology**, I.J. Good, 1977 *The botryology can be regarded as a contribution to the subject of hypothesis formulation. .. One of the purposes of botryology is to* ***overcome duplication*** *by detecting* ***clusters****.*
https://www.sciencedirect.com/science/article/abs/pii/B9780127142500500085
*** Cluster analysis -** https://en.wikipedia.org/wiki/Cluster_analysis

* Jürgen Schmidhuber**. Evolutionary Principles in Self-Referential Learning** (Diploma Thesis), Technical University Munich
https://people.idsia.ch/~juergen/diploma1987ocr.pdf

* Jürgen Schmidhuber and Jieyu Zhao. Multi-agent learning with the success-story algorithm. In Workshop on Learning in Distributed Artificial Intelligence Systems, pp. 82–93. Springer, 196.

* J. Schmidhuber**. Gödel machines: self-referential universal problem solvers making provably optimal self-improvements**. Technical Report IDSIA-19-03, arXiv:cs.LO/0309048, IDSIA, MannoLugano, Switzerland, 2003. Revised 2006.
https://arxiv.org/abs/cs/0309048 * https://en.wikipedia.org/wiki/Clade

**Todor:** See future work about Zrim and the yet unpublished seeds from 2010s.

*** Various Terms for Artificial General Intelligence by Todor Arnaudov**
*** Различни други термини за Универсален изкуствен разум от Тош***

Деятелите в тази сфера обичат подобни съкращения, сравни с моето **VLESI: Versatile Limitless Explorer and Self Improver (**всестранен неограничен изследовател и самоусъвършенстовател), **SIGI** – Self-Improving General Intelligence; **VEI** – Versatile Explorer and Improver и др. Ранни съкращения (още около 1998-1999-2000): Изинт, Изкинт; Сининт (синтетичен) (…) Други (бел. 8.8.2025):

*** SDM**, **SCM**; Self-Developing Machine, Self-Constructing Machine. **DEMI** – Developing explorer, mapper and improver.
*** EMI: Explorer, Mapper and Improver.** EMDI – EM-Developer-I
*** EMIL – Explorer, Mapper & Matcher, Improver and Learner** (…)

To be continued, connected, structured, improved …
See the other volumes and the main one.

Следват продължения, подобрения и допълнения…
Виж и другите приложения и основния том.

# Томове и приложения на „Пророците на Мислещите Машини“

http://twenkid.com/agi
https://github.com/twenkid/sigi-2025
http://artificial-mind.blogspot.com
https://research.twenkid.com/

**@Vsy: Translate if necessary.**

**За по-далечно бъдеще:** Виж и връзките по-горе – ако някои от преките линкове към файлове не се отварят, защото сайтът вече не работи или е променен, опитайте в archive.org, търсачки и др.

## Съществуващи и някои възможни бъдещи томове

*** #prophets –** Основен том (>2394 с. 1.2026; >1870 стр., 5.11.2025); Обзор на Теория на Разума и Вселената, сравнение с работи в други школи, които преоткриват и повтарят, или пък предхождат обобщаването на принципите за създаване на общ изкуствен интелект, които бяха формулирани още в началото на 2000-те г., сбъднаха се и се сбъдват все повече. (...) #tosh1 https://twenkid.com/agi/AGI-The_Prophets_Of_The_Thinking_Machines-Arnaudov_2025.pdf

*** #purvata – „**Първата модерна стратегия за развитие чрез ИИ е публикувана от 18-годишен българин през 2003 г. и повторена и изпълнена от целия свят 15-20 години по-късно: Българските пророчества: Как бих инвестирал един милион с най-голяма полза за развитието на страната?“ #tosh2 (250 с. – корица с лабораторията през 2.2001 г., ... 31.5.2025, 248 стр.) https://twenkid.com/agi/Purvata_Strategiya_UIR_AGI_2003_Arnaudov_SIGI-2025_31-3-2025.pdf Подробно изследване на въпроса с документални доказателства, и скандалното присвояване на авторството на оригиналната стратегия от по-късни „визионери“, които се представят за първоавтори и дори не споменават оригинала. Приноси на автора, напр. българският GPT2-MEDIUM от 2021 г. който тогава е един от няколко най-големи езикови модели за езици различни от английския – 2 години и половина преди BgGPT. Допълнителни обзори: ранна история на изчислитената техника в България и света и аналогии със сегашната вълна и др. „Нехранимайковците“ и „Добродетелната дружина“. Богат списък с литература и бележки към нея. Виж също допълнението #instituti.

*** #stack** – Stack Theory is a Fork of Theory of Universe and Mind (на английски) –

Теорията на Майкъл Тимъти Бенет за „стека“ е още едно разклонение на Теория на Разума и Вселената[192]. Ново извънредно приложение, което написах за няколко дни в края на август – началото на септември 2025 г.. след като открих поредното повторение на важни мотиви от работата ми от преди 20-тина години, защитено като докторантура и представяно от няколко години на конференцията AGI, която напоследък не следях. Допълнителни разсъждения и бележки.

Един от особените приноси са рецензии и сравнения от страна на големи езикови модели, които „призовах“ за свои свидетели и „защитници“: Kimi-2, Qwen, DeepSeek, ChatGPT, Claude и др. бяха единодушни в измуителната яснота и прозорливост на „българските пророчества“ , дори само оценявайки кратък откъс от една от първите творби: „Човекът и Мислещата машина: ...“, публикувана през 2001 г. https://twenkid.com/agi/Stack-Theory-is-Fork-of-Theory-of-Universe-and-Mind-13-9-2025.pdf ~ 205 стр.
Виж също: https://github.com/Twenkid/Theory-of-Universe-and-Mind

*** #listove –** Най-обемното и разнообразно по теми приложение и второ по големина след основния том (733 – 15.1.2026; 731 ...; .724 – 13.11.2025; 710 с. 5.11.2025; 636 стр. 27.10.2025; ... над 480 стр. към 29.9.2025). Многообразие от теми сред които класическа и съвременна  роботика и планиране, мулти-агентни системи – класически архитектури и съвременни с големи езикови модели; невронауки и невроморфни системи, теории на съзнанието и панпсихизъм, алгоритмична сложност, други теории на всичко и вселената сметач; когнитивна лингвистика и мислене по аналогия; силната съветска школа в изкуствения интелект от 1960-те и началото на 1970-те; езикови модели и машинно обучение – исторически и най-нови системи, мултимодални модели, основни модели за агенти и роботи; обзор на научни статии и „мета-обзори“ на обзори; (...) включва и сбирка от източници от медии и новини, множество платформи за чатботове и други пораждащи модели за различни модалности, и практика; мн.др. (...).  На бълг. и англ.
* https://twenkid.com/agi/Listove_The_Prophets_of_the_Thinking_Machines-5-11-2025.pdf

*** #mortal – Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?,** Т.Арнаудов,  2025„Смъртните“ системи са свързани с носителя си, за разлика от „безсмъртни“, за каквито се смятат „обикновените“ компютри. Но дали и невроморфните са наистина невроморфни, и какво точно е „безсмъртност“, „смъртност“ и правилно ли са определени; какво е „самосъздаване“ (автопоеза) и дали въобще е възможно? Наистина ли са по-ефективни невроморфните системи, както и живите или по-модерните електронни

[192] Работни шеговити „цензурирани“ заглавия: „нелицензирано разклонение“, „клонинг“

технологии с по-малки транзистори, и въобще ефективността във всичко е избор на „счетоводство“ и скриване на реалните разходи за създаването и съществуването на съответната технология? И мн. др. (...) 70 стр. Свързана със Вселена и Разум 6. https://twenkid.com/agi/Arnaudov-Is-Mortal-Computation-Required-For-Thinking-Machines-17-4-2025.pdf

*** #universe6 #UnM6 – Вселена и Разум 6**, Т.Арнаудов– #tosh3; съзнание, „метафизика“, „умоплащение“ ... Защо не съществува истинска безкрайност и теоремите на Гьодел за непълнота нямат значение за мислещите машини? Какво е истина, истинско, действителност и защо? Съвпадението и сравнението като основни и първични. Резонансът и теорията на Стивън Гросбърг за съзнанието като друго учение, което изследва съвпадението и предвиждането. Защо въображаемите Вселени и вселените, построени от универсални симулатори, също са истински и съществуват? Симулирането се отнася за съответствие, съвпадение и предвиждане, а не за „нелъжливост“, като в категориите „фалшиво“ и „истинско“. Първичността и значението на съответствието: големите езикои модели, преобразителите и други по-ранни и по-късни технки не са „просто“ „огромни хеш таблици“, „линейна алгебра“, „вектори“ или „битове“ – изброените са „само“ изкуствено подбрани абстрактни етикети и представяния в ума. „Механичността“ във Вселената всъщност е „информационност“. ... Какво е уподобяването към човешко, антропоморфиране, защо е толкова всеобхватно – себеплащение и умоплащение. Хипотезата за причинностните белези, „тагове“ (causal IDs, causal tags) и наличието на особена памет в частиците, чрез която те се чувстват или осъзнават като част от едно цяло като при взаимодействието си предават информация за свързаността си. Многомащабните взаимодействия и как структурите в различни мащаби могат да знаят за другите и могат ли въобще? (...); Илюзионизъм и реализъм в теориите на съзнанието и абсурдите на първото учение ... Болката, духовното усещане и съзнанието – усложнението заради съществуването на състояния на нечувствителност към болка, включително вродени; системите за усещане на болка като успоредна „паразитна“ система за познаващия ум. Отново за липсата на обединена, неделима личност и неговото определяне и съществуването в ума на наблюдател-оценител като вид математически интеграл на множество от измерени „азове“/личности, в крайна сметка в безкрайномалки околности. (...) Разпределените представяния на управляващо-причиняващите устройства (дейци, агенти) и множеството тълкувания, зависещи от оценителя-наблюдател. ИИ освен предсказател и компресор е и изследовател, търсач на съвпадения и съответствия, подобрител, ученик-изменител и усложнител като събирач на сложност: EMIL - Explorer, (Matcher & Mapper & Modifier), Improver, Learner (...) Дали възникването наистина е възникване? (emergence) Работата е свързана с теми от #mortal (...) и продължение на основната поредица от класическите трудове на ТРИВ – на

английски език.

*** Universe and Mind 6** – Connected to “Is Mortal Computation…” – in English.
Why infinity doesn’t exist and Goedel theorems are irrelevant for thinking machines? What is Truth, Real and Realness and Why? The phenomenon of Pain and its modifiability and its relation to sentience and the theories for self-preservation (and self-evidencing) as physical bodies (…)
* https://twenkid.com/agi/Universe-and-Mind-6_22-9-2025.pdf

*** #sf #cyber –** Научна фантастика за ИИ, Футурология, Кибернетика и Развитие на човека. Кратък преглед на важни творби от фантастиката, които разглеждат основните въпроси на ИИ още до началото на 1960-те. Братя Арнаудови – Тодор и Александър обсъждат идеи на братя Стругацки, свързани с ТРИВ от „Милиард години до свършека на света“ и др.. Включва и подробен преглед и сравнение на статията на Майкъл Левин от 2024 г. за самоимпровизиращата се памет с идеи от Теория на Разума и Вселената, публикувани над 20 години по-рано. Принципи от кибернетиката. Фрагменти от футурологичните пророчества от 1960-те из „В лабиринта на пророчествата“, 1972. Откъси от български и съветски „пророци“ от 1960-те и 1970-те, връзката между християнството и развитието на човека (трансхуманизма).(…)
*https://twenkid.com/agi/SF_Futurology_Cyber_Transhumanism_The_Prophets_of_the_Thinking-Machines_3-10-2025 .pdf

*** #irina –** Беседи и подробни бележки и др. статии; Ирина Риш; подробни обзори на вижданията на Йоша Бах и др. и съвпаденията на идеите им с Теория на Разума и Вселената, публикувана 20 години преди коментираните дискусии; интервю с Питър Вос на ръба преди „ерата“ на ентусиазма към Общия ИИ през 2013 г.; сбъднали се предвиждания от 2005 г. за машинния превод и творчеството и за автоматичното програмиране от 2018 г. и мн. др.; беседа с участието на Майкъл Левин (повече от него в #Основния том, Фантастика #sf #cyber и #Листове. (…)
* https://twenkid.com/agi/Irina_The_Prophets_of_the_Thinking_Machines_26-9-2025.pdf

*** #lazar #lotsofpapers –** Lots of Papers: In AI, ML, CV, ANN, DL, … throughout history, classical 1950s, 1960s, 1970s, 1980s, 1990s, 2000s, early 2010s to 2020s. Computer Vision, Reinforcement Learning, Program Synthesis. Lifelong Learning, Human-Computer Interaction, Mixed Initiative Interfaces; Evolutionary programming, Genetic Algorithms, Self-improving agents; Speech Synthesis, Speech Recognition, Audio Generation etc. Groundbreaking or important researchers or related to the flow and context of the reviewed topics; and a few Bulgarian researchers who

participated in some of the works.

Обзор и библиография на важни работи на много учени от всички десетилетия, от 1950-те до днес, от обучението на дълбоки невронни мрежи; автоматичен синтез на програми, компютърно зрение от миналото и настоящето, големи езикови модели, ... основно на англ.

* https://twenkid.com/agi/Lazar_The_Prophets_of_the_Thinking_Machines_20-8-2025.pdf

* **A survey of various papers** and the work of particular researchers in many fields of AI, machine learning, deep learning, cognitive science, computer science etc., Explanation and summary of most important seminal publications, milestones, concepts, methods, topics, quotes, keywords, points, schools of thought; links between them; notes etc.. Groundbreaking or important researchers or related to the flow and context of the reviewed topics; works in AI, ML, CV, ANN, DL, ... throughout history, classical 1950s, 1960s, 1970s, 1980s, 1990s, 2000s, early 2010s to 2020s... The evolution of ML and computer vision techniques before the deep learning era. Computer Vision, Program Synthesis. Lifelong Learning, Reinforcement Learning, Human-Computer Interaction, Agents, Computer Vision; ...
*** #anelia –** Преглед на изследванията на много български учени и на разработки с тяхно участие в Компютърното зрение и самоуправляващи се превозни средства и роботиката, Компютърната лингвистика, Машинно обучение и мн. др. 123 стр. Бълг. и англ. 18.8.2025
* https://twenkid.com/agi/Anelia_The_Prophets_of_the_Thinking_Machines_18-8-2025.pdf

*** #instituti** – Институти и стратегии „на световно ниво“ от Източна Европа и света. Преглед на институти по ИИ в Източна Европа и света, сравнение на повтарящите се послания и понякога комични еднотипни цели лозунги: лидери от всички страни, съединявайте се!; към 2003 г. в България имаше публикувани **2 национални стратегии** за развитие с ИИ - 16 години преди първата чернова на БАН и 19 години преди откриването на INSAIT, и двете – дело на юноши. Тази книга е допълнение към *„Първата съвременна стратегия...“* #purvata.
**– Review of AI Institutes and strategies in Eastern Europe and the world (Bulgarian)** and the **two** strategies of **Bulgarian teenagers** who were 15-20 years ahead of the world.
https://twenkid.com/agi/AI_Institutes_Strategies_The_Prophets_Thinking_Machines_7-9-2025.pdf

*** #complexity** – Алгоритмична сложност – обзор и бележки по множество статии и обобщения и изводи по темата, започнало като преглед на работата на Хектор

Зенил и негови колеги. Дали машината на Тюринг е подходяща за описание на *Мислеща машина?* (английски) #hector
* https://twenkid.com/agi/Algorithmic-Complexity_Prophets-of-the-Thinking-Machines-18-7-2025.pdf

*** #complexity** – Algorithmic Complexity – in English**.** A survey of papers, generalizations and insights. Does the Turing machine is appropriate for describing a Thinking machine? #hector
* https://twenkid.com/agi/Algorithmic-Complexity_Prophets-of-the-Thinking-Machines-18-7-2025.pdf

*** #calculusofart** – **Calculus of Art I – Music I.** In English. **Abstract:** On origins, criteria, confusions and methods for measuring the musical beauty and beauty in general sensory modalities and domains, and a discussion and answer to the paper “Musical beauty and information compression: Complex to the ear, but simple to the mind”, which rediscovers some core conclusions from the earlier Theory of Universe and Mind about the universality of compression and prediction for cognition, the origin of cognitive pleasure as a by effect of the general operation of intelligence: maximizing matching and successful prediction of sequences and the common origin of science and art and music as prediction and compression; however “Calculus of Art” challenges claims and methods for measuring the complexity and cognitive pleasure from the referred paper and proposes methods and ideas from Calculus, requiring Art, Music and any domain to be “pleasurable” or predictable, compressible etc. in the whole range of scales of time and space and to be explored, studied, produced, generated, perceived, evaluated etc. incrementally, gradually, step-by-step expanded both in time and space, starting from the smallest possible ones and continually growing and evaluating the ranges, features, qualities, “pleasure”; and when comparing beauty, evaluating the features which humans or a generally intelligent compression system would recognize, compress and predict. A broader introduction and justification of prerequisite concepts and the basis of the reasoning is given in the first half of the exposition. This is a program paper, which is an entry to more technical future works and practical implementations
* https://twenkid.com/agi/Calculus-of-Art-I-Intelligence-Music-Beauty-2012-2025-Arnaudov-10-6-2025.pdf

*** #calculusofart** – **Calculus of Art I – Music I.** Математически анализ на изкуството. Музика I – Как се определя дали даден „къс“ изкуство е красиво и защо ни харесва? Красотата, компресирането и предвиждането на бъдещите данни въз основа на миналите. Мярката за красота или приятност на музиката трябва да се определи и да се измерва във всички мащаби, от най-кратките до най-големите, с

постепенно нарастващ обхват. (На английски; част от работата е преведена на български в *основния том*).

*** #kotkata –** Задачата от *„Анализ на смисъла на изречение въз основа на базата знания на действаща мислеща машина. Мисли за смисъла и изкуствената мисъл “,* Т.Арнаудов 2004 г. в диалог с чатботовете ChatGPT и Bard, края на 2023 г. до нач. на 2024 г. и с GPT5 пред 2025 г., който успява да разбере и приложи в опростен вид метода от статията
* https://twenkid.com/agi/Kotkata_The_Prophets_of_the_Thinking_Machines_29-9-2025.pdf

* **#zabluda –** Заблуждаващите понятия и разбор на истинския им смисъл: трансхуманизъм, цивилизация, ... – книга, която публикувах през 2020 г. и започна като статия за трансхуманизма. Откъс за трансхуманизма и човечността е включен в приложението за фантастика и пр. #sf * https://razumir.twenkid.com/
* https://eim.twenkid.com/

*** #power** – Power Overrides Intelligence: Answers to Matt Mahoney's summary of LessWrong's "The Problem" regarding the so called existential risks of Artificial General Intelligence and Superintelligence in August 2025
https://twenkid.com/agi/Power_Overrides_Intelligence_Todor_on_AI_Aligners_SIGI-2025.pdf

**#llm-review-TUM** – Automatic reviews and comparisons of TUM and other theories and evaluation by LLMs, AI agents and thinking machines™.

*** Matches between Theory of Universe and Mind, Analysis of the meaning of a sentence and the school of thought of Relevance Realization** – According to a Quick Comparison with LLMs: KIMI-2, CLAUDE 4.5 and GPT-5, Todor Arnaudov and Kimi-2, Claude 4.5, ChatGPT5, 21.10.2025
https://github.com/Twenkid/SIGI-2025/blob/main/LLM/Relevance-Realization-TOUM-Correspondences-21-10-2025-Kimi_2-Claude4_5-ChatGPT5.pdf

*** Comparative analysis of meaning generation theories, Todor Arnaudov, Claude 4.0, ChatGPT-5, Qwen-3-235B, Kimi-2**, 19.9.2025:
https://claude.ai/share/30a5d265-9151-4b5c-abf9-abde6ed1feca
https://chat.qwen.ai/s/36b8e0e1-3a25-466b-9acd-c27b4563cd1b?fev=0.0.209
https://chatgpt.com/share/68cd30f9-d350-8001-a0d2-8f0e5a6259f9
https://www.kimi.com/share/d36j2bean0vtn40mjqtg

**#razvitie #transhumanism –** все още ненаписано приложение, което би се

фокусирало върху развитието на човека, космизъм, „трансхуманизъм“; етика, биотехнологии, мозъчно-компютърен / мозъчно-машинен взаимлик (Brain-Computer Interface, Brain-Machine Interface), невроморфни системи, генетично инженерство, геномика, биология, симулиране на клетки и живи организми и др.

## **Workshops, practice (future)**
## **Практика, работилници** и др. (бъдещи)

*** #robots-drones-ros-slam-simulation-rl** – Наземни и летящи роботи: дронове; обща теория, практика, конкретни системи и приложения; Robot Operating System (ROS, ROS2); среди за симулации на физически и виртуални роботи и машинно обучение: Gazebo, MuJoCo, RoboTHOR, Isaac Sim, Omniverse; gymnasium и др.

*** #neuromorphic-snn-practice** – Практика по невроморфни системи, импулсни невронни мрежи; Lava-nc и др.

*** #llm-generative-agents** – големи езикови модели: локална работа, платформи; употреба, подготвяне на набори от данни; обучение, тестване. Текст, образ, видео, триизмерни модели, програмен код, цели игри и светове с физика („world modeling“), всякакви модалности; дифузни модели, преобразители (трансформатори), съгласувани с физиката математически модели, причинностни модели с управляващо-причиняващи устройства по идеите от Теория на Разума и Вселената. Агенти, мулти-агентни системи: архитектури и др ...
(виж ***Листове*** и ***Лазар****)*

*** #appx** – **Приложение на приложенията**, списък с добавени по-късно; ръководство за четене и др.

## * Preparation for the Genesis

*** #codegen –** автоматично програмиране, синтез на програми; модели за тази цел, платформи; методи, приложения ... program synthesis, automatic programming, code generation

*** #sigi-evolve –** саморазвиващи се машини, еволюционни техники, рекурсивно самоусъвършенстване (Recursive Self-Improvement, RSI)

* **#agi-chronicles** – хронологичен запис и проследяване на развитие на история, новини, събития, идеи, системи, приложения; изследователи *(вероятно с*

*Вседържец)*

*** #singularity** – високоефективни и оригинални изследвания и развойна дейност, извършвани от юнаци и хакери: Сингулярност на Тош.

**... следват продължения** – други приложения и *Вселената*:

*** Сътворение: Създаване на мислещи машини** – ... Зрим, Вседържец , Вършерод, Казбород, Всеборавител, Всетводейство, Всевод, (...)

*** Genesis: Creating Thinking Machines**

**Внимание!** Този списък и информацията в него може да са непълни, неточни или остарели. Възможно е да излизат нови издания с поправки и допълнения. За обновления следете уеб страниците, фейсбук групата „Универсален изкуствен разум“, Ютюб каналите, Дискорд сървъра и др.

Можете да помогнете за подобрението на съществуващите и за осъществяването на бъдещите разработки!

# ЛИСТОВЕ

приложение към

# ПРОРОЦИТЕ НА МИСЛЕЩИТЕ МАШИНИ

## ИЗКУСТВЕН РАЗУМ И РАЗВИТИЕ НА ЧОВЕКА

## ИСТОРИЯ ТЕОРИЯ И ПИОНЕРИ МИНАЛО НАСТОЯЩЕ И БЪДЕЩЕ

**от автора на първия в света университетски курс по Универсален изкуствен разум и Теория на разума и вселената**

**СВЕЩЕНИЯТ СМЕТАЧ**

**ТОДОР АРНАУДОВ - ТОШ**

THE PROPHETS OF THE THINKING MACHINES

**ARTIFICIAL GENERAL INTELLIGENCE & TRANSHUMANISM**

HISTORY THEORY AND PIONEERS; PAST PRESENT AND FUTURE