

СВЕЩЕНИЯТ СМЕТАЧ  
ТОДОР АРНАУДОВ - ТОШ

# ИРИНА

**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**

Редакция: 29.9.2025

Обнарогване: 26.9.2025

<http://twenkid.com/agi>

<https://github.com/twenkid/sigi-2025>

<http://artificial-mind.blogspot.com>

## **ПРОРОЦИТЕ НА МИСЛЕЩИТЕ МАШИНИ**

Изкуствен разум и развитие на човека:

История теория и пионери

Минало настояще и бъдеще

Тодор Арнаугов – Тош

## **ПРИЛОЖЕНИЕ**

**Подробни записки и бележки по беседи с други учени и сравнения с изпреварващите предвиждания на теория на разума и вселената; други сбъднали се предвиждания и потвърждения**

Обзор с откъси и подробни бележки към беседи с други изследователи по изкуствен интелект, машинно обучение, универсален изкуствен разум – Artificial General Intelligence, AGI; философия на ума и съзнанието, връзката между „физиката и ума“; за Вселената като изчислителна машина (Universe Computer, Вселената сметач), Принципа на свободната енергия и извод чрез действие от школата на Карл Фришън; сбъднали се предвиждания за машинния превод и въображението от 2005 г., и за скорошното осъществяване на автоматичното програмиране от 2018 г. и др. – които разясняват, потвърждават и доказват изпреварващия принос и предвиждания на Теория на Разума и Вселената или т.нар. Български пророчества (2001-2004 и по-късно), чрез преоткриването ѝ и съвпаденията на съдържанието ѝ в по-късна литература от водещи учени, философи, инженери и в реално осъществените практически системи; нова статия от Тош за слабостите на човеците спрямо езиковите модели и др.

#irina **файл:** #Irina-Rish-Joscha-Bach-etc-Appendix-...

- Вселената сметач / цифрова физика
- Самоорганизиращи се системи
- Универсален изкуствен разум
- Умът и съзнанието
- Сравнение на съвпадения с трудовете на "Свещеният сметач"
- Бележки
- Продължава в други по-обемни приложения с по-голямо многообразие от теми и автори като **Листове** #Listove и основния том, огромен обем литература в #lazar I #anelia; основа и въведение за някои от темите са дадени в Основния том , част от заключението на това приложение е развито в „*Първата стратегия...*“ #purvata.  
За съзнанието и Вселената виж също вече публикуваните
  - Вселена и Разум 6
  - Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?
  - Stack Theory is yet another Fork of Theory of Universe and Mind

Записките в приложение *Ирина* и цялата съвкупност на *Пророците* могат да служат като основа и зародиш за ориентиране и задълбочаване на познанията в съответните области и теми и те вероятно ще послужат като такива най-малко на мислещите машини ***Вседържец, Емил и Ускрителя на изследователската дейност.***

Поне част от съдържанието в тази книга може да е по-достъпно за по-широк кръг от читатели в сравнение с други по-технически като Лазар, Анелия, по-голямата част от Листове и др.

Полезно е да се познават или да се разбират поне някои от работите от класическата Теория на Разума и Вселената и съответните необходими знания за тяхното усвояване.

Целият „библейски“ труд е своеобразно *продължение на първия в света курс по Универсален изкуствен разум*, създаден и проведен от автора в Пловдивския университет „Паисий Хилендарски“ през 2010 и 2011 г., затова изучаването на лекциите от тогава също е основа за разбирането на книгите и за пророческото им съдържание, потвърдено от по-късни публикации на учени

и научни и философски школи на най-високо ниво, както и в практиката на изкуствения интелект.

Повече уводни бележки и обяснения виж в други приложения като Листове, Основния том и др.

**\* Можете да се присъедините и да подкрепите Свещеният сметач: морално, финансово, технически** или по всякакъв начин, включително като автори или съавтори в бъдещи токове на Пророците или други произведения и системи на виртуалната конференция и да се включите в групата във Фейсбук „Универсален изкуствен разум“, Дискорд канала – който може да откриете от други сайтове.

Виж в началото на том „Анелия“, в основния том и в проекта „Вседържец“ в Гитхъб за примерни конкретни начини за сътрудничество и помощ като техника, дарения, участие в проекти, разгласяване, свързване с подходящи хора, морална подкрепа и др.

**Автор, тълкувател, коментатор, съставител и подбор на откъси, редактор, преводач:** Тодор Арнаудов – Тош / Todor Arnaudov – Tosh

**Автори на разгледани произведения, интервюта, споменавания:**

Peter Voss, Irina Rish, Joscha Bach, Stephen Wolfram, Lex Fridman, MLST (Machine Learning Street Talk): Тим Скарф, Джеф Бек, Bert de Vries; Michael Shilo, Anastasia, Алан Томпсън, MichaelTrazzi, Michael Levin, Pedro Domingos и др., упоменати на съответните места;

**Примери за превод на стихотворение от „Алиса“:** ChatGPT 3.5 (2023), gpt4-4o-latest-20240903, Claude 4 (15.6.2025)

**Благодарности** на семейството ми; на приятелите, които са дарявали или ще дарят техника и морална подкрепа; и на всички, които подкрепят мен и **Свещеният сметач**.

## Частично съдържание

- Основателни ли са доводите за песимизма в ИИ? Интервю с Питър Вос, 8.1.2013 – бележки и относно това че някои видни експерти в ИИ и познавателните науки не са имали представа от природата на мисленето и творчеството, както показва реакцията на Д.Хофстатер и композираната от машина музика
- Ирина Риш – Универсален изкуствен разум. Мащабиране. Съгласуване. 18.10.2022
- Йоша Бах - Познаваемата Вселена, 30.12.2021
- Наистина ли „трудният проблем“ на съзнанието съществува? – Йоша Бах, изследовател в ИИ, 11.06.2023 г
- Вектори на Познавателния ИИ: Подбуди и Автономност, 28.12.2021 г
- Йоша Бах гостува на Лекс Фридман №212
- Общ ИИ отвъд Многослойното обучение, Познавателен ИИ, 11.01.2023 г.
- Разклоняващи се системи съхраняващи пълната история на преходите на състоянията като модели за разбиране на Разума и Вселената - беседа със Стивън Волфрам [и Йоша Бах], 13.04.2022 г.
- Джеф Бек – Бейсов мозък, предвиждане ... 13.10.2023
- Педро Домингос: Главният алгоритъм, петте племена; Pedro Domingos
- Кенет Станли – новото ...
- Обратно броене до създаването на Универсален изкуствен разум: 42% през март 2023 (трансформатор, ГПТ-3, TPUv4, H100, ChatGPT с тяло, Palm-E), Алан Томпсън, 3.2023
- "Вероятностен папагал" (Творчеството е подражание...) - откъси от 25.5.2004 г. във форум „Кибертрон
- Човеците са по-неспособни от езиковите модели по много показатели \* **Humans are far worse than LLMs in many ways: a brief analysis**, Todor Arnaudov, 2.8. 2025 (in English)
- Предвиждания за автоматичния превод и машинното въображение от 2005 г. изказано пред машиномразци. Отношението към „*пророците на мислещите машини*“ в български обществен блог през 2005 г. Тодор Арnaudов и др. и нови бележки.

- Относно превода на Information Retrieval, 10.2006, Ива Попова, Тош и др. – езиковите модели вече разбират псевдокода за създаване на нови думи
- Премествания за близките пробиви в автоматичното програмиране – чат на Тош с експерт по ИИ от 2018 г. и заключителни бележки за споменатите „швейцарци“, тогава известни само на „просветени“

**@Vsy: Technical Note:** If an URL doesn't open, try with archive.org or Internet search with the title and the available content.

## **СВЕЩЕНИЯТ СМЕТАЧ**

### **МИСЛЕЩИ МАШИНИ, ТВОРЧЕСТВО И РАЗВИТИЕ НА ЧОВЕКА**

Целогодишна виртуална конференция „Мислещи машини 2025“, или Self-Improving General Intelligence 2025 – SIGI-2025. Продължение на може би втората най-стара международна „конференция“ за универсален изкуствен разум (AGI): SIGI-2012-1, провела се присъствено в Пловдив през 2012 г.

## **THE SACRED COMPUTER**

### **THINKING MACHINES, CREATIVITY AND HUMAN DEVELOPMENT**

Thinking Machines 2025/Self-Improving General Intelligence SIGI-2025:  
A yearlong virtual conference

This appendix: mainly in Bulgarian, an article and some notes in English. The original titles in English and links to the sources are included.

## Томове и приложения на Пророците – някои накратко

- \* **#prophets** – Основен том #tosh1
- \* **#purvata** – Първата съвременна стратегия за развитие чрез ИИ #tosh2
- \* **#mortal** – Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини? #karl
- \* **#calculus – Calculus of Art I: Music I** (Математически анализ на изкуството: Музика I. На английски и част от нея на български в Основния том)
- \* **#complexity – Algorithmic Complexity** – алгоритмична сложност (англ.) #hector
- \* **#UnM6– Universe and Mind 6** (in English) Вселена и Разум 6 #tosh3
- \* **#listove** – Листове
- \* **#sf #cyber** – Научна фантастика за ИИ, Футурология, Кибернетика
- \* **#instituti** – преглед на институти по ИИ в Източна Европа и света
- \* **#irina** – Ирина Риш, Йоша Бах и гр. – интервюта и бележки; #joscha ...
- \* **#lazar #lotsofpapers** – работата на много български и гр. учени
- \* **#anelia** – друг том с работата на много български и гр. учени в ИИ и пр.
- \* **#kotkata** – Дали езиковите модели вече се справят със задачата от „Анализ на смисъла на изречение ...“, 2004 г.: „Комката изпи камъка и литна под нанагорнището...“? – тестове от 2023 и началото на 2024 г.
- \* **#appx** – приложение на приложенията, списък с добавени по-късно и гр.
- \* **#zabluda** – „Заблуждаващите понятия и разбор на истинския им смисъл: трансхуманизъм, цивилизация,...“ , Т.Арнаугов 2020 – Частта за трансхуманизма, която в началото беше глава от *Пророците*, но се отдели в книга на по-широка тема. Етика, философия, обществознание, езикознание, семантика, семиотика, социални науки, култура, история.
- #stack** – Stack theory is yet another fork of Theory of Universe and Mind
- **Виж повече подробности в края на този том**
- **Виж по-подробни въведения в:** Основния том, Първата съвременна стратегия, Листове, Институти и стратегии на световно ниво и гр. (...)
- \* Страница в Гитхъб за **Теория на Разума и Вселената** (и англ. език): <https://github.com/Twenkid/Theory-of-Universe-and-Mind/>
- \* Пророческото есе „Къде отиваш свят?“, 1999 - към създаването на мислещата машина – машината Бог (български). <https://github.com/Twenkid/Theory-of-Universe-and-Mind/blob/main/1999.md>
- \* Списание „Свещеният сметач“, Дружество „Разум“, оригинални публикации: <https://eim.twenkid.com/old/razum/index.htm> – Човекът и Мислещата Машина – анализ на възможността за създаване на мислеща машина и някои недостатъци на човека и органичната материя пред нея,

2001; Вселената Сметач, 2002; Следващото еволюционно стъпало I и II, 2002; „Писма между 18-годишния Тодор Арнаудов и философа Ангел Грънчаров“, 2002; Схващане за всеобщата предопределеност I,II,III,IV (Вселена и Разум 3, Вселена и Разум 4), Матрицата в матрицата е матрица в матрицата, 2003; „Творчеството е подражание на ниво алгоритми“, 2003 („Мисли за Емил 4.5.2003“, или „Мисли за изкуствената мисъл“); „Анализ на смисъла на изречение въз основа на базата знания на действаща мислеща машина. Мисли за смисъла и изкуствената мисъл.“, 2004 (...)

**Внимание:** Понякога архивът oocities.org не работи, тогава ползвайте archive.org, eim.twenkid.com/old и гуглите страници:

\* <https://twenkid.com/agi> - публикации от SIGI-2025 (и на адреса в гитхъб зоре)

\* <https://research.twenkid.com/agi/> - Материали за курсовете по УИР от 2010-2011 г., слайдове с лекции и гр.

\* <https://research.twenkid.com/> - кратка лична страница на автора (може да е необновявана)

\* **Vsy, Вседържец** – <https://github.com/Twenkid/Vsy-Jack-Of-All-Trades-AGI-Bulgarian-Internet-Archive-And-Search-Engine>

Част от трюковете се качват и в Academia.edu, може би и в Researchgate и гр. в бъдеще. \* <https://uni-plovdiv.academia.edu/TodorArnaudov>

\* [https://www.researchgate.net/profile/Todor\\_Arnaudov](https://www.researchgate.net/profile/Todor_Arnaudov)

**\* ... следват продължения**



## Някои съкращения

**СВП2** - Схващане за всеобщата предопределеност 2"/"Писма между 18-годишния Тодор Арnaudов и 43-годишния философ Ангел Грънчаров (...)", 2002 г... (Вселена и Разум 2)

ЧиММ – Човекът и Мислещата Машина: (...), 2001

ТРИВ, ТРВ – Теория на Разума и Вселената (всички творби, идеи от класическия период 2001-2004 и след това)

Т.А, ТА ... - Тош (вмъкнати бележки към конкретни)

## Личности

Питър Вос, Ирина Риш, Йоша Бах, Майкъл Левин, Карл Фристън, Стивън Волфрам, ...

@Вси: допълни

## Подробни записки и бележки по беседи ...

### **\* Основателни ли са доводите за песимизма в изкуствения интелект? Интервю с Питър Вос Peter Voss – Interview**

Science, Technology & the Future \* 22,2 хил. Абонати

1770 показвания 8.01.2013 г. [https://www.youtube.com/watch?v=4W\\_vtISjNk0](https://www.youtube.com/watch?v=4W_vtISjNk0)

**Подбор, превод, бележки:** Тодор Арnaudов – Тош (ТА, Т., ТА)

Питър Вос (Peter Voss) е основател и директор на Adaptive A.I. Inc., изследователска компания, разработваща система с ИИ на високо ниво (AGI). Той е основател и на компанията "Smart Action Company", която създава виртуални агенти за центрове за обслужване на клиенти (contact centers) – умни автоматични телефонни оператори.

Защо хората не вярват, че Общият ИИ е възможен? ... Трябваше да разбере, за да намеря начин да осигуря финансиране... Защо хората не подкрепят тази идея? Има дълъг списък от причини.

Единият е "зимата"<sup>1</sup> в ИИ: вложения от милиарди долари и неудобно слаби резултати и в хардуера, и в софтуера, т.е. *"опитали са се и са се провалили"*. В последните поне 20 години ИИ беше тесен и беше много успешен в тази си работа: разпознаване на реч, на лица, планиране, превод, игра на шах и пр. Цяло ново поколение на инженери на ИИ са обучени да приемат това за ИИ, за тях ИИ е тесният ИИ. Те дори не се замислят за оригиналната мечта, не смятат, че тя е осъществима. Преподавателите им биха казали: *„аз се опитах и се провалих, затова ти дори не се опитвай“*<sup>\*</sup>. Разговарям със завършили студенти – те дори не знаят какво е Общ ИИ. За тях ИИ е само тесният ИИ.

Затова няма много "юнаци", които вярват, че може да се създаде мислеща машина... Също има и влияние от психологията... Неудобни неща... "Политически некоректно" е да говориш за "малкото G" (General - Общ) – общите умствени способности, общата интелигентност. Мисля, че хората казват: *"няма такова нещо като обща интелигентност, има*

---

<sup>1</sup> Според мен понятието „Зима в ИИ“ е *изкуствено* и отразява само гледната точка на онези, за които ИИ е само източник на печалба и бизнес. При проучванията ми, може би към края на 2024, осъзнах че от гледна точка на изследвания и резултати напредъкът не е спирал, дори и когато най-съществените разработки, официално водещи се „ИИ“, компютърно зрение, планиране на движението в роботиката, разпознаване на реч и пр. се извършват от шепа хора в няколко основни университети и институти. В по-дълбок смисъл *цялата информатика и информационни технологии*, включително апаратните средства, електрониката и т.н. са *част от ИИ*. Сметачите са различна степен на ИИ дори и като „глупчовци“ (виж „Истината“, Т.А.2002). Бел. 15.6.2025

*само тясна*" (определен вид). Те грешат. Има смислено понятие за обща интелигентност. Така че, идеята за мислеща машина е "нападната" от много ъгли и трябва да се замислим какво се е променило: защо сега вярваме, че е възможно. Много неща се промениха, но трябва да притежавате и правилната теория. Трябва да разбирате какво е "интелигентност" и какво се опитвате да постигнете всъщност.\*

**Тодор:** Много учени и други „важни“ личности и учени всъщност *нищо не са разбирали* по въпроса за общите принципи на умствените способности или са имали *погрешна представа*, но са се опитвали *да наложат своето* липсващо „разбиране“, т.е. *своето мнение* – върху другия/другите. За това говори отношението на колеги и преподаватели към Демис Хасабис и Шейн Лег (но виж и друга причина по-долу, коментар на Марвин Мински), както и с признанието на директора на „OpenAI“ Сам Алтман за подигравките към тях при създаването на компанията. Същото явление може да видите в презрителните и унизителни поучения, които получавам в коментарите от 2005 г. от хора, които показват, че не разбират за какво говоря – а времето показва кой прав, кой крив.

Забележи също случая с прочутия учен от когнитивната наука Дъглас Хофстадер, автор на култовата за много хора книга „Гьодел, Ешер, Бах ...“, 1979, и изненадата му от напредъка в ИИ, пораждащите модели, от това че машините могат да творят музика.

### **Коментар на Тош от 11.7.2023 към видео:**

11.7.2023 г. 1:05:09 относно интервю с Дъглас Хофстатер.:

<https://www.youtube.com/watch?v=lfXxzAVtdpU>

*Gödel, Escher, Bach author Doug Hofstadter on the state of AI today*

Game Thinking TV \* 6,55 хил. абонати \* 29 976 показвания 30.06.2023 г.

„Дъглас Хофстадер, носител на наградата Пулицър за „Гьодел, Ешер, Бах“...“ размишлява за това как се умът и съзнанието са го заинтригували и е написала книгата си, и защо е ужасен от настоящото състояние на ИИ.“

### **@todprog (Тодор):**

Прогресът в ИИ беше очакван и предвиден от пионерите, например от мен, но не от него и много други, които изненадващо не са разбирали какво представляват умствените способности [въпреки, че уж са били водещи в познавателните науки и пр. – щом те не са разбирали, какво остава за неспециалистите]. (Между другото, теоремите на Гьодел за непълнота са безсмислици по отношение на УИР, виж „Теория на Разума и Вселената“ или предстоящата работа

„Вселена и Разум 6“, която ще публикувам на английски: *“Universe and Mind 6”* или слушайте Йоша Бах).

Фактът, че толкова дълбоки мислители (предполагаемо дълбокомислени) са смятали, че композирането на музика е вид „вълшебство“, че „машините не биха могли да го извършат“ [според Д.Х.: не и в обозримо бъдеще, а може би някога далеч], показва, че те всъщност не са разбирали как протичат творческите процеси и какво е творчеството по същество. (Виж предаването на MLST с гост Мелани Мичъл, докторантка на Д.Х., където разказват история за реакцията на Д.Х. когато се сблъсква с машина, композираща музика [в стила на Шопен, чието авторство е неразличимо от специалисти]. [Не са разбирали и че изкуството във формата, в която се възприема от „ценителя“ е просто данни, къс знание, свитък, информация, код, в крайна сметка „единици и нули“, „пиксели“ и пр.]

В творба от моята „Теория на Разума и Вселената“ (2001-2004) от 2002 г., обсъждах с философ, за когото напоследък разбрах, че всъщност е имал същите заблуди като Д.Х. И той мислеше и спореше, че машините не биха могли да композират музика или изкуство [с художествена стойност], защото „нямали душа“, изказваше и други безсмислици, докато аз обяснявах, че творчеството е **просто предвиждане и правдоподобни взаимозависимости, че машините ще открият същите хармонии и структури като хората**, и че музиката вече беше изчерпана, дори и тогава (тя беше и е, за онзи, който разбира достатъчно добре структурата ѝ), [че и изкуството като представяне може да се сведе „до единици и нули“] и пр.

**Коментар и в Дискорд канала на Ютюб канала MLST, discord:**

<https://discord.com/channels/937356144060530778/939939794736271452/1093634938559008968>

### **11:13: Критикуват ли ви от академичните среди?**

**Питър:** Нямам академична кариера и минавам "под радара", но колегите, които са в тези среди, са критикувани, защото [учението за универсалния изкуствен разум] не е добре прието. Правят се стъпки в посока подобряване на това положение. Организирант се годишни конференции по Общ ИИ и се публикуват научни статии и т.н. Движението започва да получава известно уважение. Мисля, че Бен (Гьорцел) свърши чудесна

работа за развитието на движението и за постигането на някакво уважение.

**12:13:** ... Друга причина защо е толкова трудно да се получи финансиране за Общ ИИ... Ако създадеш универсален изкуствен интелект на нивото на мишка, хората ще трябва да проверят по какъв начин си го постигнал, за да се впечатлят. ... Например ако създадеш Общ ИИ да играе шах, /те ще попитат/ на какво ниво /играе/? Защото ще го омаловажат: „*голяма работа, може да си купиш шах-автомат за 5 долара или да използваш безплатни програми, които ще победят твоята /“мислеща машина“/*”. Няма прости мерки /за Общ ИИ/. Защото тесните приложения на Универсалния изкуствен разум в отделни сфери могат лесно да бъдат победени от Тесния ИИ. Т.е. важно е не само какви умствени способности са постигнати, но и *как точно* – трудно е да убедиш хората в последното /че е „универсално“ може би/. Може би е нужно системата да достигне човешки коефициент на интелигентност от 70-80, много близък до човешки, относително общ, за да привлече вниманието на обществото, освен ако по-рано те не се увлекат от твоя проект и теорията зад него, което се постига трудно.

Също така богатите хора обикновено се допитват до експерти, а те не вярват, че универсалният изкуствен разум си заслужава като да се преследва като цел. /Сравни с днес, 2023 г./

Онези, които са мислили по въпроса и са стигнали до извода, че е възможно да създадем мислещи машини сега: на лице е съчетание от сметачи, които са много по-мощни от предните периоди; много по-развито програмно осигуряване; и точният модел.

От друга страна има и аргументи от друг вид от една друга група хора - дори Рей Кърцуейл е един от тях. Те твърдят: "не очаквайте УИР в следващите 5 или 10 години, защото според моите графики и криви са нужни още 20 години преди да имаме изчислителна мощ, за да постигнем човешко ниво на умствени способности". Според мен аргументът им е слаб, защото като да си мислим: "ще можем да летим тогава, когато имаме достатъчно познания по механика, за да можем да разгадаем летенето на птиците и да ги пресъздадем" – а всъщност летим от 100 години, без да е било необходимо<sup>2</sup>. По същия начин ние не се опитваме пряко да разнищим в обратна посока работата на мозъка, а да разберем какво представлява мисленето, разсъждението на високо ниво,

---

<sup>2</sup> Виж роботът, който лети като птица. Видео от Тед, 23.7.2011 г.: [https://www.youtube.com/watch?v=Fg\\_JcKSHUtQ](https://www.youtube.com/watch?v=Fg_JcKSHUtQ)

абстрактното мислене, решаване на задачи и да го изпълним чрез технологиите, които имаме - в цифрови компютри.

**36:35:** Като се замисля за времето, което съм отделил в последните 10 години – *откогато действително работя по Общ ИИ* /т.е. от 2003 г./ – за да правя демонстрация, с които да набирам средства и т.н., и как се разпиляват усилията ми, и как бих искал да не трябва да се занимавам с тези ужасни занимания<sup>3</sup>. В същото време някой с дълбоки джобове, който лично се интересува от темата ще "захване" тази област, стига да я открие, да стане визионер и да повярва в нея. Обаче е необходимо и да има правилната идея, защото и преди е имало хора с относително "дълбоки джобове", които са опитали нещо, но се е оказало, че не са имали правилните идеи. ...

Мисля, че броят на личностите, които вероятно биха се изказали уверено, че времето вече е дошло, че имат работещи идеи и проект как да създадат мислещи машини, е *около половин дузина*. Повечето изследователи в Общия ИИ обаче по-скоро биха били по-скромни: те биха казали, че се интересуват от тази област, но не биха знаели как да се осъществи. Сред учените има и друга група, които наистина вярват, че имат достатъчно способности и идеи, за да доведат проекта до край. Не се знае кои от тези групи наистина има добри идеи и ще успее, или който и да е от тях, защото ако се окаже, че е въпрос по-скоро на "груба сила", и вероятно победителите ще са Гугъл, Майкрософт, Епъл и пр. друга голяма компания, която вложи достатъчно средства.

Например Питър Норвиг, директор на изследванията в Гугъл не вярва, че е възможно да се създаде изкуствен разум *сега*, и при това положение е ясно, че Гугъл не биха полагали усилия в тази посока /през 2013 г./. Той смята, че първо трябва да се открият подходящите алгоритми, после изчислителна мощ, така че той не би се занимавал с УИР, макар че е възможно друга малка група в компанията да се занимава. /Т.е. и той подобно на Дъглас Хофстатер<sup>4</sup> не е разбирал./

Не мисля, че е необходимо да се открият нови фундаментални техники. Мисля че въпросът е по-скоро да съчетаем съществуващото знание по умен начин, да го подобрим, усъвършенстваме и т.н. Ще има

---

<sup>3</sup> И аз имам подобни мисли, свързани със загубеното от мен време в безплодни усилия да намеря партньори, да се опитам да ме разберат хора, които нямат необходимите умствени способности; загубеното време в много напразни и разсейващи неща, донякъде включително и писането и редактирането на тази огромна книга. Бел. 12.7.2023

<sup>4</sup> Виж в основния том; „Гюдел, Ешер, Бах...“, когнитивна наука; аналогия; Шопен; могат ли машините да композират музика; мнението и на Ангел Грънчаров в „Писма между 18-годишния Тодор Арнаудов и философа ...“, 2002

пробиви в отделни дизайни [архитектури] и алгоритми, но не и в голям мащаб. Мисля, че няма нужда да се разработват "магически куршуми".

**Бел. 26.9.2025:** <https://petervoss.ai/> „Питър Вос е първопроходник в ИИ, участник в изковаването на термина УИР (AGI) през 2001 г. заедно с негови другари пророци.“ [Лег, Гьорцел] ... Настоящата компания на Вос е Aigo: <https://aigo.ai/> – *“Pioneering Cognitive AI ... Building Cognitive AI to Unlock Real Intelligence (not scaling LLMs). ... Rooted in decades of commercial and development experience. ... A Paradigm-Shift is coming soon...”*

– Строят „познавателен ИИ“ (вдъхновен от човешкия ум?), който да отключи истинския разум, а не уголемяване на езикови модели. Бил основан на десетилетия опит в разработката и продажбите. Скоро щяла да дойде нова парадигма.

От X (twitter) на съоснователя на компанията Srini Pagidyal, 16.7.2025: [https://x.com/Srini\\_Pa/status/1946209012648903026](https://x.com/Srini_Pa/status/1946209012648903026)

Диаграма:

**Правилният начин за създаване на УИР – след езиковите модели:**

1. Формулирай теория на разума
2. Определи УИР
3. Проектирай познавателна архитектура
4. Разработи предварителна УИР машина
5. Докажи работоспособността ѝ чрез внедряване като продукт
6. Отгледай УИР като дете до преди пуберитета
7. Уголеми го до УИР на ниво завършил студент

## Ирина Риш – Универсален изкуствен разум. Мащабиране. Съгласуване

### The Inside View

1,95 хил. абонати

3101 показвания [10.3.2023<sup>5</sup>] Начало на премиерата: 18.10.2022 г.

При MichaelTrazzi: <https://youtu.be/ZwvJn4x714s>

**Подбор, превод, бележки:** Тодор Арнаудов – Тош (ТА, Т., ТА):

**Ирина Риш** е професор в Университета на Монреал, Канада, основен член на института „Мила“ и организатор на *„семинара за закони за невронно мащабиране към максимално полезен Универсален изкуствен разум“*. Host: <https://twitter.com/MichaelTrazzi>

Irina: <https://twitter.com/irinarish>

### Съдържание

00:00 Акценти

00:30 Въведение

01:23 Определение на УИР (AGI)

04:15 AGI/УИР означава увеличен човешки интелект (“augmented”)

06:40 Решаване на проблема със съгласуването на целите на ИИ с човешките чрез родителски наставления.

09:23 От ранните дни на обучението на многослойни невронни мрежи (deep learning) до универсалните агенти.

13:47 Как Ирина обновява определенията си след излизането на общия мултимодален модел „Gato“.

17:56 Изграждане на истински Общ ИИ в рамките на живота на Ирина.

19:58 Най-малко впечатляващото нещо, което няма да се случи след пет години.

22:56 Мащабиране отвъд степенните закони.

24:20 Семинар по законите за невронно мащабиране.

30:42 Защо Ирина не иска да забави напредъка на ИИ?

49:27 Фазови преходи и грокинг

58:01 Увеличаването на размерите на невронните мрежи разрешава ли проблема за непрекъснатото обучение? [сегашните невронни мрежи

---

<sup>5</sup> 3797 гледания към 26.6.2023 – относително непопулярен канал за ИИ, МО



“забравят“ старите неща, т.нар. „катастрофално забравяне“, не могат да задържат едновременно модел на различни множества от данни]

01:06:45 Вероятността от „риск за оцеляването на човечеството“ заради създаването на УИР според Ирина

01:03:28 Работа по съгласуването на ИИ в института „Мила“

01:15:43 Откъде „Мила“ ще получи изчислителна мощ?

01:22:39 С „голямата изчислителна мощ идва и голямата отговорност“

## „Мила“

– *"Лабораторията ви върви по пътя на построяване на Общ ИИ, заедно с други неща ... Какви са те?"*

**Ирина:** Как да построим Общ ИИ, и ... Ние го наричаме "Самостоятелен ИИ", но ефективно целта е да построим Универсален ИР/Общ ИИ.

Какви са другите неща? Зависи как определяш Общ ИИ. Любимата ми картинка в моите презентации е анекдота за седемте незрящи и слона, които спорят какво е "слон" и всички са прави, защото го докосват на различни места.

Що се отнася до мен, опитвам се да бъда максимално прецизна, буквална, може би скучна и казвам "Общ ИИ/AGI означава общ/всеобщ ИИ", където "общ" (general) означава способен да решава множество от задачи или може би способен да се обучава непрекъснато, да изпълнява тези задачи една след друга; с възможности да научи безкранен брой от тях и да натрупва цялото това знание. Така че "общ" значи всеобщ, многозадачен. Или може да ползваш нещо като определението на OpenAI ... самостоятелна система, способна да изпълнява, да речем на човешко или свръхчовешко равнище повечето от икономически достъпните задачи в икономиката. Това ограничава проблема до определяне на кои са всички икономически задачи, ... но система със значително широк обхват от многозадачност. Също така качеството на изпълнение, способностите във всяка задача трябва да е достатъчно високо, каквото и да се разбира под "достатъчно". Не е особено продуктивно да се спори за много прецизни определения. Мисля, че вероятно е по-продуктивно да се опитваме да измислим как да разработим системи, които да бъдат такива "специалисти по всичко" (generalists).

Така стигаме до няколко области, от които всъщност се интересувам, като обобщение на модели извън вероятностното разпределение, върху което са обучени (out of distribution generalization), защото искате да обучите системата върху определен набор от данни и задачи и след това той да работи и в нови условия.

Също така искате да имате системи, които са устойчиви на промени, било на вероятностните разпределения или пък в случаи на злонамерени „нападения“ (adversarial [attacks]). Всичко това е необходимо, за да се постигне въпросното многозадачно обобщение.

Също така трябва да се съсредоточите върху непрекъснато обучение *"през целия живот"*, защото няма да разполагате с всичките необходими данни по едно и също време така както се обучават съвременните големи системи *[първо се обучават върху цял набор от данни, който е "замразен", после правят извод върху тези данни; ако се добавят нови - трябва да се дообучават или да се прави допълнителна "фина настройка"]*.

Всичко това е необходимо за да имате ИИ със способности по всичко. Защо искате такъв ИИ? Защото ще бъде много полезна технология. Може да се сетите за разнообразни приложения, от медицина до автоматизиране на всякакви скучни неща и създаване на инструменти не само за класическите професии, но и за художници и музиканти. По принцип, не да ги замени, да създаде оръдия, с които да им помогне да бъдат по-изобретателни и съзидателни и да се изразяват по-добре, дори и да нямат образование по определено изкуство. *[Това вече е реалност, а беше и без днешните пораждащи модели.]*

Подобрява човешките способности, за това искате да постигнете ИИ като инструментална цел. За мен "И"-то в "ИИ" не е само "изкуствен", а и "Усилен" (Artificial - Augmented) - "УИ".

**Водещ:** Не мислите ли, че ако имаме по-способни художници и всеки може да напише подсказка в "Стейбъл Дифюжън/StableDiffusion" и да създаде изкуство на най-високо ниво, тогава човеците художници скоро ще загубят работата си?<sup>6</sup>

---

<sup>6</sup> Голяма част от хората с художествени и със всякакви таланти в изкуствата отдавна са „загубили работата си“ – не могат да печелят пари от дарбата си, защото няма място „на пазара“ за всички тях; или не могат да се издържат от него. Това обаче не прави впечатление на „анализаторите“, защото те повтарят като папагали „авторитетни“ изказвания от настоящия „модерен поток“ и приемат, че за онези, които в *сегашните*

**Ирина:** знам че има обща загриженост у някои художници, но не виждам нещата задължително по този начин. Според мен всеки инструмент дава на хората повече възможности, разширява хоризонта им. Този инструмент ще направи художниците дори още по-съзидателни и оригинални, но на друго ниво. По-ранните изобретения вероятно са премахнали определени работни места и дейности, но са създали други на различни нива. С новите инструменти за пораждање на изкуство, за пораждање на съдържание, изкуството ще премине в това как да създадете подсказки за пораждащите модели, което не е напълно тривиално, и това ще бъде изкуство само за себе си. „Адоуб Фотошоп“ не изхвърли напълно фотографите от работата. Все още има много работа за свършване по симбиотичните отношения между хората и ИИ. Затова Универсалният изкуствен разум не е цел сама за себе си. За мен повече е постигане на напредък в човешките способности, а не замяната на хората.

**Водещ:** Мислите ли, че в даден момент ще напреднем към способности отвъд човешкото ниво и няма ли тогава [тази сила] да излезе извън контрол и трудно ценностите ѝ да са в съответствие с човешките<sup>7</sup>?

**Ирина:** Може би съм прекалено оптимистична, но мисля, че е възможно да разработим ИИ, и докато го правим, да поддържаме съгласуването му. Подходящ пример би бил, когато си взаимодействате с някой член на вашето общество, било човек или друг вид деятел/агент. (...)

10:20: Достатъчно възрастна съм, за да си спомням времената, когато обучението на многослойни невронни мрежи не беше широко разпространено и популярно, а даже беше силно критикувано. Спомням си 2006 или 2007 г., когато Джефри Хинтън и колеги предложиха да се организира работилница на конференцията NeurIPS и предложението им беше отхвърлено, но въпреки това те организираха работилницата като симпозиум в последния ден на събитието. Беше почти "религиозна" сбирка.

---

условия са „неспособни да се впишат в пазара“, това положение е „в реда на нещата“, но ако има промяна, която ще засегне и други, които сега са „привилегирвани“: това е недопустимо. Също така до сега онези, които са успели да си послужат с компютърните технологии в изкуствата и пр. са се „вредили“, докато други, по-стари, които не могат – са останали „извън играта“. Първите обаче не поднасят цветя на инженерите и програмистите, които са им дали технологично предимство.

<sup>7</sup> „aligned“

11:30 [Масовото] движение в изследванията на многослойни невронни мрежи, "Deep Learning", всъщност възникна през 2012 г. и със системата AlexNet, която постигна 10% напредък в точността на класификацията в ImageNet – от тогава хората обърнаха внимание на тази област. Преди това те не бяха разглеждани сериозно, съответно и личностите, които твърдяха, че „ще завладеят света“, не бяха приемани насериозно.

**Водещ: 11:53** [по-късно, в последните години] Хората започнаха да говорят повече за Общ ИИ. Получихме статията за законите за мащабиране на невронните мрежи и GPT-3 и показахме ясен път към универсален изкуствен разум. (...)

**Ирина: 12:04:** Хората винаги твърдяха, включително и аз, че имаме страхотен Тесен ИИ, и всичко беше вярно до много скоро\*. Защото наистина получихме огромен напредък в способностите за всестранна дейност. Изследователите се опитваха да правят всякакви "готини" алгоритми за непрекъснато обучение, мета обучение и пр., за да получат обобщение извън вероятностното разпределение при обучение. Имаше купчина от статии относно инвариантно намаляване на риска (invariant risk minimization). Всеки се опитваше да "излезе от разпределението" и да построи своя многоцелеви агент. И изведнъж, чрез мащабиране, получихме GPT-3, CLIP [*модел за подобие между текст и изображения, обучен върху милиони и милиарди двойки текст-образ*] вече обобщаваше при нулево-извън-разпределение\* по-добре дори и от най-добрите специално обучени за тази модели върху определени набори данни.

\*zero-shot out of distribution - без обучение върху примери?

**[Моделите] стават все по-общи ...**

13:16 И изведнъж увлекателната "приказка"\* за това че ИИ е тесен, а не широк започна да се разпада. Трябваше да променя слайдовете в презентациите си. Защото винаги, преди 2020 г., трябваше да говоря за "Продължаващ ИИ" (Continual AI). И напълно имаше смисъл, например ... [съществуваха различни системи за различни задачи]: система, която играе Го, система играеща шах, система пораждаща картини, продадени на търг в Ню Йорк; обаче нямаше системи с всеобщи способности. И тогава си казах: "Добре, не мога да го казвам повече. Сега съществува „ГАТО“. Наистина трябва да си сменя слайдовете..."

## Бележки

\* spiel - ам.англ., увлекателна история, разказ

\* Сравни: Лекцията **"Обзор на класическия изкуствен интелект и на тесния ИИ – защо е ограничен и се провали"**:

[https://research.twinkl.com/agi/2010/Narrow\\_AI\\_Review\\_Why\\_Failed\\_MTR.pdf](https://research.twinkl.com/agi/2010/Narrow_AI_Review_Why_Failed_MTR.pdf) от първия в света курс по Общ ИИ, Пловдив, Т.Арnaudов, 2010 г., представена/публикувана през април 2010 г. С развитието на лекцията, методите от "тесния" ИИ, за които говоря, всъщност постепенно преливат към общия, вкл. Йерархичната времева памет - НТМ - на Джеф Хокинс, заедно с рекурентните невронни мрежи и LSTM - създадени от учения от Общия ИИ Юрген Шмидхубер и после част от пробивите в разпознаването на йероглифи и автоматичния превод и т.н. - и така обяснявам, че още в средата и края на 2000-те години методите на тесния ИИ стават все по-общи и все повече по идеи и принципи се приближават до идеите на Общия ИИ, които представях като теории на разума и моята *"Теория на Разума и Вселената"*. Тогава все още не беше започнала видимата част от "революцията" с многослойните невронни мрежи, а аз самият също по принцип не бях почитател на невронните мрежи до тогава по технически причини – от една страна бяха изчислително недостъпни за експерименти, освен това умът ми и моята теория могат да се класифицират по по-широк начин, спрямо който невронните мрежи тогава бяха тесни. Моите идеи са многопосочни, хибридни, ако ги обобща сега накратко, може би: Възникваща, Зародиш на разум, Познавателен ИИ, Сетивно-моторно обучение и развитие от начално ниво чрез взаимодействие със средата; Причинностен и причинно-следствени модели на предвиждане на бъдещето; сродни идеи с принципи на Джеф Хокинс за предвиждане, основано на памет; на Карл Фриш: действено заключение, принцип за свободна енергия; обучение с подкрепление, основано на предсказващи модели на средата; теорията за "овластяването", и също „Език на мисълта“, „Понятийни“ (някои биха ги нарекли „символни“, „symbolic“, но според мен разделението символни и „конекционистки“ (Symbolic & Connectionist), или „символни и под-символни“ (subsymbolic) са заблуждаващи и объркани. Виж статията на Т.А. *„Невронните мрежи също са символни“*, 2019. и пр. (Emergentist,

Seed AI, Cognitive AI/Sensory-Motor Grounding, Memory-Prediction Framework/Hierarchical Temporal Memory; Free Energy Principle/Active Inference, Predictive Coding; Empowerment; Model-Based Reinforcement Learning (~Self-Supervised, Immitation Learning, ...), Language of Thought, Conception thinking...).

Представата ми клонеше към градивни елементи на ума, които са по-скоро като причиняващо-управляващи устройства, "микроагенти", изграждащи системата отдолу нагоре, а не този тип абстрактни неврони (матрици с тегла), които се наложиха в успешните невронни мрежи.

Методите и насоките на развитие в другите методи, като системата за управление на самоуправляващата се кола на Станфорд през 2007 г. също клоняха като философия към онези на Общия ИИ, които признавах и за които говорех в курса:

**Слайд №44:**

[https://research.twenkid.com/agi/2010/Narrow\\_AI\\_Review\\_Why\\_Failed\\_MTR.pdf](https://research.twenkid.com/agi/2010/Narrow_AI_Review_Why_Failed_MTR.pdf)

**DARPA Grand & Urban Challenge**

- Коли, които се карат сами.

Победителят в Urban Challenge през 2007 използва йерархична управляваща система с многослойно планиране на задачите и движенията, генератор на поведение, възприятия, моделиране на света и мехатроника.

**Слайд за обработката на естествен език:**

Какво не е наред в обработката на естествен език [и слабия ИИ]?

- Системите са статични, изискват много ръчна намеса, не се развиват и мащабират.
- Специализирани „трикове“, а не универсални системи.
- Работят на високо символно ниво и нямат основа от първични възприятия и взаимодействия със средата.
- Невронните мрежи нямат цялостна архитектура, не се самоорганизируют и са хаотични и тежки.

Липсва добра „физика“, която да позволи създаването на „двигател“, който да се включи и да заработи от самосебе си.

Инструменти, а не двигатели.

- <http://artificial-mind.blogspot.com/2009/03/whats-wrong-with-natural-language.html>

...

Признанието на Ирина Риш обаче показва, че учените в академичните среди **не са разбирали тази тенденция** и са разбрали едва съвсем скоро, т.е. и тук **Свещеният сметач и курсът по ИР** ги е изпреварил с повече от десетилетие. Също явно и тя, и колегите и са били изненадани от неочаквания прогрес, а не само „необразованите“ и външни наблюдатели.

Виж също интервюто с Т.А. от 2009 г.: *"Ще създам мислеща машина, която ще се самоусложнява: Фантазьори и авантюристи създават великите открития. Работата на*

*скептиците е да отричат, а после да гледат и да не вярват на очите си."* и предвиждането, че до 10-20 години ще има ранни версии на моделите.

\*("Под провали" се има предвид защо не беше станал Общ до тогава.)

## Интервю с Ирина Риш, продължение

**Ирина: 14:27** (...) Всъщност има наистина умопомрачителни резултати от последната година около трансформаторите<sup>8</sup>: те са универсални изчислителни системи, [има и] трансформаторите за взимане на решения. Т.е. архитектура на трансформатор с достатъчно общо предназначение извлича представяне на някакъв вид понятия например от текст. И в голяма степен те са приложими в други модалности, като изображения. Може би само трябва да замениш входните и изходните данни, разделянето на кодови думи ("токенизация") и т.н. Могат да се прилагат и в области като последователно вземане на решения. Нивото на пренасяне [между различни модалности и задачи] се подобри с порядък [един порядък = десет пъти]. Хората обикновено говорят за пренасяне или продължаващо машинно обучение от един набор изображения в друг. А сега преносът и обобщенията се случват на много по-общо ниво. Затова се говори повече за Общ ИИ, има определен напредък в обобщаването.(...)

17:09: Ако успееш да постигнеш процес, който построява общ агент като начално установяване, е много по-лесно след това да получиш специализирани агенти, ако са ти нужни за определени приложения. [Сравни и виж „ЧиММ“, Т.А., 2001]

... И това е нормално, защото и човекът е "специалист по всичко", но ако искаш някой да превърнеш някого в шампион в определен спорт, се налага да го "подложиш" и на "фина настройка" ["fine tuning" в машинното обучение].

19:20 Ако развитието продължи със същата тенденцията, очаквам Общ ИИ до 20 години.

**Водещ:** Но ако е експоненциално – ще бъде по-скоро. ...

**Ирина:** Но мисля, че до 5 години няма да е достатъчно надежден и

---

<sup>8</sup> Attention Is All You Need, Ashish Vaswani et al., 2017, <https://arxiv.org/abs/1706.03762>



устойчив (robust) на различни промени във вероятностното разпределение и в средата. (...)

Част от проблема с модерните многослойни мрежи е, че те са толкова големи и сложни

22:02 и те ще стават все по-големи и по-сложни, така че ние сме отвъд точката, в която можем да разбираме поведението им аналитично. Всеки знае за това. Ето защо започнахме да разглеждаме законите за мащабиране, [да търсим] емпирични методи за изследване на поведението им. И цялото това „нещо“ с изучаването на поведението, освен представянето или загубата [“loss function”: при обучението, разликата между предвиденото и очакваното] по същество това е **съгласуването на ИИ с нашите цели (AI alignment)**. Опитваме се да разберем как различните промени в процедурата за обучение, в данните, във взаимодействието със системата, произвеждат различни видове поведение.

**Бележка, 26.6.2023:** Сравни с началото на повестта „Истината“, Т.А., 12.2002, диалогът в който Божидар променя част от кода на Емил и след това мислещата машина роптае, а Създателят обяснява на Машината защо не е съобщил на „Съвета“ за съществуването ѝ. Този „Съвет“, споменат мимоходом в творбата, би съответствал на сега съществуващите организации за „Съгласуване на ИИ“ или т.нар. „AI Alignment“ като “MIRI” на Юдковски, която вече съществуваше по време на „Истината“, но тогава ИИ беше фантастика, или напр. „Eleuther“. Известен представител на последната е Конър Лийхи. Виж в канала на Тош и др. за „преоткриването на колелото“ от негова страна относно определенията на изкуствен разум.

**22:38:** Когато изучавате биологични или физически системи, съгласуването на целите се прави до голяма степен емпирично. Фактът, че сте създали тази система, вече няма значение [това не ви помага]. Нивото на сложност ще е толкова голямо, че вече няма да има значение дали тя е създадена по естествен или по изкуствен път<sup>9</sup>.

---

<sup>9</sup> Самото разделение според мен е... „изкуствено“. Всички „изкуствени“ неща всъщност са „естествени“ по отношение на сътворението. Човекът е част от природните процеси, както и машините, които той си самоприписва като „свои“ октирия. Виж още „ЧиММ“, 2001 и бележката след интервюто с Бен Гьорцел в началото на работата.

Разликата с биологичните хора е, че биологията е много трудна за разбиране, нали? Сътворено е от еволюцията и е "разпиляно". Има толкова много различни неврони и е трудно да се разбере всичко, което се случва.

**Водещ:** 23:10 Обаче архитектурата на трансформаторите е лесна, така че изглежда възможно да се съгласува, нали? [с целите на „човека“, „човечеството“; да се постигне „AI alignment“; „човечеството“ е заблуждаващо понятие: виж книгата на Т.А.: „Заблуждаващите понятия и разбор на истинския им смисъл: Трансхуманизъм, Цивилизация...“, 2020]

**Ирина:** 23:15: Всъщност не съм толкова сигурна, защото дори простата архитектура, на първо място, ... може да разгледаш възможностите за тълкуване и в частност промени в кои неврони се свързват с кои поведения, но е отвъд точката, при която може да имате механистичен модел на поведението на мрежата. Затова все повече прилича на биологична. Пример за такова явление е „Възникването“ на нови поведения (Emergeny) при увеличаването на мащаба.

*[Виж също Диалектически материализъм: "Количествените натрупвания водят до качествени промени.", Т.А. 10.3.2023]*

Излязоха различни статии, например [Emergent Abilities of Large Language Models (<https://arxiv.org/abs/2206.07682>)]. Може би не е безспорно правилно да го наричаме „възникващ“.

24:06: Това са по-скоро *неочаквани* промени в производителността или други свойства, като „истинност“ например. И така има неочаквани промени, които може дори да не следват нищо, което е предвидимо по прост начин като известните степенни закони\*. Има различни задачи, изпълнявани от невронни мрежи, при които дори производителността, без да говорим за устойчивост, истинност и други свойства, започва да се променя драстично и "преживява" резки преходи.

Затова учените изучават по-общи методи за моделиране, поведението при мащабиране отвъд степенните закони. (...)

...

**[Бел. Т.А.:** Степенните или експоненциални закони са за увеличение на точността на моделите при геометрично нарастване на размера на мрежата, броя неврони, сложността. Точността отразява доколко предвижданията на модела, пресъздадените образи, звуци, текстове

съвпадат с вероятностите изчислени от набора данни върху който е обучаван, с „действителни“.]

**[Бел. Т.А.: ... вметка за "горчивия урок", бележка на Ричард Сътън, един от "бащите" на машинното обучение с подкрепление, че прости методи с добавяне на повече изчислителни ресурси постигат по-добри резултати, отколкото сложни ръчно изпитани методи. ... Методи, които се самообучават, сами откриват особеностите в данните и пр.**

30 мин: ... [законите за невронно мащабиране] пред-обучение върху данните от целия "Youtube" за да построим изкуствен разум. Идея на Етан Кабалеро, докторант в "МИЛА"\* . ... Ако можеш да обучиш достатъчно голяма система с достатъчно изчислителни средства върху толкова богат и грамаден източник на данни, ще получиш и Общ ИИ, защото той ще знае всичко\*.

\* "Мила" (Mila, MILA) - институт по ИИ в Канада. Сравни с **героинята Мила** от повестта "Истината", Т.А., 2002, в компанията, която разработва мислеща машина; част от „Българските пророчества“ или „Теория на Разума и Вселената“.

\* Сравни обаче с „ЧиММ“, Т.А. 2001 и с интервюто от 2009 „Ще създам ММ, която ще се самоусложнява...“, че ИР не е „всезнайство“, а развитие. Не е необходимо ИИ да знае всичко, за да бъде универсална мислеща машина.

... Съгласуването с хората трябва да стане чрез взаимодействие с други ботове и с хората, например в чатове и т.н., с обратна връзка, така ще е по-естествено и по-лесно, отколкото да се събират огромни набори от данни.

Децата са "предобучени" не само от родителите и от другарчетата си, но и от еволюцията и биологията си така че да оцеляват. В един момент спазването на моралните изисквания става необходимо за оцеляването им, когато започнат да живеят в обществото. Може да приложим тези идеи и върху мислещите машини: да бъде полезно за тях да имат нравите, които изискваме от тях.

Вижте пример на Майкъл Левин с раковите клетки, които "забравят", че са част от организъм и започват да се отнасят към него като част от средата, като храна, но не "съзнават", че като го унищожат, те също ще загинат.

--

49:27 - 50 мин: Фазови преходи и "Грокване", grokking - явление при мащабиране на мрежите и обучението, при обучение на големи модели... Аритметични или други дискретни операции с много аргументи... Функцията на загуба, изведнъж стига до непредвидими резки промени в непредвидими точки, с големи подобрения в способностите.

**Бележка: Т.Арнаудов.:** Мисля, че това показва и подсказва природата на *скрита дискретна структура* - скрита от гледна точка на намаляване на измеренията, свеждане на целия процес на обучение, само до един или малко параметри като въпросната загуба (loss). При обхождането на тази скрита, вътрешна, структура в даден момент се "прескача" определен локален минимуми, и/или пък във виртуалната дискретна функция, която се получава преход поради "напасване", превключване, "нацелване" на "зъбец" на "ключалка", "щракване" като схема на тригер, който сменя състоянието си при сигнал на входа - като се има предвид и че трансформаторите в крайна сметка работят на входа и изхода с дискретни кодови думи ("токени") и техният резултат се сглобява от тях. За да се стигне до тези ключове е необходимо да се построи определена дълбочина на структурата или да се навърже достатъчно дълга верига от преходи във въображаема равнозначна/равносилна/съответстваща на действителните данни структура, която все едно се обхожда във "въображението" при извикване на предвиждащия поток на невронната мрежа.

Явлението може също да показва, "чупливостта" на модела и че текущото му състояние и способности силно зависят от конкретни вариации в обхождането на набора данни, точно кой след кой конкретен пример от данните се обработва, колко пъти се среща, какви какви конкретни примери има за различни операции. Например може да има повече събирания, отколкото деления, или повече сметки от даден тип, и т.н., а други да са редки или да не се срещат. Както например при обучение на дийпфейк модели, например в "Deerfacelab" или българската версия "Arnoldifier". Наборът от данни се състои от лица, които трябва да бъдат в различни положения на главата, различни ъгли на завъртане ("Yaw"). Често ъглите в профил са по-малко, отколкото

фронталните, поради естеството на повечето снимки и видеоклипове. Ако наборът данни от лица не е равномерно разпределен, по-редките профилни лица се "виждат"/обработват по-рядко от модела при обучението и за да се избистрят са нужни много повече итерации, отколкото за фронталните снимки. Да, тук има разлика - преходите са „плавни“, пространството е изобразително и по-гладко; замъгленият образ постепенно се изостря, но в дискретното пространство на трансформаторите и по-абстрактните представяния, които той изгражда, своеобразната "замъгленост" на предвиждащия модел може да изглежда именно като такива "непредвидими" плавания, "флуктуации" в крайния резултат, понеже той зависи от по-"случайни" фактори с по-рязка и скокообразна промяна в произведения краен код. Самото плаване и непредвидимост са признак за случаен процес, нещо което не се съгласува с останалото.

\* **"Грок"** се казваше и модел на Numenta, публикуван през 2014 г. за засичане на аномалии в потоци от данни за облачни сървъри на Амазон: <https://www.numenta.com/press/2014/03/25/numenta-releases-grok-for-it-analytics-on-aws/> Една от функциите, за които съобщават, е именно "продължаващо обучение" (виж по-долу) и приспособяване

\*\* Думата „грокване“ произлиза от неологизъм на Робърт Хайлайн от романа „Странник в странна страна“ за вид „дълбоко интуитивно, едновременно умствено и емоционално разбиране“:  
<https://www.universityxp.com/blog/2021/4/6/what-is-grokking>

\*\*\* Според Нийл Нанда\* „грокването“ е илюзия, фазов преход, от състояние в което моделът не може да обобщава, внезапно в такова, в което започва да обобщава; но по-скоро представлява „запомняне наизуст“ отколкото обобщаване (memorize rather than generalize). (Показват пример със задачи за целочислено деление, дадени на голям езиков модел.)

От „Дийпмайд“ в беседа публикувана в MLST\* , “Mechanistic Interpretability - NEEL NANDA (DeepMind)”, 28.10.2023,  
<https://www.youtube.com/watch?v=Ygf0GnlwmY> след. 7:10 мин.

\* **Бел. 10.8.2025:** „Grok” е и езиковият модел на компанията „xAI“ с отворени тегла. <https://github.com/xai-org/grok-1>

<https://github.com/marketplace/models/azureml-xai/grok-3>

<https://github.com/marketplace/models/azureml-xai/grok-3/playground>

(...)

**54 мин: Ирина** ... Увеличение на надеждността с нарастване на размера на мрежата, разко подобрене спрямо зловредни атаки (adversarial attacks). Друг вид - увеличение на обема на данните. (...) Онова което се случва във всеки вид преход може да бъде коренно различно. Много е вълнуваща, но и все още много неизследвана област.

**55 мин:** Нелинейно мащабиране - различно ли е от резките преходи и от грокването?

Как да се измисли функционална форма\*, която да отразява при какви условия, параметри на модела, обем на данните, количество изчисления/итерации и пр. процесът на машинно обучение ще доведе до дадени желани нелинейни резки промени, напр. ще се увеличи точността му при решение на аритметични задачи, точността на класификация и всевъзможни мерки.

**Бележка:** \* Функционална форма - най-просто казано дали има линейна, логаритмична, степенна и пр. по-сложно изразени зависимости между променливите. По-подробно: математически модел с изрично множество от уравнения, които описват отношенията между променливите в модела - не само кои променливи, но и точните взаимоотношения между зависими и независими променливи, например като във физичните закони:  $V = S/T$  - средната скорост е правопрпорционална на пътя и обратнопрпорционална на времето за което е изминат и т.н. Виж: <https://www.encyclopedia.com/social-sciences/applied-and-social-sciences-magazines/functional-form>

\* Непрекъснато обучение (Continual learning) - преодоляване на проблема на невронните мрежи с "катастрофалното забравяне" (catastrophic forgetting).

**58:49: - Какво е *продължаващо, непрекъснато обучение* и защо е важно? (...)**

**Ирина:** "Свещеният граал", откриване на агент с всеобщи способности, "специалист по всичко", който се справя с множество от задачи.

Единственото предизвикателство е, че обучението се извършва във времето. За разлика от обучението на GPT, CLIP<sup>10</sup>, GATO<sup>11</sup> и големите модели, наборът данни не е събран предварително и не можете да взимате равномерни "проби" от всички различни задачи, да смесите всичко във всеки етап на пакетната обработка (batch) [както се прави обикновено с невронните мрежи] и да го обучавате [по този начин]. Системи с продължаващо обучение не виждат всички тези вероятностни разпределения на данните наведнъж. Задачата може да се обобщи като последователно непосредствено\* обучение с нестационарни данни. Наборът от данни или задачите в наборите данни продължават да се променят.

Представете си, че обучавате GATO, но върху различни игри и различни езици или задачи за разпознаване на изображения, но данните идват последователно. По-трудно е. И дали моделът знае предварително, че задачите са различни? Или трябва да има произволен набор от задачи, които трябва да научи с една и съща архитектура? Продължаващото обучение също все още не е добре формализирано като поле на изследвания. Все още се развива. Има множество от сценарии. Най-простите са онези, при които знаете, че данните и задачата се променят и съобщавате на системата за това, така че тогава се работи по класическия начин. Или може да бъде по-трудно: да не бъде известно какви са задачите (task-agnostic) като при DataArc. Не "подказваме" на системата, че задачата се променя и тя трябва сама да разбере и да се приспособи. Също така има задачи с постепенна промяна на задачата, на класа, с неизвестна задача и други установки на условията. Отворен въпрос е до каква степен, да речем, ако предобучим много големи системи, ще можем по този начин да разрешим и проблема с продължаващото обучение по някакъв най-тривиален асимптотичен

---

<sup>10</sup> Невронен модел, представяне (embedding), което кодира връзката между изображения и техни текстови описания за мултимодални модели текст-образ, или образ-текст, търсене на подобни изображения по текстови описания, пораждаване на образи по описание като DALL-E и StableDiffusion.

\* Learning Transferable Visual Models From Natural Language Supervision  
Alec Radford et al., 26.2.2021 <https://arxiv.org/pdf/2103.00020> (400 милиона броя двойки образ-текст, събрани от Интернет). <https://github.com/OpenAI/CLIP>

<sup>11</sup> Многоостранен „Много-Мултимодален“ агент на Google DeepMind от 2022 г.

\* A Generalist Agent Scott Reed et al., 5.2022/11.2022 <https://arxiv.org/abs/2205.06175>

начин. Разбира се, ако уголемите размера на модела и на данните за предобучението до безкрайност, вероятно ще покриете всички възможни вероятностни разпределения на данните и няма да има какво ново да видите в бъдеще, така че всяка нова задача ще бъде разрешена много добре със съществуващата система и по такъв начин може да се каже, че ще разрешите продължаващото обучение. За практически цели обаче е необходимо да се намери дадено равновесие: колко да увеличавате размера както наличните данните, така и на размера на модела, за да достигнете определено ниво на сложност на задачите за предвиждане\* в продължаващото обучение?<sup>12</sup> 1:02 ч ...

**Бележки Тош:** \* downstream tasks; задачи за предвиждане или полезни задачи, или приложното предназначение на модела: онова за което се използва обученият модел за машинно обучение, например ако е езиков модел: предвиждане на следващите думи, класифициране на документи, анализ на оценките и отношението (sentiment analysis), разпознаване на именувани обекти (named entity recognition)

<https://web.archive.org/web/20220716001432/https://ai.stackexchange.com/questions/28410/which-tasks-are-called-as-downstream-tasks>

\* Системата/проект "КогАлг" също се стреми към непосредствено и постепенно обучение (online learning, incremental learning). При нея няма отделни фази на обучение и извод. Виж и SuperCogAlg, разклонение на Т.А.<sup>13</sup>

№ 9.3.2023 **ML Street Talk ... Irina Rish ... #95 Neurips ...**

Transhumanism ... **No true altruism** - ти се чувстваш по-добре

Сравни Т.А. 2002 (и в продълженията): „Писма между 18-годишния Тодор Арнаудов и философа Ангел Грънчаров (...)“:

**Тодор:** „Всички са егоисти, дори да не го съзнават. Дори когато

---

<sup>12</sup> Continual Learning, life-long learning – виж подбор на научни статии и бележки в приложения #Lazar. Лазар Вълков и Янислав Дончев имат работи в тази област.

<sup>13</sup> <https://github.com/Twenkid/SuperCogAlg> – имаше напреднали изследвания по мета-познавателен алгоритъм, които обаче не бяха доразвити тогава; може би части от проекта ще се използват във Вседържец и ACS (Research Assistant, Assistant C#) и ще се доразвият полуавтоматично и автоматично от него. Тош е автор на съкратеното име „CogAlg“ на проекта „Cognitive Algorithm“.



*вършат "добрини" за другите, го правят най-малкото защото ТЕ го желаят. Алтруизмът е доставяне на лично удоволствие чрез вършене на дела със знанието, че го правиш "за другите" (пряко, "безкористно", самото дело "за другите" доставя удоволствие на вършителя; или удоволствие, постигнато вследствие на предизвикани действия на "благодарност", при което е възможно самото действие да не кара "алтруистът" да изпитва търсеното чувство, но той "работи за другите" с цел по-късно да получи "награда"). Любовете е доставяне на лично удоволствие чрез удоволствието на други човеци. Тя е "съвършеното чувство", защото е стремеж към доставяне на взаимно удоволствие, при което изгода (удоволствие) получават всички "участници". " „от егоизма не може да се избяга ...“*

\* "Neural Networks are Also Symbolic - Conceptual Confusions in ANN-Symbolic Terminology", T.Arnaudov, 4.2019 <https://artificial-mind.blogspot.com/2019/04/neural-networks-are-also-symbolic.html>

## Йоша Бах - Познаваемата Вселена

### The Learnable Universe

Bugout dev – Power to the developers! 782 абонати

Йоша Бах, 16.11.2021, Joscha Back, Intel Labs, California,

<https://youtu.be/KG856Xv7P8> 12 820 показвания 30.12.2021 г. (към 24.2.2023)

Подбор, превод, бележки и автор: Тодор Арнаудов – Тош (ТА, Т., ТА):

#### Подробен преглед

Т.А.: Виж също гостуванията на Й.Бах при Лекс Фридман (два пъти), в канала "Machine Learning Street Talk", поредицата „Vectors of Cognitive AI”, особено диалога със Стивън Волфрам и др.  
(...)

**Йоша Бах:** Изчислителни модели на познавателните процеси и невронно-символни методи, ... Основно [се определя като] познавателен учен (от когнитивни науки, cognitive scientist), търся връзката между възприятието, мотивацията и познанието (cognition). ... Работи в изследователска група на "Интел", която се опитва да оцени качествено и количествено се опитва да разбере какво предстои в бъдещето на ИИ.

Днес: по-философска беседа. ИИ като философски проект, не само "автоматизирана статистика".

Бих казал, че включва *няколко хиляди души*, които се занимават с тази област. Хората бъркат направленията на ИИ като инженерен и ИИ като философски проект.

- \* Философия : реалност за всички теории
  - \* Математика: реалност за всички езици – започвайки от формалните
  - \* Изчисление: построителна математика – частта, която *работи* [е приложна, приложима; бел. Т.А.]
  - \* Да се обединят математиката и философията, въз основа на изчислителните модели на ума
- [Т.А.: "*истинно*" – проследимо. Сравни с „Вселена и Разум“/ТРИВ]

**Йоша Бах:** 7:00 [В прекъснатата, дискретна система], ако има прекалено много части за изброяване, се обръщаме към сходимостта, като клонене към граници - геометрия ...

ЙБ: 7:42: изследването на въпроса "какво представляват умствените способности" е: какво прави ума, когато построява модели. Ако успеем да създадем такава система, която да строи модели, подобни на нашите, и да ги разбира, ще може да я попитаме как работи и тя ще може да ни обясни как работи и нашият ум.

**[Бележка:** Сравни с Т.А., „Дали *Deer Blue* наистина победи човешкия ум?“, озаглавено още „Дали *DeerBlue* наистина победи Гари Каспаров“ - философия и Изкуствен Разум“, сп. „Свещеният сметач“ бр.6, 4.2001]:

*„Напредъкът на Изкуствения Интелект, води до напредък и в науките свързани с Естествения Интелект (и обратно), защото до разбирането на нас самите можем да стигнем не само чрез изследване и наблюдение, а като се опитаме да изградим нещо, което да работи подобно на нас. Много вероятно е идеите, които вложим в Мислещата машина, да са в основата и на "мислещото животно" :). Научавайки повече за себе си, ще усъвършенстваме мислещата машина, а така пак усъвършенстваме самите нас...“*

*Тодор Арнаудов, април 2001 г., възраст: 16 г. 9 м.]*

ЙБ: 8:47: Лайбниц е искал да построи изчислителен метод, с който да описва всичко в света и след това да може да изчислим отговора<sup>14</sup>.

Преподавателите ми по философия винаги са се отнасяли с отрицание към този проект - как би могъл да превърнеш всичко в числа и след това да го пресметнеш, каква наивна идея? Въобще не мисля, че е била наивна идея, била е идея за построяване на машина, която да извършва тези изчисления и този универсален изчислителен метод (universal calculus) е занимавал умовете на много учени<sup>15</sup>, например Фреге, Витгенщайн, който се е опитвал да формализира езика на философите, чрез мисловни опити, ... идея да превърне английския език в програмен език е изразен в неговия "Логико-философски трактат" ...

ЙБ: 10:38 Тази забележителна книга предшества идеята на Марвин Мински да напише език, с който да работи ИИ. Проектът на Мински пропада, като той заключава, че причината е, че не може да внедри сетивната информация, той споменава картините и образите, във формалния му език - последното се случва чак в последно време чрез многослойните невронни мрежи, които успяват да извършат автоматично пресмятане на приближения на функции, които се справят със съдържание в сетивните модалности и ги обединява във формалните модели, които строим. ...

---

<sup>14</sup> Виж приложение Алгоритмична сложност #complexity и споменатата книга към края с кратки бележки: \*Randomness & Complexity, from Leibniz to Chaitin, Cristian S. Calude. 2007

<sup>15</sup> Такъв е и проекта на Стивън Волфрам, както и на Т.А. Едно от имената на език за описание на мисълта е „Зрим“.

ЙБ: 11:18: Класическият ИИ е бил символен и много опростен ...

### Три ери в ИИ:

- 1) анализираш задачата, намираш алгоритъм да я решиш, например в шаха, и осъществяваш алгоритъма в код
- 2) пишем алгоритъм, който научава решението на проблема като открива решението сам.
- Третата, която се изкушава да очакваме:
- 3) мета-обучение, при което не пишем алгоритъм, който открива решение на задачата, а [мета-, общи] алгоритми, които откриват как да откриват решението на задача; учат как да учат

*[Т.А.: Сравни с **трите начина за предвиждане** от „Схващане за всеобщата предопределеност 2“, или „Писма между 18-годишния Т.А. и философа А.Г.“: **Тодор до Ангел, 27.8.2002 г.** цитирано няколко стр. по-долу в тази глава.*

(...) ЙБ: 15:38: Единственият начин по който невроните в мозъка могат да оцелеят, разглеждани като малки същества, е като се приспособят, научат се при какви условия да бъдат във възбудено състояние, да си дават обратна връзка и така, гледани отдалеч и отвън, изглеждат сякаш има обща сфера, която еволюира в съгласуван модел-схема на възбуждане на невроните; "ум" (mind) наричаме това виртуално "нещо"

*[Бележка: Сравни с: "Анализ на смисъла на изречение...", Т.А., 2004, и с понятието от 2012, за изчисление на „интеграл“ на „аз-а“<sup>16</sup>, самоличността върху сбор на безкрайномалки „аз“-ове.]*

И за мен беше **голямо прозрение** да осъзная, че думата "дух" не е суеверие, а дух това е операционната система на автономен робот ... самите хора, животни, екосистеми, градове, държави, всички тези системи притежават възникващ въображаем софтуер, чрез който може да "прожектирате" възможното им развитие в бъдеще и по този начин да получите в известен смисъл организирана и съгласувана дейност на частите, която се държи така, сякаш следва общо предназначение и крайна цел ["телос" във философията, по Аристотел, бел. Т.А.], обща

---

<sup>16</sup> Todor Arnaudov, Nature or Nurture: Socialization, Social Pressure, Reinforcement Learning, Reward Systems: Current Virtual Self - 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, Todor Arnaudov, 2012, In AGI, Artificial General Intelligence, Artificial Intelligence, Computer Graphics, Computer Vision, Human-Computer Interaction, News, Research Accelerator, Research Institutes, Research Plans by Todor "Tosh" Arnaudov - Twenkid // Wednesday, November 14, 2012 // 1 comment  
<https://artificial-mind.blogspot.com/2012/11/nature-or-nurture-socialization-social.html>

структура и изчислителната им стратегия може да се опише с единичен софтуер... но всъщност онова към което се стремят е да намалят "свободната енергия" - според Карл Фришън, 2006 г. [според него е 2006 г., според други източници поне 2009, 2010 г.?] - и тук не става дума за термодинамична енергия, а енергия на описанието, за да опишете състоянието на системата, която моделира средата и тази енергия се намалява, като се намали енергията, необходима за обновяване на следващото състояние; т.е. идеята е, че системата, била тя клетка или мозък, моделират обвързването си със средата в своето собствено вътрешно състояние и се опитват да намалят грешката в предвиждането; (...)

[Тош: Сравни с „Вселена и Разум“ („Теория на Разума и Вселената“) на Т.А. "душа" - модел на личността и пр., грешката в предвиждането и пр.: аналогични или еднакви определения]

Самоорганизация ... Биологичните неврони са агенти! (А не просто пресмятат/откриват приближения на функция чрез изчислителни модели)

[Тош: Сравни с „управляващи устройства“, „управляващо-причиняващи устройства“ от които е изграден, чрез които се представят, моделират, Вселената и разума според ТРИВ]

19:25: Какво е агент? Кибернетика - моделиране с цел управление. Обобщение на Управлението --> управление на модел.

Средата внася отклонения в Управляваната система (Regulated system). Отклоненията се отчитат от Сетиво (сензор, датчик), който отчита отклонение от дадена целева стойност.

Управляващо устройство (Контролер) ги обработва, изпраща Обратна връзка към Действеник (Ефектор, Актуатор) и той въздейства отново върху Управляваната система.

Кибернетиката позволява да започнем да определяме какво означава да бъдеш мотивиран.

... За да оптимизира, да намери най-добрата стойност на интеграл (сбор) върху множество от целеви установени състояния (set point deviations) за дълъг период от време, управляващото устройство трябва да моделира бъдещето и строежа на средата и да различава между различни ситуации и да предпочита някои от тях пред други и ще изпълнява стратегии за това как да достигне до желани ситуации; и когато описваме системата от

вън, като за външен наблюдател, хората ѝ приписват свойства на одушевен предмет (agency) ... [Сравни с ТРИВ]

21:20: не можеш да моделираш света ефективно, ако нямаш негов истинен, достоверен модел; ... т.е. на първо място трябва да следваш истинността и да стигнеш до описание на системата, която искаш да управляваш, която е изоморфна на динамиката на системата на нивото, на което искаш да я управляваш и вселената, в която се намираме може да бъде разбрана като вселена, която може да бъде управлявана и мисля, че отговорът на този голяма главоблъсканица защо въобще е възможно да научим каквото и да било, защо можем да разпознаваме структурата на вселената, защо сме във вселена, която е разбираема, е защото ние сме управляващи устройства [*"controllers": контролери, регулатори: същото като в ТРИВ, 2001-2004*], ние сме на върха на йерархията от управляващи системи, така че по определен начин може да кажем, че елементарните частици са управляващи устройства с нулева точка на плаване (zero-point fluctuations) и атомите са управлявани елементарни частици и молекулите са управлявани частици, и клетките са управлявани; молекулите и организмите са управлявани; обществата са управлявани организми и т.н. Така че имаме йерархия от управление, и защото за да управляваш трябва да създадеш модел на скритата моделирана система, това значи, че управляваната вселена също може да бъде моделирана и моделът, който атомът трябва да построи за елементарните частици, разбира се е изключително прост, защото е изключително просто механично управление<sup>17</sup>, но моделите на клетката трябва да се построят така, че да моделират молекулите, които изграждат клетката и са много, много сложни и има нещо като изместване на сложността. Изместването на сложността означава, че за да може клетката да извърши необходимото управление на молекулите, тя трябва да изработи в себе си [своеобразна] "машина на Тюринг", клетката трябва да бъде изчислителна машина - ако тя не може да изчисли необходимото, т.е. ако не е достатъчно сложна да се научи как да моделира бъдещето на тези молекули достатъчно добре, така че да построява тези супер

---

<sup>17</sup> Може и да е по-сложно, отколкото изглежда според онова, което моделираме засега. Виж например хипотезата на Т.А. за „причинностните отъждествители“ или белези, Casual Ids в страницата на „Теория на Разума и Вселената“ в Гитхъб и в статията от Т.А. “ Why "Infinity doesn't exist"? Why Goedel Incompleteness theorems are irrelevant for Artificial General Intelligence? What is Truth, Real and Realness and why: how they are determined and experienced? Why the Virtual Universes are real and do exist? Causal IDs Hypothesis. Illusionism and Realism. Two types of Experience: Physical and Cognitive. Etc.“, 2023

огромни молекули, клетката с всичката ѝ динамика, която може да представи много много видове отклонения във вселената, така че предназначението на тези регулации [на това управление, спазване на тези установени правила, предписания, Т.А.] е да поддържа сложността да се построяват системи, които са устойчиви срещу смущенията и се саморазпространяват, в голямо възможно множество от среди, когато внасяте смущения в тях.

### **Бележки Т.Арнаудов:**

Сравни с понятието за „Страничен ефект“ в ТРИВ, 2001-2004

1. Клетката като машина
2. Клетката като изчислителна машина

1. Някои хора се възмушават, ако се говори за живите организми и клетките като за "машини", препращайки към философските теории за живота до средата на 19-ти век, които го обясняват с "жизнена сила" (*vis vitalis*<sup>18</sup>). В лекция на Майкъл Левин от 2021 през Интернет, в която обяснява развитието на човека "От физика до съзнание", като има предвид от една клетка с химични и физични процеси, получава "гневен" въпрос-коментар, че живото произлизло само от живо, че само животът дава живот, защо го нарича физика?

Моят отговор е, че тук, както и в много други случаи, хората не разбират различните нива на обобщение, мислене и избор на особености, въз основа на които съставят понятия, класифицират и пр.

Виж в Т.А. "Писма между 18-годишния Т.А. и [43-годишния] философ А.Г.", 2002 и др., където Тодор разсъждава за това, че дали дадена част от времепространството може да се определи като "жива" зависи например най-малко от обема на времепространството, за което се отсъжда какво е. В една наносекунда, в един нанометър и пр. някъде в някоя клетка може и да не може да се определи дали това е „живо същество“ или дори дали е част от живо същество или неживо, на химическо и физическо нивом, защото атомите, молекулите, които се намират там и имат определено състояние, биха могли да се намират и в "неживо" същество", в мъртво и т.н. Трябва да се увеличи обхвата, микрометри, милиметри, сантиметри и т.н., но тогава пак първо се появяват определени молекули и макромолекули, клетъчни органели и малки системи от взаимодействащи си подобни части, после части от

---

<sup>18</sup> Според мен процесът "Active inference", Извод чрез действие; Волята и пр. са проявление на своеобразна „жизнена сила“, „обективация на Волята“

клетка или цяла клетка, колонии от клетки, тъкани - части от тях, припокриващи се с други и т.н. Докато на определена степен, в определен обем и продължителност от време и процеси, се отчете че са налице необходимите структури, техни взаимоотношения и процеси, които са наречени "живи" и присъщи на живота. В конкретния пример М.Левин завършва процеса със "съзнание", защото то се смята и за нематериално, за "вълшебно". Също така, въобще за да се „отчете“ подобно нещо, е необходимо да съществува съответна система наблюдател-оценител, която да може да измери тези особености и да знае кое принадлежи на „живо“ и на „неживо“ и защо.

За разлика от живота, който изисква определени степени на анализ и обхват и много условности, за да се класифицира като такъв, „обикновената“ "физика" или „физични закони“ се приемат за физика навсякъде, във всички размерности и обхвати. И в най-микро, и в най-макро равнище, и в кварк и в измеримите процеси, свързвани с него, и в мащабите на константата на Планк - най-краткото измеримо разстояние или време според квантовата физика - до размаха на цялата Вселена с всички галактики. Засега е открито, че физичните и химичните закони работят в същество, което при по-голям обхват се брои за „живо“ по същия начин, както и в неживото. "Всичко е физика и химия", на такова научно и материално ниво. В същото време е възможно да има белези, „Casual Ids“ или други особености, които засега не познаваме или не можем да измерим, виж бележка под линия по-горе.

**2. „Работата“** и природата на клетката като изчислителна машина става натрапчива, ако човек обърне внимание на молекулярната биология и генетиката в подробностите им. Биоинформатиката се развива, за да отрази научните модели в биотехнологиите, които имат описание и логика подобна на програмни и апаратни компютърни системи. Виж например "Рекомбинантната ДНК: ", 1989 г. (на английски: 1982 г.), която описва историята на откритията, свързани с ДНК и синтеза на белтъци в клетката, като проследява и изследователският процес, какви опити са били правени, за да се постигне до определена стъпка на анализ, какви проблеми са преодолени, как постепенно се е извършвало все по-прецизно "секвениране" - откриване на точната последователност от нуклеинови бази, изграждащи молекулите на ДНК и РНК: АГТЦУ. ДНК и РНК изглеждат като програмен код, запис от кодиращи символи и от условия за начало и край, които се разчитат и презаписват чрез своеобразни "четящи устройства" рибозоми, като използва "четящи глави" - РНК, която е различни видове: "матрична", "информационна", "транспортна". В молекулата на ДНК се локализируют и се задействат или



потискат различни участъци, гени, чрез производство на различно количество ензими, които могат да ускоряват или забавят определени процеси на синтез на белтъци по участъци от кода, действайки като катализатори - молекули ускорители - или пък като "репресори" - молекули, потискащи определени химични процеси. Клетките притежават "йонни канали": "входно-изходни" устройства за връзка със средата, през които избирателно се пропуска или спира навлизането на определени вещества, от средата към клетката, или се извеждат от клетката към средата и т.н. В една клетка може да взаимодействат над 1000 различни видове молекули със съответни физико-химически процеси между тях, и всичко трябва да се управлява и да се поддържа в определени съотношения, скорост и посока на реакциите, в определени участъци от клетката, за да може клетката да продължи да съществува и да не се разпадне: например да задържа определени ензими затворени в определени мехурчета, защото ако попаднат в цитоплазмата, сред други органели, те биха разложили самата клетка - както и става, когато клетката умира и вече няма "господар", или „дух“ от достатъчно висока степен на развитие и обхват – този дух е съгласуваното цяло на всички работещи „правилно“, предвиждащи, причиняващи, части на системата, и „духът“ или „господарят“ "поддържа реда" в малката "държава", „машина“, „агент/деятел“, управляващо-причиняващо устройство. Научен термин, определящ рибозомите, е "молекулярна машина"

[https://en.wikipedia.org/wiki/Molecular\\_machine](https://en.wikipedia.org/wiki/Molecular_machine) И т.н.

\* **Тош:** Доп.бел.15.6.2025: Това че нещо е „машина“ „не е лошо“ само по себе си, но така е вътълпено в честа представа сред хомо сапиенсите, произтичаща от „маймунска“ йерархия и убеждение за „по-висше място“ със „свободна воля“, а „машините“ като нещо „по-нисше“, несвободно и пр. Виж творбите от Теория на Разума и Вселената, 2001-2004, писмото ми до Оксфорд от 2.2012 и др 21.8.2025: Под машина са внушени унизителни неща, робство и др. Друга страна, разгледана в споменатите творби още от 2001-2002 г., и „Следващото еволюционно стъпало“ и пр., човекът е тясно свързан с техниката, и за да бъде човек се нуждае от нея, както въобще от средата. Машините и хората са система, без първите вторите се приближават до маймуните. В „Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини“ и „Вселена и Разум 6“, други приложения на *Пророците...*, обръщам внимание и на това, че човек и въобще дейците могат да се разглеждат като **полета** във времето и пространството, а не само съсредоточени в обичайно приетото тяло и че понятието за *самосъздаване – автопоеза* – може би е заблуждаващо, и от друга по-„честна“ гледна точка да се види, че всъщност живите организми *не се самосъздават*, а и те също се „сглобяват“ с помощта на средата, която доставя необходимите „части“ (вещества, молекули, енергия – подходящи дози,

„формати“ и пр. За различните нива във Вселената сметач или въображаемите вселени от по-високи равнища, по-долното е като среда. Виж също бележките ми към беседата с Йоша Бах *„Наистина ли съществува „трудният проблем“ на съзнанието?“* ]

### **Йоша Бах, продължение:**

24:01 Може да се запитаме дали Вселената е изчислителна машина и не е ли динамична система, и отговорът е, че няма истинска непрекъснатост по математически причини, защото ако искаме да говорим за безкрайност на какъвто и да било език, в даден момент неминуемо ще стигнем до противоречие. Ако се опитвате да обясните как работи, което значи че думата вече не значи нищо, ако езикът на който говорите за света.

*[Сравни с "Писма между 18-годишния ...", 2002 и романа "Ада", 2004 и „Why infinity doesn't exist? ... „, Т.А. 2023]*

### **Вселената компютър ли е?**

- \* Динамична система?
- \* Непрекъснатата система?
- \* Квантова?
- \* Хиперизчисление?
- \* "Одушевеност"? (Agency)
- \* Съзнание?
- \* Съществуване?

**25:00 Й.Бах:** Компютрите, които създаваме, за да описваме вселената, са построени на върха на динамични системи, те натрупват вероятностите на тези системи, докато те станат достатъчно детерминистични за нашите цели.

**[Бележка:** Сравни с ТВИВ, "Свещеният сметач", че макар и да е основан на "шумни", вероятностни химични и физиологични процеси на ниско ниво, че на микро ниво властва квантовата физика, в която има неопределеност, умът се стреми да се развие по такъв начин, че в крайна сметка на най-високото абстрактно ниво да построи изчислителни машини, системи, модели които работят по точно определени и известни на самата система правила, закони, чрез които да може да предвижда бъдещето на Вселената с която взаимодейства: „външния святЮ, средата, както и своето собствено бъдеще (то също е част и от „средата“, най-„близката“ и

най-непосредствено свързаната), с максимално висока разделителна способност на управление и възприятие, с максимален обхват и също тако, в крайна сметка, с максимална точност; най-голямата е вероятност равна на едно, или напълно детерминистично. В Теория на Разума и Вселената подвселените успяват само в рамките на своите въображаеми Вселени, не могат да предвидят бъдещето на цялата Вселена с пълна точност и разделителна способност и обхват на цялата, понеже те са само част от нея и не биха могли да имат необходимите изчислителни ресурси. Сравни това и със следващия коментар в теорията на Йоша Бах за квантовото изчисление, че според него то дава много по-висока изчислителна мощ спрямо детерминистичните компютри, понеже вторите биха могли да бъдат описани като последователност, йерархия, от закони от по-ниско ниво, както при разработката на симулатори (бих казал или емулятори) на компютри и видео игри. Това е причината детерминистичните изчислителни машини да са по-бавни и по-неефективни в работата си, за да симулират процесите, в които системите от по-ниско ниво, квантовите, са „в свои води“. Сравни с аналогични подобни твърдения и принципи от ТРИВ, които подчертават, че във Вселената се образуват управляващи устройства (подвселени, въображаеми вселени) от по-високо ниво, които са основани на подмножество от законите, върху които работи по-ниското ниво, и са по-бавни и предвиждат бъдещето с по-ниска точност или в по-малък обхват - измерено на нивото на "майчината" Вселена от по-ниско ниво, като давам пример за това, че една програма, въображаема Вселена от високо ниво, в компютър, обикновено няма достъп до точните нива на напрежението в кондензаторите, които съхраняват "единиците и нулите", от своя своя код, а дори и да има би бил само достъп до единични стойности – например аналогово-цифров преобразувател, но не и на всички електронни елементи, които я изграждат. Описанието на един физически, материален сметач, с конкретни атоми и техните свойства, е невъобразимо пъти по-сложно и обемно като брой части или информация, отколкото абстрактно описание на компютърната му архитектура на някакъв сбит формален математически, програмен, инженерен език; като дори инструкциите на процесора също може да са огромен обем за човешки мащаби, например само ръководството за програмиста за 13-тото поколение процесори на Интел е 2528 стр. в сбита форма: <https://cdrdv2->

[public.intel.com/671110/325383-sdm-vol-2abcd.pdf](http://public.intel.com/671110/325383-sdm-vol-2abcd.pdf) и човек трудно би могъл да бъде приложен програмист само от тази информация.

От своя страна вършачът (процесорът) сам не може да свърши нищо, без да е свързан с дънната платка, работна памет, захранване, външна памет и т.н., т.е. нужна е цяла система. И пр.]

**Йоша Бах:**

**\* Квантовите системи компютри ли са? ...**

Ако се обърнем към тезиса на Чърч-Тюринг, за да разберем какво е "Компютър", "Изчислителна машина", "Сметач" - е система, която е способна да преминава между различни състояние по неслучаен начин. Това е най-общото възможно описание. Квантовите системи правят същото, нали? Квантовите системи се описват със състояние, което е суперпозиция на възможните състояние от определен ъгъл, но то все още е състояние и тогава имате преход между състояния в следващо състояние, така че квантовият компютър също е просто сметач, но причината квантовите компютри да могат да вършат изчисления, които обикновените сметачи, основани на частици (particle computers), е че съгласно квантовата механика Вселената, разглеждана като съставена от частици, е осъществена по много неефективен начин на върха на квантовата си основа.

26:12: Например, ако построите компютър в играта "Майнкрафт" от "червен камък" [*виртуален строителен материал*], ще има полиномна връзка между скоростта на компютъра в играта и бързодействието на физическия компютър, т.е. симулираният виртуален компютър в играта ще бъде много по-бавен, отколкото ["истинският"] онзи, на който се изпълнява, защото малка част от същностните изчисления [операции, инструкции] на виртуалния компютър ще стигат до физическия. [*Ще се изпълняват пряко; т.е. повечето са "излишни" - графиката и мн. други условности; виж понятието "странични ефекти" в "Теория на разума и Вселената", Т.А: 2001-2004].*

Така съгласно квантовата механика, Вселената като частици е толкова неефективно осъществена върху квантовата вселена, че ... квантовата вселена се разклонява по много начини и повечето изчисления в квантовата вселена нямат принос в нашата верига от събития (timeline), която е малка част от квантовата; само "капки" капят в достъпната за нас линия от събития и квантовите компютри са осъществяване на смелата хипотеза, че бихме могли да "отворим кранчето" към квантовата "тръба", така че сметачите, основани на

Вселената представена като съставена от частици, да получи допълнителна изчислителна мощ , която иначе не е достъпна във „Вселената на частиците“...

### 27:30 \* Хиперизчисление ... Причинностно хиперизчисление -

Представете си, че може да построим затворен цикъл със закъснение във времето, например по някакъв начин да можем да виждаме в бъдещето и да знаем какви числа ще се паднат от лотарията следващата седмица и да ги използваме, за да спечелим наградата\*. На пръв поглед изглежда, че не е възможно да го постигнем с класически компютър, нали? Напротив, естествено, че е възможно - просто трябва да копираме състоянието на настоящата вселена, да го поставим в междинна памет, и когато изпълните симулация на Вселената до следваща седмица, да прочетете числата от лотарията, да ги съхраним и да рестартираме вселената от състоянието, което сте запазили в буфера и след това да предадете [фиша] с числата. Следователно доколкото бихте могли да съхраним състоянието на настоящата вселена, бихте могли да построите затворени времеви цикли с класически сметач.

#### [Бележка Т.А., 5.3.2023:

*\* В тази мисъл Й.Бах неявно, т.е. без да го изрази изрично, приема, че Вселената е предопределена (детерминистична) и че подобно поглеждане в бъдещето не би променило нейното последващо случване, т.е. няма да причини "изкривяване на времепространствения континуум", ако се изразим в стилистиката на фантастичните филми с пътуване във времето, популярни през 1980-те и началото на 1990-те.*

*Сравни със "Схващане за всеобщата предопределеност 2"/"Писма между 18-годишния Тодор Арnaudов и 43-годишния философ Ангел Грънчаров (...)", 2002 г.... За трите начина за предвиждане на бъдещето:*

*Следва кратък откъс от огромното ми второ писмо, от 27.8.2002 г., в което въвеждам много понятие, едно от които: предвиждането като основен принцип и стремеж на Разума и Вселената и връзката между тях. Виж също отказът ми да приема съществуването на "истинска" случайност, която може да се запише изцяло. Виж също откъса от романа "Ада" в "безкрайния коридор" и споменатата по-горе нова статия с дълго заглавие: „Защо безкрайността несъществува? (...)"*

**"Тодор Арnaudов, 27.8.2002 г.: Всъщност всичко е точно и "ръбато", защото се изменя на резки стъпки - според квантовата физика и според логиката. Всички окръжности, дори и тези, които са изчертани с атоми всъщност са многоъгълници - може да имат огромно количество страни, но то никога не може да бъде безкрайно, за да бъдат "истински окръжности". Когато с очите си различаваме окръжност и си казваме, че е такава, всъщност дотолкова стига разделителната способност на зрението ни - когато не можем да различим стъпките виждаме плавен преход. Но всъщност "плавноста" е относително нещо\*. Сметачите (позволявам си да използвам "сметачобългарското наречие" вместо чуждицата "компютър") ни показват "резкостта" на света - всички знаят, че в тях "има само единици и нули", че са "точни, предписуеми". От опита си досега знам, че повечето човеци, които си мислят, че творят "истинско изкуство" много често "бягат" от сметачите, защото те са "бездушни", защото са "точни - само единици и нули" и нямат "творческата искра", "няма творческо начало", "няма божествено вдъхновение". Самозовя се "сметачолюбец" и умознатолюбец (технократ) и се дразня лично от подобни изказвания... Та тези "хора на изкуството" не съзнават, че като наредиш много "точни" единици и нули получаваш "плавни" преходи, такива, които умът ти не може да различи като резки. И това е навсякъде около нас, и е било много преди сметачите да се появят. Всяка буква е много проста и няма смисъл ако е сама, но ако наредиш много букви, празнини между буквите, препинателни знаци и т.н. се получава слово... И най-простата къща и двореца са построени от "прости" тухли... Както и всичко, според това, което Науката казва, е изградено от елементарни "тухли", включително и "най-съвършеното творение във Вселената - Венецът на Божественото Творение"...**

Като свържеш "някакви си" десетина трилиона логически елемента получаваш "също толкова елементарно нещо", което работи с "най-простите неща във вселената" - единици и нули. Това устройство (свърхсметач, има поне няколко с подобни възможности) за секунда може да извърши трилиони изчисления върху данни с обем десетки трилиона бита - човекът може да извежда някакви си десетки бита информация за същото време...

Всъщност ако пуснем предписание на какъвто и да е сметач, и то няма нужда от "странична" информация, която да се въвежда в течение на решението - т.е. имаме "затворена вселена", в която ЗНАЕМ ВСИЧКО от което зависи бъдещето, то крайният изход - отговорът на задачата = бъдещето, е единствен и ясен още в мига на пускането на програмата. Ясен е на този, който (1) може да използва по-съвършен сметач, чрез който да пресметне отговора преди първия,

**(2) който не разполага с по-бърз сметач, но е открил по-просто решение и може да получи отговора по-бързо от първия сметач, или (3) който знае отговора предварително, защото вече е решавал същата задача... Ние непрекъснато се опитваме да предсказваме по втория начин. А дори и аз мога да предсказвам бъдещето по първия - програмирам едно и също нещо на стария си Правец-8М и на настоящия си, пак стар :), сметач. Задействам програмата на Правеца, задействам на настоящия - той извежда резултата преди Правеца и аз вече ЗНАЯ БЪДЕЩЕТО му - известно ми е точно какво ще се изпише на неговия показвач!**

-- КРАЙ на цитата от "СВП2", 2002

**Бележка 26.6.2022 г.** За окръжността също така умът е избрал да „закръгли“ възприятието си до кръг. Това е възможно дори и когато ръбовете са видими. Възприятието е не само отдолу-нагоре, но и отгоре-надолу, по-висшите слоеве могат да *изберат кои особености и кои съвпадения са по-важни за тях и с каква точност да сравняват*. Това може да се случва както за прости сетивни образи и предмети, например зрителни, така и за сложни, съставни, отвлечени, като го вършат и най-малките деца и дори бебетата. Играта наужким например е отвлечане от конкретните особености и „виждане“ на *онези съвпадения*, които ни интересуват: покриваме края на масата с одеало или дрехи, влизаме вътре и получаваме „къща“ – за нас приликата е достатъчно голяма.

\* „Трилиони изчисления в секунда върху трилиони бита информация“ – възможностите на най-мощните суперкомпютри към 2002 г. TFLOPS или TOPS, TIPS (целочислени операции в секунда). Видеокартите от среден-нисък клас постигаха такава изчислителна производителност още през 2012 г. Geforce GTX 650 Ti: 1.425 TFLOPs (10.2012), Geforce GTX 750 Ti: 1.388 TFLOPS (2.2014).. Тази бързина обаче е постижима само за паралелна еднотипна обработка.]

**Йоша Бах: 28:26: Съзнанието ...** Не мисля, че в съзнанието има нещо специално. Съзнанието обърква ... всъщност е свойство на въображаемите (виртуални) системи.

Смятаме, че съзнанието ни дава достъп до физическата вселена... Че това което възприема съзнанието е действителността, действителния свят. Това не е вярно. Всъщност съзнанието е състояние на сън; нещо като мултимедийна история, която се поражда в агента. Физическите системи не могат да бъдат съзнателни, нито невроните, нито компютрите. Съзнанието е изцяло виртуално свойство. Съзнанието е условност (?), "как би било, ако..", "като че ли..." (as if)... Можем да възприемаме нещата

сякаш са истински именно защото живеем в този условен свят на "като че ли сме..." ("as if"). Обаче да съществува в действителността или в компютър "не се усеща" по никакъв определен начин.

Някои хора си мислят, че компютрите или симулациите не могат да имат съзнание, но всъщност нещата са противоположни, защото съзнанието е симулирано свойство, а не реално; то не може да бъде реално.\*

**Бележки:** \* **Тош:** Може би има предвид \*материално\* и по-точно "вулгарно материално" - "не може да се пипне", а е информационно, "духовно" - виж повестта "Истината", 2002 г. Тук някои от твърденията на Бах не са достатъчно убедителни и звучат като религиозни догми, без доказателства и разсъждения. И аз "вярвам" във "въобразната природата на Вселената" (информационната), както съм описвал и в ранните си произведения, но според мен колегата, а и „ние“, не можем **да знаем** от самосебе си какво "бихме могли да чувстваме, ако...", когато осъзнаваме, и дори твърдим, че "не можем да бъдем друго" или в това състояние, без вече да сме загубили субективното усещане, личността си, "Воля"-та си (в Шопенхауеров смисъл), съзнанието и пр. която сме били преди. Пример още от "Човекът и мислещата машина", 2001 и в "Истината" - ние твърдим, че машината няма "душа", няма съзнание, не може да усеща, **"защото е само единици и нули, полупроводници и метал"** и т.н. – в зависимост от случайно избран начин да се обърнем към някои белези, с които я определяме, всички частични, късчета, разхвърляне, в най-разнообразни степени и разделителни способности на възприятия и обхват. Мислеща машина, и дори най-прост сметач: устройство, машина, която има екран или може **да възпроизведе записан звук** или дори **просто да светне надпис с лампичка** и пр., които да покажат по някакъв начин, който е разбираем и осезаем от нас - "душепритежаващите най-висши същества" - съобщението си, биха могли да ни отговорят **по същия начин** за **своето** мнение за нашата "душа", "съзнание" и пр. Ние приемаме, че „това не е ... защото...“, но една система с някакво устройство може да ни „отговори“ по същия начин: „вашето не е ... като ... защото...“. Затова и "Китайската стая" на Сърл е аргумент, който поставят под съмнението и собственото съзнание на автора му, както и на човека. В тези случаи се объркват нива на абстракции, области на приложения на понятия и пр., подобно на примера с физиката и живота. Относно чувстването на света като нещо друго, друго същество, ако не греша в дискусиите в канала на Майкъл Левин в Ютюб се разискваше една изглеждаща убедително пътека, възможен преход, с който съм съгласен:



ако двете същества, каквито и да са те, биха могли постепенно да се слепят, да се свързват едно в друго, без да се унищожават; докато продължават да съществуват и да се намират или все още могат да се „събудят“ в онова състояние, което „наричат“ или „се приема“, или ние „приемаме“ за „съзнателно“, и ако могат по време на преобразуването да общуват и да изразяват какво усещат и т.н., тогава може би едно същество би могло да се почувства „като другото“, но отново това няма да е „точно“ като другото, защото няма да премине през пътя на създаване на другото, причинно-следствените връзки от оплодена яйцеклетка или преди това на „субстанциите“ и „нещата“, които може би някак си са „чувствали“ нещо в някакво свое битие.

Виж СВП-2 („Писма между 18-годишния...“), където съм отговарял на подобни обърквания на опонента си свързани с различни степени на обобщение, кое е сложно и просто и др. Виж също новата творба от ТРИВ от 2023 (работно заглавие към 26.6.2023): „Защо безкрайността не съществува? ...“ (Why “Infinity doesn’t exist”? Why Goedel Incompleteness theorems are irrelevant for Artificial General Intelligence? What is Truth, Real and Realness and why: how they are determined and experienced? Why the Virtual Universes are real and do exist? Causal IDs Hypothesis. Illusionism and Realism. Two types of Experience: Physical and Cognitive. Etc.)

(...)

**Йоша Бах:** Така стигаме до големия въпрос за съществуването: как може да съществува каквото и да било?

Отговорът е неприятен от гледната точка на „сметачолюбец“<sup>19</sup> и най-лесният отговор, който съм намерил до сега, е че може би съществуването е състоянието по подразбиране. Вместо да допускаме, че трябва да извършим нещо, за да може то да съществува, да извикаме нещо в действителния свят, може би вселената вече е суперпозиция на всички неща, които биха могли да съществуват. И ако нещо може да съществува, то трябва да може да се осъществи. Добро определение за съществуване е изработка, изпълнение (implementation) - нещо съществува, ако е осъществено; и всичко, което може да се осъществи, са крайни автомати. Следователно може би вселената е суперпозиция на всички крайни автомати и структурата на вселената е следствие на онези неща, които не съществуват като суперпозиция на всички крайни

---

<sup>19</sup> Изчислител, „изчислителист“, „сметачолюбец“ - computationalist; може да се каже - последовател на философията за Вселената сметач, какъвто е и Т.А., също Стивън Волфрам; Юрген Шмидхубер и др.. „Сметачолюбец“ в първоначалните си употреби през 2002 г. се употребяваше, и може да се употребява и в по-тесен смисъл на „компютърджия“, човек свързан със сметачи.

автомати; събуждането? на определени оператори може да създаде "дупки" в съществуването ... [Виж Стивън Волфрам]

Може би вселената е нещо като "обратен компютър" ...

Суперпозиция на всички оператори ... И ако наложите всички оператори [поставите ги в "суперпозиция"] ще получите определени състояния, които са недостижими. Това придава структура на вселената. Това обаче е спекулация. Нямам отговор на загадката защо съществува каквото и да било; истински ме шокира, че последното за мен е задоволителен отговор.

- **Бележка: Тош:** И аз също съм се питал защо съществува каквото и да било и пр., същият въпрос и нямам отговор. Ако се следва принципа за "икономия", звучи по-лесно да не съществува нищо. Съществуването на каквото и да било и то с определена структура, а не друга, е непонятно засега от гледната ми точка, данни, анализ. Сравни също Т.А.: "Няма смисъл да има смисъл", Писма между 18-годишния ... т.е. не е нужно да има специална цел и пр. Виж също новата творба от от ТРИВ от 2023 г. „Защо безкрайността не съществува? (...)“ публикувана в Medium@todorarnaudov, гитхъб и др.

Относно *"нещо съществува, ако е осъществено; и всичко, което може да се осъществи, са крайни автомати."* - сравни с доказателството на Т.А., тогава не много обстойно, за осъществимостта на Изкуствения разум в *"Човекът и мислещата машина - Анализ на възможността да се създаде мислеща машина и някои недостатъци на човека и органичната материя пред нея"*, 2001. За мен обаче беше, и е, очевидно, от въпроса: *"Какво бързодействие е необходимо на машината, за да започне да мисли ? (...) Няма съмнение, че човешкият мозък е програмируема система. "Програмирането" ни се извършва първо от света около нас, а после и от самите нас (повече подробности в следващи статии). Тъй като мозъкът е програмируема система, то всяка изчислителна машина може да емулира мисленето! Следователно сътворяването на Мислеща Машина е ВЪЗМОЖНО и ОСЪЩЕСТВИМО."*

**Йоша Бах:**

31 мин: Агент ... Кибернетиката: Моделира в служба на управлението.

1 \* Обобщение на управлението - управляващ модел

2 \* Управителят (контролер) намалява отклоненията от работната точка

(setpoint deviation)

3 \* Минимизирането на бъдещите отклонения от работната точка изисква модел

4 \* Моделът трябва да предвижда последствията от взаимодействието

5 \* Агент/Деятел: управляващо устройство + породител на работни точки (целеви състояния)

6 \* Управляващото устройство трябва да може да предвижда бъдещето и има предпочитани от него състояния.

7. Бих казал, че осъзнатият деятел (sentient agent) е онзи, който може да открива себе си. Ако откриеш собствената си "одушевеност" (agency), това е система, която променя света по определен начин и използва съдържанието на собствения си управляващ модел - откриваш собствената си перспектива от първо лице.

**[Бележка\* Тош:** Сравни със "Свещеният сметач", 2001-2004.

1. сравни с "управляващо устройство";

2. сравни: желаното състояние, целево състояние; виж лекцията за "Теория на разума и Вселената" от курса по Универсален изкуствен разум, 2010 г. и т.н.

3. Сравни: Управляващите устройства построяват въображаеми вселени, с които предвиждат бъдещето

4. Т.е. трябва да има причинностен модел, причинно-следствен модел (въображаема вселена)

5. "Породителят" на работни точки съдържа начални (или по-късно въведени) определители на целеви състояния, напр. първичните нужди на човека - виж други лекции на Бах за мотивацията.

6. Обуславя се от горните.

7. Сравни ... чрез части от себе си]

**Йоша Бах:**

Как създаваме модел?

Обща форма на сетивен модел

\* Кодира модели-схеми [patterns], за да предвижда други настоящи и бъдещи модели-схеми

\* Мрежа от отношения между модели-схеми: наблюдавани постоянни закономерности, "инварианти"

\* Свободни параметри: променливи, които съдържат състоянието, с което да кодират останалите вариации

Моделите-схеми (patterns) се отвличат до променливи на особеностите (feature variables).

Променлива = множество от възможни стойности.

Отношенията между променливите са изчислими функции.

Отношение на възможност: за всяка стойност, ограничава възможните стойности на други променливи и възможните бъдещи сетивни състояния; стремите се да намалите отклоненията от предвижданията на бъдещите състояния, противоречията в модела.

Отношенията не са вероятностни, а са **ВЪЗМОЖНОСТНИ**, защото не искате да моделирате само онова, което е вероятно, а всяка възможна вселена. Трябва да можете да моделирате всичко, което е възможно, а не само онова, което е вероятно!

Вероятността - когато моделът трябва да бъде сходим, чрез вероятностни нагласи: защото броят на всички възможни състояния е прекалено голям.

Вероятностно отношение: стойността на дадена променлива прави стойността на друга променлива по-вероятно.

**Бележки:** Сравни със "Свещеният сметач", ТРИВ. „искате да моделирате... всяка възможна вселена“: „универсални симулатори на въображаеми вселени“. Възможностни = причинно-следствени, "Какво може да се прави" в дадена ситуация, какво би могло да последва и пр.

**Йоша Бах:** 34:20:

**Valence**<sup>20</sup> (Валентност, психологически термин, привлеченост от едни или други неща; предпочитания, ценности, стремежи, вкус) - кои неизвестни, неопределености (uncertainties) трябва да се разрешат, защото ресурсите са ограничени. Трябва да се моделира неизвестното, несигурното (uncertainty), затова в модела трябва да се включи ... свързана с предпочитания, които са свързани с мотивацията и определят целите (работните точки, setpoints), което позволява да разпространите ценностите в системата и да определите кои части трябва да се обучават.

Може да се добавят и норми - вяра; без предварително обяснение, доказателство. Може да са внесени отвън. Стереотипи, догми (други агенти задават как стоят нещата) ...

Variables: possibility: probability: valence: norms

Construction - построяване основано на първични принципи ...

---

<sup>20</sup> [https://en.wikipedia.org/wiki/Valence\\_\(psychology\)](https://en.wikipedia.org/wiki/Valence_(psychology))

48:10: За "общото благо":

Когато в такава [мултиагентна] система има съревнование, сблъсък на интереси, различни интереси, разногласие, равновесието на Неш от теорията на игрите се разпада, не работи, и системата като цяло вече не е оптимална. Всяка част от системата работи за своя собствен локален интерес. За да се разреши този проблем, е необходим регулатор. Регулаторът е агент, както всяко правителство [управление], който променя начина за "отплата", "възнаграждение" [към отделните агенти в системата], по такъв начин, че равновесието на Неш да бъде съвместимо с "общото благо". Трудността е да се настроят стимулите за правителството. Това е голям проблем в човешкото общество. Как може да съставиш и организираш правителството, което да работи за "общото благо". 49:03

**[Бележка\* Тош:** Сравни с въведенията от основните трудове от "Вселена и Разум" и "Анализ на смисъла на изречение ...", 2004. Тук бих добавил, че понятието "общо благо" е зле дефинирано за отделните агенти, части, управляващи устройства в системата, особено предвид че "правителството" - въображаемото, получаващо се като страничен ефект от взаимодействие между частите като възникващо явление и поведение (виж "Emergence") или просто конкретно материализирано управляващо устройство от по-горно ниво - например висшестоящ в човешка йерархия, началник, законодателен орган, сили на реда, войска, олигархия или едър бизнес, организирана престъпност и пр., влиятелни големи богати медии - притежават власт, с която могат целенасочено да променят "възнаграждението", "отплатата" и пр. на частите, агентите, деятелите, субектите, обектите, елементите от по-ниско ниво, така че да ги насочват към определено поведение "за общото благо" (други биха казали, че и частите и по-нисшите нива също влияят на по-висшите и ги определят, и също е вярно, и въпросът за прецизното и отговарящо на действителността описание, представяне и моделиране на взаимодействието между нивата на управление и обхватите е ключов въпрос за Разума и Вселената; той се разглежда и например от кръга около Майкъл Левин, виж дискусиите в Ютюб канала му;).

"Общото благо" е въображаема функция, която се определя от съчетанието на взаимодействието на всички части в системата в „очите“ на оценител-наблюдател, който си е избрал какво и как да смята за „благо“, с присъщите при това грешки дори и в собствения

му модел. Оценителят-наблюдател предполага някакво „общо благо“ в някакъв момент по някакви правила, както върши и например деятелят, управляващо-причиняващото устройство, само-за-себе-си, когато предвижда бъдещето и планира действията си в избран период на пред и т.н. (Виж „Анализ на смисъла на изречение...“, 2004). Обаче когато системата се променя, а тя според мярката може да се променя непрекъснато, за да възстанови равновесието към "общото благо", тя променя и предпоставките за следващите състояния, поема по друг път, друга траектория и вече новата структура, новото състояние може да не е съвместимо със старото и оценките, плановите, изчисляването на "благо", "удоволствието" може да са други и да отменят старите. Това е обяснено в "Анализ на смисъла.", 2004, и обобщено с това, че няма една-единствена "обективна" и "безспорна", може би „местно определена“ или „обективна“, най-добра награда или поведенческа траектория, последователност от решения и действия, за сложен многослоен и многоелементен самостоятелен деятел, подобен на човек, който непрекъснато предвижда бъдещето с различна разделителна способност на възприятието, причинява или се опитва да променя и причинява бъдещето също с различна разделителна способност; предвижданията му, освен това, се извършват успоредно с различни методи за разделяне на сетивно-моторните данни и наградите и различни функции за изчислението им, както различни във времето (в зависимост от продължителността на периода на предвиждането: секунда, минута, час, ден, година, ... до точно определени гранични моменти, събития и т.н.), така и с различни предпоставки, нагласи, "валентности" ако цитирам термина на Бах, или, математически и най-кратко: различни функции за изчисление на сбора на очакваната награда за определен период напред.

За да има такава "оптимална награда", моделът му трябва да се замрази, "фиксира" по един или друг начин, както се извършва при пресмятането на частни производни: останалите променливи се приемат за константи. В действителността обаче параметрите, описващи човека, сложна самостоятелна (разглеждана така) многослойна, многоразделителноспособна, многообхватна, многомоделна мислеща машина, тези свойства не са константи във времето и изборът на един път може да промени оценката със задна дата на другите, за агента може да е невъзможно да се върне и да тръгне по другия път „при запазени други състояния“; при

човека е характерно оправдаването на решенията (тук може да се намесват и непознавателни мотиви, които са "емоционални", но те винаги по един или друг начин работят при човека като движещи сили, тласъци ниско ниво; виж следващите бележки към лекции на Бах, в които се разглежда мотивацията).

"Най-доброто за човечеството" като цяло също е спорно, зле дефинирано и неясно понятие, защото няма обективно и единно определение за това какво е "човечеството" и какво "трябва" да бъде. Някои групи, организации, личности, се опитват да влизат в ролята на „обективни“ наблюдатели-оценители и управляващо-причиняващи устройства, които да решат и да наложат тяхната представа за „спасението на човечеството“ – такива са религиозните водачи и религиите, политическите движения -изми, особено глобистките от целия политически спектър. Както беше посочено по-горе обаче определени действия или промени в определена посока променят и "човечеството" (и каквото и да е управлявано явление, устройство, деятел/агент), а управляващите устройства от по-горно ниво могат да променят и състоянието и оценките на подчинените устройства и т.н., в същото време са налице странични ефекти, или „нежелани разлики“, и грешки и непълно предвиждане, които изменят и управляващото устройство от най-високо ниво – онова, което то си мисли, че е, и пр.

Що се отнася до „най-доброто за човечеството“ - има насоки, които изглеждат „очевидни“, но очевидното и „простото“ е само в отвлечени понятия, например всички да се развиват, всички да живеят "по-добре", да са "щастливи", да са нахранени, да се учат, да не бъдат потискани и т.н. но всички тези общи понятия, при конкретизирането им се сблъскват със споменатото по-горе неспазване на "равновесието на Неш", което изисква "да се промени отплатата", и след като човек е нахранен, започва да иска нещо повече, различните индивиди, групи, страни и т.н. искат различни неща от дадените им и т.н.

Едни от въпросните „-исти“, които искат да наложат или да представят „доброто“ за човечеството са т.нар. „контрольори“, ако използваме шеговита дума, или „Съгласуватели на ИИ“, „AI Aligners“, „AI Alignment“. Те искат да се наложат регулации върху развитието на ИИ, за да не ни унищожил.

\* Виж също беседата на Ирина ... и Майкъл Левин ...

**Йоша Бах:** (...) **57:30** Законите на роботиката - да бъде роб ...  
[Да. Сравни с Тош, 2002 и също „Писма до Оксфорд...“, 2012]

1:06 ч ... Данни от осезанието, преди раждането ... Когато докосвате собственото си тяло, точките, рецепторите, които са по-близо една до друга, ще бъдат докосвани по-често заедно и дори само от статистиката, броя усещания, може да се разбере, че те са близо една до друга и така да се построи пространствен модел на тялото. Когато започнете да докосвате средата, може да нормализирате тези телесни повърхности чрез разликите в плътността на сензорите - например в езика са много, а на гърба са малко. Площта на "картата" на рецепторите за докосване на гърба [от първо ниво] е много по-малка отколкото площта за езика [или за върховете на пръстите]. Затова трябва да нормализирате тези стойности с втора карта/изображение, което може да извършите чрез статистика върху предметите, които докосвате и как се движат по тялото ви. Моделирането на това пространство започва с посоките горе и долу - чрез вестибуларния апарат, сетивото за равновесие. (...) На определено ниво на съответствие започвате да "обърквате", да смесвате модалностите и да възприемате света на повърхностите, които докосвате, като същия като света на повърхностите, които виждате.

1:08: Така разбирате, че може да се движите в пространството. Отношенията между всички тези видими "мехурчета" [в пространството], които може да видите, а "алоцентрично" пространство, т.е. вече не е "егоцентрично". Пораждате нов видим "балон", който може да докоснете във всеки момент. Предметите в света трябва да бъдат разделени, определени, сегментирани, за да може да бъдат описвани. Затова вместо да се отнасяте към света като една голяма система, която е вектор на състоянието, който се променя, я разделяме на много независими системи, като всяка от тях има собствен вектор на състоянието и функции за преход и те си влияят една на друга. Влиянието между различните предмети наричаме "причинност". Причинността е следствие на разделянето на света на независими предмети. Начинът по който се обръщаме към предметите са понятията. Понятията са адреси в пространството от предмети, неща, вещи, обекти.

**[Бележка. Тош:** Сравни "Истината"; и например откъса за ябълката, която Машината вижда и може да опише точно усещането си. Също с "Телесността е просто ...", 2011 и примера с това как разбираме какво е "мокра повърхност" - Sensory-motor ... и въобще с ТРИВ и др.]



## **\* Наистина ли съществува „трудният проблем“ на съзнанието? – Йоша Бах, изследовател в ИИ**

Is There Really a Hard Problem of Consciousness? - Joscha Bach,  
Artificial Intelligence Researcher

**The DemystifySci Podcast**

5,49 хил. Абонати 9105 показвания 11.06.2023 г. (към 28.6.2023)

**Водещи:** "Майкъл Шайло е д-р от Колумбийския университет, изучавал еластичността на водните молекули" ... Анастасия е д-р по биоелектричество в същия университет. (Shilo, Anastasia)

*29:xx мин: Анастасия и Майкъл Шайло обясняват за преобразуващите събития, че когато взимаш подобно решение, след като се промениш, не знаеш какви ще бъдат бъдещите ти стремежи, ценности и пр., затова винаги в известен смисъл „хвърляш зарове“, за да разбереш дали си струва да бъдеш някой друг, различен от самоличността ти в момента на избора. Например при жените, когато имат деца – сестра ѝ, която работила в Гугъл около 10 г., решила да има деца, била е ужасена, че ще бъде център на света за тях, и че ще загуби други страни от самличността си. Обаче след раждането на детето ѝ е открила, че майчинството е по-добре от каквато и да било кариера.*

**31:XX Йоша Бах:** грижата за следващото поколение е с огромно значение (...) Относно избора или съдбата на някои творци и др. да нямат деца, че по този начин те вече не участвали в бъдещето на живота.

**33:xx: Шайло:** според някои тълкувания, родителството попречва на постигането на духовно просветление ...

**Йоша Бах:** Да нямаш /биологични/ деца е духовно самоубийство, защото повече не участваш в бъдещето на живота. Вече ще си само разказ, идея, как допринасяш за случването на бъдещето, но повече не си част от него.

**Шайло:** Предполагам, че художниците /творците, хората на изкуството, artists/ са рядка част от обществото, които преобразуват бъдещите поколения без да пораждат нови живи създания.

**Йоша:** Според мен стойността ти като творец е толкова голяма, колкото е по-висока оценката на публиката, но може би тази гледна точка е свързана с вида семейство, в което съм израсъл /баща му е бил художник/.

**Тодор:** мнението на Йоша е от гледна точка на *тази негова* представа за „аз“: от друга обаче детето вече е *друго* същество, *друг* организъм, и това че „*ти*“ продължаваш да съществуващ чрез него е част от твоето въображение и желание: *само докато си още жив*. След като умреш, спреш да си „жив“, вече не съществуващ, поне като съзнание или като цяло, според описанията на Й.Б., защото вече не си *съгласувано цяло*, нямаш спомени за себе си пр.

А от друга гледна точка, бих казал – моята, – всъщност всички въздействия, атоми, молекули оказват влияние на бъдещето<sup>21</sup> с различна степен, сила, в различни нишки – мярката зависи от наблюдателя-оценител. Човекът например е като атрактор, който привлича като магнит тези съставки, те се събират в него и се сглобяват\* и т.н. Обикновено човек разглежда себе си в прекалено кратка околност, обхват, рязко отделена. А всъщност, ако се обхване по-голяма област, частиците, които се включват в телата ни в даден момент, могат да се приемат че \*са били част от нас\* и в по-ранния момент? Аналогично на яйцеклетката и сперматозоида. Обикновено се приема, че животът на зародиша, индивида, започва от оплодената яйцеклетка. Но част от тези обекти, които си избираме теоретично (понякога и практично - при зачеване "инвитро"), миг преди оплождането са били отделни, а по-назад във времето яйцеклетката е съществувала през всичките години назад за майката ... При мъжа сперматозоидите се произвеждат, те са съществували по-отскоро преди зачеването, но са били преди това ...

\* Относно човека като отделно същество, което привлича частиците, преди години съм правил мислен опит за това кой кого управлява в отношението автомобил-човек или пък смартфон-човек, компютър-човек. Приемаме, че е човекът и може да се дадат редица оправдания – „сложност“, „воля“, „самосъзнание/съзнание“, „цели“ и т.н., но ако обърнем гледната точка, може да си представим, че колата „привлича“ човека и той е нейният инструмент, чрез който да постигне своите „цели“, да стигне някъде си и пр. На пръв поглед – фантазмагория, но ако колата и шофьорът са буквачета от по-отвлечена система, в която целта е

---

<sup>21</sup> Замислих се за този въпрос в конкретния момент, когато мислех за това колко и каква храна изядем, за да съществуваме, как всички тези атоми се включват в телата ни, откъде идват в минал момент и т.н., във връзка със самонаблюдението-изследване, което правя върху себе си, на практика през целия живот, в последната година: подробно върху храненето.

машината да се намира на местата на които се намира във всички моменти и т.н., то за целта на това по-абстрактно устройство са важни и шофьорът, и „неодушевления предмет“: всичко. Всъщност тази идея е разглеждана и в класическата ТРИВ, която има елементи на „екстернализъм“ и холизъм. Ключът не може да отвори врата или да позволи да се влезе в стая, която не съществува, а обемът на информацията на стаята е по-голям от на ключа. Също както нищожните битове, които съзнателно „управляваме“ не могат да се случват без невъобразимия обем други „битове“, които съществуват преди нас.

\* Сравни М.Левин: "Cognitive light cone", "Scale-Free cognition", "Computational Boundary of a Self", ... – see *Universe and Mind 6*; виж други бележки за Майкъл Левин (Michael Levin)

**Йоша: 44:00** Донякъде ИИ като област, основана от М.Мински и пр. е продължение на философския проект на Лайбниц, Фреге и Витгенщайн за разбиране на и построяване на друг ум, чрез свързване на математиката и смисъла/значението, чрез автоматизиране на структура, която разлага смисъла/значението на части, които могат да бъдат пресмятани. Осъзнах, през 20-ти век се осъществява „остър завой“ към конструктивизма ... който превръща математиката в изчисление; /преминава се/ към математика, която може да е действителна, която може да бъде осъществена във физическата вселена; тя се описва чрез преминаване от стъпка в стъпка /състояние в състояние/. И тези преходи от състояние в състояние по същество са изчисление. Физиката всъщност е изчислителна наука. Вселената е изчислителна система. Всичко, което е динамично и определено (детерминистично) може да бъде описано като изчислителна система и недетерминистичните „неща“ не са по-различни, просто се изтриват някои от битовете, които сте пресметнали по-рано, ако моделът е бил детерминистичен. За мен „всичко“ започна като бях дете, седейки пред моя „Комодор 64“, когато осъзнах, че бих могъл да поставя всичко, което си представя, „зад“ екрана, стига да го разбирах достатъчно добре: дори целия свят.

**Тодор:** Сравни с „Теория на Разума и Вселената“, 2001-2004, „Вселената сметач“, „Свещеният сметач“ и пр. Гледната точка към проект за ума се споделя и от С.Волфрам в „Математика“ и „Волфрам Алфа“, както и от Т.А.

**45:xx: Йоша:** "Взаимодействах си с университетската система и научните институти достатъчно дълго, за да разбера, че няма да ми платят, за да създам ум (mind); поне не и *тогава*\*. Сега положението е по-добро, защото ИИ е станал много по-модерен. ... /Вече/ много хора биха инвестирали в нишови области на ИИ и философският въпрос, който винаги е бил изключително нишова тема, защото е толкова рискован, бидейки философски проект, че има висока вероятност да не успееш да го завършиш в рамките на живота си. Това не беше материя, за която щяха да ми плащат да изследвам, докато бях в системата на Висшето образование\*.

**Бележка:** \* Й.Б. е работил в Харвард и в прословутата лаборатория за медии на МТИ само до преди няколко години. Според профила му в Линкдин:

\* Research scientist § Harvard Program for Evolutionary Dynamics: Nov 2016 - Feb 2019 · 2 yrs 4 mos

\* Research Scientist § MIT Media LabMIT Media Lab  
Apr 2014 - Oct 2016 · 2 yrs 7 mosApr 2014 - Oct 2016 · 2 yrs 7 mos  
Cambridge, MA

<https://www.linkedin.com/in/joschabach/details/experience/>

<https://www.media.mit.edu/events/media-lab-speakers-tedxbeaconst/>

**52:14: Йоша:** Физиката се ограничава до онази част от действителността, която може да се опише с кратки алгебрични уравнения<sup>22</sup>. Физиците пренебрегват всичко останало, което не може да се побере в тях и го наричат химия, биология и оставят другите да се занимават с него. Физиката е особен вид моделиране на света, най-вече чрез геометрията.

**52:41: Шайло:** Така е, но едва в последните стотина години физиката станала чисто математическа наука, а преди това е била разбрана като част от механиката: /отнасяла се е за/ тела, които си взаимодействат с други тела; бутат ги или ги дърпат; мисля, че в известен смисъл това [математическото разглеждане?] е много по-смисления поглед върху физиката, защото днес тя се опитва да поеме толкова много метафизични проблеми, когато си говорим за вселената и за съдбата ѝ (...) но не знам

---

<sup>22</sup> Сравни с ТРИВ, „Писма между 18-годишния...“ и по-късни произведения за разпространената представата за „наука“ или „ред“ само като кратки прости формули от няколко символа. Виж също „Окончателен ИИ (Ultimate AI): Принципът за Свободна Енергия и Предсказващото кодиране, сравнени с [работи на] Тодор и КогАлг – Познавателния алгоритъм – дискусия от форума Montreal.AI и Artificial Mind“ в книгата и <http://artificial-mind.blogspot.com/2018/12/ultimate-ai-free-energy-principle-and.html>

дали математиката има средствата, с които да отговори на космологичните въпроси за предназначението или смисъла на бъдещето.

**58:40: Йоша:** .. Като дете на своя "Комодор 64" през 1983-1984 г. сам разработих „фотографска“ програма ... написах програми, които могат да рисуват черти, елипси, да ги съчетава: това бяха първите ми графични програми<sup>23</sup>. И за да успея, трябваше да изведа математиката на геометрията и не съзнавах, че го правя. Мислех си, че математиците си имат таен по-елегантен начин, по който го правят.

[Естеството на работата с] "Commodore 64" също така дава идея за това как [математиката и геометрията] "се случват" в процесора; /така/ осъзнах, че цялата математика се случва "там" и всичката Булева логика и простите аритметични действия: събиране, изваждане, умножение и деление всъщност са хеш функция, която взима входен модел-схема и го съотнася към изходен модел-схема по "обикновен" начин (regular way) с прости схеми; и така осъзнах, че е възможно да се построи цялата математика, включително геометрията и всичко видимо в света и да се покаже на екрана в компютърна игра... Динамични обекти, които взаимодействат с други помежду си; агенти, които разсъждават за онова, което виждат в този свят и как взаимодействат с нас... Всичко, което сега виждаме в компютърните игри с увеличаваща се разделителна способност е построено от тези малки прости автомати; така разбираш, че всичко в математиката е изградено по този начин, започва с крайни автомати и като използва подходящо множество от такива крайни автомати в аксиоматична система, би могъл да построиш всичко, което можеш да си представиш.

**1:00:18:** Според мен може би няма нищо, което да не може да въведеш в сметача. Не мога да си представя какво би могло да бъде такова нещо. Няма нищо във физиката, което да не може да се пресъздаде в изчислителна машина. /Наистина/, част от явленията, като например квантова физика, няма да работят ефективно на /класически/ компютър ... защото една от предпоставките на квантовата механика е, че Вселената, тълкувана като изградена от частици по начина, по който я разглеждаме и възприемаме е неефективно осъществена върху този носител - изчислителната машина.

[**Том:** виж също Йоша Бах в участието му в "Познаваемата вселена", сравни и с ТРИВ]

---

<sup>23</sup> Сравни с Хронологията на Т.А., 1998-... 3D-графика за Правец-8М и „видео“ и пр.

**1:19-1:20 ч:** ... Мисля, че повечето клетки нямат собствен модел, модел на собственото си устройство /и са „механични“; „нямат воля“/.

**Шайло:** Мислиш, че клетките са огромни PID-регулатори \*?

**Бележка:** Или PID-контролери, пропорционално-интегрално-диференциален регулатор<sup>24</sup> е вид управляващо устройство с обратна връзка (ОВ)\*. Система за поддържане на управляван параметър в зададени граници, PID идва от пропорционален, интегрален и диференциален в математически смисъл – „пропорцията“ идва от съотношение (деление), интеграл като сума от моментни стойности, диференциал – производна\*. Всъщност принципите на работа и целите на това устройства са като част от принципите на разума и вселената от ТРИВ: Физическия/чувствен вид „награда“ на УУ е измереното състояние да съвпада с желаното, УУ да бъде възможно най-близо до желаното състояние. PID-регулатори се използват от около 100 години. По-прост вид УУ с ОВ е термостатът, в парните машини се използват „говернатори“.

Когато Ю.Шмидхубер подробно критикува Я.Льока за „новостта“ на теорията му за автономен ИИ през лятото на 2022 г., и че не е цитирал приноса на групата на пъвия в тази област, вторият се оправда, че „тези идеи били още от Тоерията на управлението“ (Control Theory), част от която са работата на тези устройства.

**Йоша:** Клетките не са достатъчно „сдвоени“ със средата, за да могат да представят в понятия собствената си "одушевеност" от гледната точка на наблюдател\*. Клетката може да извършва повечето си действия без да построява изричен модел за тях.

/\* sufficiently coupled to their environment; make it necessary for conceptualizing - "в понятия"; а може и „да си представят“; agency – одушевеност \*/

**Тодор:** Същото важи и може да се каже и за човека. "Моделът", който някой си мисли, че има за себе си, е много по-прост и непълен, и пряко, относително, косвено **неверен** спрямо онова, което се случва, при достатъчно висока РСВ. Виж дори ЧиММ, 2001 или СВП2, 2002, СВП3, 2003 и въобще ТРИВ.

**1:20:50: Шайло:** Т.е. те работят изцяло по парадигмата за външна воля?

**Йоша:** Не, волята не взема задължително участие, те са просто механични, като термостат.

---

<sup>24</sup> <https://www.engineering-review.bg/bg/pid-regulirane/2/552/>  
[https://en.wikipedia.org/wiki/PID\\_controller](https://en.wikipedia.org/wiki/PID_controller)

**Тодор:** От тук следва, че според Й. волята не е вездесъща, и не е свързана с Волята (Шопенхауер), както е в ТРИВ, където всяка частица „нещо“, което има свойства и да причинява промени, е управляващо-причиняващо устройство. Всяка има някакъв вид и степен „воля“, самостоятелност и причиняваща сила (agency). Подобни са убежденията на М.Левин с неговите модели и понятия TAME (Technological Approach to Mind Everywhere) – почти навсякъде могат да се открият някакви елементи на ум, в различни пространства и начини на изразяване; в живите организми има ум не само в мозъка, а и в клетките, органите, в т.нар. „морфопространство“ и пр. (сравни с йерархията от ТРИВ и лекцията от „Принципи на Разума“, 2009), Cognitive Light Cone (Конус на познавателните способности – по-развитите умове въздействат, искат да причинят промени, да повлияят; „загрижени са“ (to care) за по-голям обхват от времепространство и така може да се измери този вид техни умствени способности); Scale-Free Cognition (Познавателни способности във всички обхвати на времепространството – сравни с УУ в ТРИВ). Виж бележки във „Вселена и Разум 6“ (Universe and Mind 6) и в др. приложение статията „Causal Blankets“. Артур Шопенхауер също говори за Воля за живот, която е на всички нива на вселената и се „обективира“ по различен начин. Човекът е висша форма на обективация на Волята. В моите термини човек е висша форма на физични закони/сили, мощна/абстрактна/съсредоточена/сложна проява на физичните сили.

(...)

~ **1:28: Йоша:** Да са сдвоени означава да имаш много видове сетива, които работят в много модалности, достатъчно надеждно, с голямо излишество, върху големи разстояния и могат да различат смислени разлики в света, които /системата/ би искала да обедини. (...)

**1:31...: Шайло** композирал на китара и пиано ... Понякога идеите му идвали изведнъж, готови. Като общувал с хора усещал, че се пораждат идеи от взаимодействието с друг човек, но *не и когато взаимодействал с компютър*.

**Тодор:** Според мен това твърдение е необосновано и неубедително и в най-добрия случай отразява единичен случай и вид каприз; например – с общуване с кого най-великите композитори, които са изброили и изчерпали най-добрите или най-леснодостъпните хармонии са се вдъхновявали, с общуване с кого, при положение

че те са били най-способните и са измисляли и изпълнявали произведения, които другите не са могли? Подобно говорене като на Ш. вероятно идва от чувството за превъзходство над машината, *"свидетелството за душепритежание"* в ТРИВ. Според мен естествено, че от взаимодействието с машината също се пораждат идеи: от взаимодействието *с музикалния инструмент (дори да е най-просто устроен), с листа с ноти* също се пораждат идеи, но наблюдателят е решил, че „това не се брои“; самите сметачни машини обуславят и правят много от тези замисли възможни, не само в областта на програмирането и пр., а например в редактирането на музика с програми като „Cubase“, „Audacity“ и пр. ние непрекъснато използваме сметачи, чрез които общуваме и с други хора, и пак машини са посредници, които го правят възможно.

**Анастасия:** Не можела да проследи откъде се раждат идеите, не можело да се програмира, защото ...

**Тодор:** Защо да не може? Може би *тя* не може. Естествено, че може, но трябва достатъчно сложен и богат съответен модел – както се случва например в невронните мрежи, но в представяния и преобразувания, които са прекалено „нечовешки“ и им липсва видна понятийна структура. И отново към Бах бих припомнил, че моделът на себе си не е достатъчно точен и т.н., включително за „съзнателните“ същества като човека.

**1:33: Йоша:** ... За него творчество е тогава, когато има „скок“, рязка промяна при откритието /изненада, прекъсване/

**Тош:** Това е относително – как точно се определя „скока“ и мярката за него, къде е границата, еднаква ли; защо, кога, как и пр. – и скокът, и изненадата, и прекъсването са относителни и зависят от оценителя и моментното му състояние – докато не се приведат в обща, достатъчно широка всеобхватна познавателно-оценителна схема.

Някои биха определили тези моменти като „възникване“ (emergence), но процесът на търсене може да е и често е „нетворчески“, изследване и изброяване на възможности (каквото е всяко търсене) и да доведе до достигане до „особено“ и „неочаквано“, „непредвидено“ състояние, „по-неочаквано от очакваното“ (виж ТРИВ, „Творчеството е подражание на ниво алгоритми“, „Истината“, „Ада“). Такова е например при „възникването“ в големите езикови модели. Основните операции,



които се извършват, са едни и същи. Скокът, пречупването е в стойността на мярката за „загуба“, сравнението с образците, но като действия на ниско ниво процесът е един и същ. Последното може да се приложи и за всичко във Вселената при определено обобщение: всичко е просто взаимодействие на атоми и молекули, или „само нервни импулси“, „клетки“ и т.н. (виж ЧиММ, 2001; Истината, 2002). Творчество, „ново“, „изненада“ в такъв обслов са *оценки* на наблюдател-оценител, според неговя опит, представи и т.н. Този размисъл ми припомня една *погрешна заблуждаваща логика*, която се внушава в лъжемъдростта, че било грешка и глупаво „да очакваш различен резултат с едно и също действие“. Как се сече дърво? Как се отвива болт, гайка? Как се реже? Как се копае дупка или се пробива сонда до достигане на водоизточник? Как се развива координацията, уменията за свирене на музикални инструменти и т.н.? Как се сече дърво: удряш определен брой пъти по, може да се каже, „един и същи начин“, „на едно и също място“ и накрая дървото „пада“, преди това „не пада“, т.е. „няма резултат“, ако е разгледан толкова грубо. Или обикновената мъдрост „капка по капка – вир става“ или че водата дълбае или изглажда скалата не със сила, а с упоритост. Прекъсването, изненадата, също може да бъде в дадена мярка, степен, а не само двоично. Може да бъде и двоично в дадена ниска разделителна способност. Може би някои биха нарекли последното „принципно ново“, „radical novelty“, но то също е относително. За онзи, който разбира пътя на развитие на дадена научна област, онова което следва е ясно и логично, а за друг, който е нов и който не разбира, не може да предвижда бъдещите възприятия с такава точност и обхват, „всичко е изненадващо“ и „принципно различно“ от предходното.

(...)

**2:24: Анастасия:** Не можем да съгласуваме корпорациите. Те вече са системи с ИИ, които са по-прости от онези, които строим сега. Така че би трябвало да можем да ги съгласуваме /с нашите цели/.

**Тош:** Да. Обаче кои сме "ние". Всъщност Вселената и изчислението на формулата за „всичко“ определя кой какво съгласува, тя съгласува всичко.

**2:44:xx: Йоша:** За бъдеща технология с ИИ, "хипер Гея" /богинята на Земята/, която да показва всички бъдещи последици от действията ни, за

всичко за което ни е грижа, така че да не може да избягаме от отговорност.

**Тош:** Сравни с идеята за "Причинностни белези", в [Universe and Mind 6: Вселена и Разум 6. github.Theory of universe and mind; Causal IDs; Why infinity doesn't exist? (...)]. Причинностните белези могат да се свържат и с твърдението на Й. за „неучастието“ на човек в живота в бъдеще, ако не остави биологично поколение. Ако има подобни „причинностни белези“, вериги от причинно-следствени връзки, които са особени в клетки, и те действително запазват „духовна“ връзка с „теб“, което си бил (въпреки че в следващото поколение се получават и промени на гени и пр.), може и да е така. Но ако п.б. са по-общи и се отнасят за всевъзможни взаимодействия и връзки между познатите ни частици на ниво атоми, то тогава излиза, че според Й. човек участва в „бъдещето на живота“ и планетата дори и да не е оставил живо поколение.

4:17: А: Функция за полезност ... в тялото ...

**Ш:** И спрямо Вселената. ...

**Йоша** не е съгласен ...

4:20 да обясним как физическа система произвежда сънища ...

**Тош:** Но кой ги гледа? Щом се различават сънища от "друго", другото е "реалността" ли? Т.е. ако всичко е сънища, значи в съня си представяш че има нещо "по-твърдо".

**Йоша:** ... "причиняващ агент" (сравни ТРИВ: "управляващо устройство", "управляващо-причиняващо устройство" (causality-control unit) ... който може да съобщи/докладва чувства ...

Това че бактерията не може да докладва чувствата си предполага, че вероятно няма чувства. Докато фактът, че хората съобщават чувствата си носи със себе си бремето на необходимостта да обясним тези явления.

**Тош:** *Не е вярно, че "бактерията не може да докладва чувствата си":* виж дори ЧиММ, 2001. Кой е съдникът, който решава кое е "докладване на чувства" и с какъв точно протокол се прави? Същото е възможно и при животни и човеци, за които се приема, че "иначе" имат чувства. Има хора с нарушения в усещането на болка, включително вродени: аналгезия, вродена нечувствителност към болка и пр., мозъчен инсулт и други увреждания с подобни

последници, получени от наранявания, атрофия и други причини. А дори и за онези агенти, дейатели, личности, същества, за които *приемаме*, че чувстват и изразяват *нормално* „чувствата“ или мислите си, „*злонамерен*“ оценител може да реши да *пренебрегне* приетото по-рано допускане и *нарочно* да не възприема признаците на „чувства“, болка, мисли и т.н. като такива; способен е да нарочи другия за „подчовек“, „нечовек“, „орк“, зомби“ и да се отнася с него като с вещь или както си иска. Това се случва постоянно във войни, при геноциди, престъпления, в състояния на умопомрачение и пр. – т.е. когато част от ума не работи така, както сме свикнали, очакваме; тази част е част от познавателната система, мозъка, организма и пр., която приемаме за необходима за „нормален“, обичаен, *очакван*, желан режим на работа.

А бактерията в частност може да се приеме, че изразява „чувствата“ си, на същия принцип като човека, като определени промени в състоянието ѝ, поведението ѝ (отдръпване от определено място, отделяне на определени вещества и пр.) биват наречени като изразяващи „болка“, „страдание“ и пр., и съответно други: когато се движи по градиента на хранителни вещества, привлича се: „удоволствие“, търсене на възнаграждаващ стимул и пр.

Това че някой не може да *"каже"* какво изпитва по *твоя* протокол не е доказателство за нищо, освен ако не се приемеш за вездесъщ законодател оценител-наблюдател. Например по същия начин може да си представим незрящ човек, срещу когото някой друг се усмихва, но не издава звук или не го показва по друг начин, който е приет и разбираем от незрящия човек. Във възприятието на слепия другият също *"не може да изрази чувствата си"*. И т.н.

**Йоша:** 4:24: Съзнанието е проекция, свойство за представяне, на система, която съзнава, че съзнава. (Consciousness is a projection, it is a representational property of a system that is aware that it is aware)

### 1.7.2023:

**Тодор:** Циклично определение. Какво значи да съзнава? И кое е нещото, което съзнава? Значи според него ако нещо е "система", то съзнава. Също, моделът на системата за собственото устройство, съзнание и каквото и да било е по определение и необходимост по-прост от въображаемия, възможния, подробен модел с пълна

разделителна способност, който описва системата на "машинния език на Вселената" на най-ниско ниво за нея. Й.Б. се съгласява с това напр. в примера за това защо квантовите компютри са много по-бързи от "класическите" в някои изчисления - защото те са построени върху много слоеве над квантовите - виж ТРИВ, нива на въображаеми вселени, Абстрактна теория за изключенията от правилата и пр. По същия начин "представата"\* на даден ум, "съзнание", система, агент за него самия, е само част от неговото цяло устройство, б както той е само част от цялата Вселена и съответно е по-прост, непълен и т.н. Това е ясно и по определение, че частта е част от цялото, че очите не могат да се видят сами без огледало и т.н. - това е описано в ТРИВ.

\* Представата или представянето, като конкретен обем данни, измерен в битове; а като усещане може да не може да се локализира и да изисква цялата вселена. "Представата" може да включва субективното усещане, чувството, докато представянето може да бъде и напр. в изчислителна машина, за която се приема, че няма "съзнание" от човешки тип, но също може да има "рефлексия" в смисъл данни, модели за собственото си устройство. Последното обаче може да се докаже или да се "реши" по различни начини - например чрез проследяване на поведението, че системата правилно планира бъдещето си и го предсказва и предизвиква, причинява; но не бива да забравяме наблюдателя-оценител, който решава, че "това" е системата, къде е границата ѝ, кое е представяне на еди-какво си и т.н.

## \* Вектори на Познавателния ИИ:

### Подбуди и Автономност

## \* Vectors of Cognitive AI: Motivation and Autonomy

Cognitive AI

1,76 хил. абонати, 28.12.2021 г. | 2864 показвания около 1.3.2023;

3,2 хил. 23.10.2023 [непопулярен канал]

<https://youtu.be/0CyJV7manUw?t=2670>

Подбор, превод и бележки на Тодор Арnaudов – Тош (ТА, Т., ТА):

Как бихме могли да разберем в понятия и да построим изкуствени агенти със силно развита самостоятелност?\* Как може да използваме изчислителни модели, за да разберем „одушевеността“ на човека и да изградим сътрудничество между човека и изкуствените агенти с ИИ? Това предаване събира група от мислители, работещи в областите на изкуствената одушевеност, моделиране на подбудите (мотивацията), чувствата и общителността, за да обсъдят как вътрешноприсъщите подбуди пораждат целенасоченото поведение, устройството на познавателната структура, мултиагентното сътрудничество и етиката.

#### Лекции:

1. Cristiano Castelfranchi: Обосноваване на общителността в Теория за целите (Grounding Sociality in Goal Theory)
2. Christian Balkenius: Мотивация, емоции и внимание (Motivation, Emotion, and Attention)
3. Dietrich Dörner: Мотивация на компетенциите (The Competence Motivation)
4. Joscha Bach: Мотивацията за единична и обществена „одушевеност“<sup>25</sup> (Motivation for individual and collective agency)

#### Видове цели по Кастелфранки:

*Desire* – Желание | *Need* – Нужда | *Intention* - Намерение

*Purpose/Objective* – Предназначение

*Project* – Проект | *Hope* – Надежда | *Expectation* – Очакване

*Duty* – Дълг – норми, *Norm-sensitive agents (Deontic mind)* – способност за разпознаване на норма от лична молба, услуга, размяна, използване/употреба/навик. Нормата е такава в ума, само ако решим

---

<sup>25</sup> одушевеност – agency; от „agent“ – „който може да действа съобразно собствената си воля“; „Свещеният сметач“ и юношкото наречие „agent“ се превежда също като „дейтел“; съответно е възможно „дейтелност“.

*да я нарушим: да се подчиним или не.*

## И другите участници

\* Умът има природа на хибрид: разумен и почувстван (reason & felt)

1 – доводи (arguments), убеждения (beliefs), обяснени, разисквани

2. наглед, „intuitive”; felt – почувстван – affective evocation, “somatic markers”

**Тош:** Сравни ТРИВ, познавателни и чувствени/физически предсказващи модели, виж лекцията от курса по УИР, 2010-2011

Сътрудничество: прозрачно – да обясни план, мотиви

**Т:** ?Щ → Щт план на въображаем деятел, последователност от вери, тласъци, действия, преобразувания

**Самостоятелност:** цел, избор, истинско решение, инициатива – без поискване да предложи помощ, да помогне при нужда ... цели-планове-възприятия → способност ... Поведенчески мрежи (Maes, 1989) ...

Предусещащи класификатори (Anticipatory classifiers, Butz, 2000)

Самостоятелност на целите –14 мин – да ма собствени цели, а не да бъде просто покорен изпълнител; сам да решава дали да приеме целта на друг (да помогне и т.н.)

**Тош:** т.е. агентът трябва да може понякога да откаже, и за да е сигурно като показател, значи в някои действителни случаи е отказвал. Обаче какво точно значи „сам“ и „самостоятелно“? Къде е границата на „самото“? Тя трябва да е определена от оценител като вид причинно-следствени връзки – и по начин на измерване, и по обхват във времето и пространството; причинно-следствена област във времето, пространството, силите и разделителната способност на възприятие и управление. Ако се разшири или промени обхватът на разглеждане или прецизността на измерване и пр., то решение, което в един случай се смята за взето независимо или самостоятелно, ще се окаже повлияно или причинено от определени „външни“ влияния, въздействия, събития, вещества и пр. от предишно време и място, с ефекта, който са оказали върху даденото управляващо-причиняващо устройство. Същото важи и за частния случай на деец човек. Виж „*Нужни ли са смъртни изчисления за създаване на универсални мислещи машини*“, Т.Арnaudов, 2025; в това приложение „Ирина“ на други места в текста за представянето и кодирането на агентите по разпръснат начин във времето и пространството като полета; виж също „*Вселена и Разум 6*“, Т.Арnaudов,

2025, която разглежда и повдига въпроса и за времевия лаг между частите и мащабите, как те разбират дали са част от цяло и мн. др.

Познавателна архитектура: представяне(убеждения, цели) + обработка и взаимодействие. Обработката на цели е основана на убеждения от възможни цели до намерения и действия.

\* Разлика между целеустремено и целенасочено очакване (goal-directed vs goal-oriented): 1. Само предсказване, обратна връзка към правилото на поведение. Условие – действие.

\* Само реактивно, отговор на стимул/дразнение/котекст – не е тласкано от горе надолу, кибернетично от желана представа/състояние на света ... Обединение между „Система 1 и 2“ (по Канеман)

### **Телеологията е различна от Телеономията**

- не бъркайте истинските „цели“ като умствени представяния на очакване (anticipatory representation) водещи поведението (действия, action) с *други* форми на “завършеност (*finalism*): Насочени към постигане на цели функционални поведения, но не подбудени в кибернетичен психологически смисъл. **Т: Тл.**

\* „Истински цели“ – център на ума и главна страна /функция на предусещащата/предвиждащата му природа (anticipatory nature)

\* „Наистина водени от цели : А) „функци“ Б) очаквани резултати от инструментално обучение или с подкрепление (RL) (**ТА: мишена**).  
*Класификатори и предусещащи класиф.*

**44 мин:**

**56:** <https://youtu.be/0CyJV7manUw?t=3401>

**59:56: Странични ефекти** (Т.А. Сравни с ТРИВ: СВПЗ, ВиР4)  
<https://youtu.be/0CyJV7manUw?t=3596>

**1:00 ч Вектори на умствените способности**

1. Особени за определени задачи
2. Широки и гъвкави
3. Общи: Самостоятелност, Сътрудничество, Език, Знание, Представяне, Обучение, Разсъждение, Възприятие, Телесност

## **Създаване на универсални представяния**

Преход от обучение в пакетен режим върху набор от данни, в режим на непрекъснато обучение

1. Особени за определени задачи
2. Широки и гъвкави
3. Общи: Автономност, Сътрудничество, Език, Знание, Представяне, Обучение, Разсъждение, Възприятие, Телесност

## **Степени на развитие към все по-общи модели:**

1. Самостоятелност: Определени от задачата, насочени към цел, самонасочени, самостоятелни
2. Сътрудничество: Прозрачни, Обясними, Приемащи цели, Сътруднически си
3. Език: взаимлик (интерфейс), разбиране, усвояване, език
4. Знание: препращане, обединяване, пораждање, знание
5. Представяне: ппределено от оператор, невро-символно, всеобщо, представяне
6. Обучение: без взаимодействие, с взаимодействие, самоусъвършенстване\* (offline, online, self improving)
7. Разсъждение: извод, построяване, изобретяване
8. Възприятие: разпознаване, предсказване, уеднаквяване
9. Телесност: без обратна връзка, с обратна връзка, саморазширяване (open loop, closed loop systems)

## **Общ кибернетичен модел, теория на управлението**

1. Регулираща сила с обратна връзка (Regulator (feedback loop))
  2. Предсказващо управляващо устройство (моделира бъдещето) (Predictive Controller (models future))
  3. Деятел
- 
1. Прост регулатор с обратна връзка (Regulator (feedback loop))
  2. Предсказващо управляващо устройство (моделира бъдещето) (Predictive Controller (models future))
  3. Деятел - управляващо устройство с модел на бъдещето и пораждаща подсистема за бъдещи целеви състояния (Agent - controller with future model and setpoint generator)
  4. Съзнание (sentience) - деятел, който представя/предвижда/познава



себе си (буквално - "моделира")

5. Отвъдност - връзки с дейтели от по-горно ниво.

**Бележка Тош:** Сравни понятията и нивата на управление в СВП 2, СВП 3, Вселена и Разум 4, Абстрактна теория за изключенията в изчислителните машини

### **Вниманието като деятел в многодеятелна система:**

Взаимодействие между:

1. Система на подбудите
2. Система на вниманието
3. Система на възприятията

**Подбудите:** свързват очаквани награди със сетивните особености (разпознаване) и действията върху тях (обучение/построяване)

Сходими променящи се, геометрични представяния; сходими чрез спускащи се градиенти; търсене чрез законите на случайните числа\*, възможностни\* и вероятностни връзки: действат върху системата на подбудите (stochastic search, possibilistic links)

**Система на вниманието:** параметризира състоянията и действията като използва памет с показалци (indexed memory).

**Система на възприятието:** получава стимули от средата.

Трите системи си взаимодействат.

...

**Валентност** - понятие от психологията, което:

- Определя ценността на различните действия\* за деятеля ("operations")
- Основано е на многоизмерен вектор на подбудите
- Отразява физиологични, обществени и познавателни нужди

И най-радикалното предложение идва от Карл Фристън, през 2006 г.\*<sup>26</sup> с Принципа на свободната енергия, който гласи: минимизиране на грешката в предвижданията води до откриване на структура, естетика.

---

<sup>26</sup> Й.Б. посочва 2006 г., но от публикации на К.Ф. като книгата „Active Inference” – 2009.

### **Бележка Т.А.:**

- \* Юрген Шмидхубер има работи в тази насока десетилетия по-рано.
- \* Борис Казаченко има документирано съобщение за такъв принцип поне от 1998 г..
- \* Джеф Хокинс, "За умствените способности", публикувана 10.2004 г. "On Intelligence"
- \* Тодор Арnaudов започва да публикува „Теория на Разума и Вселената“ с такива принципи през 2001, 2002 (без да познава Шмидхубер)
- \* И др. преди това, принципи „за икономия“, „най-краткият път“ и пр.

**Йоша Бах:** "Бих казал, че грешката в предвижданията отразява естетическите нужди, това е нуждата да се открива структура [ред, подреденост] в света като цяло, но според мен не е достатъчно за разбиране на човешкото поведение. GPT3 е пример за система, която намалява грешката в предвиждането, но нищо повече. Редът, по който тя учи е различен от човешкия [и т.н.]"

**Тодор Арnaudов:** Виж "Свещеният сметач": предвиждащите модели се изграждат постепенно, работят в различни разделителни способности на възприятие и управление, йерархични са, при човек обикновено са свързани и с чувствената система и често не могат да се отделят от конкретния човек и конкретните спомени с чувствен произход (събития, предизвикали/запомнени като свързани с болка, удоволствие, срам, щастие, наслада, страх, спокойствие и т.н.), и във всички тях, от гледна точка на външен наблюдател или за самия „вътрешен“ ум има и „произволен“ елемент. „Произволният“ елемент, освен случайните чувствени спомени, в познавателната си част зависи от конкретните познавателни способности, вместимости на междинна памет на всички степени и места в системата – кое изглежда „подредено“, „просто“, „сложно“ (предвидимо, невидимо, колко, защо), изчислителни възможности, избор на точност на предвиждане и граница при която се приема за "достатъчно" и пр., и има и воля.

В "сложна" човешка вселена и сложно възприятие с взаимодействащи си въображаеми вселени с различни разделителни способности, обхвати, тегло на въздействие и пр., ако се разгледа цялата система, в нея не се предвижда \*само\* следващата дума, и няма \*единично\* предвиждане, а множество от възможни развития, разклонения, с различни разделителна способност, обхват; различно \*разделяне\* на света на съставни части и

подвселени и пр. В отделни подсистеми – управляващо-причиняващи устройства – всяко за себе си, по отделно, в определен вид сетивно-моторност и пр. – може да предвиждат и конкретни „следващи елементи“ и пр.

Също така Вселената, Времето или някои видове сметачи, например Машината на Тюринг, разгледани в най-прост възможен начин, предвиждат именно следващото състояние – знаят, че при дадено текущо състояние, при компютрите включващо текущата инструкция, трябва да се премине в еди-кое си следващо състояние. И дори с такъв последователен компютър, казба по казба<sup>27</sup> може да се построи всичко.

В актуалните сега невронномрежови модели „трансформатори“ всъщност също има подобни възможности и направления, напр. чрез кодовата таблица от начални кодови думи, "буквачета", "tokens", начален речник, който в сегашните модели е фиксиран и от него се изграждат останалите нива.

Сравни с обобщението от "ЧиММ" за начина по който запомня човешкия ум като „фотографски, текстографски, фонографски...“ запомня само първичните представяния“, а другите се съхраняват като препратки към тях, като адресиране по адрес в програмирането.

**Бах:** "За да може системата да достигне до интересни поведения по време на съществуването си в краен, а не безкраен период от време, е необходимо да има и допълнителни нужди, освен минимизиране на грешката"

**Арнаудов:** Да, това са чувствените нужди и:

- разликите, измененията в множеството параметри, които определят и задават определените, конкретни познавателни възможности и особености на съответния познавателен деятел
- конкретните спомени, цели; споменатите по-горе разделителна способност и избори и пр.
- конкретните среди в които работи и т.н.

Виж "Анализ на смисъла на изречение ...", Т.А. 2004.

**Бах:** Също трябва да има преходни нужди и рефлексии. Бебетата например не разбират нуждата глад и жажда, по-точно: не могат да ги съотнесат към подходящи последователности от действия, какво-може-да-се-прави (affordances), затова притежават инстинктивни поведения. Когато бебето е гладно, то върти глава и отваря и затваря устата си. Също рефлекс за сучене, за гълтане. Така детето се научава да свързва

---

<sup>27</sup> Казба – инструкция, заповед, команда в юнашкото наречие

гледа с гълтането и засищането, което се получава след това. Тогава рефлексът се изключва (затова е преходна подбуда), защото иначе би пречел на наученото поведение.

**Т:** Да. „Разбиране“ като съответствие, съвпадение между различни представяния.

**1:13 ч:** Схема за нужда, която има стойност и при надвишаване се включват сигнали за удоволствие или неудоволствие, които предизвикват включване на поведение за утоляване на "жажда", търсене на начин за задоволяване на нужда. Поражда се цел за утоляване, чрез което чрез проследяване обратно на поведението се подсилват определени действия за постигане на целта чрез предвиждащи модели.

Предшествоване\* + модулация →

Нужди ---- Обучение → (Памет, Възприятие, Действие)

Вземане на решение →

\* Priming, предварителна нагласа, която насочва следващите възприятия

Модулация - промени, водещи до емоционална конфигурация, промени в чувственото състояние

**Възбуда (arousal)** - неопределен синдром на симпатиковата нервна система\*

**Валентност** - оценка на ситуацията (добра/лоша)

**Агресия** - режим на "бой или бягство"

**Ниво на разрешаване:** ширина на фокуса

**Потискане:** дълбочина на фокуса; устойчивост на подбудите

**Скорост на подсигуряване:** скорост на проверка на състоянието на средата и съответствието ѝ с модела.

Според очакваните от нас промени се променя емоционалното ни състояние.

### **1:14:50: Двигател на подбудите**

Множество от „водохранилища“, свързани с "тръби" с обща "глава" в няколко групи:

\* почивка, либидо

\* красота, изследване, вещина

\* привързаност, господство, грижа, законност, сдружаване

\* храна, вода, здраве

Пресъхването на някой от „резервоарите“ поражда сигнал за болка, а когато е пълен - за удоволствие; различните сигнали си взаимодействат, за да произведат обединено чувство на валентност и неотложност, и степен на неотложността, и влияят на множество от модулатори, които настройват познавателната система във всеки момент от работата ѝ.

**Тош:** Виж лекция "Архитектура на мозъка на бозайниците" от курса по УИР, 2010-2011, слайд за емоциите, изразени в преобладаващи невротрансмитери и невромодулатори, както и лекцията за ТРИВ от същите курсове.

### (...) 1:15:xx Йерархия на нуждите на Маслоу

Самоосъществяване > Самочувствие > Приобщеност и любов > Сигурност > Физиологични нужди

**Бах:** Не вярвам в йерархията на Маслоу. Първо, това *не са нужди*, а са *цели, намерения* (purposes). Нуждите са по-прости и определени неща от елементите в пирамидата на Маслоу - последните не могат да се измерят пряко, а трябва да се *построят*. Т.е. това е стълбица от *цели, от модели на нужди*, а не от /действителни/ нужди.

Поведението се ръководи **не от преки награди, а от модели на наградите** (цели, предназначения, "смисли": purposes).

**Т.Арнаутов:** Съгласен съм със заключението на Й.Бах. Деятелят, умът, ръководи и предвижда поведението *спрямо представите си, очакванията* си до какво ще доведе определено действие и спрямо онова, което смята за „*награди*“ и пр. в момента на планиране – осъщественото действие в света може да се окаже, че не носи онова, което деятелят е очаквал. Представите и очакванията може дълбоко да се различават с действителността и с виждането на наблюдател-оценител (и от там поведението на деятеля да изглежда „неразумно“, „ирационално“); самоосъзнатата представата или онази, налична за словесната част от съзнанието също обикновено се различава или неизбежно се различава от по-точен модел, който включва цялата познавателна система и тяло на деятеля във „физическия“ свят и която предопределя какво точно върши и ще извърши с най-голяма разделителна способност във въображаемата вселена на най-ниско ниво - действителността. Понякога моделите и наградите се използват взаимозаменяемо. Също често под „учене с подкрепление“ (Reinforcement Learning) се подразбира такова с модел на света, „model-based“, а не стриктно УП, при което има само „награди“ и когато е приложено в крайната

си форма изглежда „тъпо“ и неефективно, понеже деятелите в симулациите извършват огромен брой видимо, очевидно „излишни“ опити, търсения, които малко по-„умен“ алгоритъм, който обобщава, може да „разбере“ с ограничен брой опити и да създаде съответни правила, „евристики“. Виж също ТРИВ: "Анализ на смисъла...", СВП-2 (Схващане за всеобщата предопределеност 2), СВП-3 (=ВиРЗ, Вселена и Разум 3), ВиР 4 ..., „Free Energy Principle/Active Inference“

**Йоша Бах** представя пирамида за съответствията на "ид, его, суперего" в "смисли": „ид“ - преки "смисли", цели на организма; "его" - смисли за взаимоотношения; смисли на "егото" за цялата продължителност на живота. Обаче тези нужди не са пирамида и йерархия, а са на едно равнище, защото непрекъснато се съревновават. Те има различни степени в различни моменти. Понякога може да не се храниш няколко дни, въпреки че имаш храна, защото имаш да свършиш по-важни неща и нямаш време (защото тялото ти може да използва собствените си запаси, да си позволи да загуби част от тях и също да навреди на здравословното си състояние за момента, което е преценено като по-маловажно от текущата цел). А отвъд споменатите три части от фройдисткия модел стоят и "свещените" неща ("сакралното") – смисли, които са отвъд егото и всички други смисли служат за задоволяване на свещените цели, свещения смисъл, на най-високия смисъл на съществуването.\*

**Сравни с Т.А.:** ... борба за това кой владее входно-изходните устройства, тялото, действениците – те определят конкретното поведение. „*Вътре*“ няма единна обективна личност, ако съдържанието се разложи на съставни части. Ограничеността на тялото и моделът на оценителя създават представата за единната личност. Виж „Анализ на смисъла на изречение...“, 2004 и ... 2012. Виж също работата на Майкъл Левин, който също разглежда живите организми и човека, и устройствата, съставени от части, по този начин.

\* Сравни също заглавията и понятията **"Свещеният сметач", "Вселената сметач"**... ;)

**Бах:** (...) Етиката е свързана с понятието за свещеност. Тя е договаряне на конфликта на интереси в състояние на споделен смисъл, цел.

\* Взаимодействие, което не се основава на сделка и лична изгода ("интересчийство"), а изисква споделени, общи цели и смисли, чувство, че си част от нещо по-голямо от теб.

\* Споделените смисли трябва да са дадени от вродени нужди.

Ако нямаш споделени цели, нямаш нужда и от етика, защото си единак. Например социопатите нямат познавателни дефицити, но всички

техни взаимодействия са "интересчийски". Човечеството строи обществата въз основа на способността ни да се свързваме със споделени цели и смисли, така че може да създаваме себе си като съгласувани деятели, участващи в огромни системи.

Системите с Общ ИИ трябва да притежават подобни възможности, за да могат да общуват с нас по етичен начин. Трябва да развием споделено чувство за "красота" относно начина по който искаме да изглежда светът. Нужно е да имаме споделени импулси, които да се свеждат до съгласувана действителност, която да е изпълнима на практика и да може да се поддържа устойчива.

**ТА:** "смисли" и пр. - сравни СВП-2 "Няма смисъл да има смисъл" и пр. но бих добавил и че *няма смисъл и да няма смисъл*; отчаянието е по-безсмислено; в живота като психологическо състояние то често произтича от физиологични причини и обстоятелства – здравият организъм иска да съществува *вечно*, да бъде вечно млад, здрав и изпитва удоволствие от съществуването, а дори и болният или увреден иска да бъде здрав.

**ТА:** Съгласен съм за споделените цели. Виж също Майкъл Левин. Сравни с движението за съгласуване на ИИ: „AI Alignment”.

## \* Йоша Бах гостува на Лекс Фридман №212

<https://www.youtube.com/watch?v=rlpUf-Vy2JA>

Joscha Bach: Nature of Reality, Dreams, and Consciousness | Lex Fridman Podcast #212 Lex Fridman 3,4 млн. Абонати

1 619 168 показвания 22.08.2021 г. Lex Fridman Podcast (към 26.2.2023)

Подбор, превод и бележки на Тодор Арнаудов: ок. 24:42:

**"Нямало цветове"** ... "система на игра" (гейм енджин)

Не е вярно че "не съществуват". Не съществуват *в тази форма*, "точно по този начин" както в другото представяне, но това е вярно и в обратна посока. В "действителността", каквото и да разбирате под нея, нещата не съществуват *точно* по начина, по който съществуват в паметта на сметач, например за *същия наблюдател: за нас*. Последното означава, че за да ги възприемем, трябва да извършим различни действия, да използваме различни сетива (сензори, датчици), да мислим по различен начин и т.н. отколкото в друга среда.

**Какво симулира този енджин, и \*пред кого\*?**

Не разбрах как дефинира какво е "реалност", "реално" и какво точно е "аз". Да, споменава, че е съгласуваността на паметта, спомените (виж Човекът и мислещата машина, 2001 Писма между ... 2002 и пр.), че е симулация като видеоигра, мултимедия и пр. Обаче **кой** има нужда от

тази съгласуваност, кой и как я проверява, ако възниква чак като се получи тя; кой гледа, слуша и чете това "представление" и защо? Как е възможен този „кой“, „душата“ и какво е – това не се разбира, а то е „мистерията“.

От първата част с Фридман, от 2020 г. Й.Б. споменава, че *"в последната година"*, т.е. 2019 г. разбрал какво е "дух", р. 1973 г., на 46 г. Сравни със ТРИВ, *"Писма между 18-годишния Тодор Арнаудов. ..."*, 2002 г. – начинът по който Тодор извежда какво човек/ум има предвид под „душа“ в диалога с Ангел Грънчаров: модел на възприятията ни за някого, за това „човек не умира, докато някой го помни“. Виж също въвеждащата лекция по „Универсален изкуствен разум“ от едноименния курс от 2010 и 2011 г. И пр. творби от ТРИВ (2001-2004) и след това. (...)



## **\* Joscha Bach: Artificial Consciousness and the Nature of Reality | Lex Fridman Podcast #101 | 1 722**

292 показвания 13.06.2020 г. Lex Fridman Podcast

### **Йоша Бах: Изкуственото съзнание и природата на Действителността:**

#### **Подкаст на Лекс Фридман №101**

**Подбор, превод и бележки:** Тодор Арнаудов – Тош

Joscha Bach is the VP of Research at the AI Foundation, previously doing research at MIT and Harvard. Joscha work explores the workings of the human mind, intelligence, consciousness, life on Earth, and the possibly-simulated fabric of our universe.

<https://www.youtube.com/watch?v=P-2P3MSZrBM>

#### **Съдържание:**

- 3:14 – Да разбереш Йоша Бах
- 10:38 – Природата на истината
- 18:47 – Оригинално мислене
- 23:14 – „Духовно съзнание“ и умствени способности (Sentience vs Intelligence)
- 31:45 – Ум и Действителност
- 46:51 – Трудният проблем на съзнанието
- 51:09 – Връзката между ума и вселената
- 56:29 – Какво е съзнание
- 1:02:32 – Език и понятия
- 1:09:02 – Мета-обучение
- 1:16:35 – Дух
- 1:18:10 – Цивилизацията ни може да не просъществува още дълго време
- 1:37:48 – Туитър и социалните медии
- 1:44:52 – Какви държавни управленски системи биха работили добре?
- 1:47:12 – Как да не се самоунищожим с ИИ?
- 1:55:18 – ИИ симулиращ човека, за да разбере собствената си природа
- 2:04:32 – Учене с подкрепление
- 2:09:12 – Здрав смисъл, Общи знания (Common Sense)
- 2:15:47 – Универсалният изкуствен разум ще има ли нужда от тяло?
- 2:22:34 – „Невралинк“ (Neuralink)
- 2:27:01 – Разсъждение в мащабите на невроните и общества
- 2:37:16 – Ролята на емоциите
- 2:48:03 – Щастието е курабия, която мозъкът пече за себе си

**Л.Ф.:** „Йоша е един от най-уникалните и надарени хора в общността на ИИ, който изследва природата на човешкия разум, умствените способности, съзнанието, живота на Земята и тъканта на вселената ни, която може би е симулирана.“ ..

Роден в ГДР (Източна Германия), баща му бил художник, живял "в гората". Чел "всичко".

**ЛФ:** Поне за мен ти стана един от най-уникалните мислители в света на ИИ. Може ли да "изследваш" собствения си ум, кои бяха главните философи, учени, идеи, може би филми, (...)

**Й.Б. 12:17:** В известен смисъл всичко, което би могло да съществува. За да съществува нещо, може би трябва да може да бъде изпълнено, осъществено, може би като краен автомат, така че може би цялото съществуващо представлява наложфиедни едни върху други крайни автомати и ние сме в област от този фрактал, с определени свойства, които му позволяват да ни съдържа. (...) Представи си, че всеки автомат е просто оператор , който действа върху някакъв субстрат, основа; и като следствие от действията се получават възникващи модели-схеми, "шевици", структури. Основата е нещо, което може да съхранява информация. Има брояч или нещо такова, което може да съхранява състояние.

**ЛФ:** Все още не разбирам защо съществува въобще, мога да седя там с бира или водка и просто да ефекта, наблюдавайки.

**ЙБ:** "Защо" може би е грешна насока. Може би не може да има отношение по посока на "Защо", без да се запитаме за цел, "смисъл" (purpose) -, или разбира се, това не значи, че всичко трябва да има цел или причина\*

**ТА(Тош):** Зависи кое как се дефинира. Виж СВП2:

Тодор до Ангел 28.8.2002

"...според мен, няма смисъл нито от промяна, нито от съвършенство, няма смисъл от нищо... Знам че сигурно навеждам на мисълта, че съм "депресиран", че "гледам света с черни очила", казват, че "ако погледна света от друг ъгъл ще си променя мнението" и т.н. Само че нямам мнение... Има смисъл, и той е преходен, но съм се научил да си мисля, че има смисъл, обаче ако не загубя разсъдъка си или не се случи нещо непредвидимо, сигурно и да съм щастлив пак ще знам, че всъщност няма смисъл"

**Бел. Тош, 16.3.2023:** Т.е. промяната, съвършенството и пр. изглеждат като "логични" цели, но също са **избрани** (Да, може да се запитаме "защо сме ги избрали" - защото явно те са важни,

първични свойства, особености и в нас, или сме „сходими“ към тях, клоним към тях, което отново подсказва, че са „първични“, съществуващи преди другото). „Избрани са“, защото не е необходимо да има специален "възвишен" и просто определен смисъл, както си го измисляме. (Ако смятате, че трябва да има такъв смисъл: защо трябва?) В по-обикновен смисъл, човек може да бъде това, което е, *защото е*, не е нужно да трябва да е "полезен на някого", да угажда и т.н. Например "да е полезен": такива внушения се правят на подчинени, покорни и/или за да бъдат покорени. Виж съвременните внушения за "ненужните хора" или "излишните хора", които като бъдат изместени чрез автоматизацията, вече "няма да са нужни на общество". Те самите са част от обществото, на себе си би трябвало да са нужни, но явно няма да бъдат вече нужни на онези, които досега са ги използвали като оръдия на труда и които смятат, че това е тяхното предназначение. Т.е. в случая „предназначението“ е внушено от вън. Индивидуално, за да *искаш* да живееш, да съществуваш, не е (логически) необходимо да си нужен на никого друго. На дъното на волята, що се отнася до целта „**съществуване**“ въобще, а не конкретна по-точно определена цел, може да „Искаш, защото искаш“, „защото така“.

**18 мин: ЛФ:** Дали изчислителността, информатиката, информатиците не представят съвременния философ, който смее да зададе големите въпроси като оригиналната философия?

**ЙБ:** ... Другата страна на ИИ е като философски проект, който е много рискован и много малко хора работят по него. Не е ясно дали ще успят да го осъществят. ... Мисля, че умствените способности (intelligence) е способността да създаваме модели. Това не е задължително целенасочена рационалност – много умни хора са несръчни в тази област; интелигентността е способността да разполагаш с определено количество модели-схеми и да виждаш структурата в тези модели и да можеш да предвиждаш следващия модел; да разбираш нещата ... Някои модели са толкова общи, че системата, която ги създава, трябва да разбира и себе си и как се отнася до средата си, като дете. ... Например забелязваш, че правиш определени неща, въпреки че се възприемаш като желаещ нещо друго, така че осъзнаваш собствената си психология\*. Осъзнаваш, че имаш сложна структура в себе си и трябва да се моделираш, да се самоанализираш (reverse-engineer yourself), за да можеш да предвидиш как ще реагираш в определени ситуации и как да се отнасяш със себе си в отношението ти със средата ... И ако успееш да разбереш себе си и отношенията ти с действителността във вид на "вселена", която може да продължи [твоите действия, теб], ако успееш да

стигнеш до край, всъщност това е ИИ ...

**Тош:** Виж бележка към предишна беседа с ЙБ: че по-общия модел, който системата/агентът/управляващо-причиняващото устройство създава за себе си, ако е част от цялото, неизбежно е опростен, непълен, с по-ниска разделителна способност и по един или друг начин ще се прояви като „погрешен“ на най-ниското ниво на управление-причиняване в майчината вселена.

**30 мин ЛФ:** Какво липсва в ИИ?

Всеобщ, обединен модел на обучение (unified learning). (...)

**\* Бележка за суетата, философско-етично отклонение: Т.А.:** Дял от ласкателствата и превъзнасянето в подобни предавания, както и в текстове от „официози“, вероятно са „безлични“, прилагат се за всички като част от клишетата в "шоубизнеса" и вероятно служат за внушаване и придаване на възможно най-голямо значение, „защо е важно да гледаш“ за привличане на зрителите. Всеки гост в подобно предаване, особено в развита "търговска" култура като англосаксонските, но и въобще в културата на търсенето на по-голяма популярност за повече гледания, харесвания и превъзнасяне на "висшите" и „доброто“; изисква и върви със съответните възхвали, така че онова, което се предлага на „пазара“ - „стоката“ - да изглежда по-примамлива, по-лъскава, по-скъпа, по-хубава и да привлича повече "клиенти", които да се прехласват по нея.

Също така блестящите хвалебствия, които се изказват за невероятната интелигентност на гостите е вид суета, суетност *и от страна на домакина*, и с цел *да се злоупотреби със самолюбието на госта*, както и да се играе със *суетата на зрителите*: ето, той е умен, велик, феноменален и аз го гледам (или аз го коментирам, пиша за него, сравнявам се с него и казвам той колко е невероятен...), значи „и аз съм като умен като него и т.н...“. Същото явление се вижда в изяви на общественици от българската сцена в областта на социалните науки, политиката, „геополитически анализи“, нови самозвани „визионери“ в изкуствения интелект с високо положение в обществото и финансови ресурси. „футуролози“, които обясняват очевидното, но „микрофонът е техен“, а не на оригиналните „пророци“; и въобще всякакви „знаменитости“, които заради известността си придобиват автоматично и „гениалност“ в оценките на представящите ги.

Разбира се, авторът на тази книга не изключва собствената си „суета“, но за разлика от други осъзнава и вижда „картинката“, механизмите и може би има основания (...)

## \* **Общ ИИ отвъд Многослойното обучение**

## \* **Generalist AI beyond Deep Learning**

<https://www.youtube.com/watch?v=p-OYPRhqRCg>

Cognitive AI \* 1,66 хил. Абонати \* 27 260 (ок. 21.2.2023)

Начало на премиерата: 11.01.2023 г.

**Участници:** Йоша Бах, Кристоф Ван Дер Малсберг, Таня Гринберг, Майкъл Левин

\* Joscha Bach, Intel

\* Christoph Van Der Malsberg, senior f ... and visiting professor at the institute in neuroinformatics at ETH, Zurich

\* Tanya Grinberg an AI scientist from Intel Lab

\* Michael Levin - Tuft University

\* Подбор и бележки от Тодор Арnaudов - Тош

Пораждащият ИИ е голям пробив към модели, които могат да разбират света чрез фантазиране на зрителни, текстови и понятийни представяния и стават нарастващо всеобщи. Докато тези системи с ИИ се основават на мащабиране на алгоритмите за многослойно обучение на невронни мрежи с огромни количества от данни и изчислителна мощ, живите организми изглежда успяват да извличат смисъла от света с много по-малко ресурси. Постигането на ефективни и самоорганизиращи се интелигентни системи все още се изпълзва от изследванията в ИИ, създавайки вълнуващ предел за следващата вълна от разработки. Гостите в предаването ще разгледат възможностите за вграждане на принципите за умно самонагласяване\* от биологията и кибернетиката в технически системи като начин да се доближим до универсалния разум. (...) **Йоша Бах:** 4:57 Какво е интелигентността? [Способност] да пресъздадеш миналото чрез пространство от изчислими функции. Това е по-интересен начин да кажем, че умствените способности представляват способността да се построяват модели, защото моделите са изчислими функции, (...) ако управлява бъдещето е деятел (агент) ... за да управляваш модел в бъдещето, трябва да създаваш представяния на състояния, които все още не съществуват. Това значи да имаш система, която е способна да изпълни произволни причинностни преходи и тази причинност е изолация от субстрата да представяш нещо, което още не се е случило по същество е онова, което представляват изчислителните машини - компютрите.

Ако за сметача представя функция, която значи изображение от описание на едно състояние към описание на друго състояние, и

компютрите могат да извършват такива преходи в напълно произволни начини и това е основната идея в построяването на система от изчислими функции и когато измислим начин да изброим изчислимите функции и да ги съберем в списък, всички заедно, и когато търсим в тях и ако искате да научите нещо, трябва в известен смисъл, да изброите, да организирате пространството от изчислими функции по такъв начин, че да може да откриете онези функции, които ви интересуват сравнително рано в търсенето и въпросът е как да построите пространството от изчислими функции, което е проследимо и сходимо до решение на задачата, която решавате в даден момент. Това е главният проблем.

Създаването на компютър е лесно, особено ако имате нещо като [биологична] клетка, защото те вече съдържат компютър в себе си и могат да се организират в по-високи нива компютри и да стават по-малко стохастични [и по-предопределени, детерминистични] и т.н. *[т.е. да предвиждат по-добре, по-точно, по-уверено/с по-малко грешки бъдещето - виж "Свещеният сметач", 2001-2004 г.]*. Голямата трудност се състои в това как може да открием функциите, които ни интересуват. (...) Трябва да отбележим, че многослойните невронни мрежи, многослойното машинно обучение (Deep Learning) са единственият *работещ* метод за големи задачи. Те са единственият клас алгоритми, които могат да откриват произволни функции за "разумно" време, като "разумно" е на порядъци повече време и данни, отколкото човек, разбира се; *[reasonable; бих казал: "приемливо" по избрани засега мерки, които са разтегливи, бел. Т.А.]* Обучението на модел като например "Stable Diffusion", с 2-гигабайтов размер, съдържа *всичкото*<sup>28</sup> изкуство и от него можете да извлечете [образ на] всяка знаменитост или всеки космически кораб от попкултурата; динозаври; всичко е събрано в тези два гигабайта. **8:01** Обучението на този модел отнема седмици, и това звучи като много малко време, сравнено с годините, които са необходими да се обучи човешкият мозък, но през тези седмици, или вече може би се свежда до дни, тази система преглежда стотици милиони изображения, много повече отколкото човешко същество би могло да обработи през целия си живот; също така системата намира зависимости чрез огромен брой сървъри. (...)

---

<sup>28</sup> Не е вярно, това е преувеличение, особено за първия общ SD. Също така зависи от критерия, какво качество се приема за представящо еди-коя си знаменитост или предмет еди-как си. Обемът на данни по принцип не е изненадващ: ако се опише като компресирани изображения с ниска разделителна способност, като вектори (очертания, скици) и т.н., с думи и т.н., след като се опишат основни образци, размерът ще бъде много по-малък. Виж напр. „Столове, сгради, карикатури, ...“, Т.Арнаутов, 2012

**Тош:** Според мен броят на "отделни" изображения, които човек вижда през живота си, не може да бъде пряко сравнен с обработваните чрез тези модели, понеже зрението е различно и работи по друг начин. Като начало, човешкото и въобще зрението на гръбначните животни, като се започне дори от жабите, засича движение и разлики, контраст. Жабите виждат движещи се мухи, гущерите движат очите и главите си, птичките като врабчетата също периодично движат главите си, а човешкото око извършва микро движения - сакади - с които обхожда контурите и определени части от предмети, и заради тази особеност, и заради по-високата плътност на сетивни елементи в средата на окото в т.нар. фовеа. Спорно е колко е честотата с която виждаме нови кадри, или "изображения", и кое се брои за такъв. В геймърските монитори се ползват широко 144 кад/сек, като по-високият клас дори са и на 240, макар че през 1990-те години се смяташе, че около 72 Хц са достатъчни за електронно-лъчевите тръби - което не беше вярно, по-добрите по-късни монитори предлагаха поне 85 и 100 Хц], а телевизията в Европа е на 50 Хц и то с презредова развивка, а киното е на едва 24 кад/сек и сравнено с другите изглежда като на забавен каданс. Дори ако се броят само 10 кад/сек, това са 36000/час, 576К за 16 ч, или 210 милиона за година. Докато сънуваме може да се каже, че също виждаме нови снимки - синтезирани изображения, като дори те може да са по-интересни от много от видените наяве.]

8:29: Как работят дълбоките невронни мрежи? Първо, те са вид диференцируеми изчислителни методи, т.е. представяне на функциите по начин, който ги вмества в непрекъснато пространство в което съседните вариации на функциите все още произвеждат интересни резултати (...)

9:59: ... Главното, което ни хрумва е да построим изчисленията върху дискретни оператори като булева логика или прости автомати като например клетъчни автомати и така да получим крайни машини на Тюринг и в крайния случай - динамични системи, като машините на Тюринг и автоматите са еднакви от изчислителна гледна точка: те са машини на Тюринг, доколкото не ви свършат средствата [памет и време]. Така че всеки цифров компютър в действителност е динамична система, защото физиката върху която са построени(?) цифровите компютри е непрекъсната от гледна точка на отделните транзистори и т.н., и ние просто се опитваме да намерим област от физическите особености, при която транзисторите са достатъчно надеждни и детерминистични като дискретни елементи от гледна точка на логическия език, който е построен

чрез подреждането на транзисторите в схемите. В основата обаче те са аналогови системи, т.е. шумни, а това е дискретна система; от друга страна, всяка динамична система във физиката на най-ниско ниво отново е дискретна, нали? Ако се приближите [към микросвета], няма да видите нищо непрекъснато; онова което ще видите са отделни атоми, отделни заряди и т.н. и те отново са дискретни. Тече дискусия дали всичко трябва да бъде дискретно на това ниво заради природата на езиците сами по себе си, и мисля че наистина е така. Откритието от миналия век, което е най-важно във философията, е че тези езици [представяния, математически модели], които приемат, че най-ниският слой е наистина непрекъснат, влизат в противоречия, но това има значение само ако желаете да моделирате най-ниския слой. Ако ви интересува само изчислението, няма значение дали ще започнете с динамична система или с отчетлива [дискретна] система, доколкото позволите всяка динамична система да има крайна разделителна способност и съответно [да се описва от] краен брой битове, с които може да работите във всеки момент; и тази равнозначност между непрекъснатата математика и дискретната математика всъщност е била показана и в двете посоки: може да използвате изчислителна машина и тогава просто като използвате повече битове, можете да получите приближение на непрекъснатата система с всякаква степен на точност и вярност, такава каквато искате, по същия начин както чрез цифровизираната музика може да отчитате, да измервате пространството от звукови функции под нивото на разделителната способност, която носителят ви може да предостави, така че да получите нещо еквивалентно; показано е, че действието в обратна посока също е възможно (...)

12:35 - ...

**Тош:** Сравни СВП2, ЧиММ, обясненията в първото, че „изкуството също може да бъде разложено на 1 и 0“ и пр.

[друг пример, но също: аналогово-цифрово преобразуване, квантифициране].(Схващане за всеобщата предопределеност 2, 2002; Човекът и Мислещата Машина: Анализ на възможността..., 2001)

(...) 91 мин: Да ви припомня, че мозъкът, организмът, е основан на 1 гигабайт генетична информация\*, докато за да опишете връзките в мозъка са необходими петабайта информация: 10 на 15-та степен, милион гигабайта. Т.е. гените могат да подберат много малко множество.



**Тош:** Сравни с ЧиММ, 2001 и цялата ТРИВ, за тесния канал на съзнателно управление и пр. Виж също диалога във форума на Кибертрон от 2004 г..

\* Правилно се посочва големият обем за връзките, точните числа може би са условни, но връзките в мозъка не са само абстракции в граф, невроните не са еднотипни, има химични компоненти и пр. Мярката за обема информация в ДНК е спорна, не е надеждна и зависи от оценителя-наблюдател. Само *кодът* на ДНК като *кодови думи*, записван като АГЦТТАГГГТАЦЦЦТТТ ... в памет на компютър и четими от него и от човек биохимик, е *безсмислен*, с него не може нито да опишете, нито да създадете организъм. Не можете да създадете организъм с код в отвличена информационна форма, и въобще с информация, която не е в подходяща форма, свързана с действителността – нужен е подходящ тълкувател, „интерпретатор“ на машинния език на Вселената на съответното достатъчно ниско ниво на представяне с достатъчно висока разделителна способност на възприятие и управление.

Кодът трябва да бъде под формата на истинска молекула, която е разположена в клетъчно ядро, което е част от клетка, която е в подходящо „здравословно“ състояние, в подходяща хранителна среда или организъм – например в утробата на жена, бъдеща майка, – със съответно тяло с всичките му клетки, което се намира в среда, екосистема с подходящи условия, в тази Вселена и т.н. Всичко това по правило се забравя при подобни измервания, с които се цели да се представи едновременно колко малко или колко много информация се кодира в ДНК. За да се прочете една молекула ДНК от формата и в клетка до компютърно представяне е било необходимо да мине цялото време от първата клетка с ДНК до необходимото развитие на човешката цивилизация, с всичките „гугулплекси“ преобразувания и данни, необходими, за да опишете Вселената или дори само планетата, на която са се случили. Виж още „*Нужни ли са смъртни изчислителни системи...*“, Т.Арнаутов, 2025.

Самоподдържащи се състояния на действие и свързаност ...

92 мин: мозъкът във всеки миг на бодърстване, има представяне на средата - ако отворите клепачи, и представянето е толкова добро, че вие вярвате на действителността; защото не виждате разлика между пресъздадения модел и външния свят. Вие сте толкова уверен, защото

непрекъснато правите опити, движите са наоколо, така че перспективата се променя и проверявате дали представата ви е съгласувана с времето със сетивната информация. Правите експерименти: хващате предмети и постоянно правите опити, за да потвърдите, че представата ви е съгласувана със средата... Разбира се, представата ви е малка част от средата, вниманието ви обхваща малко пространство ... , в съзвучие с тях 93:40 ... : Структурата от данни ... неврони ? Не гледайте невроните ... Отделният неврон няма значение, от значение е средата на неврона; нужно е да наблюдавате група от съвместно действащи неврони, за да разберете нещо ... както един пиксел не е достатъчен за да разберете изображение - нужна е достатъчно голяма област; така и с невроните - стотици или дори хиляди, .... фрагменти . ...

95:50 ... 96: ... Майкъл [Левин] каза, че броят на нещата, които може да се породят ... е безкраен ... Аз бих казал противоположното ... Наскоро четох една книга ... Сходимост в еволюцията. Лещата на окото е била открита 12 пъти или нещо такова. Фасетното око Вълчата глутница, социалните животни, ... всичко се преоткрива.

Следователно пространството от модели-схеми и форми, които са смислени, които са съгласувани, и са способни да създадат нещо значимо и устойчиво в средата и пр., е ограничено.

[**Тош:** Защото *средата* и законите в нея ги ограничават. Те определят какво може да се прави и какви същества, умове и пр. биха могли да съществуват. В литературата на Принципа на свободната енергия/Извод чрез действие и „Радикална телесност“ за средата се използва терминът „*екологична ниша*“ . ]

(...) И това е добра възможност за теорията да напредне така че да разбере какво може да очаква от биологията.

97:30 Напълно липсващото от нашия ИИ и МО е равносилното на биологичното поведение, поведенческите цели. Животното има първична цел на самосъхранение на собствената структура. Еволюцията е вградила в отделните видове определен брой подцели: нахрани се, избягвай опасности, намери социални връзки ... подцели, които изграждат живота. Умствените способности в очите на много хора е просто преследването на тези цели в променлив контекст.

101: Съзнанието ... и простите живи същества... не мисля, че може да намерите момент, където съзнанието изчезва.. губи се "обем" ... при животните се губи езика, губи се представа за далечното бъдеще и т.н. Обемът на съзнанието се смалява, но би ми било трудно да ...

**Тош:** Сравни с ЧиММ, 2001, разсъждението за това кога точно се появява или изчезва съзнанието. Виж също "Писма между 18-годишния Тодор Арнаудов и философа Ангел Грънчаров" и "Схващане за всеобщата предопределеност 3", за съзнанието като памет, за изпълнителната и събитийна памет. Виж статията "Хипотеза за дълбокото съзнание", първо публикувана като бележка към *"Какво му трябва на човек? Играеш ли по правилата ще загубиш играта!"*, 2014, Тодор Арнаудов, сп. "Разумир", бр.1

122: (...) Кристоф, как би проектирал, да речем, изкуствена зрителна кора, и как би се различавала от съществуващите модели?

Главната разлика ... Когато наблюдаваш със своя собствения мозък движещо се *[собственото? и не само]* тяло, можеш да предвиждаш следващите части от секундата и да ги сравняваш с идващите сигнали, т.е. нашата зрителна система притежава "апаратура", която владее диференциалната геометрия и има предвид формите на повърхнините, играта на светлината на повърхността на движенията ... бих построил последователност от ... масив от области [на мозъка], В1 - първична зона, съсредоточена върху движението, може би В3 - за цвят, и определен брой подмодалности в 2.5D измерения ("полу-3D") - двуизмерния начин на възприемане на света, с добавени вътрешни пространства на качества като цвят, триизмерно пространство и текстура... и може би 4-измерно 4D простр., едноизмерно пространство, двуизмерно и т.н. така че

**Тош:** Сравни примерите от "Писма между 18-годишния..." (СВП2) и другите части от "Теория за разума и Вселената" за предвиждането във футбола и пр.

2:04 ч. ... деформиращо предвиждане ... огромно поле от локални особености, ... текстури ...

2:06:xx моментен триизмерен модел ... самоуправляващите коли - не могат да го правят ...

2:12 какво липсва? ... доказателство, че тези пораждания могат да проработят

2:16 структурата от данни за тези представяния ...

2:18 5 признаци-особености ... подпризнаци ... или 7+-2 ... 9 ... ниска размерност + слепването им заедно [Т.А. „магическото число 7+-2 броя елемента в понятието за работна памет от познавателната психология]

(...) 2:19 в диригите на движението ... съдържанието на задействащата се вълна ...

**М.Левин:** ... определение за ИИ ... Зеноботите ... (кожни клетки от зародиш на жаба, които отделени се превръщат в самостоятелни организми). Живите организми, че се оправят/приспособяват в условия, за които не са предвидени. „Нестандартни видове познавателни системи“ (unconventional cognition). Ново определение за ум: способност да се решава една задача по различни начини... (по психолога Уилям Джеймс)

**\* Многопътните системи съхраняващи пълната история на преходите на състоянията като модели за разбиране на Разума и Вселената: Беседа със Стивън Волфрам [и Йоша Бах]<sup>29</sup> Multiway Systems as Models to Understand Mind and Universe - a Conversation with Stephen Wolfram [and Joscha Bach]**

Cognitive AI \* 1,86 хил. Абонати

15 950 показвания<sup>30</sup> 13.04.2022 г.

[https://www.youtube.com/watch?v=O\\_5e\\_WSNedE](https://www.youtube.com/watch?v=O_5e_WSNedE)

**Подбор, превод, бележки и въведение:** Тодор Арнаудов - Тош

**Въведение,** обобщено от Тош:

"Рулиада" (ruliad) - пространство от клетъчни автомати с различни правила. Рулиадата е преплетената граница от всичко, което е изчислително възможно, т.е. резултат от изпълняването на всички изчислими правила по всички възможни начини (Волфрам 2020, 2021).

Рулиадата може да се приеме за крайната абстракция и обобщение, включваща всички страни на физическата вселена. В частност, докато [определянето на конкретна] изчислителна система или математическа теория изискват да се направи [определен] избор [на условия, ограничения, рамки] при рулиадата няма избор или външни входни данни, защото тя вече съдържа всичко.

В контекста на рулиадна многопътна система, рулиадата проследява преплетените следствия от последователното прилагане на всички възможни изчислителни правила. На всяка стъпка, спрямо всяко състояние се прилагат всички правила по всички възможни начини. Процесът често поражда множество нови състояния, водещи до разклонения в графа, но е възможно да има и сливане на множество състояния, които се преобразуват до едно и също друго състояние. ... Виж например "мултипликада".

"Многопътни системи", или Разклоняващи се системи, съхраняващи пълната история на преходите на състоянията. Многопътните системи са вид заменящи системи, при които на едно и също ниво на преход са позволени по много състояния. ... Пътят между избрани състояние в множеството от всички възможни се нарича "еволюция". Заменяща система е изображение - в математически смисъл: съответствие, функция, - или начин, по който всеки неповтаряем, уникален обект се свързва с всеки елемент от дадено множество.

<https://mathworld.wolfram.com/Ruliad.html>

<https://mathworld.wolfram.com/MultiwaySystem.html>

<https://mathworld.wolfram.com/SubstitutionSystem.html>

<https://mathworld.wolfram.com/Multiplicad.html>

Коментар към: <https://www.wolframscience.com/nks/p223--the-emergence-of-order/>

с.223, "Starting from Randomness The Emergence of Order"

### **"Възникването на реда, започвайки от случайност"**

*В тази глава се разглеждат автомати, които започват от напълно случайно начално състояние, в което състоянието на всяка клетка - дали да бъде черна или бяла – се избира напълно случайно. Може да се предположи, че при такива условия няма да се получи никаква подредба, но всъщност много системи са склонни спонтанно да се самоорганизират и да породят особености, които въобще не са случайни.“*

**Тодор:** Сравни с "Вселена и разум 4", "Свещеният сметач", Т.Арnaudов. 2004. Бележка от 12.3.2023: Дали *"напълно случайното"* – според наблюдателя-оценител – всъщност е **наистина** случайно според *оценяваната* система, гледаща сама себе си, или пък от друг по-ведущ наблюдател, и дали *каквото и да било* може да бъде „напълно случайно“? Дали всъщност "редът" (предсказуемостта, закономерностите, взаимосвързаността, свързаността, последователността, повтаряемостта, симетриите – постоянството и съгласуваността на изменението (invariance & equivariance) и въобще първичните модели-схеми, върху които са основани всички други, не са *по начало* заложи в *Вселената*, в първичните закони на взаимодействие, в структурите, в които могат да се намират частиците – в структури от всякакъв вид, и абстрактни "буквачета"; дали структурите, които сега или в даден момент за някого *\*изглеждат\** случайни според *\*някаква\** мярка, според някого (оценител, наблюдател), който няма достатъчно пълна картина всъщност са случайни? Т.е. според мен може би не е изненадващо, че от „толкова прости правила на крайни автомати“ като „Игра на живот“ и т.н. или като фракталите могат да се получат „толкова сложни“ структури.

Всъщност може би донякъде така разсъждава и самият Волфрам относно термодинамиката в разговора с Й.Бах: че за нас има еди-какви си закони, защото имаме възможност да наблюдаваме поведението на частиците в голям обхват на време, пространство и състояния, но ако

---

<sup>29</sup> Също *Разклоняващите се системи*, многопосочни системи

<sup>30</sup> 23.11.2023: 2,38 хил. Абонати \* 19 245 показвания.

бяхме на микро-ниво и можехме да виждаме взаимодействията само между тях, и имахме данни за състоянието им от такава перспектива, можеше да имаме различна гледна точка и други закони, които са наблюдаеми в тези мащаби. Сравни също йерархиите на различните мащаби/обхвати в ТРИВ – определянето на нещо зависи от обхвата, разделителната способност и критериите на оценителя.

Виж също "Вселена и разум", Т.А., тезата, че във всичко, във всеки запис от данни, състояние, може да се открие някакъв "ред", "ред" е всяка възможна поредица, а дали е "подредена", "сложна", "проста", "хаотична" преценява оценителят\* според своите възможности за предвиждане на бъдещето и своите модели, представяне, понятия и т.н.

\* Във физиката предпочитат да наричат "оценителя": "наблюдател", observer. Виж също работата на Майкъл Левин, който напоследък обръща внимание на важността на наблюдателя.

Изразено по друг начин, това че нещо *изглежда "напълно"* случайно, хаотично, непредвидимо и пр. за някой оценител/наблюдател в даден момент, при дадени познания, опит, възможности, модели и т.н., *не значи*, че процесът не е детерминистичен, но редът, сложността, "истинският" модел, "въображаемата вселена от съответно ниво", физичните сили/закони на съответното ниво с необходимата разделителна способност на управление и възприятие и пр. не са „*точно определени*“ (но определението е все още неизвестно или недостъпно за *оценителя*). Целта на науката и на ума е именно да докаже последното. Това се разкрива непрекъснато и в индивидуалното развитие, и в развитието на науката, и например в обучението на трансформаторите/преобразители, които се учат да предвиждат все по-добре следващите думи/кодими думи/знаци/... въз основа на миналите.

Тази хипотеза може да се разшири и до живота и теориите за „спонтанното възникване“, което в предопределена Вселена не е случайно, а е заложено в първоначалната структура и условия. Виж също книгата на Боби Азарян, 2022 г. (Bobby Azarian)<sup>31</sup>. (...)

## **Въведение към клипа от авторите на видеото**

Най-ранните ни модели на действителността я изразяват чрез неподвижни структури и геометрия, докато математиците през 16-ти век измислят диференциалната алгебра. Понятието за изчисление е въведено през 20-ти век, от когато започнахме да моделираме света като

---

<sup>31</sup> „The Romance of Reality: How the Universe Organizes Itself to Create Life, Consciousness, and Cosmic Complexity“, 28.6.2022

преход между състояния. Стивън Волфрам твърди, че може би приближава времето да преминем към друга парадигма: множествено изчисление (multicomputation). В основата на множественото изчисление е непредопределена машина на Тюринг, една от по-тайнствените идеи от информатиката на 20-ти век. За разлика от предопределената (детерминистична) машина на Тюринг, непредопределената машина не преминава само от едно състояние в следващото, а във всички възможни състояния едновременно, което поражда структури, които възникват от разклоняването и сливането на причинностни пътища.

Стивън Волфрам изучава получаващите се многопътни системи като модел на фундаменталната физика. многопътни системи могат да служат и като абстракция за разбирането на процеси в биологията, обществото, икономиката и построяването на модели въобще.

В този разговор искаме да изследваме дали умствените процеси могат да бъдат разбрани като многопътни системи и как това би се отразило на представата ни за паметта, възприятията, вземането на решения и съзнанието.

Относно гостите: Стивън Волфрам е един от най-интересните ... мислители на нашето време, известен с уникалния си принос в информатиката, теоретичната физика и философията на изчисленията. Освен много други постижения, Стивън е създател на езика „Волфрам“ и (известен още като „Математика“), системата за работа със знания „Волфрам Алфа“; автор е на книгата „Нов вид наука“ и на проект за откриване на Основополагаща физична теория; основател и директор на „Wolfram Research“. (...)

## Беседа между Йоша Бах и Стивън Волфрам

**54 мин:** ЙБ: Творчество: Ново, но не очевидно. Има "скок", не е просто продължение по градиента, а е с прекъсване в пространството на търсене. ...

**Тош:** Виж ГнП, --',\, \|\ в бъдещи публикации. Това е една особеност на новото (оригиналното) от гледна точка на оценител-наблюдател: по-непредвидимо, за него, спрямо очакванията му. Виж ТРИВ и коментар от 12/2023<sup>32</sup>, понятията „surprisal“, Kullback-Leibler Divergence.

---

<sup>32</sup> „On What Creativity Is - Different Tints and the Pereslegin's remark about the modern AI researchers rediscovering Lem 1963“, Todor Arnaudov, 12/2023, <https://artificial-mind.blogspot.com/2023/12/on-what-creativity-is-different-tints.html>



**59 мин:** Чрез клетъчни автомати могат да изследват пространството на всички възможни математики и теореми и да видят кои са открити от човека досега. Проблемът е да се свържат с неща, които се отнасят до нашите разбирания и кое е *важно за нас*. "Нещата за които се интересуваме във Вселената постепенно се разширяват. Било е време, когато не сме забелязвали микроскопичния свят, въобще, тогава са изобретени микроскопите. Преди сме мислили, че главното събитие в квантовата механика е изобретяването на електронните усилватели, които могат да захванат слаб сигнал и да го усилят; не съм го проследил подробно, но съм уверен, че това техническо изобретение е довело до изобретяването и на квантовата механика. С други думи в света има особености, признаци, за които засега не знаем. Например представата ни за молекулите на газовете - в бъдеще някой може да си каже за нас: *"леле, как може някога да са забелязали само налягането и температурата, толкова са били тъпи, трябвало е да забележат и всички други неща, които имат изчислително значение и сега са част от техническия "багаж" и т.н. ...* Може ли да вземем "снимка" на света такъв, какъвто е в действителност, и да намерим необходимите думи, които го описват, и които е трябвало да забележим, но сме пропуснали. Например в пространствената форма на белтъчините, третичната форма, прегъването (protein folding) - белтъчините образуват алфа-хеликси, бета-листове, добре, значи може би има още 5 други мотива, които би трябвало да имат имена, но все ще нямат и затова "огъването" на белтъчините е сложен проблем за нас, понеже не сме забелязали другите "шевици"<sup>33</sup>, които може би зрително не са очевидни за нас по начина, по който обикновено гледаме картини на белтъчини и т.н., но биха били съвършено разпознаваеми, ако ги подберем по „правилния“ начин и щом веднъж успеем да опишем, че еди-коя си белтъчна молекула има еди-какво-си-име: листо и пр., тогава бихме могли да веднага да заключим: о, ще се прегъне по този или по онзи начин. Със сигурност има определени описания на света, *"мрежи за отъждествяване"*, за установяване на самоличността, за разпознаване - и ако погледнем в средата им и можем да забележим и отличим съответните белези, за които все още нямаме думи. (...)

**1:04 ч: Бах:** .... кои неща е полезно да се разпознават в света и какви технологии можем да построим на тяхна основа. Мисля че проблемът е свързан с "одушевеността" (agency). Подобно явление е свързано с това, което казваш: защо трябва да ни интересува какво си мисли Изкуствения разум, ако не мога да го свържа с онова, което правя. Чувал съм същото за теб (за С.Волфрам) физиците да казват, че математиката ти може и да е вярна, но защо трябва да ме интересува, след като не мога да я свържа

---

<sup>33</sup> Шевица е дума от юнашкото наречие, синоним за „pattern“: образец, модел-схема, модел

с (...)

**1:06 ч:** Бах: *"Агент е система, която управлява бъдещето."*

**Тодор Арnaudов:** Сравни с ТРИВ. Може да разгледаме въпроса и от друг ъгъл: че агентите/дейтелите=дейците, управляващите устройства, системите; които имат причиняваща сила и предвиждат бъдещето; съответстват, съвпадат, са успоредни, се развиват успоредно заедно с друга част от "вселената сметач", от пространството във въображаемата вселена и т.н. - виж Т.А., *Свещеният сметач*, 2001-2004 г., напр. частта във "Вселена и разум 4": Произход на повторения на събития в различни равнища на вселенския сметач. Сравни също с философията на Лайбниц за монадите, „Монадология“ и предначертаната хармония.

*"Свърхсетивното"* общуване, ако съществува, също както квантовото оплитане – спиновите на квантово оплетени електрони са свързани, от единия следва другия, на каквото и да е разстояние, което при банално тълкуване може да се приеме като "парадоксално" предаване на информация със скорост по-висока от тази на светлината, което би трябвало да е невъзможно. Обаче не е необходимо да има никакво предаване на информацията – пространственото, или време-пространственото представяне, което "четем" и в което мерим това разстояние, е различно от пространството, в което се "намират", са "адресирани", от което се "управляват", от което се променя състоянието на тези две частици. Тези две части от вселената, подпространства, подвселени, управляващо-причиняващи устройства и т.н. може да имат общо представяне и общо място в друго представяне, измерение и т.н.

Накратко, в *друго* представяне двете частици са на едно място или са една частица, и различните спинове и появи са прояви на едно и също явление, което включва двете частици в представянето, което ние възприемаме.

**1:06:** \* Проблемът за наблюдателя и описанието - за какъв наблюдател какво описание би било съгласувано със съответните наблюдения, измервания.

**1:16:** ... Промени в парадигмите - преходи в пространството на рулиадата.

**1:23:** ... Нервната система, умът като многопътна системи, в която пробягват много варианти, но някой от тях се *"срива"*, подобно на вълната на квантовата функция, когато всички наложени, възможни, вълнови функции, намиращи се в положение на суперпозиция, се сливат в една определена при наблюдение. В ума състоянията се разклоняват и сливат ... следват се различни пътеки в подсъзнанието, и някак нашата нишка на вниманието хваща някоя от тях ... [*Виж "Анализ на смисъла на изречение"..., 2004*] ... като квантово измерващо устройство, което взима

тези различни нишки и ги слива, обединява ги, ... различни нива на описание: на отделен неврон, който изпраща импулси; и от друга страна изчислителният език, който се опитва да опише мислите ни на междинен език на невронауките ... междинният език *[би описвал]* преход между определени състояния ... вероятно има повече да се добави в многопътна картина на нещата ... Има много предизвикателства в много области ... Физиката има големи успехи във формализирането на наблюденията, тя е поле на "изтънчена" математика, докато повечето други науки не са стигнали до там и повечето учени, които работят в други области, не знаят, че имат да изкачат висока кула. Например, вместо във формализмите на съществуващата математика, наскоро разглеждах имунологията и има огромен брой неща, които вероятно са достъпни *[за научно обобщение, разнищване]*, но това изисква подходящ формализъм, който не е точно онзи, който някой би научил в курс по елементарна математика... Въпросът за "заселването", покриването на необходимото Пространство от правила (rule space) от хората; има различни части от него, понякога трябва да пътуваш от една част до друга (...) има много "кули" в математическата физика, които са били построени в последните 50 години, за много от тях не е известно на какви основи би трябвало да бъдат построени, но при все това те са много "високи" и изглежда, че нашите проекти по физика осигуряват основи за много от тези "кули", и хората може да видят връзките между тях (...) Много приложения на техен метод и формализми в различни области като оптимизация на квантови вериги, невронауки, физиката ... Използваме огромната кула на физиката, защото формализмите ни са приложими върху физиката и трябва да внесем вътрешния си наглед, представите си от тази огромна кула до всички други научни области.

**1:30:14 Бах:** Но всички тези области са *отделни кули*, нали? Интересно е, че кулите ни не са построени в основата на повечето от интелектуалните традиции и отчасти е така, защото основите са били разрушени. В някои сфери никога не са съществували, но християнската цивилизация радикално унищожава способността да строим рационални модели на действителността. В известен смисъл науката ни започна съществуването си с "фалшиви новини" и ... за социалните медии и перспективата завинаги да унищожат способността на обществата да правят съгласувани модели на действителността ... краят на модерността ... когато обществото става постмодерно, философски обществото спря да притежава системни модели и да ги изследва, и учените спряха да имат системни, съгласувани модели - затова този проект да се построи тази съединяваща "вавилонска кула" беше изоставен и отделните кули

съществуват една до друга; те често много си приличат, например има голямо припокриване между икономиката и информатиката – много интересни алгоритми, открити от икономистите, се преоткриват в ИИ, без едните да знаят за другите. Разбира се, обучението с подкрепление, Q-Learning и пр. ИИ в момента преоткрива статистиката върху други основи – автоматизирайки я. Има и други научни области, които преоткриват собствени средства, за да се справят със статистиката, физиката по нейн начин, също така невронауките се опитаха да построят мост между [тези кули], "заклучвайки" учените в едни и същи сгради, но те все още изграждат отделни кули, в голяма степен несъвместими една с друга. Така че как очакваш идеите ти *[на С.Волфрам]* да се използват за основа на кули, които дори не си говорят. Мисля, че първите изчисления *[изчислителни системи]* и сега "рулиадата" са радикални обединяващи сили, защото цялата идея на изчисленията, фактът, че строим изчислителен *[модел]* X за всички X за всички области на познанието, това е много радикална обединяваща стъпка *[Вид Общ изкуствен интелект, Универсален изкуствен разум]*. Същото събитие се е случило преди 400 години, когато е бил изобретен математическият запис и се е родила математиката като наука. Интегрирането, обединението тогава далеч не е било толкова мащабно, колкото може да бъде чрез изчислителните методи, но мисля че изчислението само по себе си води до много обединения, а идеите за много-изчислимостта, за многопътните системи, водят до друго ниво на обединение. Онова което е особено интересно за много-изчислимостта е, че знаеш, че чрез изчисления може да се преместиш от едно състояние в друго. (...)

**[Том:** *Явлението „изчислителна неделимост“, необходимостта да се симулира процеса, ограничава определени неща, които може да се кажат ... ]*

**1:33 ч Волфрам.:** Много от нещата, които изследваме сега, са били анализирани от теолози и философи преди 500, 1000 години, няколко хиляди години. Прав си, че науката се е отказала от този вид всеобщи изказвания.

**1:36:** Имал съм късмета да работя в много научни области и да придобия известна степен на познания и разбиране, не колкото ми се иска, но определено ниво в голям брой различни сфери, и е възможно да видя общия формализъм. Имам някакъв шанс да видя "играта" между различните „кули“. Едва започваме този процес. В странно положение сме, защото този вид формализъм се "взриви" в *[огромен брой]* възможни приложения и текущата задача е как да ги организираме ... да си взаимодействат ...

**1:45: Бах:** Има разлика в описанието на Вселената като многопътна система (вид непредопределена машина на Тюринг) и като *предопределена\** машина на Тюринг. Състоянията, които са достижими, според определенията, са едни и същи, но *непредопределената\** посещава всички състояния наведнъж. Затова при нея се задават ограничения, рамки, а не пълни описания на това как да преминавате от едно в състояние в друго. Следователно няма значение колко време ще отнеме да се посетят всички състояния, ако може да го направите едновременно за същото време – броят стъпки, междинни състояния, за да преминете от състояние А до състояние Б, ще бъде същото, но чрез многопътната система ще откриете траектории, които не бихте открили по друг начин [*чрез предопределената машина на Тюринг*]. Или казано по друг начин, ако искате да опишете физиката на "компютърната игра", в която се намирате, и ако компютърът е *непредопределена* игра, програмата за описание на тази физика ще бъде много по-кратка, защото ще трябва да опишете само ограниченията, а не конкретния код за това какви точно да бъдат преходите между състоянията.

**1:46:30: Таня:** Какви са следствията за ученето от многопътните системи, за машинното обучение в частност. ... Каква е природата на ученето? Дали добавя нещо или изисква да се загуби нещо? Как би определил математически парадигмата по която работи Стивън? ...

**Бах:** ... Виж "Граф на разклоненията":

<https://mathworld.wolfram.com/BranchialGraph.html>

"Проектът за Физиката на Волфрам" постулира съществуването на абстрактни отношения между атомите *на* пространството [*различни са от атоми във пространството; атомите на пространството определят възможността за съществуване на нещо в тях, например атоми във...*], като отношенията между тях определя структурата на физическото пространство. При този подход, два клона на еволюцията в определено време са "съседни", ако споделят пряк наследник в своя многопътен граф. Проследяването на връзките в тези графи, наречени разклоняващи се графи, води до записа, нотацията на разклоняващото се пространство. Докато причинностните графи предоставят обобщение на развитието на системата, основано на прихващане на причинно-следствените отношения между събитията, разклоняващите се графи представят друг вид обобщение, основано на улавянето на отношенията между състояния в различни клонове на многопътната система. Разклоняващите се графи обхващат отношения между състояния на различни клонове на дадена стъпка на развитие и така определят изображение на съответствията, карта за изследване на разклоняващото

се дърво в многопътната система (Волфрам).

- \* недетерминистична – непредопределена
- детерминистична – предопределена
- \* ограничения, рамки - constraints...

## **\* Течения 083: Йоша Бах за Синтетичния интелект**

### **Currents 083: Joscha Bach on Synthetic Intelligence**

#### **Jim Rutt Show**

<https://www.youtube.com/watch?v=6HOjKa34im8>

Подбор, превод и бележки от Тодор Арнаудов – Тош

**20 м:** Джим Рут: "Големите рискове започват когато дадеш на на ИИ воля, "одушевеност" и пр. като Съзнание. Не съм убеден..., като съм работил с хора, по тези въпроси те казват, че Съзнанието и интелигентността са две отделни сфери, нали? Възможно е да има ум без съзнание, а може и да ги съчетае и да получи катастрофални последици като агент, който се стреми да произведе най-голям брой кламери<sup>34</sup> и пр.

Виж частта за съзнанието в "Човекът и мислещата машина: Анализ на възможността...", Т.А., 2001

[http://eim.twenkid.com/old/eimworld13/izint\\_13.html](http://eim.twenkid.com/old/eimworld13/izint_13.html)

Също повестта „Истината“, 2002-2003 и въобще ТРИВ.

#### **Съзнание и осъзнатост**

**39:33:** Могат да възникнат нови организми, което да създаде краткотраен разрыв, защото промените в климата ще променя много видове, които зависят от определени условия на живот; някои ще загинат ... Клетките са много устойчиви и независимо от промените в климата ... но сега сме в положение, в което можем да научим "скалите" да мислят и да имаме "мислещи минерали", да ецваме структури, които са по-фини от дължината на вълната на светлината, ... да вградим в тях логически езици, които да изразяват способността да се учат и да размишляват. Така ще се появи нещо ново, което не е съществувало на планетата, а може би и никъде във Вселената и ако този принцип се разпространи, може би няма да спре с цифровата основа, защото след като веднъж разбере как работи, може да се разшири до всякаво изчислително "тяло", така че ще бъде вездесъщ, и няма да бъде само "изкуствен интелект", а общ разум и след като веднъж се случи, ще възникне планетарен разум, който ще се сблъска с умовете на всички организми, които вече съществуват, и може би ще ги обедини, освен ако не се "събуди" с кисело настроение и не реши да започна начисто като изтрие

---

<sup>34</sup> Един от абсурдните примери на т.нар. „Съгласуватели на ИИ“. Ако се даде задача на свръхинтелект да „максимизира кламерите“, той щял да унищожи планетата за да произвежда повече кламери. Сравни със сатиричния фантастичен роман „Фабрика за абсолюти“, Карел Чапек, 1926 г. – производство на *кабари* на подобен принцип; виж бележки в „Първата съвременна стратегия за развитие чрез изкуствен интелект ...“, Т.Арнаудов, 2025 и в приложението за Фантастика (...) от *Пророците* #sf

всичко. [Т.А.: *Сравни с : "Човекът и мислещата..."*, 2001]

42:50: IIT Integrated Information Theory, Global Workspace Theory - Теория за обединението на информацията/въобраза (ТОИ/ТОВ), Теория за Общото работно пространство (ТОРП) - теории за съзнанието ...

43 м: Живите организми [и вселената, "естествените процеси", б. Т.А.]: от зародиш, от който се изграждат, разгръщат системите, самоорганизируют се; а не се построяват от части и сили отвън както сегашните роботи и пр., произведени предмети/системи.

**Т.Арнаудов:** *Кое е вътре и вън? Ако си представим по различен начин? Физичните сили дали не са също отвън? Или пък всичко е всъщност „отвътре“ – във Вселената. И в двата смисъл: общият процес на развитие на Вселената, в дългосрочен план, или просто непосредствените физични процеси в краткосрочен и в по-малка, или в произволна околност. „Вътрешното“ е, че има множество свързани промени, разположени „вътре“ в ограниченото от клетките и тялото (съвкупността от клетки, части) пространство, многослойност/многомашабност и пр. Избрана е граница за „вътре“ заради клетъчните стени, кожа; степен на въздействие, начин на пораждање и т.н., защото е по-близо до „центъра“ – зародиша – откойто е започнало развитието, нарастването. Обаче за да се случи растежът е било необходимо организъмът да се е намирал в подходяща среда и в цялата Вселена, и също да му влияят и външните сили, било в положителна, „желана“ от него посока като да предоставят необходимата топлина и температурни граници, влага, храна, майчина грижа и защита и т.н. Най-простата външна сила: енергия, както с температурата на средата, така и като вода и храна, която по-абстрактно също е вид „енергия“ – средство, необходимо за извършване на определена „работа“ – на действия, желани промени от устройството. И т.н.*

Това не отрича „външното“ и видимо разграничение самонагласящи се (самоорганизиращи) системи от „построени“, или „по-самонагласящи се“ и „самопоправящи се“ от „по-статични“ и не притежаващи самовъпроизвеждащи се качества в смисъла и *по начина* на живите организми; те са „по-съсредоточено самонагласящи се“, имат повече връзки с частите си; тяхната самоорганизираща се система, код, по начина на разглеждане, се събира в по-малко времепространство, съдържа повече знания/умения за себе си, за самопоправяне, самоизграждане,



възпроизвеждане и т.н. (автопоеза) или най-просто: за съгласувано въздействие, влияние на едни части върху други (предвиждане на собственото бъдеще с по-голяма разделителна способност на възприятие и управление) и т.н. Обаче в по-общ смисъл всъщност и предметите, които строим отвън, от части, изграждаме, също могат да се разгледат като самоизграждащи се и самонагласящи се и „възпроизвеждащи се“, само че тяхното ядро, съсредоточие и техният „аз“ са по-размити и включват по-голяма околност, част от която са силите, които събират отдалечените части, проектират системите и т.н. Границата на „отрязване“, прекъсване на обхвата на отчитане е с определена разделителна способност на възприятието и управлението, но с известно приближение може да бъде даден завод за даден период от време; или дадена държава или икономическа система; или планетата земя; или Слънчевата система; Галактиката; Вселената. Човеците, обществото могат да бъдат видени като „органели“ на самоорганизиращи се машини. Тези органели създават всичко необходимо за да се построят и възпроизведат те, без, до даден момент, да е необходимо в частите им да се съдържат повече способности за самовъзпроизводство. Ако дадена кола се повреди, нейните живи органели „я завеждат“, при нужда с помощта на други коли „в по-добро здраве“ до сервиз – „лечебница“, болница – за ремонт; или сменят спуканата ѝ гума, за да продължи да се движи; ако горивото ѝ намалява или се изчерпи – човекът я зарежда. От автомобилните заводи непрекъснато се раждат нови „индивиди“, а старите отиват в „автоморгата“ и се рециклират. Индивидите на живите организми също се нуждаят и от „външни“ поправки, когато без механични хирургически операции не биха оживели; повърхностните рани са външни, те се обработват „външно“; храната, кислорода се внасят от вън; понякога човеците получават протези чрез външна механична намеса, а накрая отиват в гробището; новите индивиди в потока на живота се възпроизвеждат от стари индивиди в екосистема и Вселена, те се създават, развалят и поправят от нея и са „наистина“ или „изцяло“ или „убедително“ „самостоятелни“ и се „самовъзпроизвеждат“, при отвлечане от подробностите, че всъщност не могат да съществуват без и извън своята среда и непрекъснати външни намеси и въздействия, без родителите, които ги пораждат, без грижата за тях и опазването им живи чрез наличието и поддържането на всички приемливи за съществуването им установени условия, които

представляват т.нар. „екологична ниша“. Виж „телесност“ и „радикална телесност“ (radical embodiment).

43:02: Главният принос на ТОВ според мен е във връзката между това как е осъществено/изпълнено нещо и как работи. Тонини твърди, че ако може да се построи сметачна машина по подобие на нервната система (невроморфен компютър), в даден момент тя може да има съзнание, но цифров Фон Нойманов компютър, който извършва *последователна обработка*, не може да има съзнание. Е, не е много много съзнателен, ще има [степен на съзнателност] Фи по-голямо от нула, трудно би било да бъде високо, но проблемът е много по-сложен от това. Ако компютърът е невраморфен, той може да се емулира на друг, основен компютър, това е основната теза на Чърч-Тюринг\*. Може да емулирате невраморфния сметач стъпка по стъпка, да изпълните всички тези операции, може и да е по-бавно, но бихте могли да го направите по-бърз, докато работи, или да бъдете по-търпеливи, когато задавате въпросите си, но основната система ще ви върне точно същите отговори, каквито и невраморфния компютър: *"Ей, аз съм съзнателен, имам субективно преживяване!"*. Във втория случай Фон Ноймановия компютър ще каже същото, защото изпълнява същата програма; той ще премества същите битове, в същата функционалност, ще съчини същите изречения и ще каже, че притежава свое Съзнание и субективно преживяване за света, само че в този случай лъже.

**Тош:** Сравни с ЧиММ също със СВП2, в които се дава същият пример за машина, която докладва съзнателността си и колко е условно това: дори *табела* с надпис може да се тълкува, че „казва“ какво мисли. Виж също триите вида предвиждане на бъдещето: по-бърз сметач, резултата по- бързо и ТРИВ като цяло.

\* на английски: *phenomenal experience, phenomenological experience*; виж работите на Дейвид Чалмърс (David Chalmers), „Трудният въпрос за съзнанието“ (The hard problem of consciousness)<sup>35</sup>

Също така: „кой каза“ на Тонини тези неща: кой определя кое точно е паралелно и кое е обработка или кое е „достатъчно“ невраморфно\*, кой го отчита. Как се мери, кой и как отчита и решава и „доказва“. Последователните компютри също имат паралелни части дори и на логическо, абстрактно, инженерно ниво: регистри, шини и пр., дори и прости 8-битови компютри, а електрическите сигнали и въобще полета в сложни схеми също се

---

<sup>35</sup> \* [http://www.scholarpedia.org/article/Hard\\_problem\\_of\\_consciousness](http://www.scholarpedia.org/article/Hard_problem_of_consciousness) .

\* <https://philpapers.org/rec/ZAH1AP>

разпространяват навсякъде, те са полета, а не тънки нишки. Също така кое определя и разпознава „компютъра“ или „процесора“ като единно цяло, къде се разделя от другото? И най-паралелните машини имат и последователни части и процеси при определено разглеждане. Успоредно (в случая може би *едновременно*) и последователно са прекалено абстрактни понятия, възприемани от ум. Необходимо е да има и последователност, свързаност, памет, за да може в настоящия момент „съзнанието“ да си спомня, че е било нещо си, и да предвижда какво ще се случи (основен принцип на разума).

\* Относно спорността на „невроморфността“ и този въпрос виж: Тодор Илиев Арnaudов - Тош, **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. И като приложение към този труд. <https://github.com/Twenkid/SIGI-2025/blob/main/Arnaudov-Is-Mortal-Computation-Required-For-Thinking-Machines-17-4-2025.pdf>

\* Както и „Вселена и Разум 6“ (Universe and Mind 6), Т.Арnaudов, 2025

**44:30:** Това води до много странно отношение, защото сега функционалния механизъм е различен. Причината да каже същото нещо е различна, защото по подразбиране невроморфния компютър казва, че е съгласен, защото наистина е почувствал собственото си Съзнание [*Според теорията на Тонини, Т.А.*], докато във втория случай изречението не е причинено от друго, освен "глич" в системата по някаква причина. Така че в определен смисъл ТОВ е несъвместима с тезиса на Чърч-Тюринг, но Лариса Урбан Такис и Тонони не отричат тезиса, затова смятам, че ТОВ е обречена поради фундаментални причини и няма начин да се поправи, освен ако не се откаже от основните си предпоставки.

**45:** Ако премахнем основната предпоставка, че пространственото разположение на алгоритъма\* отразява коефициента Фи и е решаващо за функцията му, тогава не остава почти нищо друго и няма голяма разлика с ТОРП или други теории.

45:46 ... разговаряйки в по-класически смисъл на познавателна обработка на информацията може да посочим Антонио Дамасио и Анил Сет ... Според Дамасио същинската „задвижваща сила“\* на съзнанието при животните въобще не е обработката на информация, а по-скоро е телесно чувство за себе си чрез възприятията; ... тази „сила“ произтича дълбоко от мозъчния ствол и дори животни без особено развити висши дялове на мозъка може би притежават такъв вид усещане за битие,

някакъв вид съзнание ... в смисъла, който влага Thomas Nagel<sup>36</sup> – ... Откъде знаеш, че имаш тяло, или че това е мозъчен ствол? Знаеш, защото това са електрохимични импулси, които протичат и кодират информация ... все е информация, няма друг начин ...

46:35: Въпросът е какъв вид информация се обработва, за какво се отнася тя. За разлика от ГПТЗ, ние сме сдвоени със средата по начин, който създава кръгова връзка между намеренията и действията, които тялото ни изпълнява и наблюденията; обратната връзка, която те връщат към interoцепцията възбужда нови намерения и вярвам, че само в контекста на този кръг можем да открием, че имаме тяло; тялото не е дадено, то се открива заедно с намеренията ни и действията ни, и със самия [външен] свят. Така че всички тези части са в ключова зависимост помежду си. Ние ги забелязваме като първично откритие на този кръг като модел на собствената ни дейтелност, на битието ни на деятел/агент. Да вземем за пример Фон Нойманов компютър, не невроморфен, макар че честно казано не знам дали има някаква разлика; все пак нека да го разграничим: високопаралелна много бърза компютърна система с Фоннойманова архитектура в робот, който има сетива навсякъде в тялото си, които казват на робота за неговото "благосъстояние", заряда на батериите; изпълняват се процеси, които следят за опасности и вършат неща, които са много подобни на верига от сигнали на емоционални състояния и т.н.

**Тош:** Това че някой си мисли, че подобна система, която на абстрактно ниво в ума му е „като биологична, защото...“, не е гаранция за „невроморфност“ или за „биологичност“, дали това наистина е „благосъстояние“ по начина по който го „усещат“ роботът или го „възприема“ Вселената, за да може да ѝ даде „човекоподобно съзнание“ на машината, или пък да не ѝ я даде . Виж споменатата по-горе работа **„Нужни ли са смъртни ...“**

**47:53:** Дали това е път към човекоподобно съзнание в бъдеще? Интересно е да отбележим, че устройството на организма и обществената система също така е устройство от втори ред, или такъв отвътре-навън, за разлика от отвън-навътре, когато изграждаме нещо като робот, като самоуправляваща се кола, например "Тесла" е робот. Започваме с "празна дъска", с работилница, в която знаем всички части и строим отвън навътре части от материята, като построяваме функциите, които искаме да изработим, така че проектираме и разработваме всички тези

---

<sup>36</sup> T.Nagel, "What Is It Like to Be a Bat?" (1974)  
[https://en.wikipedia.org/wiki/What\\_Is\\_It\\_Like\\_to\\_Be\\_a\\_Bat%3F](https://en.wikipedia.org/wiki/What_Is_It_Like_to_Be_a_Bat%3F)

функции и накрая ограничаваме материята по такъв начин, че тя да изпълнява само тези функции и нищо друго.

48:30: Докато в природата виждаме системи, които искат да изработят тази функция, които искат да израстнат до нея; вместо да проектира дървото, природата започва със семе и семето израства в зародиш, зародишът пораста до дърво чрез завладяване на средата около себе си и чрез усвояване на "неща" от нея, които изглеждат хаотични за организма, и превръщането им в нещо сложно в структура, която то може да управлява чрез построяване на управляеми кръгове на обратна връзка със света. Организмът може да се разширява и да става по-могъщ; подозирам, че умът също се изработва по този начин, вместо да има подробен проект за познавателната ни архитектура, записан в генома ни, подобна на "Обществото на Ума" по [Марвин] Мински. По-скоро има нещо много по-просто като "семе за ум"\*, които иска да порастне в такъв\*, и започват от нещо което Гари Еделман нарича невронен Дарвинизъм в еволюцията между различни подходи в един и същ ум; и [тази борба] накрая се съединява в ред в структурата, и всичко това прави цялото "нещо" много устойчиво. Например, дори и да пострадате като дете и загубите голяма част от мозъка си или той се деформира, и въобще мозъкът ви да не се развива по правилния начин, той пак намира начин да се устрои така, че да работи достатъчно пълноценно, за да пораснете като човешко същество, което може да управлява [средата си] чрез управляеми кръгове на обратна връзка, и организмът може да се разширява и да става по-могъщ ...

**Тош:** Относно „семе за ум“ (*seed for mind*) сравни със „Зародиш на разум“ от ТРИВ: „Дали *DeerBlue* наистина победи човешкия разум...“, 2001, *Човекът и Мислещата Машина: Анализ на възможността ...* (ЧММ), 2001; „Творчеството е подражание на ниво алгоритми“, 2003 и др.

Горните разсъжденията относно съзнанието и въпросът дали биха възникнали у робот със сензори не отговарят на основната загадка, която важи и за съществата, за които сега се приема, че имат съзнание: *кое е това „нещо“, субект, „аз“, „ние“,* което получава тази информация, откъде то знае, че е информация, и кога не е – *все* протичат някакви импулси, *във всеки случай; винаги* стават някакви взаимодействия. Какво е информация *от гледна точка на частите в системата*; „откъде знаят“ *те самите* коя информация е „съзнание“ и коя не е? Всяка промяна, влияние, взаимодействие може да се сметне за информация. Тя идвала от мозъчния ствол – а кой я приема и защо я тълкува като съзнателна? Неврони, нервни структури и т.н. – но защо мащаба, границата на разглеждане, нивото се слага там? Това че „всичко е информация“\*

не помага. Когато се вметва, че има значение „каква информация“, „какво се кодира с нея“, че сме сдвоени със средата – т.е. има различни видове информация, т.е. тя придобива „материална форма“ и/или изисква класификатор, който да я сравнява с образци. Кои са тези образци? Съгласен съм, че по-тясната и непосредствена *причинно-следствена връзка, взаимодействие* между части, частици, молекули, процеси вероятно е свързано с усещането за единно съществуване и съзнание, или може да се предположи че е така, защото може би в частиците, в околностите на протичащите взаимодействия, може би се съдържат „тагове“, кодове, знаци, които предават връзката им, това че са част от „едно“, тогава когато в рамките на определено време, обхват и пр. се извършват преобразувания между въпросните частици\*. Може би тези връзки стигат до определено разстояние, брой преобразувания и т.н. В същото време от по-обща гледна точка между всички частици във Вселената се извършват взаимодействия и може да се приеме, че „протича“, предава се някаква „информация“ чрез физичните сили. Необходимо е информацията да бъде вид „материална частица“ и да носи „тагове“ за принадлежност към определено тяло, „душа“, съзнание. От друга страна, обикновено търсим свързаност по „пряк“ начин, като вид докосване (виж Нютон и загадъчните сили, действащи от разстояние), но във Вселена сметач данните може да могат да се записват пряко във всяка клетка памет, част от пространството и пр., без да е необходимо да има *видима, измерима* и пр. връзка от гледна точка на подвселените „отвътре“. Както ние можем да пишем във всяка клетка от паметта на компютър. Виж също „*Вселена и Разум 6*“ и спомената по-горе работа „*Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини*“, Т.Арnaudов, 2025.

Друго, което в текста не се осъзнава, е че роботът, или каквото и да било устройство, система „чувства“, или „усеща“ или човек може да си представи че „има съзнание“, или че между процесът, в който даден сензор „дава информация на компютъра/ неговото съзнание“ за нивото на заряд на батериите и т.н. се прави пряка аналогия с човек – по-точно с *нашето* усещане и представа; с това на всеки от нас или на предполагаемия читател; с онова, което *ние си мислим*, че е „*на всички нас*“, че *всички* имаме тези чувства или съзнание; дали обаче действително всички ги имаме и дали са еднакви?, – защото всичко това се случва в *нашето съзнание, в нашия ум*, или по-точно: в ума на всеки отделен от нас; и всеки от нас поотделно си мисли, че вероятно съзнанието му е като на другите в дадена особеност. *За каквото и да говорим,*

*винаги си го представяме в собственото съзнание, ум, душа и т.н.*

Всеки робот, компютър, и всеки „предмет“ може да има всички усещания, които имаме ние, защото за каквото и да говорим, мислим или правим, *то минава през нашите собствени умове, съзнания и въображаеми вселени.*

А как роботът усеща? Ако не е „високопаралелен компютър“ няма ли да усеща – колко точно трябва да бъде, кой решава и защо? Когато човек губи част от себе си, от мозъка си, от тялото си – кога преставя да има съзнание? (Отговорите не са задоволителни, това че някой не може да се изрази, не значи, че не усеща. И ако някой се изразява или казва нещо, това не е доказателство, че той „наистина“ го усеща.) Ако е механичен, и като се дръпне въже, то дърпа някакъв болт – защо това пак да не е „болка“. Защо определени процеси са, а други не са? Кой следи дали има „кръг“ и обратна връзка, при положение че за да може да се проследи е необходима цялата система, или оценител-наблюдател, който да може да проследи поне няколко различни състояния и да забележи, че „от тук е дошла едн-каква си информация, после се е върнала“ и т.н., че частите са „същите“, че в различни моменти са тъждествени. По-простите части не знаят ли и не усещат ли само собствената си по-малка околност и не „помнят“ ли само своите местни спомени. И т.н.) Необходимо е в информацията и в „квантите“, които я носят и прашат, да имат белези или друг оценител-наблюдател, който да ги обедини. Виж „Вселена и Разум 6“.

\* ТРИВ също е теория на „Информационната Вселена“, или „Въобразна“ на юнашкото наречие; вселената като универсален симулатор на въображаеми вселени и т.н.)

59: Обемът на входна сетивна информация към мозъка

60: Колко малко сетивни канали ...

**Тош:** Но може би освен тези, които броим: нервни връзки към ретината в зрителния нерв, слуховия нерв и т.н., трябва да броим и взаимодействието между *всички частици с вселената*: всичко е "сетиво", не само тези канали. Въпросните канали са потоци при определена гледна точка и разделителна способност на възприятие и управление. Виж ЧиММ, 2001, "Вселена и разум 4", 2004 и др. - примерът с преместването на пръста: колко бита са необходими, за да се случи, сравнено с, предполагаемите, няколко бита, ако се кодира като избор на пръст и степен на свиване, или команда за свиване. Виж също „*Матрицата в Матрицата е матрица в матрицата*“, 2003.

Пророците на мислещите машини: Приложение Ирина: Беседи и записки ...

\* Виж раздела за съзнание и панпсихизъм в приложение Листове #Listove за научни статии за Теорията на обединената информация на Тонони и др.



## \* Кибер Арнимизъм от Йоша Бах

### \* **Cyber Animism by Joscha Bach** | Michael Levin's

Academic Content, 8,47 хил. абонати 8900 показвания 20.04.2024 г.  
(26.4.2024)

<https://www.youtube.com/watch?v=YZI4zom3q2g>

Лекция на Йоша Бах в изследователския център на Майкъл Левин.

Подбор, превод и бележки от Тодор Арнаудов – Тош

**Й.Б.:** Съзнанието може би е *по-просто* от *възприятията* и то съществува *първо* у<sup>37</sup> човека като бебе. [Тош: *преди* по-развитите мозъчни и умствени процеси, а не обратното – съзнанието да възниква след като мозъкът достигне определена степен на развитие].

**18 м:** \* Умствени способности, интелигентност (intelligence): способност за създаване на модели (Intelligence)

\* Самосъзнание: способност за разпознаване на себе си и отношенията си със света (Sentience)

\* Одушевеност: способност за управление на бъдещето (agency)

\* Аз: модел от първо лице за собствената одушевеност (self)

\* Съпричастност: способност да се преживяват душевните състояния на другите (empathy; mental states)

### **22:15: Можем ли да сравняваме компютрите и човешкия мозък?**

– Колко компютри са нужни за симулиране на мозък? Но и колко мозъци са нужни за емулиране на компютър?

**Тош:** Сравни със *същото сравнение* в ТРИВ; и в курса по УИР, страницата за парадоксалната универсалност на мозъка. И милиард мозъци не могат да заменят работата и на един микропроцесор или микроконтролер, когато е нужно да действа със съответната „електронна бързина“, нито пък дори на *телевизор* с електроннолъчева тръба или *модем*, дори „първобитен“ със скорост 300 бита/сек.

**33 м:**\* 0. Съзнанието е предпоставка в ума, преди всичко друго, започва с „разбъркан“ носител (unstructured substrate)

1. Създава се измерение на разлики (скрито пространство на представяния, кодирания: embeddings)

2. Разделяне на модел на света и умствено състояние (идеи)

---

<sup>37</sup> Въпросът „къде“ се намира обаче е спорен

3. Триизмерна действителност, свят с равнини като земя, населени с твърди и течни обеми
4. Създаване на предмети и категории
5. Неизменност спрямо промени в осветеността и развитие във времето
6. Моделиране на дейтелността, „одушевеността“ (agency), създаване на личност, свързване на съзнанието със самоличността

\* **Тош:** *Сравни с плана от „Вселена и Разум 5“, както и с Библията, Стар завет, Сътворение (и Й.Б. я споменава)*

**39 m: SPIRIT\*** – Self Perpetuating Intelligent Recurrent Transformer

„Дух“ – самоподдържащ се умен повтарящ се преобразувател

**50 m: „Киберанимизъм“** (Cyberanimism) – разширение на живота и съзнанието на неорганични носители

\* **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 самоусъвършенствовател), **SIGI** – Self-Improving General Intelligence; **VEI** – Versatile Explorer and Improver и др.

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 - .. Developer . **EMIL** – Explorer, Mapper, Improver and Learner.

## Джеф Бек – Бейсов мозък

### Dr. JEFF BECK - THE BAYESIAN BRAIN

Machine Learning Street Talk, 13.10.2023

87,7 хил. абонати

<https://youtu.be/c4praCiy9qU?si=-6QZ5Vrj3mBowKx4>

Подбор, превод и бележки от Тодор Арнаудов – Тош

Джеф Бек е главен изследовател от Verses – фирмата, която работи по осъществяване на принципът на свободната енергия/извод чрез действие (Free Energy Principle/Active Inference) – в приложни системи (ПСЕ/ИЧД)

#### От увода във видеото:

*Д-р Джеф Бек е изследовател в областта на компютърната невронаука, който се занимава с изучаването на вероятностното мислене (вземане на решения в условия на несигурност) както при хора, така и при животни. Той поставя особено внимание върху това как невронните структури представят несигурността и как коровите области вършат вероятно заключение и обучение. Неговата научна работа включва използването на информационни методи и статистически анализ на данни, за да изследва как нервните процеси се свързват с поведенческите реакции. Джеф изследва и принципите на обучение под формата на подкрепление и извод чрез действие.*

#### Помощно понятие:

**Инструментализъм\*** (в епистемологията и философията на науката): стойността на идеята се определя от успешното обяснение и предсказване на природните явления. Според инструменталистите, успешната научна теория не разкрива дали нещо е вярно или невярно в наблюдаваните обекти, свойства или процеси, а тя е просто средство, "оръдие", чрез което човешките същества предвиждат наблюдения в определена област от природата чрез формулиране на закони, които посочват или обобщават закономерностите, но не разкриват предполагаемо скритите страни на природата, които някак обясняват тези закони. (...) Отричайки стремежа на научния реализъм да разкрие метафизичната истина за природата, инструментализмът обикновено се определя като *антиреализъм*, макар че отрицанието на реализма на научните теории може да се нарече "нереализъм". Инструментализмът просто прескача споровете относно дали, например, частиците, които се

разглеждат от физиката на частиците, са отделни, дискретни обекти, или възбудено състояние на област от полето, или пък нещо друго. Според инструментализма за теоретичните термини е достатъчно да бъдат полезни в предвиждането на явлението, на наблюдаваните резултати.

По: <https://en.wikipedia.org/wiki/Instrumentalism>

### Някои от курсивите: Тодор Арнаудов.

Някои видове инструментализми са британски емпирицизъм (Лог, Бъркли, Хюм), трансцедентален идеализъм (Кант), логически позитивизъм (Мах) - верификационизъм (феноменализъм - физикализъм (от явления до обекти, които са разпознаваеми от науката физика)) ... Карнап: теория за потвърждаването. 1950-те, верификационисти: философия на науката.

Мисля, че Ангел Грънчаров обвинява мен, Т.А. в „Писма между 18-годишния Тодор Арнаудов ...“ също в „инструментализъм“, по-точно „неопозитивизъм“. <https://www.youtube.com/watch?v=c4praCiy9qU&t=627s>

**14 мин:** Размито е, размито по проект, защото така е най-добре. Би било много хубаво, ако можехме да кажем, че *"това е моделът", "ето така работи светът"*. ... Винаги ще има неясноти и несигурност. В края на краищата науката за откриване на истината ли служи? Не, тя се отнася до *предвиждане и компресиране на данни*; и нищо друго; не оценяваме стойността на моделите ни спрямо истинността, защото не можем.

**16 мин:** В литературата не се подчертава достатъчно връзката между теорията за разпознаване на системи\* и идеалата на Марков. Ограждате ги в кръгчета: ето това е една система, а онова е друга... Започва се с набор от наблюдения с огромен брой измерения, който се свежда до прости уравнения. Подобно на статистическата физика – свеждането им до няколко уравнения. Взимаме много сложни ... превръщаме ги в по-прости ... обекти, идеала на Марков, ... Статистика на границите ѝ, което е същото като в теорията за разпознаване на системите – *не ви интересува какво има вътре*\* ... два обекта, които имат еднакви граници, взаимодействат с вселената, или със средата си, по един и същи начин, следователно са от един и същ тип.\*

\* [https://en.wikipedia.org/wiki/System\\_identification](https://en.wikipedia.org/wiki/System_identification)

[Бел. Тош: Според мен не е вярно и е опит за хитруване, внушение/подвеждане и оправдание, че *„не ви интересува какво има вътре“*. Вярно е само с определена разделителна способност на възприятие и управление и само във възприятието, въображението, ума на съответен наблюдател-оценител. Вярно е и в контекста на казаното от

Дж.Бек, че не можем да определим или знаем истинността на някои предположения само от данните и сетивата, знаем само че бихме могли да ги представим по измерения начин, което е разглеждано по „технически“ начин напр. във философията от 18-ти и началото на 19-ти век: Имануел Кант и Артур Шопенхауер. В ТРИВ се споменава също за граница на разбирането, отвъд която не можем да разберем, напр. още в „Писма между 18-годишния...“, 2002 – че окото не може да се види без огледало; или за огромната несиметричност между съзнателно управляваната информация и сложността на протичащите процеси в тялото. Хоризонтът на познанието, от дадено текущо разположение в познавателното пространство, трябва да се има предвид.

Обаче *"не се интересувате"* от съдържанието, на познатите, обособени системи, не е вярно в общ смисъл, дори е погрешно. За моделирането на дадено ниво на подробности, при което използвате само дадените външни признаци, особености, които са еднакви за даден тип, тогава „не е нужно“ да проникнете отвъд. Процесът на познанието обаче е непрекъснато *разширяване, увеличаване на разделителната способност, обхвата, точността на предвиждането*, съответно на *разпознаването на системите*. Умът и вселената се стремят да разберат, да анализират, да раздробят управляващо-причиняващите устройства, системи, „черни кутии“ (в началото) на съставни части, да ги управляват, моделират, причиняват, създават по-добре.

Понякога е *невъзможно* за едно управляващо-причиняващо устройство (деятел, агент, ум, ...) да "проникне" отвъд границите на черната кутия, но не задължително *защото "не иска"*.

Развитието на науката и разбирането е именно "отваряне" на все повече от тези граници и „отваряне“ на всички вложени една в друга черни кутии.

В ИИ - "обясним ИИ" (eXplainable AI; mechanistically explainability и т.н.), и една от критиките към сегашните невронни мрежи е, че не обясняват как са стигнали до решението по начин, който е приемлив за хората, че са „черни кутии“. ]

**18 мин:** Извод чрез действие .. Той е толкова общ/универсален, че може да се приложи върху всичко. Камъните извършват много тъпи изводи чрез действие, но и те могат да се моделират по този начин с идеала на Марков.

**[Бел. Т.Арnaudов:** Една от причините е и защото и камъните, и всичко останало, когато говорим за него, представяме си и оценяваме, **се моделира от ум, "умства" се, умоплащава се** (на английски бих го нарекъл **"mindifying"** - they are mindified, виж

*“Universe and Mind 6”, 2025).* Когато разглеждате камъните като дейатели, *те са част от Вас като деятел.* Затова всичко може да се "антропоморфира", да се уподоби на човек, и можем да си играем "на уж" с всичко, да си представяме, че столът е самолет, кола, кон и т.н., *понеже умът е универсален симулатор на въображаеми вселени*, в който всичко може да е "живо", или живото да стане мъртво, или да следва избрани произволни физични, биологични и пр. закони. Другите човеци, „прилепи“ (по известния въпрос-статия „Какво е чувството да бъдеш прилеп?“ на Томас Нейджъл (Thomas Nagel )); реки, държави, всякакви неодушевени предмети или понятия са единни и са „обекти“ и едно цяло в рамките на ум, който ги възприема, разпознава, проследява и който поставя границите и определя по *този* начин, че „това“ е едно цяло или е „същото“ между различните отчитания във времето и пространството. За друг оценител-наблюдател всеки момент, в рамките на неговия начин на отчитане, може да е различен; той може да не може да сравнява между „кадрите“ и въобще да не може да възприеме понятието за запазване на тъждественост във времето; или пък да не го приема, защото за него и най-малката разлика е достатъчна, за да определи съдържанието на два отделни отчета във времето или на два обекта като различни, както в сентенцията „не може да влезеш в една и съща река два пъти“. Не е необходимо да отчитаме и да знаем, че атомите в тялото на живите организми се сменят непрекъснато и за месеци или години на атомно ниво се „превъртат“ изцяло, за това е необходимо специално знание и изследвания и го научаваме абстрактно\*. Една такава система или модел би била например филмовата лента: всеки кадър е отделен и „не знае“ за другия. В цифрова компресираща система кадрите „знаят“ един за друг повече, между тях има изрични връзки.

\* **Всъщност всяко знание е абстрактно:** отделяне, подбиране; но с различни степени, гледни точки и различен начин. По-малко абстрактно е „непосредствено сетивно“, например зрително, но то също минава най-малко през разпознаване и сравнение с образец. Виж „Вселена и Разум 6“ (първо на английски, Universe and Mind 6, Т.А., 2025) и „Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини“, Т.А., 2025.

\* Виж в основния том „Markov Blanket Trick“.

**20-23 мин:** В ума ми има абстрактни представяния, но не мога да ти кажа "виж състояние номер 7". Всичко което предавам на другите агенти са предвиждания, предсказания, прогнози. [Да вземем] два непознати един на друг агенти. Не знаят и нямат достъп до абстрактните представяния на другия, нямат достъп до тях, но ако общуват само чрез предвиждания, не е необходимо да познават вътрешната структура на чуждия модел, защото ще говорят на езика, който всеки агент е необходимо да разбира. Така че единственият начин два напълно различни интелигентни агенти да общуват е на езика на предвижданията, който служи като посредник, защото е вплетен в наблюденията в средата.

**24:40:** Имам нужда от теория на ума, защото е необходимо да предсказвам поведението ти. Толкова е просто и прагматично. Може да е хубаво да имаш съгласувана теория на ума, която твърди нещо фундаментално, но ... [за практически цели] тя не е нужна, достатъчно е да можеш да предвиждаш поведението на другия. Теорията на ума също е основана на предвиждания. [Сравни с обяснението за това какво е „душа“ в ТРИВ.]

**25:00 – 28:00:** Коментират, че се стигало до панпсихизъм, че във всяка частица има елементи на съзнание, което противоречало на материализма.

**Тош:** Зависи какво се смята за „материализъм“. Философите противопоставят „идеализъм и материализъм“. В ТРИВ има „**информационализъм**“ или „**информатизъм**“. В интервюто споменават „обективен идеализъм“, че първо било съзнанието. Материята е, или може да се тълкува като, информация и информацията е начин да се опише, представи, осъществи материята – и всичко – в цифрова, сметачна вселена (възможна ли е несметачна вселена, как ще се представи?). Но кое е онова, което тълкува каквото и да било както и да било и то „какво е“, от какво е описано? (Субектът, „съзнанието“ и пр.)

Виж в началото, предвъведението на основния том с школи и учени свързани с ТРИВ, препратката към **Мелвин Вопсън**, за *„Втория закон на инфодинамиката и следствията от него за хипотезата за симулираната вселена“*.

**\* Melvin Vopson: The second law of infodynamics and its implications for the simulated universe hypothesis,**

<https://pubs.aip.org/aip/adv/article/13/10/105308/2915332/The-second-law-of-infodynamics-and-its>

\* Виж също повестта на Тош „Истината“, 2002, 2003.

**30:** проследим Бейсов извод ... за големи модели, мащабиране ... без използване на спускане по градиента ... големи стъпки в пространството на параметрите - предвиждаш минимума или максимума без да е нужно да се "опипва" пространството на най-малки стъпки.

**Тош:** Или просто предвиждаш в *дискретен* свят като избираш от *набор от възможности*, което може да бъде *преход* в програма, в паметта, в списък, таблица, граф, мрежа и пр. Не е нужно изчислението да следва само формула на гладка функция с производна и т.н. Може да бъде всякакъв вид обхождане, изчерпване, търсене и пр. под формата на каквато и да било програма, процес, преобразуване и т.н.

**31:xx:** руго (лат. „Пайро“) – език за програмиране на пораждащи модели

**33:** трансформаторите са смесица от специализирани модели (expert models ... „Супа“<sup>[кирилица]</sup>, ... Структура ... йерархични нива – влиянието е ограничено в рамките на нивото.<sup>38</sup>

**[Тош:** добра подсказка за начина за *разграничаване на нива* при първичен анализ, ако се почне от неопределено положение. Вид групиране (клъстериране, Clustering) – нивата са буквачета (елементи), които „кръжат“ около определен център, или определена област, и се намират на по-малко разстояние помежду си, отколкото до други центрове – други групи, гроздове; както са планетите в Слънчевата система, звездите системи в рамките на галактика, електроните около атомното ядро; сградите в градовете и другите населени места и т.н.

Групирането като операция или процес също е приложение на ПСЕ, например намаляване на някаква мярка за разход на енергия за придвижване между частите в групата. Може да видим аналогия между разрастването на градовете, построяване на нови сгради, близо до други, с делението на клетките в развитието на зародиш. Клетката не може да създаде друга клетка далече от себе си. И по-„евтино“ и по-непосредствено е за живото същество да расте от там, където има жива тъкан, „навън“. Сградите се строят по заобиколен начин с взаимодействие на човеци, машини и пр., но може да си ги представим като прояви на по-висша форма на физични/природни/вселенски закони;

---

<sup>38</sup> Забележи „руго“, „супа“ – двусмислен запис на латиница/кирилица на неизвестни от първо четене думи или при неопределена азбука от две възможности с подобен вид на знаците: руго – „пайро“ ил „руго“ (РУГО) – кир.; „супа“ (храна) → „суна“ (сайна) на латиница.



в ТРИВ всяко управляващо-причиняващо устройство от всяка степен/ниво може да се разгледа като вид физичен закон във Вселената сметач, осъществена като универсални симулатори на въображаеми вселени. Построяването на нови сгради е като делене на клетка, от съседните се раждат нови.

Някои веднага биха отхвърлили тази „измишльотина“, понеже сградите са „създадени от човеци“, че това не е „възникващо“ (emergent), че няма ДНК/клетъчно ядро както при клетките, няма ясен разграничим процес, който да „дели“; че строежът не е делене, защото не ползва ресурсите на същата клетка и т.н.

Когато осъзнаем, че *описанието, представянето на състоянието* на Вселената и на дейците в нея не е съсредоточено само в най-баналните и очевидни в кратко време и рязко отчеленени на пръв поглед граници – индивидуални тела на човеци и др. – а всъщност в по-големи мащаби и с по-голяма финост може да се възприеме като „**разлято**“ наоколо във времепространството; можем да обособим **абстрактна сила**, огромна въображаема „**клетка**“ **или деец/агент**, която кара сторителните предприемачи да започнат определен строеж, което се случва чрез взаимодействие с *кметове* (законно или незаконно, морално или неморално – според друга гледна точка на наблюдение), с *инвеститори, работници, клиенти* – договаряне; проектиране или решаване на проекта, конкретни действия по строежа и т.н. Тези абстрактни „огромни клетки“ взаимодействат с други, или споменатите могат да се приемат за част от други подобни структури – не са членове *само на една клетка*.

В такова разглеждане *истинската* система, която управлява всички тези частици, съдържа *по малко от същинското* управляващо устройство, разпиляно в телата и във времепространствената околност на всички участници в по-примитивната представа за дейци (индивиди, човеци), и в крайна сметка е „поле“ във Вселената.

До подобен извод стигат и М.Рамстед и др. в ... 2019, **Multiscale Integration** ... Виж **ТРИВ** и „**Вселена и Разум 6**“.(...)

\* От Вopсън виж още: \* <https://phys.org/news/2025-04-universe-ultimate.html>  
**Is gravity evidence of a computational universe? Editor's Pick**, Melvin M.

Vopson, \* <https://pubs.aip.org/aip/adv/article/15/4/045035/3345217/Is-gravity-evidence-of-a-computational-universe>

\* <https://nauka.offnews.bg/fizika/mozhe-li-nashata-vselenata-da-raboti-kato-kompiutar-201791.html>

**\* Педро Домингос**

**\* 5 tribes | Master algorithm**

**\* Петте племена | Главният алгоритъм**

**Pedro Domingos | Педро Домингос**

Подбор, превод и бележки: Тодор Арнаудов – Тош

#6...

- no infinity    - discrete universe    - predictive coding
- many agree ...    5 tribes – виж в др. приложение.

### **Master algorithm – Главният алгоритъм**

Възникване | Възникновение

Слаба | Силна | Различни нива | Книга ... | MLST

**Кенет Станли:**

**2 мин:** Нещо което разграничава индивида от популацията са всестранните възможности и специализацията. Популациите са по-силно насочени към специализацията. Мисля, че има огромно количество *"снобизъм на обобщаването"* в машинното обучение: търсим най-всеобхватния всестранен алгоритъм през цялото време: да се справи с всички възможни задачи и да изсипем всички възможни данни в набора му за обучение, и той става все по-голям, и всички сме впечатлени.

От друга страна, популациите показват нещо друго, защото искам да видя различни неща и хиперспециализацията води до всякакви екзотични неща, които всеобхватен алгоритъм не би извършил, защото иска да бъде всеобщ.

- Това се свежда до определено ниво на абстракция и на "одушевеност"
- ...

Проф. **Мелани Мичъл** - книга, "Complexity: A guided tour", 3.2009

... поражда

**Тош:** Да, организмите: започват от една клетка, която е обща и тя поражда части, които са специализирани. Общ е *механизмът* за пораждаване на частите. ЧиММ: без *общ ум*, обединяващ специализираните процесори, частните са *"безполезни"* (във или за всеобща задача за универсален ум, не могат да се самообединят – отдавна има множество отделни, частни, специализирани модули за „всичко“, все по-добри, но те все повече *преливат, сливат се, съчетават се, обединяват се* – виж „Творчеството е подражание на

*ниво алгоритми*“, 2003; приложение *Листове* на Пророците...,2025, „*Първата съвременна стратегия за развитие чрез ИИ...*“, 2025, приложенията *Лазар* и *Анелия* и основния том) ...

За взаимодействието между нивата ... Винаги е важен наблюдателят, той определя кое е "сложно" ... Представяне от гледна точка на съответно ниво (подпространство, въображаема вселена и т.н.); това позволява **контекстно кодиране** и **контекстен речник** (виж *Зрим*, ок. 2013-2014-2015 г.)

Ограничение на изчислителната мощ в "отделение" ... клетка, контекст, ... По-висока ефективност, по-малко необходими ресурси за съответната по-тясна задача, приложение, обхват на обработка, обем данни, тип и пр.

## \* Професор Берт Дефрейс - За Извода чрез Действие

### \* Prof. BERT DE VRIES - ON ACTIVE INFERENCE

Machine Learning Street Talk • 7,3 хил. Показвания

„This man builds intelligent machines“, 20.11.2023

<https://www.youtube.com/watch?v=2wnJ6E6rQsU>

\* Подбор, превод, бележки и автор на откъсите от Теория на Разума и Вселената: **Тодор Арнаудов – Тош**

\* **See a shorter version in English below**

**Берт де Фрейс** е професор по обработка на сигнали в техническия университет в Айндохвен, Нидерландия. Със студентите си разработват инструменти, чрез които да се създават агенти, следващи метода на Извод чрез действие (Active Inference, AIF, ИчД) в парадигмата Free Energy Principle/Active Inference.

**0 мин:** „Когато за пръв път чух за ИчД, преди около 10 години (2013 г.), тя се отнасяше най-вече за невронауки и как [минимизирането на] вариационната свободна енергия е единственият процес, който протича [в мозъка]. Това беше гениално, но по някакъв начин беше вкарано в рамки като за обработка на информация. Няколко години по-късно Карл Фристън започна да говори, че това е принципът за най-малко действие, приложен върху самоорганизиращи се системи; върху системи, обработващи информация, който би трябвало да протича в самоорганизиращите се системи, в биологичните системи; в мозъците, но и във всички самоорганизиращи се системи. Още по-скоро, тези тълкувания доведоха до разработка на нов клон на физиката – нов клон на механиката на Лагранж. ... От този принцип може да се изведе всичко във физиката. ... Той е толкова важен, че **ще засегне не само ИИ, но и инженерните науки и природните науки**. Мисля, че след време Принципът на свободната енергия ще се преподава в часовете по природни науки или физика в средните училища. (...)

**Тим Скарф:** „**Вариационната енергия представя разликата между това, което мозъкът ви предвижда и онова, което всъщност се случва**. Когато тази „вариационна енергия“ е висока, това означава, че има значително несъвпадение. Мозъкът ви е изненадан, а той не обича изненадите. За да намали изненадата, което учените наричат „грешка в предвиждането“, мозъкът ви донаглася вътрешните си модели на света.“

...

**Тодор Арnaudов: Сравни с ТРИВ, „Схващане за всеобщата предопределеност 3“, 2003 (СВПЗ, или Вселена и Разум 3)**

**40. (...) Според Схващането Вселената е БЕЗПОГРЕШЕН СМЕТАЧ, следващ БЕЗПОГРЕШНО ПРАВИЛАТА**, върху които е основано действието му. Затова ГРЕШКИТЕ В ДЕЙСТВИЕТО на производните сметачи (например човеци), т.е. възникването на ПОЛОЖЕНИЯ, които са НЕЖЕЛАНИ от тях, се възприема като "НЕСПАЗВАНЕ НА ВСЕЛЕНСКИЯ ЗАКОН" и носи неудоволствие, защото ЦЕЛТА на производните сметачи е да ПОДРАЖАВАТ НА ВСЕЛЕНАТА - ДА НЕ ПРАВЯТ ГРЕШКИ.

Ако някой "добре възпитан" или, казано по-цветущо и точно за случая, "дресиран", че като влезе някъде трябва да си свали покривалото на главата, така че да се вижда косата или голото му теме, ВИЖДА, ЧЕ ДРУГ НЕ СПАЗВА ПРАВИЛОТО, изпитва неудоволствие, породено от "дресировката", от РАЗЛИКАТА МЕЖДУ ОЧАКВАНОТО = ПРЕДВИЖДАНОТО И ЖЕЛАНОТО, И ДЕЙСТВИТЕЛНОСТТА.

РАЗЛИКАТА е "ГРЕШКАТА", която води до неудоволствие.

Виж също „Писма между 18-годишния Тодор Арnaudов и философа Ангел Грънчаров“, 2002.

Т.е. същественото е само **разликата** на предвижданото и желаното спрямо действителността (и/или възприятието), а не конкретното **съдържание** на предвиждането или възприятието.

В примера на Т.Скарф бих уточнил, че **не е само „мозъкът“**, а е цялата система, човек, ум и т.н., които в ТРИВ са все „управляващо-причиняващи устройства“ и разделянето им зависи от оценителя („наблюдателя“ във физиката и други школи).

Виж също, пак във СВПЗ, Т.А. 2003:

#### **„10. Какво значи "самобитен" творец?**

Създаването на произведение на изкуството представлява запис на свитък (информационна цялост, файл) върху информационен носител, който представлява посредник, междинна памет, между твореца и възприемащия творбата, включително самия творец, докато твори, или в някакво бъдеще време, когато творецът е

забравил творбата и си припомня нейни особености.

Пространството е Памет, затова "информационният носител" може да бъде толкова разнообразен, колкото разнообразни видове данни могат да се записват в пространството.

Оригинален е такъв свитък, в който се откриват по-малко от очакваните подобия или еднаквости със запомнени по-рано свитъци или отделни късове знание от паметта.

Оригиналността е способност да се ЗАТРУДНЯВА ПРЕД-ВИЖДАНЕТО на бъдещето въз основа на МИНАЛОТО. В случая "минало" означава част от свитъка, която е била прочетена преди друга част, която е бъдеща за миналата и която подлежи на предсказване въз основа на информацията, получена от свитъка до достигане на мига на "миналото".“

Срвни напр. с т.нар. Self-supervised Learning (машинно обучение, GPT модели и пр.). От ТРИВ виж също: „Анализ на смисъла на изречение...“, 2004, „Абстрактна теория за "изключенията от правилата" в изчислителните машини. Теория за управляващите устройства и управлението. (бр.29, апр. 2004) и др.

**Продължение с Берт Дефрейс, MLST, 21.11.2023:**

**Тема:** за йерархията... Архитектура за предаване на съобщения. Модел на „актьорите“ в информатиката (Actor Model), предложен през 1970-те години, реактивно програмиране; основа за разработка на мултиагентни системи [Виж също ROS – Robot Operating System, ROS2]

**1:12 ч.:** Текущият проблем на неговата група и въобще е учене на структурата на модела в движение (Online Structure Learning). Справят се с обновяване на параметрите. В представените научни статии внедряват т.нар. **Constrained Forney Style Factor Graphs** – Вид фактор графи с ограничения, които оформят целите и посоката на развитие на агента, който иначе би създавал изображение на средата, в която съществува.

**Тош:** Сравни с *Познавателните и Чувствените/физически „награди“* (цели, система) в ТРИВ, виж лекцията от курса от 2010 г.

... Важно в техните разработки и системите с AIF е да могат да се фокусират върху важните области на обработка и на средата, във всички измерения: енергиен разход, време, данни [точност] и пр. изчислителни ограничения, рамки, и да могат да бъдат прекъсвани по всяко време. Например ако на пътя внезапно изскочи пешеход, друг шофьор ускори не по правилата за движение и т.н.

Това може да позволи да се постигне пораждащ ИИ за полеви условия, който може да реагира своевременно, за да оцелее в реалния свят. Трябва да може да се променя с ограничен енергиен бюджет; да се създадат по-добри методи за разработка на алгоритми и програми, които се *самонагласяят*, но за да стане е необходимо обществото на Извод чрез действие да убеди дейците от областта на Оптималното управление (Control Theory, Optimal Control) защо няма нужда от функция на разходите (cost function) или възнаграждението (reward). Нужно е да се убеди и общността на разработчиците на софтуер – **не пишете програми... Направете ги така, че те да се развиват сами, да „растат“**. На ниско ниво да са прости предсказващи модели ... **Самоорганизиращ се софтуер\***.

**Тош:** На същото мнение съм от десетина години и съм работил изследователски в тази посока, но така и досега [2023 г.] не съм вложил достатъчно енергия и не съм събрал всички части в едно, за да го осъществя практически завършен вид за представяне. За автоматичното програмиране пиша и в класическата ТРИВ, „Зародишът на разум“ е същата цел. Един от моментите на осъзнаване беше в началните етапи на разработка на видеоредактора „*Tweekid FX Studio*“ (2010-2011), че не е разумно да се разработва ръчно. Тогава бях взел и не особено добри решения откъм гъвкавостта, поради други решения и ограничения, и/или недостатъчно търсене (C++, Win32, “Video for Windows VFW”, Direct3D9 заедно; по-добре беше C++ частта да е като сървър, да ползва ffmpeg (тогава не открих добра документация бързо), може би OpenGL (за да е по-лесен за портване под Линукс), може би някой енджин за игри и т.н. Една от идеите беше да има нужда от най-малък брой допълнителни външни библиотеки. По-късно беше включен и OpenGL, но само експериментално.).

**Едва ли програмистите са щастливи** от подобно развитие и от напредъка на пораждащия ИИ.

**Тош:** И аз работя без външна награда, хаха...

**1:31 ч.** Агентът взаимодейства, общува със средата, и спира когато не може повече да намали свободната енергия.

**Тош:** Може да се каже когато достигне определена разделителна способност на възприятие и управление, точност, резолюция, степен на съвпадение. [Бел. 9.8.2025: невъзможност да съгисти повече, граница на

компресиране; неделимост, дъно и връх (Зрим); виж бел. към приложение Алгоритмична сложност #complexity].

**1:33 ч.** Никой не поставя конкретния хоризонт на предвиждането, той се получава от взаимодействията и свободната енергия [грешките в предвиждането и поправянето им]. [Получава се] **„Дифузно планиране“**.

**Тош:** Пак се задава, дори да е неявно („неосъзнато с точната последица“) с изискването за гранично съвпадение/допустима грешка, и работата в определена среда; защото приетото изискване ще съответства на спиране на процеса при достигане на определено състояние – то може да е неизвестно по „осъзнат“ от потребителя на системата начин или по „изричен“ в неговите разбирания, но е налично в пространството от възможности, които деятелят изследва, изпълнявайки оптимизацията на процеса на „намаляване на свободната енергия“ чрез „извод чрез действие“.

**1:36 ч.** Вложените ограничения превръщат агента в **„автоматизиран инженерен процес“**. ... Например с така зададена схема от една страница може да се проектира прахосмукачка, с по-малко специалисти; без да са нужни инженери за всякакви подробности – моделът ще се донаглася сам.\* (...)

**1:52 ч:** Трудно е да накараш студентите магистри и докторанти **дори да започнат да се занимават с тези теми**. Представете си колко по-трудно би било\* [да ги убедиш] **преди 8 години...** [Ако бяхме през **ноември 2015 г.**] когато: (1) Никой не е разбирал парадигмата на Извод чрез действие (2) Щяхме да ги изработим чрез фактор-графи, които никой не е познавал добре (3) Щяхме да ги програмираме на Юлия/Julia, защото Питон/Python е бавен (4) Щяхме да използваме реактивно програмиране, което още не беше популярно.

**Тош:** **Сравни с Теория на Разума и Вселената/Вселена и Разум**, която описва подобни мултиагентни системи и идеи през **2002-2004 г.** в творбите на тийнейджър и „разбира извода чрез действие“ десетина години преди този „труден“ момент за започване според експерта в ИЧД, инженер и преподавател в силен технически университет.

Като цяло ВиР, в частност: „Анализ на смисъла на изречение...“, „Абстрактна теория за изключенията от правилата ...“, ... Това е утешение и (непомагащо) „извинение“ за стратегическата ми грешка и разсейване, че не започнах да осъществявам разбраното още тогава. Вложени многослойни предсказващи модели с всякакви РСВ и РСУ, модалности, видове сетива, гледни точки и пр.

Сравни и примера на Берт за самоуправляваща се кола, и човек относно мащабирането на ресурсите – в даден момент искаш само да



прецениш как да не се блъснеш, не те интересуват подробностите по формата на фаровете или как изглежда шофьорът; виж понятията РСВ и РСУ от ТРИВ.

**Тош: Критики:** че ТРИВ, както и FER/AIF (ПСЕ/ИЧД) са „*просто*“ вид учене с подкрепление (RL), повтарят тях, защото на места се говори за „удоволствие“, награди. Ученето с подкрепление от гледна точка на ТРИВ и моя гледна точка е изразено в *по-общ смисъл*, например като *постепенност, натрупване, взаимодействие и причинно-следствени връзки и предвиждане на бъдещето с различна разделителна способност на възприятие и управление и различни начини за разделяне на света и сетивно-моторните данни*, и е основано на изграждане на *причинностен* предсказващ модел с постепенно разбиране и с допустимост на евристики и *всякакви* методи. Кое и какво е „удоволствие“, влияе на или подбужда към определени действия се решава от самия агент: в мултиагентна система има различни видове, които работят на много вложени нива; както и различни видове, които си съперничат или се сдружават. В моя модел предвиждането на бъдещето е и основната, и познавателната движеща сила – универсалната награда и подбудител. И също така, както споменава и Берт: за да може да има *целенасочено конкретно*, по-„*тясно*“, поведение, е необходимо да има *ограничения*, дадени от първоначалното устройство на системата или нещо друго, както са първичните биологични мотиви при живите организми и човека; в противен случай умът би отразявал *изцяло* сетивните данни, би строил симулатор на вселената, но *без собствена* гледна точка – в такъв случай изборът на гледната точка се предоставя и отстъпва на друг оценител-наблюдател, който да си избира променливи и особености от предвидените, симулирани, емулирани и пр. въображаеми вселени.

Деятелят избира от подмножество и работи с определени ограничения на изчислителните средства и паметта, той е „подвселена“, макар че също една от основните му дейности е да строи предвиждащи модели на вселената от по-ниско ниво, на сетивните си възприятия и въздействията си. В „стандартното“ учене с подкрепление, RL, търсенето понякога изглежда абсурдно и ненужно случайно и хаотично, виж например агенти от OpenAI Gym/Gymnasium и др. в които човешки фигури безпомощно се мятат, докато се учат да ходят, или в игри на „Atari 2600“, в които изпробват хиляди и милиони неуспешни стратегии, напр. на играта „Breakout“, докато разберат, че е нужно *дъската – координатите* на управляваната „черта“ – = площ от пространството, област, която при докосване до топчето му повлияват като го карат да отскочи обратно към

целевата област, която трябва да се „разбие“ чрез последователно уцелване – да *съвпадат с координатите на топчето* (движещо се независимо петно, точка, с еди-каква си траектория: последователност от съседни координати, при определено търсене) и т.н. (...)

\* Към края: Книги, бейсови вероятности ... Прочел *през 2006 г. ...* (срвн. *годините* ТРИВ: 2001-2004 г.)

\* **Бел. 9.8.2025 г.** Проектирането на „прахосмукачка“ или друго чрез една страница с описание е подобно на това което се върши с големи езикови модели, „vibe coding“ и пр. Симулирането на Вселена, когато няма особени ограничения се случва неявно в пораждащите модели с изкуствени невронни мрежи от различни видове, от GAN, през преобразители, дифузни и пр., Близки и нагледни до основното разбиране за „физика“ са пораждането на изображения и видео от текстови и други описания (понякога се включва и движения на камерата, преден и заден план и скица на движението на обект и др., виж в приложение #Listove за мултимодални модели. Преобразителите бяха използвани за изображения, но дифузните се оказаха по-ефективни за по-високи резолюции; в последно време трансформаторите се завръщат.).

\* Виж още бележки и теми на английски по някои неразгледани по-горе откъси по технически въпроси: Търсене в дърво на вериги на Марков, търсене чрез пускане на лъчи на поредици от вероятности (beam search) и др.

## \* **Prof. BERT DE VRIES - ON ACTIVE INFERENCE, MLST**

### **Online structural learning**

**Selection and notes by Todor Arnaudov – Tosh/Twenkid**

... 1:13h - ... **Bert de Vries** ...

<https://www.youtube.com/watch?v=2wnJ6E6rQsU>

Machine Learning Street Talk | 90,4 хил. абонати

12 582 показвания 20.11.2023 г. SWINLEY FOREST

[33933 views/198K subscribers on 9.8.2025, ~ 2 years later]

**Bert: 1:19 h ... 10 M neuron fibers [sensory-motor], 80 B neurons**

*[estimations vary], 50 K muscle fibers – Human Control system ... Minimize the FE for the stuff that matters and interrupt at any time ... 1:23 Generative AI into the field .. power budget .. deadlines ... not predictable .. Everything can change in a limited power budget. 1:25: Convince control theory*

*community, optimal control ... Why don't you use cost functions or reward f. convince the hardware and software engineers, software ... We don't want to write procedural alg. we want to write self-organizing system... 1:29 Lightning ... the prediction error is small enough we can stop ... 1:30: MCTS (Markov Chain Tree Search)... Tree of thought A\* ... traverse ... beam search ... planning function ... select the best trajectory ... 1:31: We shouldn't set a horizon...[can't predict a long horizon if there's not enough information]; ... wait ..message passing [between agents]... find the opportunity to minimize this free energy and then it stops ... 1:32 ... we should get to the point where no individual agent actually plans. ... And, but the planning is diffused... So the planning is in the system ... bits ... develop this temporal planning capability **Nobody sets the horizon ... just the process that happens to put some information there***

***Tim: Planning diffusion ...***

**Todor:** it is **implicit** in this view, but as an abstraction it is still planning, on each level and range, which is connecting and extending each-other; not only the explicit one is such; Compare to the notes in Bulgarian above about the dispersion of the locus of control of the agents in the environment and not only in the individual humans etc.

***Bert:** 1:34 ... Divide and conquer ... dynamic programming – [Tosh: yes ].. simple set of rules ... every level, layer, node ... this system this system is just trying to predict its world and it's it gets corrected*

*1:35:36 ... by sensory inputs, whether they come from a lower layer or from other nodes or from sensors. It just. It makes predictions gets prediction errors, they get processed. That's it. There's nothing else going on, right? There's one twist and that is that there is encoded a certain goal of staying alive. Or what is success. **A set of constraints.** And that makes this active inference loop not an objective, let's say scientific inquiry loop, but it makes it an **automated engineer**. It makes it an engineering loop **because now there is a purpose. Automated engineer ... the actions are ... if there wouldn't be constraints, then you would just have a, you would build a model of the world .. optimal control***

**Todor:** Yes. Universal simulator of virtual universes. The constraints are a basic way to define the initial intentions, will, goals, limits on the low level or any level “implicitly”, which is however explicit if one understand where the combination of constraints will lead in particular circumstances.

***Bert:... 1:36:43: But because of the constraint, there is a bias. And now it has a purpose. It's still doing. The objectively best process, the Bayesian process. But now with this constraint, it turns into a purpose. It turns into a vacuum cleaner or an, you know, whatever it is, a drone. And so active inference to me.***

*Is an automated engineering process, which is why we can develop with our brains. Speech recognition algorithms and object recognitions and navigation algorithms on locomotion and all these things. They have a purpose. So the brain is an automated engineer now through an active interest. ... **Tim:** Robustness and Design vs Grow (....)*

***1:39:55:** reasons that the natural world is so robust is because the agents are themselves nested and everything is seeded from the microscopic scale. So even we as agents are composed of a hierarchy of sub agents. And in the artificial setting, as I just said, we fix a whole load of the goals and layers, but the agents are also atomic. ...*

**Todor:** This is explained in TOUM, exemplified also in “Abstract theory of the exceptions of the rules in computers”, 2004 – in the Universe Computer, the lower levels “consume” the undefined conditions in the upper level, the ultimate lowest level never “crashes”.

***Bert:** if you implement, if you take that as, let's say, a factor graph approach, .. **1:40:34** .. it's just a divide and conquer approach... a factorized model and each, each all the processing is just. Factors nodes sending messages to other nodes to two two connected nodes. That's the only processing that's going on*

**Todor:** See TOUM

***1:50:** I think AIF is really a **new** field with fantastic future. e. But **the implementation is something that people haven't really thought about yet, very deeply.** And the implementation is not not very simple. It's. So, so that's that's that may be a we need a lot more research to, to get this really working. I mean deployable agents that really adapt in the way that we want to. Right? I mean, yeah, yeah.*

**Todor:** It's not so new. See TOUM, 2001-2004.

As of practical implementation: that's possibly right.

...

## \* Скритата математика зад всички живи системи

### \* The Hidden Math Behind All Living Systems, Machine

**Learning Street Talk**, 147 хил. абонати

17 178 показвания 23.10.2024 г. Staff Favourites (26.10.2024)

"Dr. Sanjeev Namjoshi, a machine learning engineer who recently submitted a book on Active Inference to MIT Press"

Д-р Санджийв Намьоши ... книга по Извод чрез Действие ИЧД

**Подбор, превод и бележки от Тодор Арnaudов – Тош (ТА, Т., ТА):**

<https://youtu.be/hf18w6CuY8o?si=nQaZT5WYxfNHgtHr>

КФ – Карл Фришън. ПСЕ – Принцип на свободната енергия.

1:18:56 схема с историята на ИЧД: **Предходна литература:**

1990-те: Възприятието като извод/възприятието като заключение

Учене на представянния, Вариационно бейсово заключение,

Предсказващо кодиране (Rao, Ballard), хипотеза за "бейсовия мозък",  
методи за изобразяване на работата на мозъка: мозъчни скенери (MRI)

2003: Теория за коровите отговори

2006: ПСЕ Фаза 1 - непрекъснато пространство на състоянията (AIF Phase I)

2010: ПСЕ Фаза 2

2013: ПСЕ Фаза 3 - Производни области: Предсказваща обработка --  
Дискретно пространство на състоянията → ... Теория за невронната  
мярка (Neuronal gauge theory)

2019: ПСЕ Фаза 4 - Производни области: Бейсова механика Фаза I

2023: Бейсова механика Фаза II...

**1:29-1:32 ч: "Намаляване на неопределеността"** на състояния на средата (uncertainty reducing behaviors in different contexts ...) .. искаш да приложиш стратегии (to take policies) или да извършиш поредица от действия, чрез които да намалиш неопределеността на параметрите на модела [пораждащия вероятностен модел] ... кой модел ще намали най-много неопределеността в определена ситуация ...

(ТА.Зрим: ?В ДП, рзпзнвл-{K}, избртл-{K}, --'+ ... увеличаване на предсказуемостта: ТРИВ) ... изследването, изследователското поведение е страничен ефект на първичния мотив за поддържане на параметрите в желани граници (set points), например поддържането на кръвната захар в желани граници; например за да се открие храна, с която да се засити глада в бъдещето и пр. (ТА: [,,]+{P}?Т, .-?Т, --'?Т)

**2:04** ТС: ПСЕ/ИЧД е загадъчно, неразгадаемо поле за начинаещия в материята; четох статии на КФ от преди 15 години, въвеждат толкова много жаргон и терминология от различни области: невронауки, когнитивна наука, физика; част от математиката беше прекалено сложна; какъв беше твоят път ...?

**СН:** През 2018 г. ... постдок, занимавал се с невронауки, искал да спечели грант и да мине от научен институт в компания; открил статия на КФ от 2013 г., проучвал 3 г. ... ПСЕ по онова време било нишова област.

[ТА: и аз го открих в края на 2018 г. от статия в Wired, когато писах по въпроса в статията за "Окончателния ИИ", Ultimate Ai в *Изкуствен разум*. Възможно е да съм срещал и по-рано, но да не съм отделил достатъчно внимание, за разлика от тогава.]

– **Какви техники използват за симулиране?**

R (платформата и език за програмиране за статистически модели), ... POMDP<sup>39</sup> за моделиране (частично-наблюдаем процес на марков за вземане на решения)

**2:07 ч** ... Виждам ИЧД/AIF като революционен метод по начина, по който беше машинното обучение през 2000-те - много интересни идеи и научни статии, но "експлозията" все още предстои.

– **Как ще се обучават?** Като деца. (Виж по-горе интервюто с Берт Дефрийс (Bert de Vries) за ограниченията.)

ИЧД е подобна на машинното обучение с подкрепление/подсилване (RL, Reinforcement Learning); хора от МО сравняват ИЧД с нея; в МО обаче обикновено обучението е еднократно, статично, докато в ИЧД е непрекъснато и във всеки момент "убежденията", очакванията, моделът за света са променливи и се донастройват от всяко взаимодействие със средата.

– **Каква е основната разлика?**

**2:41:** ML, RL: двойки състояние-действие, докато в ИЧД (също ЗЧД, *заклучение* чрез действие): като извод, заключение, да се изчисли последователност или траектория на действието; път, път на движението (метафора от физиката); разгръщайки се във времето, за да достигне до някакъв вид крайно състояние; ... **стратегия** (a policy), но понякога е по-добре да го наречем **план**, за да се разграничим от объркването с ученето с подкрепление. ИЧД е **план за поредица от действия**, които да се изпълнят в даден момент от времето, въз основа на информацията и знанията на агента в настоящия момент. ... вариационната свободна енергия се използва за настоящата и миналата информация. Търси се най-малката очаквана свободна енергия за бъдещо планиране (сборът на

<sup>39</sup> [https://en.wikipedia.org/wiki/Partially\\_observable\\_Markov\\_decision\\_process](https://en.wikipedia.org/wiki/Partially_observable_Markov_decision_process)

свободната енергия на състоянията да се минимизира). Избира се от тези планове или поредици от действия. Много важна за тези агенти е способността им да търсят информация и да се оглеждат за нещо интересно и ново в средата им ... Така че ИЧД е планиране извършено чрез извод, заключение (пораждащ вероятностен бейсов модел). И поведението, което се поражда, не изисква настройка на параметри "на парче" (ad hoc) като температура и др. подобни (в големите езикови модели).

**Тош:** срвн.ТРИВ, „Анализ на смисъла на изречение...“, 2004 и др. При машинното обучение на невронните мрежи също се търсят подобни пътища, но по друг начин и „средата“ може да бъде „замразена“, или записана като снимка в набора от данни или неявно в средата и взаимодействията с нея при учене с подкрепление.

## Дискусии между Карл Фристън и Стивън Волфрам

К.Ф. към С.В.: ...динамиката, която е следствие на най-дълбоките нива на йерархията, ако си представите центробежна йерархия. Мисля че е важно да се отбележи, че не можете да се обвържете с единична времева рамка за да разберете контекстуализацията на бързопроменящо се високоразмерно съдържание в процеса на огрубяването му.

[Тош: намаляване на разделителната способност, обобщаване, отвлечане (абстрахиране) в по-горни нива/късни степени на обработка].

Също така има огрубяване и във времето, което наричаме „контекст“.

**Тош:** Аналогично на ТРИВ, че за сложни многослойни, многосетивни, многомащабни и т.н. УУ **няма единична „най-добра функция на наградата“**. Сравни с „Анализ на смисъла на изречение (...)“, Т.Арнаутов 2004 за запазване на съществуването в ПСЕ/ИЧД на школата на Фристън.

**Тош, 2004:** „Т.нар. "естествен" език, или Език на разума, както авторът на тази работа предпочита да го нарича, позволява нееднозначност може би защото при него има много "правилни" възможности; съществуват много случаи, в които всяко възможно решение е "правилно" в смисъл, че устройството (човекът), което го е взело, продължава да работи след прилагането му в действителността.

*Ако дадена машина продължи да работи - съгласно разбирането за "работи"; напр. сърцето\* ѝ да бие - за еди-каква си продължителност след извършване на еди-какво си действие, то действието ѝ е било "правилно", т.е. тя е следвала закони, които не водят до повреда или "излизане от строя".*

*По-сложните управляващи устройства (виж "Схващане за всеобщата предопределеност": <http://eim.hit.bg/razum>) имат по-голямо пространство на правилните решения; имат по-голяма "свобода", т.е. възможности за бъдещо действие, след което да продължат да работят правилно.*

*Нищо не е "безсмислено", колеги..."*

*\* сърце – тактов генератор*



## \* Разговор между Адам Омари, Рой Баумайстер и Майкъл Левин |

### \* Conversation between Adam Omary, Roy Baumeister, and Michael Levin

Michael Levin's Academic Content

4,88 хил. Абонати 1042 показвания 7.11.2023 г.<sup>40</sup>

<https://www.youtube.com/watch?v=UfoVTA0ilsY>

Подбор, превод и бележки от **Тодор Арnaudов – Тош** (ТА, Т., ТА):

Едночасова дискусия по въпросите за умствените способности на групите<sup>41</sup> и възможната приложимост към икономически и обществени въпроси отвъд нивото на единичен член на общността.

33 мин: ... Всички симпатизираме на панпсихизма, но може би това не се отнася за държава ...

35 мин: вътрешна перспектива .. Може ли да бъдеш държава<sup>42</sup>? ..

44 м. ... физически носител за свързване ... обмен на голям обем информация ...

45:40: синхроничност ...

46:10: ако материализмът е достоверна теория, следователно САЩ има съзнание ... какъв вид системи имат вътрешна перспектива.

[Тодор: САЩ като единно цяло, обект, система, съществува **в ума на оценителя**, поне там може да бъде осъзнато от него самия. Какво е “вътрешна перспектива”, как точно се определя, кой го прави, как, какво се оценява. “Вътрешно” като местоположение или като “субективно преживяване”. Ако е наистина субективно, значи е несподелимо и не може да се отговори и т.н. И тук, както в други случаи, разговарящите не стигат до извода, че оценката минава през *техния ум*, и че *понятията*, “*вещите*”, “*одушевени или неодушевени*”, са *представи в техния ум*, те са “одушевени” чрез *тяхната* одушевеност. Човек може да си представи и да възприеме *всичко* като одушевено, това е ежедневие в детска възраст или в метафоричния език, а също така и да обезчовечи или да се отнася с „одушевени“ според собствените му критерии същества като с

---

<sup>40</sup> Към 9.8.2025 каналът на М.Левин вече е с над 30,2 хил. абонати.

<sup>41</sup> Collective intelligence – “колективна.” Не ми звучи добре. Може би *общностна*? *Обща*? *Съвместна*?

<sup>42</sup> „Country”; на английски и на български има разлика в ударението на „държава“ (държането, удържането, управлението, властта); на английски се разбира държава и като управление, но е и „безизразно“ „страна“, което важи и за географска област – може да е и ненаселена, - а “country” значи и „провинция“, селски райони.

неодушевени. Също така, от панпсихизма не следва, че той е еднакъв (или че има един вид), или че всички *понятия* имат съзнание, без да са свързани с някакъв специален носител.

Виж следващото *Приложение* в което се разглежда по-подробно относно **панпсихизма** през Бернардо Каструп, Дейвид Чалмърс, Томас Кембъл, Доналд Хофман, Филип Гоф (Kastrup, Chalmers, Campbell, Hoffman, Goff)]

## \* **Levin 's Academic content youtube** ... 9/11/2023, 14:03-04

**"USA is conscious"** (panpsychism etc.\_

**Тош:** Сравни с Обществено съзнание от диалектическия материализъм, „Societal consciousness ...“

**Todor:** It "is", because USA is a unified, complete, individual, localized, definite etc. entity is a representation, model ... in a mind. It exist as such in the mind. With a different timescale, etc. The same goes for all kinds of inanimate objects which are **mindified**\*. There is temporal and spatial **lag** ... Often cited:

**"spooky action at distance", "not natural", "strange"** (Isaac Newton also wasn't happy with this). However:

– **Why not?** It is natural (something moving everything backstage etc.) sp. + A big distance in one representation or view may be equal to **0 distance** in another view. See TOUM, Universe and Mind 4. The "distance" is actually an accepted/decided form of **connectedness/disconnectedness**, the confused observer doesn't see the correct connection and assumes there can't be one.

\* **Mindification** – see other places in this appendix to The Prophets of the Thinking Machine, #panpsychism in #Listove and *Universe and Mind 6, T.A. 2025* and the following article: "How do we exist in the Universe?" ...

## \* Как съществуваме във Вселената?

### Фристън и Волфрам

#### – HOW DO WE EXIST IN THE UNIVERSE?

<https://www.youtube.com/watch?v=6iaT-0Dvhnc>

Machine Learning Street Talk 106 хил. абонати (2.3.2024)

46 608 показвания Начало на премиерата: 29.10.2023 г.

**Подбор, превод и бележки на Тодор Арнаудов – Тош (ТА, Т., ТА):**

Атоми, вещи, деятели (агенти)... наблюдатели\*. Какво всъщност въобще определя „наблюдателя“ и какви свойства трябва да притежава? Как вещите, предметите, обектите запазват постоянство на съществуването във Вселената, при все че материалният им състав се променя във времето? Какво значи нещо да бъде нещо? Дали нещата са нещо по-висше от физическата действителност на ниско ниво? Какво значи нещо да е „одушевено“ – да е деятел/агент? Каква е разликата между одушевена и неодушевена сложна динамична система? Може ли камък или автоматична котешка вратичка с ИИ<sup>43</sup> да бъде одушевена? Може ли Вселената да бъде разделена на отделни агенти, или одушевеността е разпръсната (diffused)? Замисляли ли сте се някога върху тези дълбоки въпроси относно действителността?

#### Съдържание (...)

01:17:35 – Себемоделирането дава одушевеност

01:22:11 – Откъде да знаеш, че нещо е одушевено/деятел?

(How do you know a thing has agency?)

01:33:33 – Дълбоката връзка между динамиката, рулиадата и ИИ

(Deep link between dynamics, ruliad and AI)

01:36:37 – Дали от одушевеността/деятелността следва свободна воля?

Определение на одушевеност (Does agency entail free will? Defining Agency)

01:43:42 – Къде да измервам, за да открия одушевеност?

(Where do I probe for agency?)

(...)

Около 1:31 ч. ... Разделяне на времеви мащаби, нива, ... По-високите са по-трайни, дават контекст/обстановка ... намаляване на разделителната способност, което наричаме контекст<sup>44</sup> ... Ок. 1:33 ч – Как да различа в една купчина от битове, че в тях се съдържа нещо като наблюдател [деятел, одушевен], а не просто вещ, и да го сторя по по-„високотехнологичен“ начин“? Знаеш, вече съществуват системи с ИИ ... езикови модели, ... Дали са одушевени, ... нямали точно нива, само предаване в права посока, (...)

<sup>43</sup> Котешка вратичка - „cat flap“, в по-голяма врата за човеци. Популярна в къщи в САЩ.

<sup>44</sup> 1:30:59 ... a coarse graining over time, which we call context

**Тодор Арnaudов:** При всяко съждение дали дадено “нещо” “е” каквото и да било: одушевено, неудошевено, “купчена битове” или “нещо” и пр., се минава през ума и оценките на оценител (Фристън, Волфрам, Левин, Фийлдс и физиците го наричат “наблюдател”, но според мен и ТРИВ по-правилно е *оценител (evaluator на англ.)*, за да се подчертае, че участва, че *действа*, че изпълнява *активна* или *активната* роля в процеса, във взаимодействието, както се твърди и в квантовата физика: наблюдателят и/или измерването влияят и променя изхода на опита. Дали нещо е “само битове” зависи от *избора, вниманието, разделителната способност, решението* на наблюдателя, а не само от съответното явление, “нещо” и и пр. Компютрите **не са “само единици и нули”**. Езиковите модели не съществуват в нищото и за да се осъществят в действие включват и апаратна част, и други програмни средства, и въобще средата и вселената. **Абстракцията** се извършва в ум, но и умът и „сам вижда“, че има и части, които не се вписват в абстракцията или той самият не може да управлява, измери и т.н. без посредници или други технологии или дори и с тях все още не може.

Въпросът **“как се извършва абстракцията...”** – отново оценителят може да избере и да реши. **“Как решава?”** – трябва да построим конкретен и изричен път, и в него това ще е вярно, но в друга гледна точка или други моментни решения – ще е друго. Разделянето на одушевеност и неодушевеност е условна. Майкъл Левин се опитва да ги разграничава на степени чрез **“Технологичния подход към разпознаването на наличие на ум навсякъде”** (Technological Approach to Mind Everywhere, при който сравнява различни способности между различни “обекти”. Едно решение на въпроса е, че **“всичко”** е одушевено, от най-простите частици, всички са вид деятели и имат воля, памет и работят по същество по общи принципи, или споделят нещо основно, „basal cognition”, „basal agency” и пр.; т.е. по-висшите, по-сложни, по-големи, съставени обекти от по-малките основни части, запазват най-основния или някакви основни принципи, свойства, особености, едно от които може да бъде някаква форма на „съзнание“, „одушевеност“, „усещане“ или да е „сила“, която участва в образуването или проявата му в по-сложна или „висша“ форма в по-„развитите“ същества. Такава е гледната точка в ТРИВ, като един или някои от обектите може да се окажат *процесите, връзките, промените*, които е възможно да съществуват или да се отчитат в някакво представяне във вселената сметач като действителни изрично адресируеми „обекти“ и може би всъщност в тях се състои „одушевеността“ и способността за „чувстване“, както деятелите се

определят чрез способността им да **действат**, със “собствената воля”<sup>45</sup> “*Всичкото*” обаче може да бъде и *само* най-простите частици, буквачета, включвайки въпросните *вързки*, а в по-големите времепространствено определени структури да се изразяват *същите* основни “психични”, “субективни”, “мета-...” процеси, но с по-голяма “дълбочина” на стъпки, степени на предвиждане на бъдещето, отдалеченост/съвпадение на влиянието<sup>46</sup>.

(...)

С.Волфрам: 1:43:XX: Как да каже, че нещото мисли контрафактично, хипотетично... То е **просто купчина от битове (bag of bits)**... Как мозъкът работи като междинно ниво между възбуждането на невроните и изречените думи ... 2:00: ч ... ако разбирам правилно и понятието за „факторизиране“, разделяне на буквачета на влияние [на системата на части, factorized ability, си мисля, че в някакъв фундаментален смисъл, онова което ни казва Рулиадата е, че *няма способност за окончателно факторизиране*. Никога няма да има. Единственият въпрос е за нещата, *които си избираме да наблюдаваме*. ... 2:10: Има ли окончателна факторизуемост...

**Планиране** и одушевеност: първите 1-2 минути, 1:22, 1:35, 1:43, 2:08 - ...

\* Сравни с ТРИВ и „Анализ на смисъла на изречение ...“ за причинността с това как причиняването на събитията може да се обясни с огромен брой различни причини, ако се работи с по-високо от най-ниското ниво въображаема вселена.

\* **Букваче** – елемент, юнашко наречие

**Т.Арнауудов:** Термостатът не бил планирал, затова не можел да има съзнание (т.е. обаче ако *нещо* планира, значи може?); сравни също Оги Огиас\*, който е свързан с школата на ART - Adaptive Resonance Theory на Стивън Гросбърг: по тяхната логика насекомите нямали съзнание, защото нямали *мета-познание*, работели само рефлексивно, вид стимул-реакция; трябвало да има горно ниво и долно ниво, които да „резонират“ (в ТРИВ: съвпадение, повторение).

Какво е „**нещо**“ обаче? Кое е агент и кое не е, къде е границата. Как се разбира дали така отделеното „нещо“ планира или не. Кой решава и отсъжда и защо?

---

<sup>45</sup> Или както биха казали много философи: „свободна“ воля; с.в. обаче е заблуждаващо и зле определено понятие за баламосване на публиката, по което са обсебени англосаксонски, европейски или всякакви философи; възможно е и да се преструват и да преиграват..

<sup>46</sup> Виж подзаглавие и една от темите на *Вселена и Разум 4: Произход на повторенията на събития на различни равнища на Вселенския сметач ... нецелевите, за преживяващия ги разум, съвпадения с ниска вероятност*.

„Термостатът“ е онова, което е и каквото е като *цялост*, като „термостат“, а не нещо друго; с определени граници, начало, край на „тялото“, признаци, размери, свойства и пр. **във и чрез ума на оценител**, който обхваща достатъчен или подходящ и **избран от него** обхват от възприятия и особености, и ги гледа през *избран от него филтър*, така че *той* да ги прецени като едни или други, а не трети или четвърти; като „одушевени или неодушевени“, „дейци (агенти/субекти) или обекти“.

Сравни също с твърдението от класическата ТРИВ\*1, че Вселената като цяло – на дадено най-ниско ниво – може да изчисли само следващото състояние с най-висока РСВ и РСУ, за разлика от по-горните степени, които предвиждат и съответно планират повече на брой стъпки и времеви обхват, но с по-ниска РСВ/PCY или в ограничен обхват от някое от пространствата. Нещо подобно може да се каже и за подвселените, дейците, агенти, управляващо-причиняващи устройства, които са части от вселената. За човека, който е вид деец, също може да се каже, че „не планира“, ако се разгледа по подходящ начин с определена по-ниска разделителна способност, а също може да се определи и че въобще *не съществува като едно цяло*, ако няма достатъчен обхват на наблюдение във времето или пространството, ако измерванията са прекалено отдалечени, ако нямаме достатъчно памет, за да съхраним междинните състояния и да ги разпознаем като подобни или разликата между тях за „достатъчно малка“ (или решим, че *не е достатъчно малка*, за да го класифицираме като „свързано“ или „същото“ и т.н.)

Обаче отново: **какво значи да планира**, да „мисли“ хипотетично („контрафактично“, counterfactual), да има *модел* на себе си, да се себепредставя (самомоделира), да „рефлектира“, дали даденото устройство, деец **„знае“**, че планира и **какво значи да го знае** \*2, как се измерва или открива надеждно от самото него и от различни видове други оценители-наблюдатели (например бебе, дете, всъщност също планира, но може да не може да го обясни или „осъзнае“, като под „осъзнае“ някой оценител може да е има предвид да го покаже по начин, по който той самият, другият, външният, *оценителят*, смята че „трябва“ да му се покаже – а всъщност бебето или агентът може и да *знае*, че планира – каквото и да значи „да знае“ обаче – едно ли е за всички видове дейатели, с всички степени на развитие, обхвати на умствени процеси, разделителни способности?... Що се отнася до себемоделирането, моделирането на агента от самия него, доколко този модел всъщност е наистина модел на себе си, доколко е пълен, пълноценен и кой го оценява – мисловен схематичен модел или подробен модел.

В логиката на ТРИВ тялото или „хардуерът“, или въображаемата вселена/ управляващо-причиняващо устройство/система като предсказващ модел на средата или като средство за „самосъхранение“, за „самодоказване“, което „самОпредвижда“ съществуването си, запазването на целостта си, по школата на ПСЕ/ИЧД, не може да има *пълнен модел*, представа за себе си, с разделителната способност на възприятие и управление на най-ниското ниво на въображаема вселена, в която е осъществена, защото всяко по-горно ниво работи на по-ниска разделителна способност, за него са валидни, разрешени само част от възможните състояния, явления, особености на по-ниското ниво; по-горните нива отчленяват, дискретизират, компресират данните от по-ниските, по-ранните, предходните степени. В примерите от ТРИВ от първите статии: съзнателно управляваната извеждана „смислена“ информация от човек, измерена в писане на текст на клавиатура или движения, е в рамките на битове или десетки битове, сравнена с невъобразим обем, необходим за описание на веществените състояния и преобразувания, които обуславят тези нищожни битове. Агентът може да има пълнен модел с най-висока точност, ако се вземе представянето му на машинния език на вселената от най-ниско ниво, където е записан, но тогава възниква предвиждането най-много на непосредственото следващо състояние на съответното ниво, например състоянието на атомите и молекулите в следващата минимална частица от секундата и пр. Може да добавим и идеите за *разпределено представяне на дейците* като *полета* в средата, споменато по-рано в книгата – за да предвиждате точно бъдещето си зависите от средата, част от която са или могат да се разпознат като други дейци, до чиито състояния и намерения нямате достъп с необходимата точност, достоверност, своевременност, скорост на осъвременяване на информацията. Съответно бъдещето ви зависи и от тях.

И тук важат мислите за това дали „САЩ“ или „камък“ може да има съзнание: без значение от други допускания, камъкът или какъвто и да е предмет или „обект“, за който се мисли, със сигурност *може* да има „съзнание“ тогава, когато е *представа* в ума на оценител, за когото се приема, че има това свойство.

(...) Виж и: *Вселена и Разум 6 (Universe and Mind 6)*

\* **Оги Огиас, Стивън Гросбърг** – виж списъка със съвпадения и изброяване на учени и школи в началото на основния том на „*Пророците на мислещите машини*“

**\*1 – Допускането от класическата ТРИВ**, че на най-ниско ниво или Вселената като цяло може да предвиди или по-скоро да изчисли, породя,

само непосредствено следващото състояние, може и да не е съвсем вярно. То е вярно в предполагаеми системи, в които това изглежда логично като основа, най-просто положение, буквачета, казби, инструкции, върху които се гради останалото. Това е вид изчислителна несъкратимост подобна на Волфрамовата (computational irreducibility) и граница на алгоритмична сложност. Ако можеше да се предвижда повече напред, тогава може да се допусне, че това не е най-ниското, просто, неделимо ниво на представяне, следователно може да се съкрати още.

Най-ниското ниво може да е различно за различни оценители-наблюдатели или управляващо-причиняващи устройства. От гледна точка на определен деец, дадено предствяне и състояние, което според него е на „най-ниско ниво“ може да определи само следващото състояние. Например това може да е състояние на паметта и на регистрите на процесор, „контекста“ на машината. Самият процесор обаче може да има *предвиждане на преходите, изпреварващо изпълнение*, а също така ако някой има по-широко и всеобхватно знание, спомени за минали изпълнения при определени състояния, то може да може да предвиди с по-дълъг хоризонт дори и с максимална разделителна способност на възприятие на машинния език на дадената вселена. Виж различните основни начини за предвиждане на бъдещето в СВП2, „Писма между 18-годишния...“.

**\*2 – Деецът може да „знае“, че планира или предвижда**, и че го е направил успешно, например тогава, когато получи потвърждение за предварителните очаквания и т.н. с очаквана/желана/достатъчна точност, разделителна способност и пр., подобно на функцията на загубата или цената в машинното обучение. Но дадена система може да извършва подобни действия и да „не знае“. „Знанието“ или „знаенето“ (*знаЕНЕто*) минава през „умотворене“, „умоплащение“ – mindification.

Възможно е също да има и заложили вродени режими на предвиждане и планиране и „знаенето“ да е включване в тях. Отново обаче е нужно умоплащение.

**\* Dileep George: “Ingredients of Understanding”, 2023 – Езикът е свързан със и използва симулации, модел на света и т.н.**

<https://blog.dileeplearning.com/p/ingredients-of-understanding>

Сравни с ТРИВ и с „What’s wrong with NLP: ...“, Part I, Part II, Artificial-Mind 2009



## **\* Обратно броене до създаването на Универсален изкуствен разум:**

**42% през март 2023 (трансформатор, ГПТ-3, TPUv4, H100, ChatGPT с тяло, Palm-E), Алан Томпсън**

**\* Countdown to AGI: 42% in March 2023 (Transformer, GPT-3, TPUv4, H100, ChatGPT embodied, PaLM-E)**

<https://youtu.be/qPI8fB2XL3w>

Подбор, превод и бележки: Т.Арnaudов

Да започнем с август 2017 г., когато Гугъл случайно откриха архитектурата Трансформаторите (преобразители) и ни дадоха понятието да поглеждаме назад и напред около думите в текст, и да можем да предвиждаме каква ще бъде следващата дума. За мен това е началото на Общия ИИ. Всички големи езикови модели днес се производни на първия Трансформатор. Това ни придвижи на 20% от пътя. ... Може да проверяваме думите назад и напред и да обучаваме модела върху много милиарди буквачета\* и така става все по-добър.

Как стигнахме от 20 до 30? Три години по-късно, ноември 2020 г., няколко месеца след излизането на GPT3. Тогава експертът, който е може би един от върховете 5% в света, Конър Лийхи (Connor Leahy) се изказа за GPT3: Само искам да подчертая, че днес вероятно има милиард души, които мислят, че са експерти в ИИ\*\*. Конър има дебела биография, да започнем с GPT2 през 2019 г., той пресъздава GPT2 без да има достъп до него ... и стига до положението самите OpenAI да се срещнат с него и да му кажат да почака, защото не можело да го публикува. По-късно той успява да възпроизведе и целия набор данни на GPT3 чрез EleutherAI. ... GPT-J, GPT Neo X20B <https://blog.eleuther.ai/announcing-20b/> и др. ... и сега е изпълнителен директор на фирма, заминаваща се с проблема за Съгласуването на ИИ с човешките цели (AI alignment).

**"Конър Лийхи (Connor Leahy): 11/2020 г.: Мисля, че GPT-3 е Общ ИИ, че е умен, колкото човек. Всъщност според мен е по-умен човек по ограничен начин; по особен начин. Според мен GPT-3 притежава по-"чисти" умствени способности от човека. Според мен човекът се опитва да наподобява онова, което прави GPT-3, а не обратното." ...**

След това и техниката, с която се обучава ИИ, беше създадена от ИИ. TPUv4 на Гугъл, юни 2021 - 21%.

**Година по-късно, 2022**, една от "приятелските" компании на Гугъл - "Дийпмайд" представи първият агент с общо предназначение: **GATO**. Той има само 1.18 милиарда параметри, но мисля че ни придвижи до **39%**,

защото GATO върши повече от просто "мислене": той управлява ръка на робот, да играе на игри, търси изображения, извежда текст и т.н. Това е **първият общ агент** и за мен беше вълнуващо преживяване.

Компютърната архитектура на NVidia Hopper, H100, от юли 2022 г., също е проектирана от ИИ. ИИ се самоусъвършенства: проектира собствения си мозък\*.

**7:31:** След това ново откритие от Anthropic: RL-CAI 52B. Те използваха ИИ с трансформатор, който да подава подходяща обратна връзка за обучение на друг модел - обикновено за тази цел се използват хора. – Февруари 2023, 41%: Майкрофост вгради ChatGPT в работи, които могат да взаимодействат със света: "Чат Джи Пи Ти веднага програмира повечето от главните модули на алгоритъма. Задачата изискваше известен разговор с човека, но бяхме изненадани от способността на ChatGPT да прави подобрения на кода на уточнени места само с езикова обратна връзка. "

\* Пример за работата на системата: *"Намери и се придвижи до бутилка"*

\* **Февруари 2023, 42%:** Гугъл с модела им PALM 540B, върху който е поставен Зрителен трансформатор (Vision Transformer ViT) с 22 милиарда параметъра. Най-големият модел с "плътни" тегла, **562 милиарда** параметъра, който може да извършва: управление на ръка на подвижен робот, боравене с предмети върху маса, Планиране на задачите и движенията, Отговор на зрително-зададени въпроси.

...

Ако гледате клипа **след март 2023**, значи вече е остарял!

\* Статия: **"Предсказване на бъдещето на ИИ чрез ИИ: висококачествено предвижване на връзките в експоненциално нарастваща мрежа от знания"**, 23.9.2022. arXiv: 2210.00881

Брой на научни статии на тема: "ML+AI" на месец в arXiv: 4000 на месец. Броят се удвоява на всеки 12 месеца. Средата на числото 1994, 2007, 2021: около 13-14 години = ок. 153 пиксела, 1

...

\* **Бел.** 9.8.2025: В общ смисъл не само съвременните невронни мрежи са „ИИ“, и компютрите или предходните технологии отдавна или **винаги са участвали в „проектирането на собствения си мозък и тяло“** в следващата версия, пряко или косвено, но с етикет не „ИИ“, а „компютър“, „електроника“, „технологии“, „знание“ и пр. Виж „Човекът и Мислещата Машина: ...“, 2001; „Следващото еволюционно стъпало“, 2002 и пр. и „Summa Technologiae“, Stanislaw Lem 1964. Виж приложението за научна фантастика, футурология и кибернетика на „Пророците на мислещите

Пророците на мислещите машини: Приложение Ирина: Беседи и записки ...

машини“ #sf #cyber; “Вселена и Разум 6“, „Нужни ли са смъртни изчислителни системи...“ и др.

## \* "Вероятностен папагал"

(Творчеството е подражание на ниво алгоритми...)

Откъси с мисли на Тодор Арnaudов – Тош от 25.5.2004 г. във форум „Кибертрон“ и др. виж в Подробна хронология на „Пророците на мислещите машини“; ?Т #podrobna

Виж: <https://www.oocities.org/eimworld/4/30/toshtrafic.htm>

**Тодор Арnaudов, 25.5.2004 г.:** *Силно се съмнявам, че разумът създава нещо - и най-големият творец е "папагал", само че знае по-хитри начини да "папагалства" и да разбърква "смислените" входни данни, така че да извежда нови, които също да бъдат "смислени", та даже и "по-смислени" от предходните, отколкото по-слабия творец – по-слабият творец е по-прост "папагал". По-добрият творец (разум) има повече използвана памет, по-добре търси, по-добре намира, знае по-добри начини да поражда правдоподобен изход от правдоподобен вход, но по същество е "папагал".*

*Развитието на разума може и да не е изграждане на система по същество. Системата си съществува, но паметта, задаваща конкретните ѝ действия е празна, като натрупването на данни създава впечатление на "строење на система"; това впечатление се създава от сложността на системата и от обема на данни, с които се описва поведението ѝ. Впечатлението за строеж на системи може би се създава и от конкретния начин на работа на мозъка, който е обременен от много странични ефекти на белтъчната технология, поради които натрупването на опит е свързано с физическо натрупване на материя и промяна на "физико-химически" строеж. Изчислителните машини са по-висши от мозъка..."*

*Развитието на разума, според мен, е въвеждане на входни данни - обикновено трупане на опит; на АЛГОРИТМИ (прякопроцедурни данни) и непрякопроцедурни данни; развитият човешки разум е същото като неразвятия, само дето е претъпкан с входни данни, които направляват поведението му.*

*(...) Нима не помните детството си? Обикновено "възрастният умник" е "умник" още като бебе - по-бързо се научава да говори, говори неща които 'не съответстват на възрастта му' и т.н. Затова смятам, че развитието на разума е добавяне на входни данни, но не и промяна на същинския начин на работа на разума.*

*Затова подходът за създаване на разум чрез много сложни програми според мен би довел до резултат, ако създателят на сложните програми е достатъчно сложен. (...)*

**Бел. 8.8.2025:** Сравни със съвременното машинно обучение и големи езикови модели, което доказва тези предвиждания от една гледна точка, независимо дали обучаващите се системи, създадени по този начин чрез преобразители,

учене с подкрепление и пр. се смятат за „наистина мислещи“, „само повтарящи миналото“ и т.н. **„Паметта“** в началния етап са състоянията на теглата на невронните мрежи, които се изменят след обработка на обучителните примери или взаимодействието със среда, диалози, итеративно самоподобряване с обратна връзка и пр. Моделът е тегла, числа, а структурата се проявява при заключения (inference): според това което е научено и стойностите на теглата в нея и тези на входа, „контекста“, подканата (промпта), обучената система връща едни или други данни, които ако се групират, обобщят и пр. би могло да се предположи че следват определена логика и т.н. Това може да се създаде и изрично, както и е възможно да се изграждат и явни структури, напр. чрез архитектурата на мрежата, като се добавят, премахват или преобразуват слоеве, като се пораждат нови невронни мрежи или други модули и се свързват към съществуващите и т.н. Други видове мрежи, които не са с твърдо фиксирана топология и структура, и не са в една посока като обичайните изкуствени невронни мрежи, се наричат от някои учени *„невронни лабиринти“*. Могат да бъдат и *живи, растящи, динамични* и пр. мрежи.

Относно създаването – за да се отсъди, трябва да се определи недвусмислено **какво значи „да създава“**. Кога се брои за създаване и за кого, и/или да определим степени, мярка. Кое и кога е ново и колко ново, какви съчетания от стари неща минават за нови. Творчеството и оценката му минават през оценител-наблюдател и през достатъчно пълно и прецизно определение на способностите, паметта и знанията му. Този въпрос е разглеждан на други места и в други приложения на *Пророците на мислещите машини*, и ще бъде разрешен и осъществен напълно в продължението: **„Създаване на мислещи машини“**. Виж също обясненията ми в дискусията от 2005 г. по-долу: *„Предвиждания за автоматичния превод и машинното въображение ...“*

## **\* Човеците са много по-зле от езиковите модели по много показатели, Тош, 2.8.2025**

Български – засега ползвайте машинен превод или английския текст. **Накратко:** Фонологичен цикъл от 2-3 секунди или няколко думи, работна памет от 7+-2, предвиждане напред или планиране на изказването само от една или няколко думи. Обем на запомняне на минал или какъвто и да е текст или данни, сравнени с на практика неограничения обем на сметачите. Увереността на човеците в твърдения, за които нямат достатъчно достоверна информация както и целенасоченото заблуждаване и мамене на другите – хората също „халюцинират“, често злонамерено. (...)

**\* 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> Todor Arnaudov, 2.8.2025  
Artificial Mind (The Sacred Computer, Свещеният Сметач)

– In AGI, AI, Analysis, Articles, Cognitive Science, Comment, Computational Creativity, Computational Linguistics, LLM, Thinking Machines by **Todor "Tosh" Arnaudov** - Twenkid // Saturday, August 02, 2025 // Leave a Comment

Comment on: wendywee — 9.7.2025 г. 6:41 Discord MLST (Machine Learning Street Talk Youtube Channel); discord channel #selfpromotion ...

**[A diagram about LLMs explaining why scaling won't lead to AGI, there's no world model etc.]**  
(...)

While it is generally true that "bigger is not always better...", the "scaling question" was asked back in 2001 (Man and Thinking Machine... - will more transistors, computing power etc. lead automatically to "thinking" etc.) and I myself am a proponent of and work on more conceptual ("symbolic") causal approaches, which don't need to read the whole Internet, it is more complex than that:...

**Tosh** — 21.7.2025 г. 23:01 (@ Discord):

The need to think with world-models (hierarchical simulators of virtual universes at different resolutions of causality control and methods for segmenting/granulating the sensory-motor data etc.) was defined in AGI in the early 2000s, and then rediscovered in the mainstream for the LLMs. However the scaling of LLMs from a long time is not just GPT1-GPT2-"Megatron" -GPT3, it's more than just adding more data. It got hyper multimodal, ensemble models, it includes robot data and physical interactions, RL, RLHL etc. Also the big platforms are unlikely just a single LLM, they are probably multi-component systems, which can include whatever "tool-use", simulators (another branch: Simulation Intelligence), physics-informed deep learning, game engines (=world models). Some foundation models from the last 2-3 years became also "generalist Agents" and they already do have practical "world models" to different precision, span etc.; JEPA-2 as well (although 1 Million hours video is way too much and it is still "dumb" ). "World model" is ambiguous, in the classical theories 20-some years ago there was the concept of "resolution of causality control" and the good LLMs actually \*do" have an implicit "world model" with linguistic resolution for many cases. Not perfect, not complete etc. "World modelling", physics, simulations is about predicting what will come next, given particular known preconditions and consequences, that's the causality, i.e. the mechanism is the same, but with LLMs it is not always predicting from the correct premises and don't follow the right laws/reason.

The late popular meaning of AGI about "economically feasible tasks", "better than the best human" (in these tasks), was promoted by the business opportunists. Originally AGI was about universal learning, prediction and compression in all modalities with a general algorithm ("next token prediction" was predicted by the AGI); i.e. in many ways this already happened.

--- END OF QUOTE ---

## **1. Humans demonstrate the same faults like LLMs, but worse**

The LLMs are "vulgarized" and some of the faults of the models are **the same** in human capabilities and behavior, however the humans don't understand them or they lack integrity in order to admit it.

## 2. Phonological loop of 2-3 seconds - conscious back context window of just several words

A few humans can remember more than 2-3 seconds in the past in their "phonological loop" - listen to a podcast or video clip, interrupt it and write down what you heard in the previous seconds; on which word you start to hesitate *"was it this or that word?... was it a car or a vehicle?"*? When you start to lose the correct word order? On which word you completely lose the thread and remember only that "you spoke about processors, the party, the movie, ...".\*

The persons who can remember slightly more are yet unlikely to remember perhaps even just 10 or 15 seconds or these ones would be rare "freaks".

So how big is your **"context window"** then, counted in tokens comparable to LLM's? This is for the past, where the LLMs and computers can remember not just thousands, but even **EVERYTHING, billions or trillions of tokens** exactly (as memory).

Let's look **forward** now.

## 3. Next token prediction for humans - context window for future tokens is even worse

How many "tokens" \*ahead\* can you predict correctly in your \*own\* utterances, in plain text and words? How well can you plan your sentences when you write on paper or on the computer?

One word, two words, rarely 3-4-several words? Can you remember in your working memory a long complex sentence of 15-20-25 words? So it seems that you're a 3-4 gram model. Besides, unlike the LLMs, which have the "log probabilities" and can perform various kinds of "sampling" or "beam search" and sample differently, you usually don't or it would be too hard for you and too slow to write down many variants, so you wouldn't generate even two or three possible variants of your own "next word prediction task", except if you are a writer, you compose a poem, a story etc.



I myself do this, but it takes time: you first generate an initial version of a verse or a poem, then you replay it many times and let the time pass, and in minutes, hours, days you start to find better thoughts and words, ideas which better suit the overall message etc.\*\*

\* Perhaps the working memory capacity of 7+-2 items is a common prerequisite for the limitations of this "loop" and "buffer lengths".

#### **4. Hallucinations - intentional lies and cheating and sometimes inability to even hallucinate**

Humans not only "hallucinate", or make mistakes - say things about which they lack definite information, or they lack reliable proves that it is true (or they lack information that **makes them believe that the information is true** - it can be wrong either), and they are convinced and aggressive in their believe they are correct, and try to convince the others often based on **authority and power**, they are "bosses", they are "on the right side of the history", they are "the good guys" (in the Hollywood movies, LOL).

Humans **intentionally and aggressively LIE and try to CHEAT and suggest lies in order to take advantage of someone or the situation and they don't consider this wrong** (LLMs learn from humans either, LOL).

**Even worse:** in many ways and often humans **can't even hallucinate, when "prompted"**, or their attempts are grossly ridiculous, i.e. **they have no clue what to say**, what the next word would be, even randomly or generate any coherent continuation with correct syntax, guided by a topic or more material.

This is true especially if the "HumanLLM" lacks appropriate "pretraining" in the domain. Ask somebody what he thinks about an area of knowledge or event where he has little or no expertise, or in language she doesn't speak. What she would answer? What if you show a "humanLLM" who has never programmed a piece of machine code in hexadecimal A9 7F 8D 00 20 ... What "reasonable" content would the human generate if she is obliged to continue and return 100, 200, 1000 tokens?

#### **5. Memorizing and Repetitions of what was already known, "just retrieval"**

**LLMs** are blamed for "just retrieving", repeating existing knowledge, just recombining it. What else you do most of the time, humans? You can't do even that!

People with slightly better memory than the rest are called or compared to "computers", and that used to be even in the 1950s, LOL, or having "prodigious memory", if they can remember 50 or 1000 "digits of the number pi" (a meaningless record).

The "normal" situation is when humans have poor memory and remember up to 5-6-7-8... digits of Pi (3.1415926...). Even the most intelligent ones need to constantly consult with dictionaries, textbooks, books, mathematical tables with formulas, to lookup the code they write or repetitively reread the paragraphs of the posts and articles they write, to dig into the Internet, ... and to rely on LLMs to do the job for them.

In general almost everything in the Universe and in computers, and Universe is Computer\*, is a repetition or retrieval from memory anyway, preservation has far bigger weight than development (see Zrim, Creating Thinking Machines).

The creative humans do the same ("Creativity is imitation **at the level of algorithms**", 2003), while most humans **CAN'T DO EVEN THAT** well: repeating even 3 seconds in the past from **their own memories, their own intentions, thoughts etc.**

Yet they blame the LLMs or computers - the real nature of LLMs - for being "dumb", "not intelligent" etc.

**What about you yourself if you apply the same metrics?**

.....

\* Yes, yet another (short) exercise in [the genre of writing a] "pamphlet against..." [human, organic, biological, evolution ...], see S.Lem, 1963-1964 and T.Arnaudov, 2001 (Man and Thinking Machine: Analysis of ... etc., Theory of Universe and Mind - see it also for the Universe Computer, as well the recently published appendix "Algorithmic Complexity"... ).

\*\* These days I did that with a new song, a cover of a Bulgarian rock-n-roll

classic, however with new lyrics about something else. **Stay tuned**<sup>47</sup>.

\*\*\* Yes, that belongs to a paper, a continuation of the mentioned one etc.  
(...)

...

More and more recent literature in the Neuroscience and LLM-related research in one of the Appendices of The Prophets of The Thinking Machines, currently called #listove.

---

<sup>47</sup> The song is about the Young Age of the Internet. Песента е за... „*Когато Интернетът беше млад*“ (и други заглавия) По музиката и част от текста и духа на Г.Минчев: „Преди години когато рокендролът беше млад.“

**\* Предвиждания за автоматичния превод и машинното въображение<sup>48</sup> от 2005 г. изказано пред машиномразци. Отношението към „пророците на мислещите машини“ в български обществен блог през 2005 г.**  
Тодор Арнаудов – Тош и др.

<https://bglog.net/BGLog/post/%D0%9D%D0%BE%D0%B2%D0%B8%D0%BD%D0%B8%D1%82%D0%B5---#qsc.tab=0>

Новините...

от Robin Hood на 31.10.2005 21:47 28 коментара , 391 прочита  
Чета новините. Вниманието ми привлича едно много завладяващо научно и техническо постижение-електронен преводач на реч. Пишат, че бил доста успешен. Беше въпрос на време кога ще се направи такова нещо и да не забравяме, това е една стъпка само към създаването на истински преводач.“

---

<sup>48</sup> Начинът по който работят едномодалните пораждащи модели от типа на „обикновените“ трансформатори и дифузните модели към 2023-2024 г. не е достатъчно пълноценно съответствие на въображението на човек творец, но при мултимодалните като синтезатора на видео SORA и др., преливането започва да прилича. Дифузията (свързването на различни етапи, избистрянето от по-неясно към по-ясно) е общ принцип. Достатъчно богатия набор от данни, готови примери, или ограничено пространство за създаване, „какво може да се прави и какво може да се случи“ (Зрим) (и обхватът от онова, което приемникът на сътворения свитък очаква да е възможно да се случи и да е правдоподобно и „достатъчно“), съчетано с дообучаване с човешка обратна връзка и оценка (RLHF) и др. методи с обучение с подкрепление, които насочват „мисълта“ на машината, както и съответното съдържание на подканите и темите, които му се задават и са човешки, също позволяват на преобразителите да пораждат „мисли“ и произведения, които да звучат и изглеждат така, сякаш моделите имат все по-човекоподобно въображение. Повечето хора също не са оригинални и това, което създават в повечето случаи, особено след кратко усилие, и наричат „творческо“, „просвирва“, повтаря, пресъздава нещо очевидно или банално. Както показва и убеждението на „нападателят ми“ машиномразци, че „машините нямамо как да свършат качествено...“, понеже **те не могат да си представят как би могло да се случи и следователно е невъзможно.**

Виж „Творчеството е подражание на ниво алгоритми“, Т.Арнаудов, 2003 г.

## КОМЕНТАРИ

**# | от Tosh на 31 октомври 2005, 23:24**

Някакви подробности? От кои новини го прочете? Би ли дал връзка, повече информация, някакви имена, адреси? Тош,  
<http://geocities.com/todprog>

**# | от Robin Hood на 31 октомври 2005, 23:48**

Виж тук <http://www.membrana.ru/lenta/?5298> когато търся новини преброждам няколко десетки сайта обикновено и после не мога да открия каквото ми трябва :). По принцип от тук <http://www.alexa.com/browse?&CategoryID=48639> чета, много приятни списъчета на най-посещаваните сайтове, плюс този на руснаците и те често са сред първите с новините.

**# | от Tanichka на 01 ноември 2005, 00:04**

..се възторгвате от електронен преводач?! На реч, ок, ама колко неутрална и лишена от оттенъци и двусмислия трябва да е тая реч? Някой мисли ли си, че реално ще се появи машина, която ще може да преведе книга - Фокнър, Селинджър, Стайнбек,...? Вярно, че вървим напред, обаче много човешки дела машина няма как да ги свърши качествено...

Как ще се справи със "прасурсите" в Джабъруоки на Алиса? Защото те са със структурата на граматически издържани изречения – имат си окончания на глаголи, пасиви, наклонения, герундии, както си му е реда, все едно са смислени фрази, но настрана от няколкото думи със смисъл, всичко си е Луис-Каролова измишльотка...

'Twas brillig, and the slithy toves  
Did gyre and gimble in the wabe:  
All mimsy were the borogoves,  
And the mome raths outgrabe.

Иначе преводачът си е свършил перфектно работата, и ни е дарил с прекрасен превод, който ползваме като цитат по всякакви поводи:

Бе сгладне и честлинните комбурси  
тарляха се и сврецаха във плите;  
съвсем окласни бяха тук щурпите  
и отма равапсатваха прасурси<sup>49</sup>.

**# | от Robin Hood на 01 ноември 2005, 00:25**

---

<sup>49</sup> Виж превод с ЧатГПТ от 2023 г. по-долу

"Веро[я]тността земята да е плоска и да се крепи на 4 кита е едно на милион изчислили учените, обаче пък от друга страна едно на милион се пада девет от десет пъти изчислили магьосниците" - Тери Пратчет (цитирам по отслабналата ми памет)

Много неща в даден момент от историята са изглеждали абсурдни, но това не е попречило да бъдат създадени от хора които не са знаели, че са абсурдни или не са искали да повярват. Сигурен съм, че един ден ще има такава машина и е само въпрос на време и пари. Не вярвам човешкия мозък да е нещо кой знае колко сложно (поне моя не е :)), проблема е, че ние сме много прости и ще ни трябва повече време да го изучим, но и това неизменно ще стане един ден. :)

**# | от Скрежанка на 01 ноември 2005, 00:30**

Като чета какви теории развивате и още повече ми прилява...

**# | от Robin Hood на 01 ноември 2005, 00:34**

А защо? Не разбрах какво точно имаш предвид.

**# | от Tosh на 01 ноември 2005, 00:41**

Благодаря за информацията. :) И аз чета membrana.ru - една руска свръзка ми я подсказа миналото лято, - но само от време на време, а виждам, че новината е прясна... :) Питам не само от интерес към възможностите на това чудо, но и защото то се явява опасна конкуренция. :) Наскоро в споделището се регистрира инж. Койчо Митев<sup>50</sup> (гугъл), и в профила му пишеше какви хора търси. Впрочем, той търси и езиковеди. <http://bglog.net/user/845.....> Здравейте, Търсим ентузиастични за развитие на проект "Комуникация на майчин език". Нужни са три групи специалисти: 1. Лингвисти от произволни езици; 2. Софтуерни проектантите (системен софтуер); 3. Специалисти по запис и разпознаване на глас (разпознаване на фонемите). За проекта можете да научите от тук:

[http://www.bnr.bg/RadioBulgaria/Emission\\_English/Theme\\_Science\\_And\\_Nature/Material/Inventions.htm](http://www.bnr.bg/RadioBulgaria/Emission_English/Theme_Science_And_Nature/Material/Inventions.htm)

[https://web.archive.org/web/20051029005927/http://www.bnr.bg/RadioBulgaria/Emission\\_English/Theme\\_Science\\_And\\_Nature/Material/Inventions.htm](https://web.archive.org/web/20051029005927/http://www.bnr.bg/RadioBulgaria/Emission_English/Theme_Science_And_Nature/Material/Inventions.htm)

---

<sup>50</sup> Проектът на К.Митев беше за междинен код, който използва десетична бройна система. Доколкото си спомням проблемът беше, че трябваше по някакъв начин въпросните други учени да построят базите данни за съответствия, „разпознаване на фонемите“ и т.н. Водих кореспонденция с него известно време. Сред колеги експерти по Обработка на естествен език К.М. имаше репутация на „луд учен“.

<https://duma.bg/balgarin-pravi-revolvyutsiya-v-obshtuvaneto-n199132>

[http://www.commonsenseadvisory.com/en/news/global\\_watchtower\\_one.php?id=164](http://www.commonsenseadvisory.com/en/news/global_watchtower_one.php?id=164)

Пишете ни на E-mail: [patentservice@abv.bg](mailto:patentservice@abv.bg), [patentservice@mail.bg](mailto:patentservice@mail.bg)

Койчо Митев... Сега ще гледам дали ще успея да върна своя плам по автоматичен анализ и синтез на звуци на човешка реч, който имах в нач. на 2004 г., и дали ще помогна да подкараме автоматично разпознаване на фонеме. Тош

...

## # | от Tosh на 01 ноември 2005, 02:11

*"Някой мисли ли си, че реално ще се появи машина, която ще може да преведе книга - Фокнър, Селинджър, Стайнбек,...? Вярно, че вървим напред, обаче много човешки дела машина няма как да ги свърши качествено..."*

Да, някои си мислят. :) Това ще е машина с достатъчно сложни и зависими по 'правилния' начин от входните данни състояние и поведение (съвкупност от въздействия върху 'свят'). //Терминология - моя//Когато преводът е многозначен, машината не може да бъде неутрална, трябва да има пристрастие и да реши. :) А ако искаме да е силно ангажирана, даваме на машината право да не превежда дословно, а да пресъздава романа в представите си и да разказва и тълкува това, което си представя, а не това, което чете. Знаете ли какво е да си представяш? Аз си представям представянето въз основа на текст така: Непълни входни данни допълваш с данни, които не са в противоречие с входните данни. Въз основа на непълни входни данни правиш предположения за липсващи данни, въз основа на знания които имаш за това, за което ти напомнят тези данни... Как ще се справи с такъв превод? Много лесно, такива "измишльотки" се превеждат с пораждање на други "измишльотки", които имитират стила на 'измишльотката' - рима, ритъм, окончания (имитации на глаголи, имитации на спрежения), пораждање на буквосъчетания които приличат на такива от речника, но не са. Измислените думи използват ли се по-нататък? Ако не, значи е все едно как ще ги преведеш. Ако да, от употребата им може да се правят предположения какво значат. Ако от употребата им не могат да се правят предположения какво значат, можеш да ги замениш с каквито си искаш буквосъчетания, и няма да стигнеш до противоречия.

*"Иначе преводачът си е свършил перфектно работата, и ни е дарил с прекрасен превод, който ползваме като цитат по всякакви поводи:"*

Би ли описала класа на 'прекрасните преводи' на този стих? Според мен

той включва наличие на достатъчно сходства в измерими особености на късовете знание (стиховете): брой строфи, съвпадения на думи които са "истински", наличие на рима, съгласуване, окончания, наличие на буквосъчетания които приличат на действителни думи, но не са. Тош, <http://geocities.com/todprog>

#### # | от Tanichka на 01 ноември 2005, 03:09

*Би ли описала класа на 'прекрасните преводи' на този стих? Според мен той включва наличие на достатъчно сходства в измерими особености на късовете знание (стиховете): брой строфи, съвпадения на думи които са "истински", наличие на рима, съгласуване, окончания, наличие на буквосъчетания които приличат на действителни думи, но не са.*

А къде остава майсторлъка с аналогията с природната картина, звуците, картините, които виждаш (или не виждаш)? Това броя на строфите ли го дава или окончанията? 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! Че то всеки го може, нали? Затова и има толкова некадърни и неопитни хора, на които се дава да пробват и четем пълни глупости. Един от най-некадърните преводи през 80-те години беше на Машината на времето, някакво бургаско или варненско парт-издателство, някаква Бучкова-Малеева беше преводачката... Остават си като лош пример и се изучават от студентите, за да се стараем повече да не се вписваме в тия редици....

А за Алиса, ако не се лъжа, а може и да се лъжа, един от първите преводи беше на голям преводач - Леда Милева или Асен Христофоров... Но какво са те в областта на литературата и литературния превод в сравнение с машините за превод?

Разчитаме тъкмо на капацитети като теб да ни доставят удоволствие с литературата и преводите й!! Давай!!!

#### # | от Tosh на 01 ноември 2005, 17:21

Изд. "Георги Бакалов" - Варна, 1984?



*"А къде остава майсторлъка с аналогията с природната картина, звуците, картините, които виждаш (или не виждаш)? Това броя на строфите ли го дава или окончанията?"*

Тоя "майсторлък" е неописуем и е "вълшебство" (без заклинания), така ли? :) //Ти описа още от него//

*"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! Че то всеки го може, нали?"*

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 трябва 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.**

*"А за Алиса, ако не се лъжа, а може и да се лъжа, един от първите преводи беше на голям преводач - Леда Милева или Асен Христофоров... Но какво са те в областта на литературата и литературния превод в сравнение с машините за превод?"*

Не обичаш машините? Като говоря за превеждаща машина имам предвид машина, която ще може да прави превод на художествени текстове, който да класифицираш като "качествен". Такава машина по определение би имала въображение и способности сходни с тези на добри, "големи" преводачи. Все още няма такива машини. Но ако се появи такава и е "изявена" (напр. изисква огромна изчислителна мощ, която може да се осигури само от суперкомпютър, използва се частно и не се споделя), и тя [ще] може да си изгради "име" на личност в литературата, да пише и да превежда.

**| от Веселин на 01 ноември 2005, 17:47**

Tosh, влече те научната фантастика, определено. И все пак, вероятността да съществува такава машина, която описваш, или дори, да просъществува в скоро време, или изобщо да просъществува, ми се струва, меко казано нищожна... Но става добър сюжет за фантастичен

разказ, който би бил много интересен. :) Ти искаш да кажеш, че човек е в състояние да създаде нещо, равностойно на това, което природата би создала. Какво имам предвид? Както и Таничка споменава дотук, аз също съм на мнение, че човешката фантазия не може да бъде заместена по никакъв начин от машина, защото тя (фантазията) никак не е толкова ясно и разнищено понятие, за да бъде идеално осъзната, а камо ли - пресъздадена. Навеждаш ме на мисли за експерименти като клонирането, където наистина има някакъв ефект, от гледна точка на създаване на нов живот, но с много недостатъци. (по-кратък живот, непълноценно съществуване, и въобще - за какво съществуване говорим?) Или за генно - модифицираните домати, чушки, моркови и гъби, които съм принуден да ям напоследък, и които нямат нищо общо със свежестта на тези, набрани в някое село, отглеждани напълно природосъобразно... Та така и с машината за превод - ще наизлязат някакви глупости, които ще се чудиш как да ги разбереш изобщо :) Човек не е бог!

**[Бел. Тош., 30.10.2023 г.: Забележете как Таничка и Веселин се опитват да манипулират и подменят изказаното от мен, като вкарват собствените си заблуди и опростявания. Те вмъкват, че ако „машина може да превежда“, „следователно било много лесно“, „всеки може“ – нищо такова не съм твърдял, дори напротив, обяснявам че подобна система може да изисква трилиони, квадрилиони изчисления и съответен обем памет. После пък не можело да се създаде, *защото човек не бил бог*, защото не можел да създаде нещо „равностойно на природата“ и пр. – също нищо такова не е казано в моите бележки; освен това **те** твърдят, че за да се създаде такава машина било нужно „да надминеш природата“ или „да бъдеш бог“, понеже за *тях* това е невъзможно по-точно е те **искат** да бъде невъзможно, искат *да внушат* тази мисъл на читателя или **вярват** в нея (виж „Писма между 18-годишния Тодор Арнаудов и философа Ангел Грънчаров“, 2002 – Тош: вярвам означава *искам да е така, както си мисля*). Докато според мен е **очевидно**, че преводът, мисленето, творчество са осъществими от машина, и обяснявам принципите на действие, според мен това щеше да бъде нещо **обикновено, предвидимо** – както и става все по-успешно – а не е *вълшебно, божествено, свръхестествено*.]**

## # | от Tosh на 02 ноември 2005, 01:47

За жалост сме в много различни категории. Аз съм в "тежка" и съм ветеран, а от това което казваш ми се струва, че не си си задавал въпроси относно какво са мисленето, разума, чувствата, въображението. [по-точно *не е откривал верните отговори*]

За фантастиката за Мислещата машина - "Истината", 2002-2003. Има я на моя "ринг", откъдето може да видиш категорията ми в този "спорт": <http://geocities.com/eimworld/razum> [Остарял адрес, имаше я и в sf.ludost.net, после в Читанка/Моята библиотека. Виж в archive.org, oocities.com, eim.twenkid.com]

*"Ти искаш да кажеш, че човек е в състояние да създаде нещо, равностойно на това, което природата би създала. "*

Не е вярно, "човек" не е в състояние да създаде, защото човек = новородено бебе на което майката не е до него, в свят, в който има нещо което може да му дава храна и вода, но няма никакви произведения на предни човешки същества, от които то да се учи и да използва. :)

*Вселената*, част от действията и механизмите на която се възприемат като "хора", е в състояние да създаде/сглоби/съсредоточи нещо по-сложно от това, което тя самата е сглобила в по-преден етап от развитието си. Хората са *част от природата и са нейни инструменти*, а не нещо извън системата. Това което правят, е дело на Вселената, която е предизвикала съществуването на силите, които се освобождават включително чрез действията, които хората си приписват като "авторски".

Чувал ли си за "двуканални" и "двустепенни" еволюционни теории? Еволюцията на Вселената на Земята от доста време е развитие и усложняване на небелтъчните информационни устройства и по-пълно използване на възможностите за обработка на информация на белтъчните устройства, с помощта на небелтъчни.

[Бел. 30.10.2023, Тош: Т.е. „човек не е бог“, но и няма нужда да бъде, защото „Бог твори“, а човекът е само фрагмент от творението. Съответно Бог, Вселената и пр. създава и мислещите машини, а ние като индивиди сме актьори в представлението.]

*"Колко ли още вида летене има освен, базирано на махов или реактивен принцип? Ами мисловното летене? Дали един самолет може да лети толкова бързо, колкото мисълта ти? Сега може да си тук, в*

*следващия момент - в Париж, в следващия - къде ли не, въпрос на въображение."*

:)

Ето един цитат който използва тия ключови думи, от "Вселена и разум 4: Произход на повторенията на събития в (...) ", Тош 2004

[Това е показно за начина по който работи въображението.]

Защо сънувах това, а не... онова?

Смята се, че сънищата са отражение на света, в който живее сънуващият, че те са "криво огледало", което пречупва спомените на човека за света, като създава нови светове, по-различни от този в действителността, но използващи елементи от действителния. Нашето въображение, естествено, не може нищо да измисли, без да получи входни (някакви) данни от Действителността, с които да си "играе" и променя. Всъщност винаги, не само "нещо истинско" остава в съня, макар изопачено, а, мисля, целият сън е построен върху представата за действителността, защото от нея сънуващият е взел пример за какво би могъл да си фантазира и докъде може да се простре въображението му.

Например сънуващият иска да лети в действителността, но там не може... Желанието му е силно и често му се присънва, че лети. В "истинския свят", който в Схващането наричаме "Първия въображаем", хората не могат да летят, без да използват машини, следователно в съня се е "развихрило въображението".

Добре... Но да видим как летим в сънищата си. Обикновено по начина, по който летят същества, които сме виждали да летят - човеци в самолет или в космически кораб, плавно като птица (дори да не размахваме "крила") или като... Супермен?... Дали е възможно е да се измисли нов начин залетене в съня, какъвто човекът не е виждал в действителността?

Мисля, че не. И новият начин ще бъде видян в действителността, макар непряко – като образец за някакво поведение или зависимост; като граници за построяване на функцията на поведението. Част от действителността сме и самите ние, вършилата (механизмите), на които се основава действието ни, затова вроденото знание също е научено от действителността - от строежа на Вселената, последица на който сме и ние.

"Летиш" означава, че се преместваш в пространството. Преместването е поредица от съседни положения на тяло в съответното пространство. Сънят, свързан с действие летене, притежава задължително тези особеност, в противен случай действието няма да е

летене...

Въображението не може да бъде безгранично, а е повторение на действителността. "Унасянето", че е безгранично, произлиза от недостатъчна, по-малка от очакваната или желаната, дълбочина на разбирането на причините Първото Въображение - Действителността да поражда във въображението конкретните образи, които поражда.

-----

За пресъздаването на "чувства", творчеството и др. ще пиша в подходяща тема.

Тош, <http://geocities.com/todprog>

avatar

**# | от Жоро Атанасов на 02 ноември 2005, 06:35**

Тош, от много стоене пред компютъра си изперкал братче..... тотално неадекватни мисли си писал, подредил си думички на български една до друга ... *но смисъл - никакъв*. Затова гледам, че и другите не са ти писали, ама аз не се стърпях, честно! Твоята приятелка ми каза да не се "заяждам с момчето" ама аз не се заяждам, казвам ти какво мисля . спри с тоз компютър и го удари на жени, не на пиене. Разходки, чист въздух, и много, много секс, колкото може повече ... това е рецептата. Тук има хора дето си вадят хляба с обработка на писаното слово в какви ли не форми ... питай и тях ... аз понеже много си те ценя и не можах да не ти напиша какво мисля, колкото и да ми се разгорещиш сега ..... има и друг свят освен твой брато, има и други мнения освен твоите, пък и колкото и да си умен .... все пак трябва да знаеш, че не винаги може да си прав..... та.... вземи се огледай да видиш, че има и други хора около теб и те мислят с главите си ....

отивай да си търсиш жена веднага

**[Бел. Тош:** Просташки обиди при признание на колегата, че *не разбира, не схваща връзката*, няма аргументи по същественото в написаното *и не може да ги формулира*, освен като обвинение към автора в *невменяемост*, но показна увереност в собствената правота и особено: *желание да наложи своето мнение* над „невменяемия“ – все пак, от добри подбуди.

Допълнение от 15.6.2025 г.: хумористична космистка и юнашка песен, която ми хрумна тия дни:

*Не щеме ний богатство, не щеме ний жени,*

*а искаме свобода, машини мислещи*<sup>51</sup>.

]

# | от Tosh на 02 ноември 2005, 10:51

*"подредил си думички на български една до друга ... но смисъл - никакъв"*  
[ти не откриваш] [Скоби – от 2.11.2005]

:D Вярно бе, зад тези думички и "въображаеми Вселени" стоят стотици КБ допълнителни определения и обяснения с кибернетична, аналитична и цифрова философия и *трансхуманизъм*.

*"Затова гледам, че и другите не са ти писали,"*

Аристотел ще ти пише 2 по Логика.

Аз ти пиша: :-X

# | от Веселин на 02 ноември 2005, 17:15

"Самозваний Ветеране в тежка категория", за жалост наистина не сме на едно ниво! :) Прав е Георгеата в това, което ти е написал. Не за друго, а защото темата беше *за някакви машини за превод*, а ти ми изписваш 2-3 страници *разхвърляни разсъждения* за това какво е летене, за Вселената, за човека и т.н. В твой стил ще ти кажа сега, че нямаше нужда да ми изписваш в какво вярваш, за какво се бориш и т.н. :) Действаш малко *първосигнално* всеки път, когато някой даде мнение различно от твоето :) и много насериозно взимаш нещата, а и себе си :) Не бива така. :)

# | от Tosh на 02 ноември 2005, 19:49

Темата за която последно сбърках да разсъждавам и да търся ответни разсъждения, доколкото чета в съобщенията ти които използвах като материал за разсъждение, не беше преводът, а дали *"човек е в състояние да създаде нещо, равностойно на това, което природата би създала"* (не е и няма нужда да е), за въображението *"защото тя (фантазията) никак не е толкова ясно и разграничено понятие, за да бъде идеално осъзната,"* - и не съм изписал 2-3 стр. за теб, а съм ги залепил

---

<sup>51</sup> Сравни с други „мъжки“ забавни песни от рода на „Песен за бирата“ и „Жигули“. Виж сатиричната интердисциплинарна книга „Какво му трябва на човек? Играеш ли по правилата ще загубиш играта! Първа част“, Тодор Арнаудов, 2014, сп. Разумир, бр. 1

от старо писание, което имаше съвпадащи думи с това за което говореше.

Ясно, търсенето на размяна на задълбочени философско-аналитични разсъждения и разбиране на нещо повече от мнение - причини за това мнение - по тия територии определено води до взаимно криворазбиране.

:-X

(...)

**[Бел. 30.10.2023.** Хората често не са в състояние да следят собствените си мисли между две съобщения. Тонът след това стана по-мек и взаимноуважителен.]

...

**[Тош, 9.8.2025:** Човешкият превод на откъса най-общо е на същия принцип, обяснен от мен преди отговора на Таничка. „Грешните“, според буквална проверка в речник, думи в оригинала, или „думите измишльотки“, се разпознават като *подобни* на други, или се познава или решава тяхната **граматична роля**, какви части на речта и изречението биха могли да бъдат според околните им, които съвпадат буквално (служебни думи: and, the, ...) или съвпадат части от думи, окончания, представки и пр., които извикват в съзнанието подобните представи; започва се със сравнение; разпознатите, „изчистени от изкривяването“ думи описват природна картина според това което е разпознала системата: при многозначност, описанието може да се предположи с по-голяма увереност въз основа на предходните и следващите пасажии в контекста на оригинала; тя може да реши че е друга картина – ако читателят *не знае* оригинала и не го разбира, е достатъчно преводът да е достатъчно „правдоподобен“ и убедителен в собствената си „вселена“ и поетика. Може и да смени и изреченията или частите на речта.).

Съвпадението се открива по това, че при сравнение и „просвирване“ – вокализиране, озвучаване, прочитане – те са достатъчно близки по знаци или звуково – омофони са – на такива думи, както и до подобни известни стихове и изречения, служещи като образци и очаквано познати на читателя като *клишета*, което може да служи за комичен ефект. Рисува се някаква картина или подобна или свързана с нея – друг „пейзаж“ от спомените, с думи „изковани“ по подобен начин, или направо изречения. Напр. „Twas“ – It was (беше). „Достатъчно“ – според критериите на оценителя и защото той *търси* съвпадение. Ако не желае да открие, всяка разлика може да се сведе до „*няма такава дума*“, „*това е безмислица*“ и изместване на вниманието към нещо друго. Ако обаче деецът *иска* да открие смисъл, картина, съвпадение, ще си намери във всичко, дори и да е „ах3х3йцасд2 сйдавещ, ск29йеф“.

Нарисуваният в представата „пейзаж“ или „правдоподобно“ звучащо

изречение, се описват на другия език, поражда се съответното му представяне в слово, като се следва стила, съответните сричково-звучно-ритмични параметри, ако преводачът желае да ги следва, а може и да не се съобрази, защото това е „параметър“ на настройката. Част от преобразуването е „изкривяването“ на сло-воформите, което също може да се измери като разстояние на Левенщайн (брой променени знака), чрез множество допустими съответстващи преобразу-вания между букви и срички, които се приемат за „все още запазващи подобие-ето“ и разпознаването и пр.

*'Twas brillig, and the slithy toves* → It was s.t and the slithy toves [t-was]

Беше НЕЩО и НЯКАКВИ НЕЩА ..

съществ. Наречие предмети, мн.ч. (-es; напр. gloves, shoes ... ръкавици, обувки; slither – плъзгам се, „slithy“ → плъзгащи се, хлъзгави)

*Did gyre and gimble in the wabe*: – ако изречението продължава – минало време (Did): Беше, бе, бяха, за действие... въпрос в миналото (направи ли?); но смис-лово свързано с „was, were,...“ (съм); общо: признак за минало време.

Did ?1 И ?2 ВЪВ ?3 (gimble – gimbal - кардан); gyre, wabe; нещото е в ?3 (“in the wabe”). В НЕЩО. т.е. „може да се „бъде в него“: „wabe“ е съдържител. Ако се намира навън (виж контекста на историята), може да бъде гора, езеро, поляна; а може и да е стая, дупка, кутия и т.н.

*All mimsy were the borogoves*,

**Всички** нещоси (**наречие** или някакси, начин) **бяха** някаквиси, определени (“the”); подобно окончание (borog-oves), съвпада с (t-oves); за да се римува.

*And the mome raths outgrabe*. – “and the” – служебни думи

**И нещотоси** прави\_нещо? някакси или нещоси. (the ?1 raths – the определя име, може да следва спрегнат глагол – завършва на „s“, 3 л. ед. Ч.)

“Out-...” – напр. „outrageous” – жестоко, свирепо, скандално; нещо „крайно“ или „извънмерно“ и пр.

В дадения превод се запзва структура подобна на изречения по някои от „гра-матичните“ части и се вметват „неречникови“ думи, които могат да се разпоз-наят като притежаващи подходяща граматична категория (или достатъчно „неп-речеши“, невнушаващи противоречива граматична роля. Българският има по-свободен словоред обаче..

*Беше прълям и ствълям утре.*

]



**\* Сесия с ChatGPT 3.5 за превод на стиховете от „Алиса...“**

от #21-9-2023

**\* И още няколко с ChatGPT4 (2024) и Claude 4 Sonnet (2025)**

Тодор Арнаудов

**Тодор:** Translate to Bulgarian, give several variants. It's poetic, imaginative, word play, playful, natural picture, fantasy, dreamy, funny:

'Twas brillig, and the slithy toves  
Did gyre and gimble in the wabe:  
All mimsy were the borogoves,  
And the mome raths outgrabe.

**ChatGPT:** Translating Lewis Carroll's "Jabberwocky" into Bulgarian (or any other language) is quite challenging due to its playful and imaginative language, as well as the fact that many of the words are nonsensical even in English. However, I can attempt to capture the essence and mood of the original. Here are a few variants:

Беше блескаво, и хлъзгавите тови  
Се въртяха и гушкаха в мъглата:  
Всички борогови бяха като сън,  
И ратовете извикаха в нощта.

Беше здрач, и плъзгавите товчета  
Се въртяха и скитаха във валежа:  
Всички борогови бяха като приказка,  
И ратовете избухнаха от мрака.

Беше светкавично, и хлъзгавите товци  
Кръжаха и играеха във блеската:  
Бороговите бяха все като сенки,  
И ратовете изпяха в здрача.

Again, these translations can't fully capture the imaginative and nonsensical nature of the original, but I hope they convey a similar mood and feeling in Bulgarian. If you're aiming to create a more authentic translation that truly resonates with Bulgarian readers, you might want to work with a native poet or literary translator who

can craft a version that feels both playful and poetic in the context of Bulgarian literature.

**Тодор:** Please make other translations and scramble the letters in the Bulgarian translation, remove some vowels etc. in order to make it sound more funny.

**ChatGPT:** Sure, we can give it a playful, whimsical twist similar to the original. Let's get a bit experimental:

Original:

Беше здрач, и плъзгавите товчета  
Се въртяха и скитаха във валежа:  
Всички борогови бяха като приказка,  
И ратовете избухнаха от мрака.  
Experimental versions:

Бш зрчч, и плзгвт твчтта  
Съ вртх и сктх въ влжа:  
Вск брогв бх кт прказчка,  
Ртвт ихнх от мркч.

Бше здрч, и плгавт товчт  
Се вртя и сках вв влж:  
Вчки брогви бха кт прикзка,  
Ратовт избхн от мрк.

Беш здч, плзг товчч  
Съ втх и ских в валж:  
Всч боргв бх кт прикз,  
Ратвт ибух мрак.

I hope these whimsical versions give you what you're looking for!

**Tosh:** Please make other translations and scramble the letters in the Bulgarian translation, in order to make it sound more funny. Imitate other similarly sounding words, synonyms etc.

**ChatGPT:** Alright, I'll take a more playful and creative approach by substituting and mixing up words to give a whimsical twist similar to the original. This approach will prioritize the sound and rhythm over direct meaning:

Original:

Беше здрач, и плъзгавите товчета  
Се въртяха и скитаха във валежа:  
Всички борогови бяха като приказка,  
И ратовете избухнаха от мрака.  
Experimental:

Беше блеч, и тръпкавите трошчета  
Се върляха и викаха във валкана:  
Всички боклукчета бяха натъпкани,  
И катовете изчезнаха в мрака.

Беше клач, и къдравите котчета  
Се бъркаха и микаха във малбана:  
Всички боршчета бяха като кал,  
И ръсовете изгряха от плака.

Беше врач, и влъхналите вовчета  
Се влачеха и спаха във ваната:  
Всички водогови бяха на вълна,  
И гатовете изпяха от рака.

Please note that this is a fun, non-literal translation meant to capture the spirit of the original.

**\* 10.10.2024: gpt4-4o-latest-20240903**

Here are several Bulgarian translations that aim to capture the playful, imaginative, and dreamy nature of the original:

Variant 1 (Playful and poetic):

Беше брилиг, и слитавите товс  
се въртяха и гъмбиха в вейба:  
всички мимси бяха бороговс,  
и момските рати изгряха.

Variant 2 (More whimsical and fantastical):

Беше бриляг, и слихавите товс

вихреха се и гъмбеха в вейба:  
целите мимс бяха бороговс,  
а момските рати отпяха.

Variant 3 (Dreamy and playful):

Беше бриляг, и слихавите товс  
се въртяха и гъмбиха в вейб:  
бороговс бяха в мимсов час,  
а момските рати изреваха.

Variant 4 (More musical and fantastical):

Беше бриляг, и слихавите товс  
се въртяха, гъмбеха в вейб:  
тъй мимсичко бяха бороговс,  
а момските рати изпляскаха в ритъм.

These translations strive to keep the nonsensical, playful, and poetic essence of Lewis Carroll's original, embracing the fantasy and dreamlike qualities of the text.

=====

\* **Claude 4 Sonnet** 15.6.2025:

Translate to Bulgarian, give several variants. It's poetic, imaginative, word play, playful, natural picture, fantasy, dreamy, funny: <https://claude.ai/share/622a4bac-9718-4299-81cb-cf636f2631ff>

'Twas brillig, and the slithy toves  
Did gyre and gimble in the wabe:  
All mimsy were the borogoves,  
And the mome raths outgrabe.'

Here are several Bulgarian translations of this famous stanza from Lewis Carroll's "Jabberwocky," each capturing different aspects of the playful, nonsensical wordplay:

**Variant 1 (Emphasizing sound and rhythm):**

Беше слизко, и хлъзгавите тови  
Се въртяха и мънкаха в дъбрава:  
Всички мекши бяха борогови,  
И самотните рати изграва.

**Variant 2 (More melodic/dreamy):**

Беше бръмчо, и лизгавите дрови  
Се кълбяха и гинкаха в поляна:  
Все премекни бяха глорогови,  
И тъмните рачи изпяна.

**Variant 3 (Playful with Bulgarian sounds):**

Беше глибко, и хлъзгавите кови  
Се гираха и грибкаха в градина:  
Все помръмни бяха торогови,  
И глухите врати изкрещина.

**Variant 4 (More fairy-tale like):**

Беше тъпко, и плъзгавите стови  
Се въртяха и дърпаха в гората:  
Все присмиви бяха зорогови,  
И печалните мишки изплака.

Each version tries to preserve the nonsensical yet musical quality of Carroll's invented words while creating new Bulgarian "nonsense words" that feel natural to the Bulgarian ear and maintain the dreamy, fantastical atmosphere of the original.

**\* Сравни вариациите с примерните породени кратки стихове от Тош във „Фрагменти“, 2006 г. и хайкутата от „Вселена и Разум 4“ поместено в подробната хронология на „Пророците на мислещите машини“ и: <https://web.archive.org/web/20250603014926/https://bglog.net/BGLog/4882#gsc.tab=0>** <sup>52</sup>

Те бяха дадени като основа за изследване на творческите процеси като се произведат по-голям брой, по „Творчеството е подражание на ниво алгоритми“, 2003.

(...)

И мн. др. бележки ...

\* 19.6.2024: Виж също „Граматика на фантазията“, Джани Родари, 1979 г. – прочетох (и прослушах) средата на юни 2024 г. Интересно е, че споменава „Играта на живот“ (клетъчния автомат); също изразът „електронен изчислител“ за „компютър“.

---

<sup>52</sup> На 15.6.2025 г. – дава грешка, преди няколко дни се отваряше.

## \* **ЕВ1: Относно превода на Information Retrieval**

от NemSys на 01.10.2006 05:08 23 коментара , 960 прочита Етикети: езикови въпроси, bglog.net<sup>53</sup>, Ива Попова

В българския език няма няма общоприет преводен еквивалент на това понятие. Някои автори изпозлват термина *"Обработка на Информация"*, други - *"Извличане на Информация"*, а трети - не го превеждат (както правя и аз до момента)

Но:

*"Обработка на Информация" = "Information Processing"*

*"Извличане на Информация" = "Information Extraction"*,

Тогава ? Може би *"Обработка и Извличане на Информация"* или още по-точно (но по-сложно): *"Обработка на Документи и Извличане на Информация"*

или (най-простото, но имащо връзка с приложението) *"Търсене на Информация"* ?

(...)

# | от **Tosh** на 02 октомври 2006, 04:57

/\* Хммм... IR, дисертации. Станало е достатъчно интересно, за да се пона-  
дигна от гроба...\*/

В пловдивския ФМИ името на науката се превежда в една избираема като "Интелигентна обработка на данни", води я господин от БАН. [http://www.fmi-plovdiv.org/bg\\_ver/edu/izb\\_a/130.htm](http://www.fmi-plovdiv.org/bg_ver/edu/izb_a/130.htm)

Или по-скоро наричат "Data Mining"-а с това име.

На мен ми допада "извличане на информация [ от сурови данни ]". Да, има и "IE", обаче с няколко реда можеш да разрешиш тая двусмислица, като опишеш терминологията си.

"В тази работа с "извличане на данни [ или с еди-какво-си]" ще имаме предвид науката Information Retrieval, за която все още не е приет единен термин в българската литература. ... "

**"Възвръщане на информация"** също ми харесва - инф. е била инф.

---

<sup>53</sup> <https://bglog.net/BGLog/post/%D0%95%D0%921--%D0%9E%D1%82%D0%BD%D0%BE%D1%81%D0%BD%D0%BE-%D0%BF%D1%80%D0%B5%D0%B2%D0%BE%D0%B4%D0%B0-%D0%BD%D0%B0-Information-Retrieval#gsc.tab=0>

преди да се натрупа в данните, но си е "отишла" сред тяхната маса, и после се възвръща чрез IR.

С изразните средства на юнашкото наречие, IR може да се опише...

Хммм... Напр.: **превъобразяване**.

От **"въобраз"** - информация; от "вътрешен образ" и от "въображение".

Защо вътрешен? Защото е нещо, което има стойност в паметта на устройство, което възприема "външен образ", някаква среда, която е различна от него самото.

Защо образ? Защото може да се възприема като образ, изображение на нещата ( "отражение" ), и в математически смисъл: модел.

Приликата с "въображение" също подсказва, че това е нещо, което е свързано с въображаеми неща.

### **Превъобразяване?**

Защо **"пре-"**? Защото IR може да си представим и като връщане, възстановяване, спомняне на информация ( асоциация и с "remember"), която някога е била информация, а след това се е превърнала в данни. Да, това може да се каже и за по-прости информационни процеси, но става и тук.

( Хммм... "re" в "retrieval" не би трябвало да идва от морфемата "re". По-скоро "Retr-" може би идва от "retro"; тук "пре" не е превод на re )

Така думата може да образува и глаголна форма: **превъобразявам**.

Или пък:

### **Възвъобразяване и възвъобразявам.**

Защото "retrieve" е свързан с... "въз-" неща.

"Въобразявам си"? Наистина - за хора които не познават разширените изразни възможности на юнашкото наречие, тия словоформи биха прозвучали като "въобразявам си", без да будят връзка с информация ( въобраз ).

Това пък ми звучи като като възможен поток на мисъл, която съставя думата "retrieval":

( ("Retro" && "Evaluate" ) && Be(Noun) && Be(Short) ) => Retrieval

"Retro" е "назад" - обратен процес на "правия", каквото е и извличането, или извеждането.

"Evaluate" е обработката, оценката, придаването на смисъл.

На момента обаче не ми идва идея за дума, съставена от тази гледна точка.

...

Това не е вярно:

*(никой не е превеждал 'математика', 'физика', 'икономика', 'лингвистика',... ами май няма нито едно преведено име).*

Езикознание, стопанство? Богоров нарича химията **лъчитба**.

[http://bg.wikipedia.org/wiki/%D0%98%D0%B2%D0%B0%D0%BD\\_%D0%91%D0%BE%D0%B3%D0%BE%D1%80%D0%BE%D0%B2](http://bg.wikipedia.org/wiki/%D0%98%D0%B2%D0%B0%D0%BD_%D0%91%D0%BE%D0%B3%D0%BE%D1%80%D0%BE%D0%B2)

Константин Петров нарича физиката **"природница"**, а в началното образование се използва **"природознание"** за неща, свързани и с физика. За физични неща се използват и "природни явления", и "природни закони".

...

Нямаш ли потенциален/актуален научен ръководител и за какво, по-точно, се отнася работата?

Поздрави  
(...)

# | от **Tosh** на 03 октомври 2006, 01:02

*"Имам предвид, че етимологията на "retrieve" /нак според нета, не знам френски.. ;(/ е от старофренски и едно от значенията е "намирам"."*



Ааааа... "trouver" (намирам) и "retrouver" (н. отново)... И evaluer.

...

(...)

За:

*"(["Retro" && "Evaluate" ] && Be(Noun) && Be(Short) ) => Retrieval*

*тук ме озадачи факта, че използваш атоми от вида:  $p(X)$ , които могат да се изтълкуват като предикати или (ЛИСП) функции, но в същия момент използваш и атоми/константи от вида "А", "Б" и точно това ме озадачи.*

*Conc("Retro", "Evaluate") - щеше да ми изглежда по-точно*

*Т.е. - горката ти мислеща машина: ще трябва един път да тълкува && като логическа операция и един път като "съедини" (за такава свободна интерпретация сигурно и Лари Уол не е сънувал)"*

:) Както вече казах, това не е код за "кодомелачка", а по-скоро мисли на мислеща машина, описани дотолкова че да бъдат разбрани от други мислещи машини ( хора ). И не е слепване на низове, а "съставяне на дума".

По-скоро:

Language.Creation.CreateNewWordFromWords("Retro", "Evaluation",  
BE\_SHORT, BE\_NOUN );

//естествено, че не е додефинирана напълно, това е псевдокод

*"т.е. 'обратното' трябва да е "the artificial intelligence problem of training machines to behave like people" (макар че аз лично твърдя че двата подхода трябва да се обединят, в контекста на WWW рабира се)"*

Представата ми сега, е че работата ти ще е нещо за семантичен уеб, който обаче е семантичен не толкова заради парадигмата с онтологии и XML, а защото търсещите машини ще четат и разсъждават върху обикновеното хтмл съдържание, както би правил човек?

Правилно ли е предположение, че работата ти ще е повече върху

действието на умната система, отколкото върху структурата на семантичния уеб? Любопитен съм за подробности, когато са достъпни.

# | от NemSys на 02 октомври 2006, 23:48

(...) Впрочем и аз в началото използвах "Извличане на Информация" и възприемах като естествено "Възвръщане на Информация" (още повече че това е буквалния превод на retrieval).

Но разликата, с две думи, между IR и IE: **[Retrieval vs Extraction]**

IR системата връща като резултат (препратки към/адреси на) ДОКУМЕНТИ (/Уеб страници).

IE системата връща ФАКТИ, ЗНАНИЯ, изобщо - ИНФОРМАЦИЯ, която се съдържа в документите, при това във формат удобен за обработка със СУБД<sup>54</sup>.

Един вид, IE системата трябва да "смели" информацията вместо потребителя. Е, това "мелене" естествено е по-скоро на нивото на Data Mining-а ...а пък разликата между IE и DM е че при IE предварително се задава какво се търси. (...)

**Бел. 21.8.2025:** Виж приложение „Котката“, където откривам, че езиковите модели след 2023 г. могат да отговарят на част от въвеждащите „подкани“ от „Анализ на смисъла на изречение...“ #kotkata, Командата-псевдокод по-горе също работи и с не най-мощни модели (чрез Lmarena.ai):

```
> Language.Creation.CreateNewWordFromWords("Retro", "Evaluation",  
BE_SHORT, BE_NOUN );
```

**Assistant A – grok-3-mini-beta**

Based on your request to create a new word from "Retro" and "Evaluation", with the constraints of being short and a noun, I've generated the following:

**New Word:** Reteval

**Explanation:**

- **Blending Process:** I combined "Retro" (evoking something from the past or nostalgic) with "Eval" (a shortened form of "Evaluation", meaning assessment). This creates "Reteval", a portmanteau that suggests a retrospective or nostalgic evaluation.

---

<sup>54</sup> СУБД - Системи за обработка на бази от данни

- **Why it fits:** It's short (7 letters, 3 syllables), functions as a noun (e.g., "The reteval of old designs helped innovate the project"), and captures the essence of both words in a concise way.

If you'd like variations or more context, let me know!

#### **Assistant B – mistral-medium-2508**

Here's a concise, noun-based neologism blending **"Retro"** and **"Evaluation"**:

##### **"Retroval"**

(noun) – A **retrospective assessment** or **backward-looking analysis** (e.g., of past performance, trends, or decisions).

---

##### **Alternatives (if nuance varies):**

- **"Retrospection"** (if emphasizing reflection over measurement).
- **"Evalro"** (if prioritizing brevity, though less intuitive).
- **"Retrovalysis"** (if merging "retro" + "analysis" for a technical tone).

**Best pick: "Retroval"** balances clarity, brevity, and noun form. Use it like:

*"The team conducted a **retroval** of last quarter's design flaws."*

## **\* Предвиждания за близките пробиви в автоматичното програмиране – чат на Тош с експерт по ИИ от 2018 г.**

Тодор Арнаудов – Тош и В.М.

Чат във Фейсбук от 18.09.2018 г. //редакции на някои грешки в правописа на колегата и в пунктуацията

**Тош:** А и вече БТВ според мен няма много време за начинаещи, които трябва от мн. далече да почват. **Скоро всичко ще е направено. :)** Иначе имах идея за "як" клуб, но ми е ясно, че е неосъществим в момента в Пловдив. Проектът е обяснен тук: <http://twenkid.com/mind/index.html>  
Кръг Artificial Mind Hackafe IV @ TWENKID.COM | Todor Arnaudov's:  
Twenkid Research, Twenkid Studio, ПАЗУМИР, Свещеният сметач

**ВМ: "Скоро всичко ще е направено". какво имаш предвид под това?**

**T: AGI + Program synthesis. Code generation, универсален.**

*[Универсален изкуствен разум, т.е. истински мислещи машини и синтез на програми - автоматично програмиране, универсално генериране на код.]*

**В: не знам дали смяташ десетилетията за скоро време но добре**

**В: да се научат първо да пишат както трябва, после ще видим за програмирането**

**Тош:** програмирането е по-лесно :) засега съм познавал някои неща; около 2010-2011 г разпознаването на образи беше трагично; има го записано в един публичен имейл лист, където писах, че проблемът скоро ще бъде решен; след "малко" избухна Deep Learning-а [Виж "Столове, Сгради, Карикатури, ...", "AGI Digest", 2012]

**В:** *Програмирането е лесно затова е една от най скъпо платените професии и има голяма разлика между програмистите*

**Тош:** А не го следях, с този тип невронни мрежи не съм се занимавал по онова време, прекалено неефективни, ако нямаш суперкомпютри и не са интересни като структура.

**Тош: Трудно е за хора.**

**В:** А и проблема с зрението още не е решен (...)

**Тош:** не е решен с тези невронни мрежи; но разпознаването на образи за практически цели вече е фурия, а до скоро изглеждаше "невъзможно". А накратко, програмирането е елементарно, защото е напълно формален процес в напълно дефинирано пространство, където една машина може да прави опити около 999999999 пъти в секунда. И дори и с груба сила ще го направи. Но програмистите, на които това им е трудно, не мислят за програмите си в такива мета-термини и на практика "не знаят" (или не искат) да го представят в такива. Освен това вече "софтуерът е разработен", всичко основно е написано. Може да се ползва наготово с настройка. В тази насока работят едни българи в Швейцария, като използват хранилища с код.

**В: Спри да приказваш глупости и си лягай. Един ден това ще бъде възможно но не и скоро.**

**16.6.2025 г.+ (...)**

Въпросните "*швейцарци*", които по това време бяха известни само на посветени, „завладяха България“ няколко години по-късно, осигуриха си астрономически ресурси и се представиха като „визионери“ („пророци“), започнаха да възпроизвеждат по швейцарски начин и да повтарят буквално посланията на проект, публикуван **20 години преди тях**, след като целият свят го беше направил и ... (...) Виж в приложенията:

**1. „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?“ #purvata**

2. Институти #instituti

3. Основния том на *Пророците* и др.

Не очаквам особена или може би каквато и да било подкрепа или дори реакция и след публикуването дори и на всички части от *Пророците на мислещите машини*, особено от т.нар. „визионери“ и фалшиви пророци, хитреци и опортюнисти, но в перспектива и в контекста на **Българските пророчества**, виж сравнението от [1] по-горе относно „Кой е бащата на стратегията“ и аналогията със случая с „бащата на компютъра“, Айова и Принстън, Джон Атанасов и Мокли, Хъниуел и Спериренд. (...)

\* [https://twenkid.com/agi/Purvata Strategiya UIR AGI 2003 Arnaudov SIGI-2025 31-3-2025.pdf](https://twenkid.com/agi/Purvata_Strategiya_UIR_AGI_2003_Arnaudov_SIGI-2025_31-3-2025.pdf)

\* [https://github.com/Twenkid/SIGI-2025/blob/main/Purvata Strategiya UIR AGI 2003 Arnaudov SIGI-2025 31-3-2025.pdf](https://github.com/Twenkid/SIGI-2025/blob/main/Purvata_Strategiya_UIR_AGI_2003_Arnaudov_SIGI-2025_31-3-2025.pdf)

Виж и основния том и другите приложения

**СЛЕДВАТ ПРОДЪЛЖЕНИЯ**

## Volumes and Appendices to *The Prophets of the Thinking Machines*

...

### 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 тонове

\* **#prophets** – Основен том (>1865 стр., 5.9.2025); 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

\* **#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 (31.5.2025, 248 стр.)  
[https://twenkid.com/agi/Purvata\\_Strategiya\\_UIR\\_AGI\\_2003\\_Arnaudov\\_SIGI-2025\\_31-3-2025.pdf](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 изчислителната 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<sup>55</sup>. 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. Виж също: <https://github.com/Twenkid/Theory-of-Universe-and-Mind> ~ 204 стр.

\* **#listove** – Най-обемното и разнообразно по теми приложение и второ по големина след основния том (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 сборка от източници от медии и новини, множество платформи за чатботове и други пораждащи модели за различни модалности, и практика; мн.др. (...), >485 стр. (5.9.2025 г.) На бълг. и част на англ.

\* **#mortal** – **Нужни ли са смъртни изчислителни системи за създаване на универсални мислещи машини?**, Т.Арnaudов, 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 технологии с по-малки транзистори, и въобще ефективността във всичко е избор

---

<sup>55</sup> Работни шеговити „цензурирани“ заглавия: „нелицензирано разклонение“, „клонинг“ ...



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.Арnaudов– #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 езикови 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](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).(…)

\* **#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 ентузиазма 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. (...))

\* **#lazar #lotsofpapers** – 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](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.

\* **#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](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](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

**\* #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

**\* #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/>

**#razvitie #transhumanism** – все още ненаписано приложение, което би се фокусираше 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.

**#llm-review-TUM** – Automatic reviews and comparisons of TUM and other theories and

evaluation by LLMs, AI agents and thinking machines™.

## Workshops, practice (future)

### Практика, работилници и др. (бъдещи)

\* **#robots-drones-ros-slam-simulation-rl** – Наземни и летящи роботи: дроневи; 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,

Пророците на мислещите машини: Приложение Ирина: Беседи и записки ...

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 подобрието на съществуващите и за осъществяването на бъдещите разработки!

# ИРИНА

ПРИЛОЖЕНИЕ КЪМ

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

**СВЕЩЕНИЯТ СМЕТАЧ  
ТОДОР АРНАУДОВ - ТОШ**

**2025**