

КОД **ИБ** ЕКАТЕРИНБУРГ · 15 ОКТЯБРЯ 2026

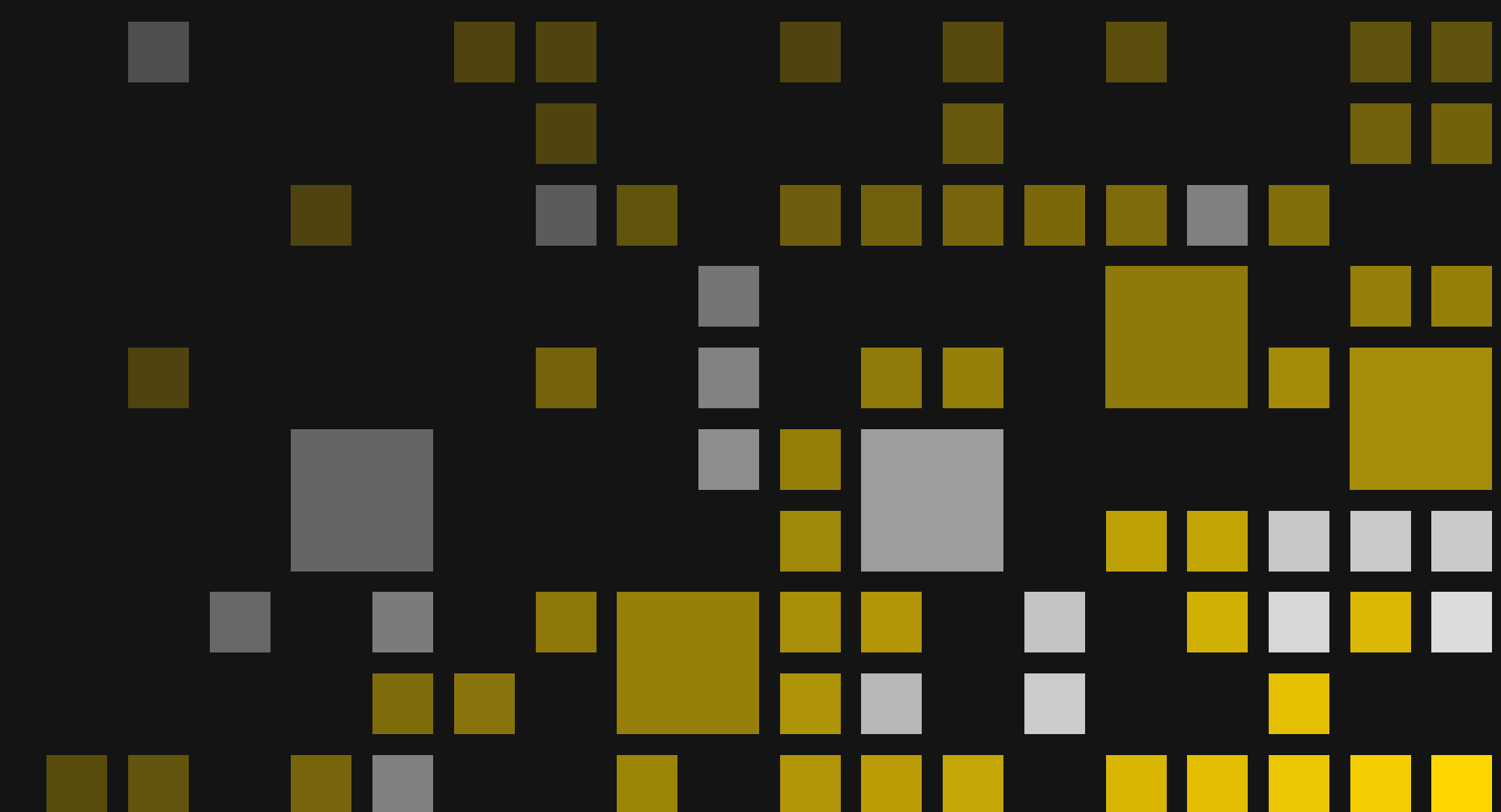
15:40–16:00

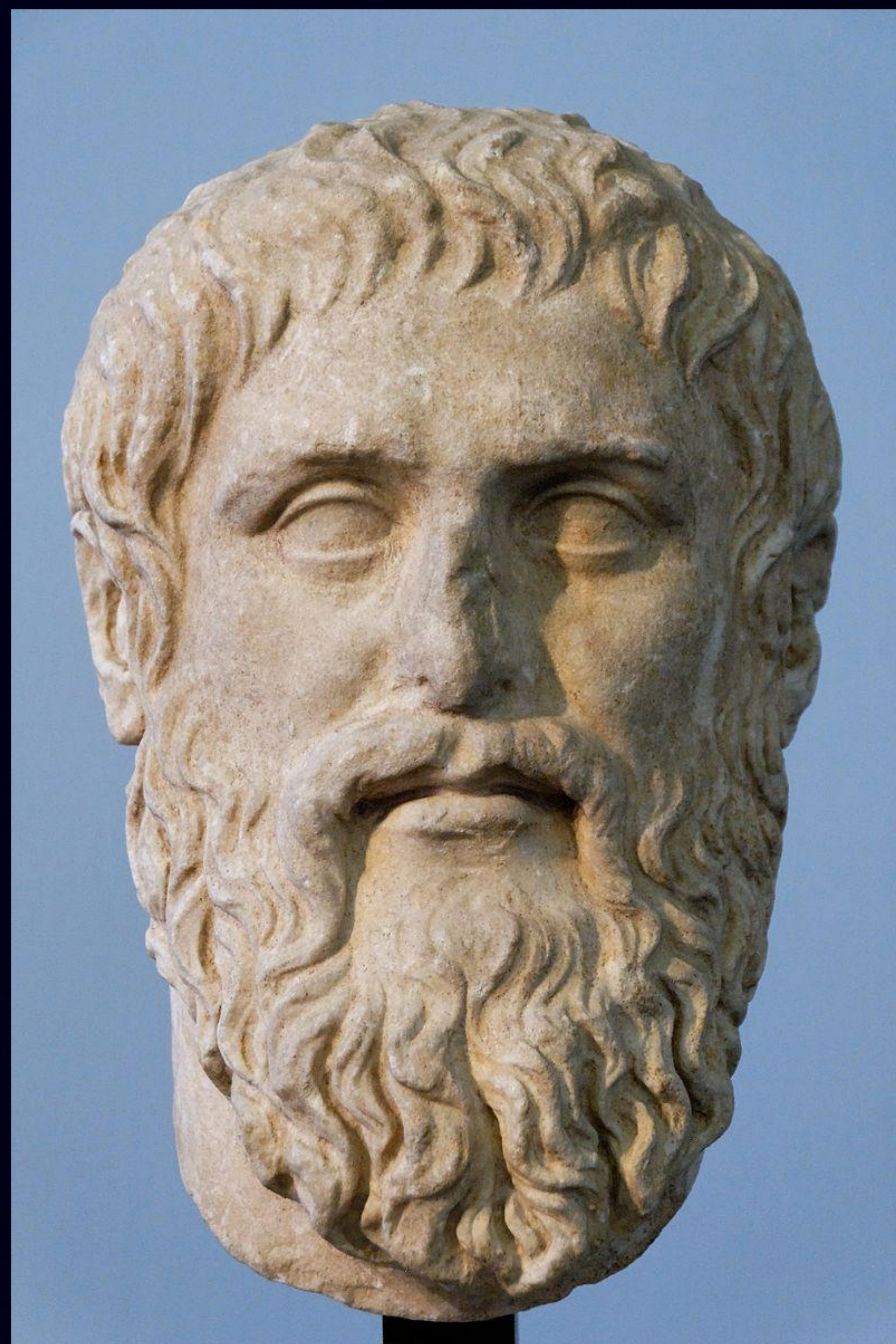
ЗАЛ 2

Съедят ли электроовцы людей? К чему приведёт бум нейросетей и заменит ли ИИ человека?

Сергей Марков

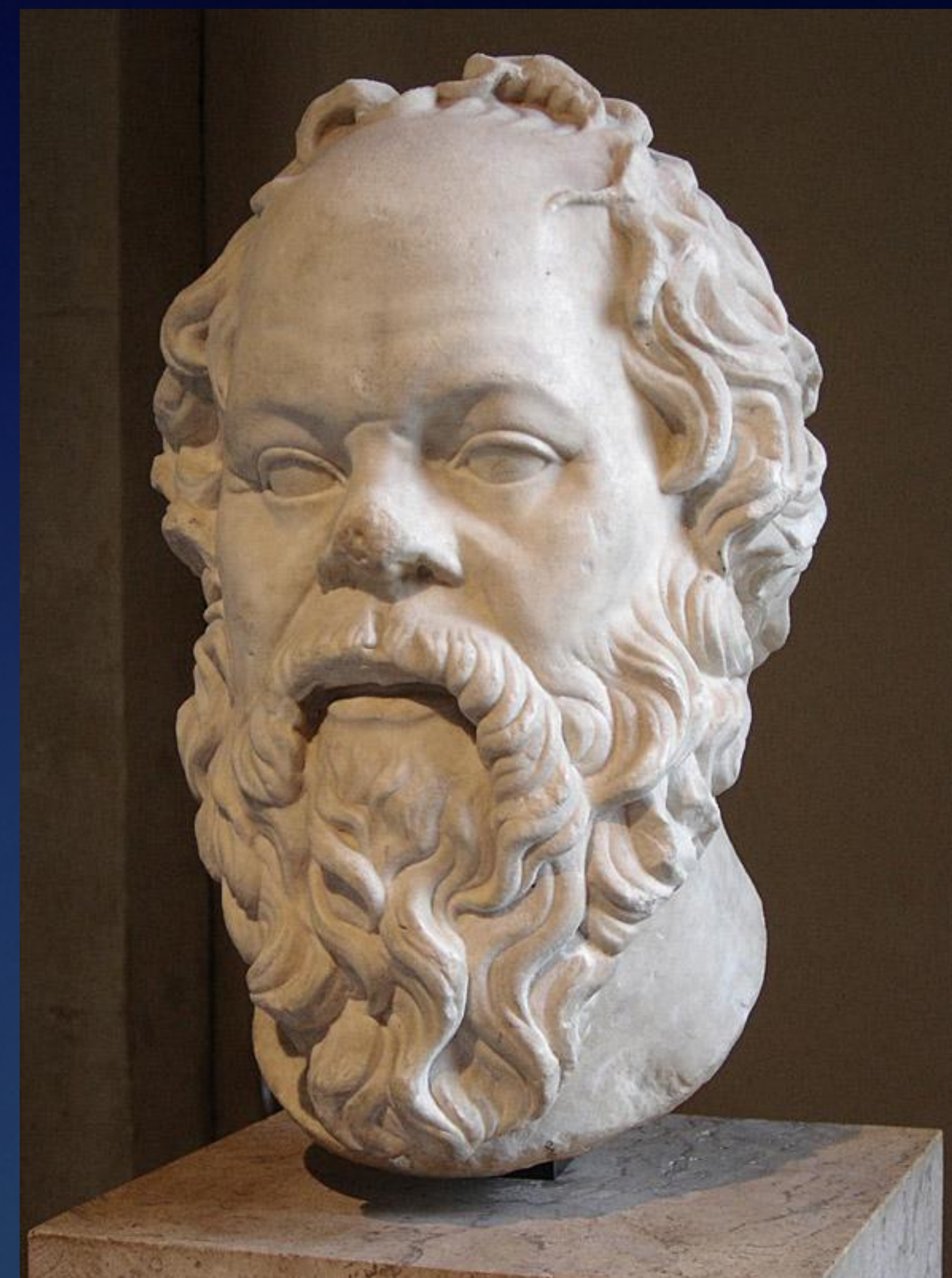
Директор по развитию технологий ИИ, Сбер





«...И со мной, Менон, точно так же: здесь я делю нужду моих сограждан и упрекаю себя в том, что вообще знать не знаю, что же такое добродетель. А если я этого не знаю, то откуда мне знать, как ее достичь? Разве, по-твоему, возможно, вообще не зная, кто такой Менон, знать, красив ли он, богат ли, знатен ли или же совсем наоборот?»

Платон, «Менон», 71 а–b



Что такое ИИ?



Автор термина «искусственный интеллект», Джон Маккарти (John McCarthy, 1927—2011), американский информатик, предложил его на границе 1955 и 1956 годов. Определение Маккарти звучало следующим образом: «Искусственный интеллект — наука и инженерная деятельность, направленная на создание умных (intelligent) машин»



Что такое ИИ?

1956 Dartmouth Conference: The Founding Fathers of AI



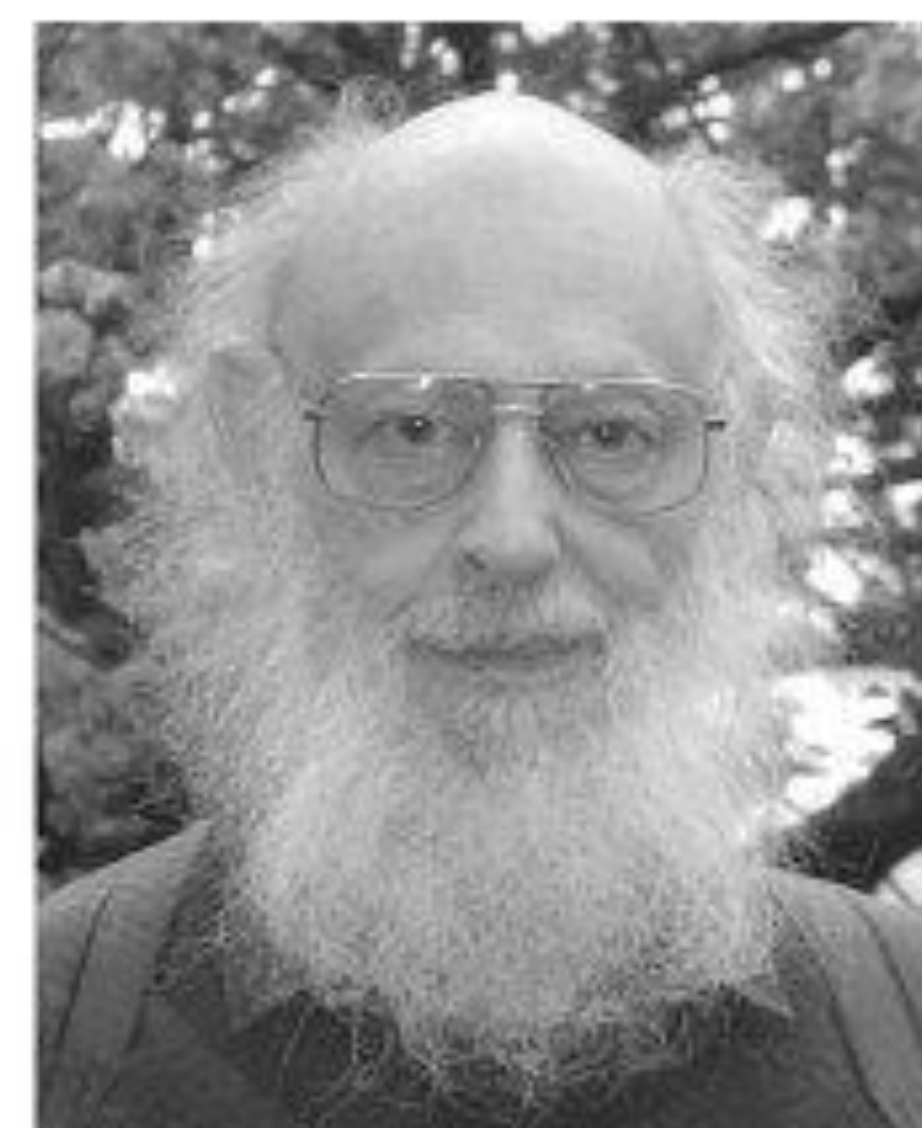
John McCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



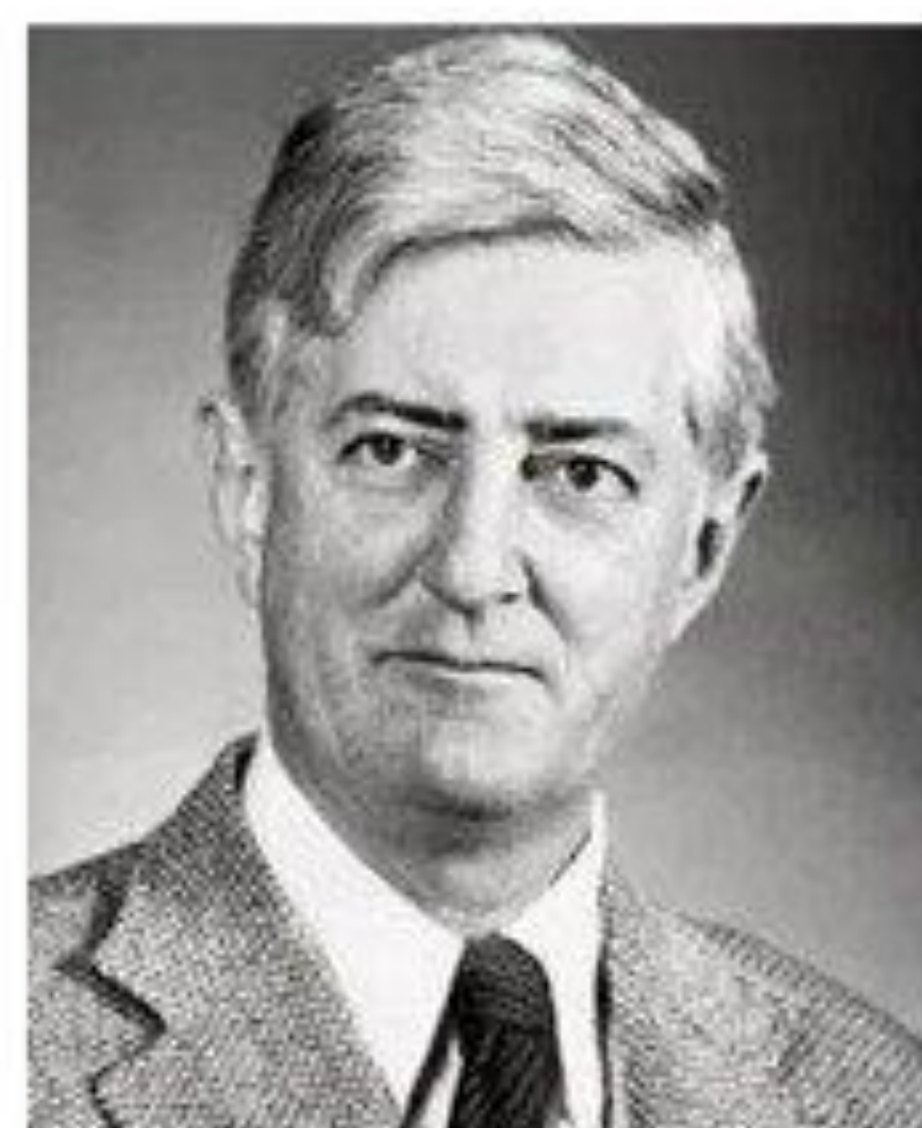
Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



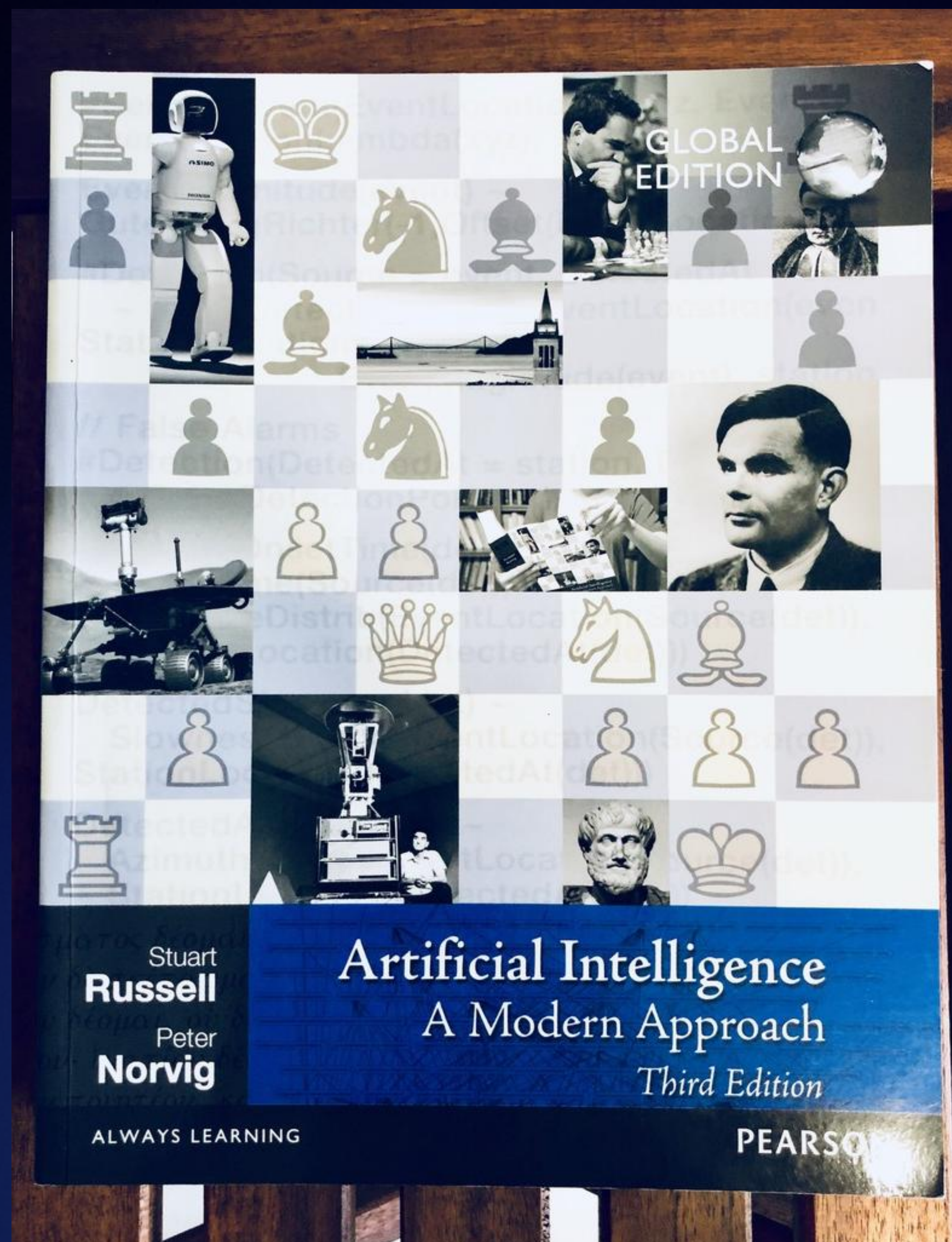
Nathaniel Rochester



Trenchard More

Что такое ИИ?

Стюарт Рассел (Stuart J. Russell, род. в 1962) и Питер Норвиг (Peter Norvig, род. в 1956) в своей классической работе «Искусственный интеллект: современный подход» (Artificial Intelligence: A Modern Approach) выделяют четыре различных исторических подхода к определению интересующего нас понятия, но в целом они сходятся в одном: **искусственный интеллект — это дисциплина, ставящая своей целью создание систем, способных решать интеллектуальные задачи, то есть такие задачи, которые люди традиционно решают при помощи собственного интеллекта. Или, как выразился Рэймонд Курцвейл (Raymond Kurzweil, род. в 1948): «Искусство создания машин, выполняющих функции, требующие интеллекта в случаях, когда их выполняют люди».**



Цель ИИ: расширение возможностей человека



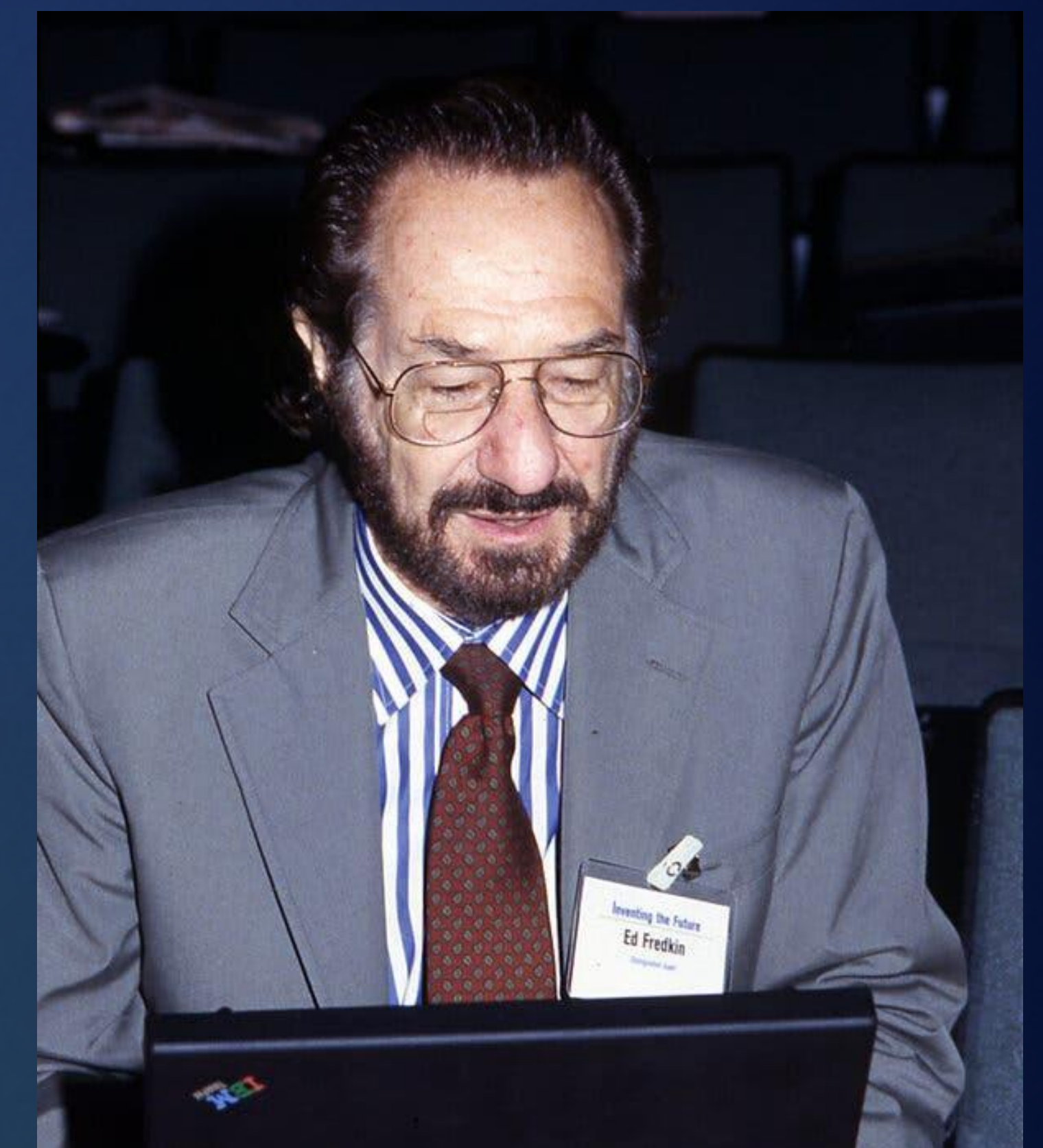
Под «усилением человеческого интеллекта» мы подразумеваем увеличение способности человека подходить к сложной проблемной ситуации, **достигать понимания, соответствующего его потребностям, и находить решения задач.** Под возросшими способностями в этом отношении понимается сочетание следующего: более быстрое понимание, лучшее понимание, возможность достичь необходимого уровня понимания в ситуации, которая раньше была слишком сложной, более быстрые решения, лучшие решения и возможность поиска решений проблем, которые раньше казались неразрешимыми. А к «сложным ситуациям» мы относим профессиональные проблемы дипломатов, руководителей, социологов, ученых-биологов, физиков, адвокатов, дизайнеров...

Дуглас Энгельбарт (1925—2013)

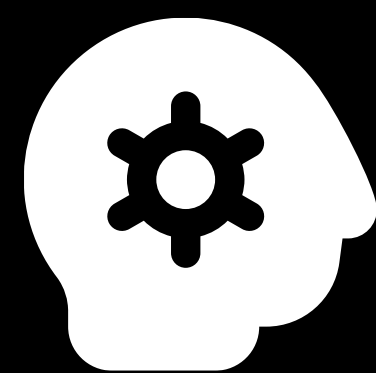
Получив благодаря симбиотической связи доступ к возможностям, которые мы теперь знаем, как реализовать в компьютерах, мы сможем <...> невиданными темпами достичь недостижимых ранее высот человеческого творчества. **Как партнеры с интеллектуальными системами, которые дополняют и компенсируют наши исправимые недостатки и в полной мере используют присущую нам креативность, мы могли бы развивать все области науки и искусства.**

Имея вычислительные крылья, мы могли бы взлететь навстречу Фениксу

Эдвард Фредкин (1934—2023)



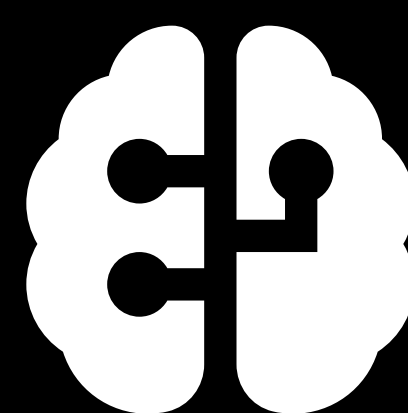
ИИ и развитие общества



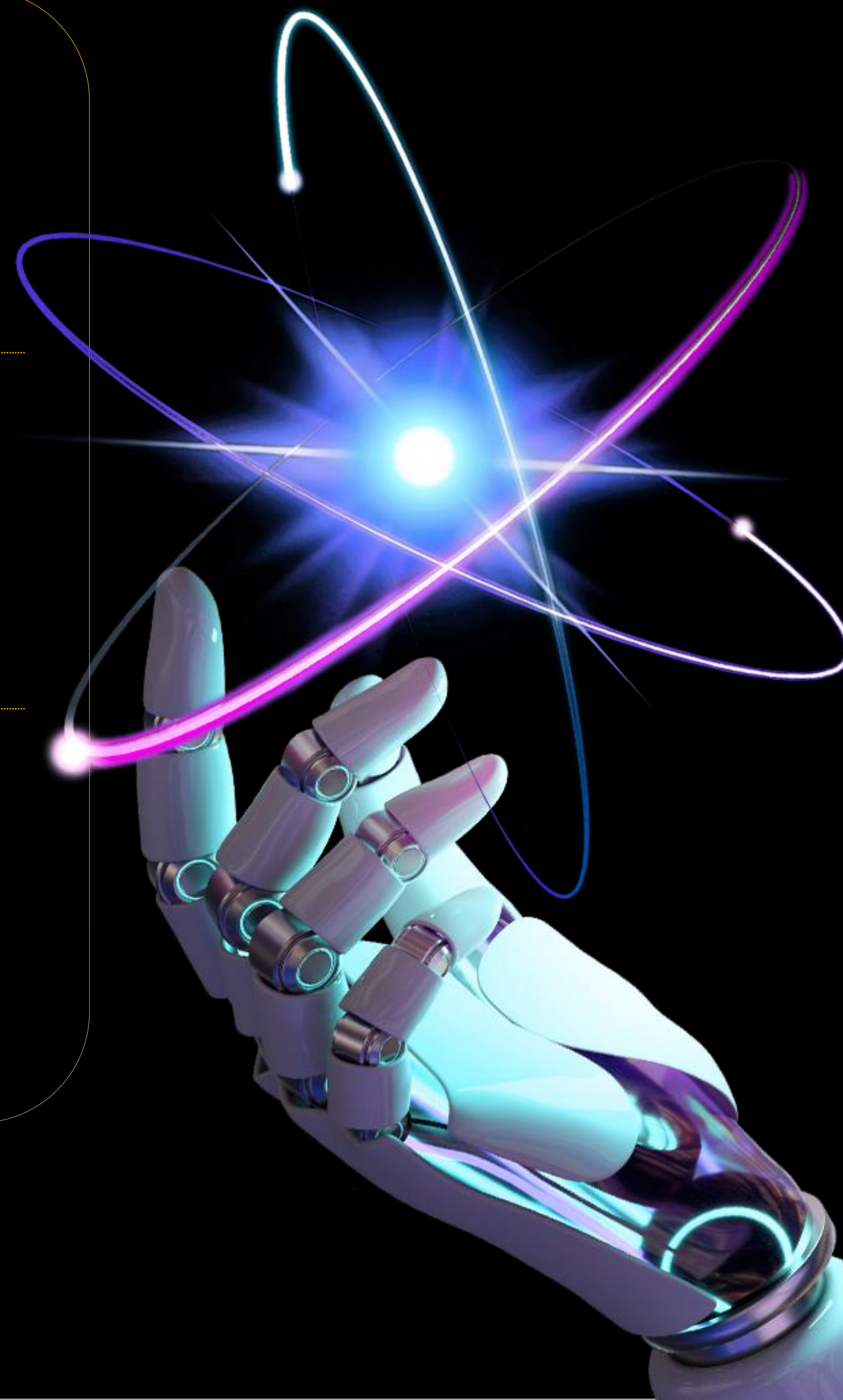
ИИ — область науки и технологии, занимающаяся автоматизацией решения интеллектуальных задач



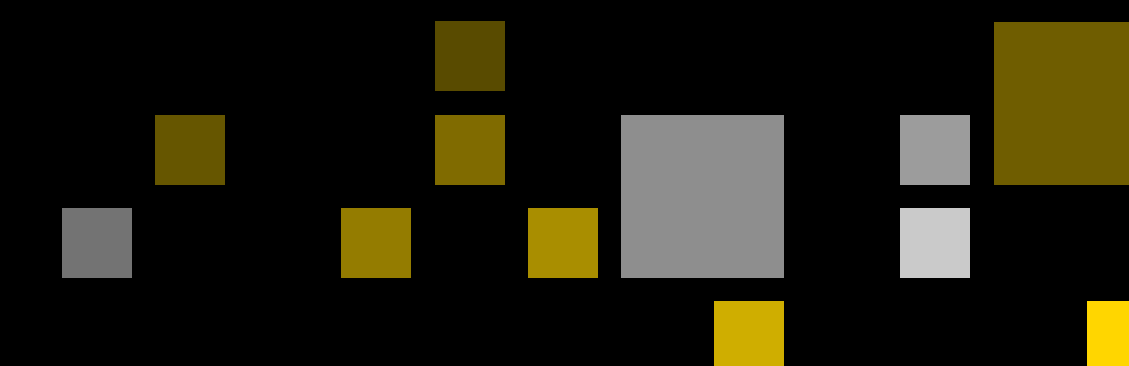
Растущее технологическое могущество человечества требует соразмерного роста качества решений



ИИ — набор технологий, которые позволяют усилить человеческий интеллект



**ИИ раздвигает
границы возможности
нашего интеллекта**



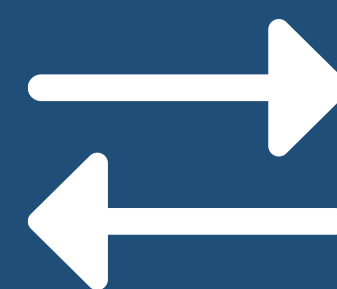
Некоторые барьеры человеческого разума



Скорость



Надежность



Ограниченный
параллелизм



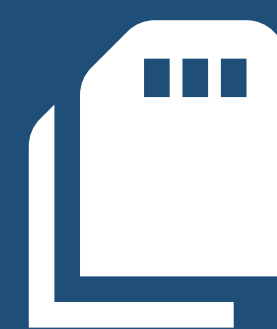
Коммуникативные
ограничения



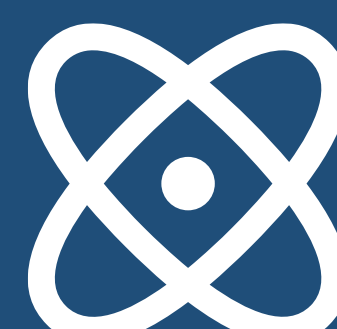
Барьеры физической
хрупкости



Порог сложности
анализируемых систем



Ограниченность
памяти



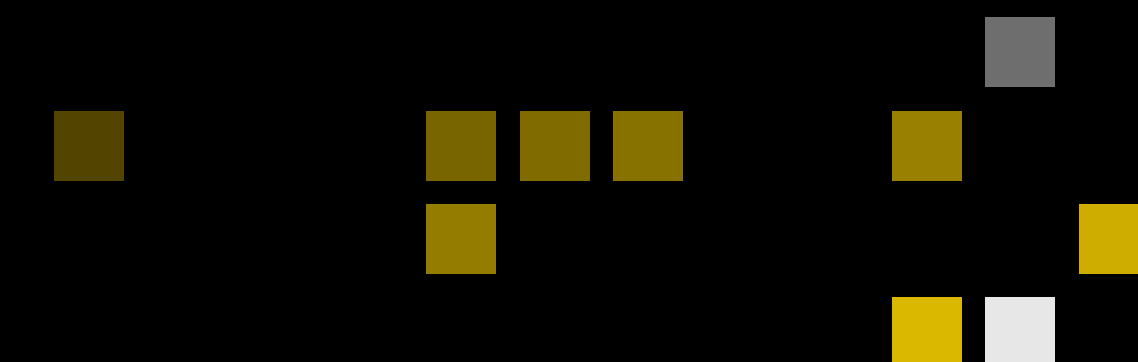
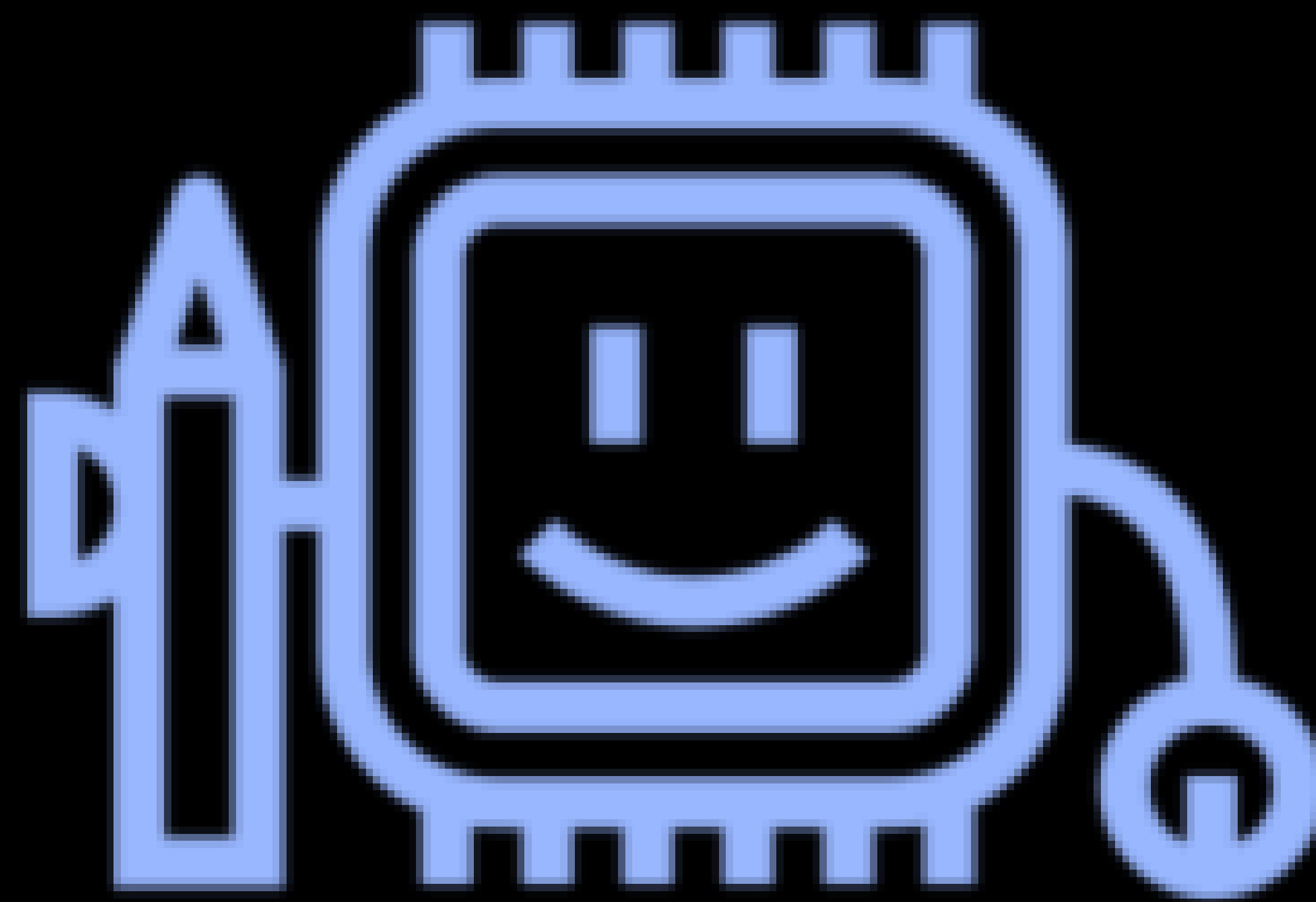
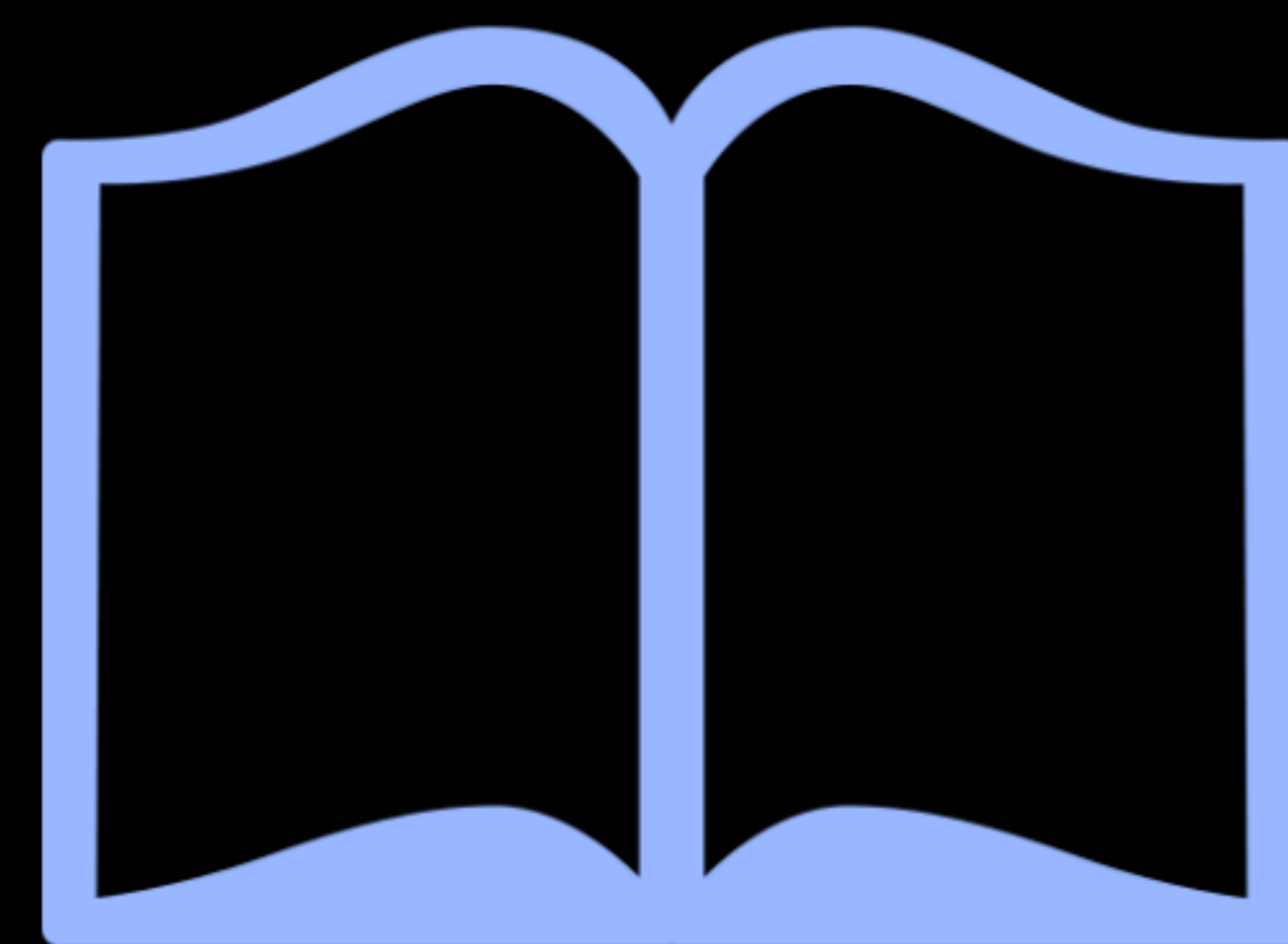
Ограниченные
экспертные
навыки и знания



Барьеры мотивации



Барьеры
масштабирования
и репликации



Новая и новейшая история ИИ



1943+: первые модели искусственных нейросетей

1940-50-е: Появление машинного обучения и нейронных сетей, первые эксперименты

1960-70-е: нейрокомпьютеры, Перцептрон Розенблатта и ADALINE Уидроу, основы теории нейронных сетей, первые прикладные проекты (SRI), ML в играх (Шашечная программа Сэмьюэла, MENACE и т. д.), первые глубокие сети (Ивахненко)

1970-е: Эксперименты с локально-связными сетями (Неокогнитрон Фукусимы), первые реализации метода обратного распространения ошибки (Келли, Брайсон, Галушкин, Вербос)

1986+: Первая революция глубокого обучения — backprop + Differential Activation Deep Networks (статья Хинтона и др. в Nature)

1987: TDNN, CNN

1997: LSTM

1990-е — ранние 2010-е: создание первых больших открытых датасетов для ML (MNIST, CIFAR, ImageNet и др.)

2003: словарные вложения (эмбединги)

2012: AlexNet и взрывное развитие глубоких сетей

2013: революция W2V

2013+: Вторая революция глубокого обучения — механизм внимания, а затем — трансформеры, фундаментальные модели, развитие NAS, xAI, MoE, SSL, RL, генеративная революция

2014: механизм внимания

2017: трансформер

2018-2020: BERT, GPT-3 и другие предобученные модели — эффективное масштабирование до сотен миллиардов параметров, вторая революция в NLP

2020-2022: DALL•E, CLIP, GLIDE, T5, Perceiver, WuDao, Imagen, JukeBox, GATO, BLOOM, ChatGPT, VALL-E, многозадачность и мультимодальность, PLP

2023+: расцвет генеративного ИИ, «эпоха ChatGPT»

Лето ИИ



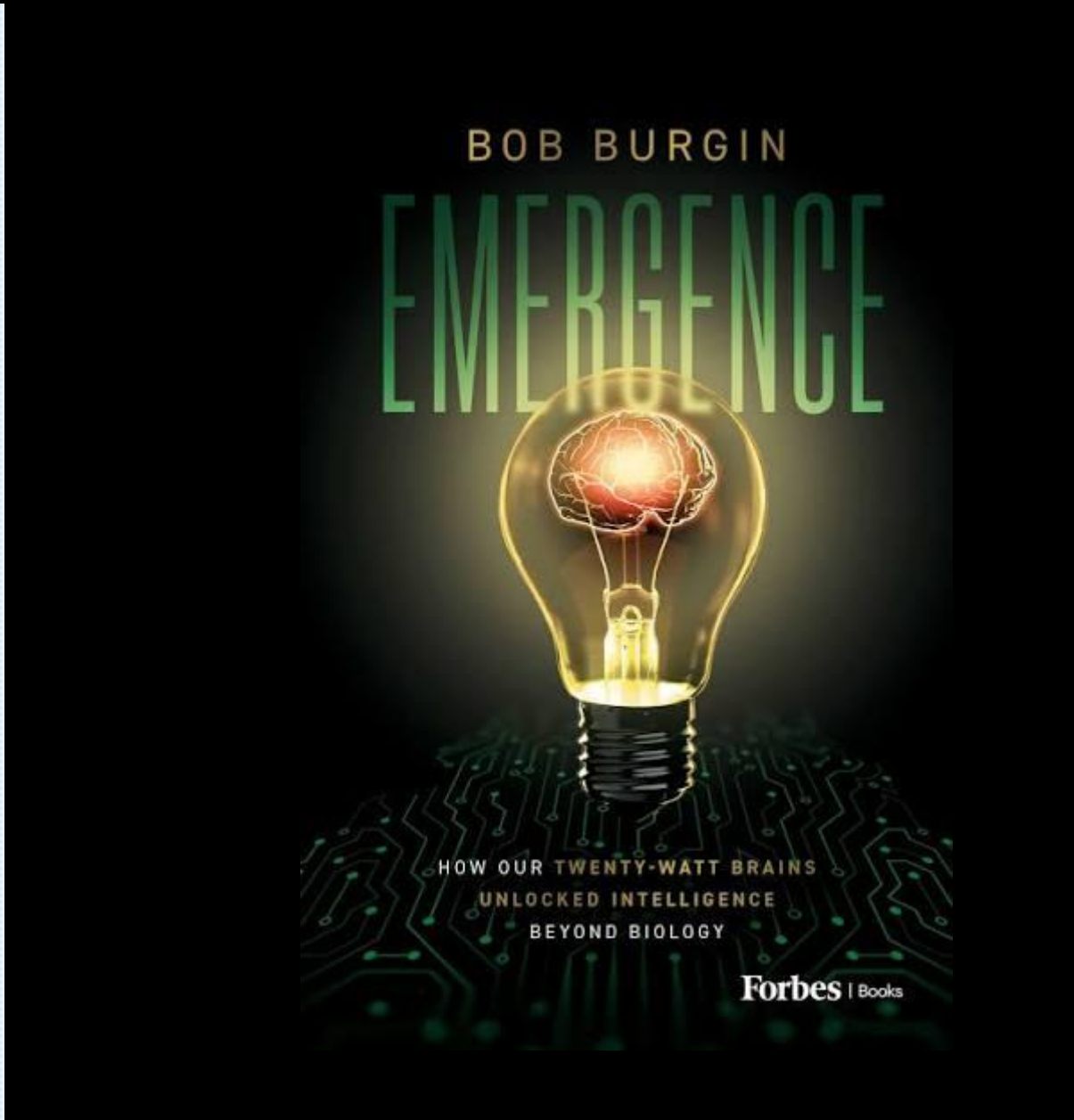
Technology, IT etc.

GOFAI

means

Good Old Fashioned Artificial Intelligence

by acronymsandslang.com



The Bitter Lesson

Rich Sutton

March 13, 2019

The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin. The ultimate reason for this is Moore's law, or rather its generalization of continued exponentially falling cost per unit of computation. Most AI research has been conducted as if the computation available to the agent were constant (in which case leveraging human knowledge would be one of the only ways to improve performance) but, over a slightly longer time than a typical research project, massively more computation inevitably becomes available. Seeking an improvement that makes a difference in the shorter term, researchers seek to leverage their human knowledge of the domain, but the only thing that matters in the long run is the leveraging of computation. These two need not run counter to each other, but in practice they tend to. Time spent on one is time not spent on the other. There are psychological commitments to investment in one approach or the other. And the human-knowledge approach tends to complicate methods in ways that make them less suited to taking advantage of general methods leveraging computation. There were many examples of AI researchers' belated learning of this bitter lesson, and it is instructive to review some of the most prominent.

In computer chess, the methods that defeated the world champion, Kasparov, in 1997, were based on massive, deep search. At the time, this was looked upon with dismay by the majority of computer-chess researchers who had pursued methods that leveraged human understanding of the special structure of chess. When a simpler, search-based approach with special hardware and software proved vastly more effective, these human-knowledge-based chess researchers were not good losers. They said that "brute force" search may have won this time, but it was not a general strategy, and anyway it was not how people played chess. These researchers wanted methods based on human input to win and were disappointed when they did not.

A similar pattern of research progress was seen in computer Go, only delayed by a further 20 years. Enormous initial efforts went into avoiding search by taking advantage of human knowledge, or of the special features of the game, but all those efforts proved irrelevant, or worse, once search was applied effectively at scale. Also important was the use of learning by self play to learn a value function (as it was in many other games and even in chess, although learning did not play a big role in the 1997 program that first beat a world champion). Learning by self play, and learning in general, is like search in that it enables massive computation to be brought to bear. Search and learning are the two most important classes of techniques for utilizing massive amounts of computation in AI research. In computer Go, as in computer chess, researchers' initial effort was directed towards utilizing human understanding (so that less search was needed) and only much later was much greater success had by embracing search and learning.

In speech recognition, there was an early competition, sponsored by DARPA, in the 1970s. Entrants included a host of special methods that took advantage of human knowledge---knowledge of words, of phonemes, of the human vocal tract, etc. On the other side were newer methods that were more statistical in nature and did much more computation, based on hidden Markov models (HMMs). Again, the statistical methods won out over the human-knowledge-based methods. This led to a major change in all of natural language processing, gradually over decades, where statistics and computation came to dominate the field. The recent rise of deep learning in speech recognition is the most recent step in this consistent direction. Deep learning methods rely even less on human knowledge, and use even more computation, together with learning on huge training sets, to produce dramatically better speech recognition systems. As in the games, researchers always tried to make systems that worked the way the researchers thought their own minds worked---they tried to put that knowledge in their systems---but it proved ultimately counterproductive, and a colossal waste of researcher's time, when, through Moore's law, massive computation became available and a means was found to put it to good use.

In computer vision, there has been a similar pattern. Early methods conceived of vision as searching for edges, or generalized cylinders, or in terms of SIFT features. But today all this is discarded. Modern deep-learning neural networks use only the notions of convolution and certain kinds of invariances, and perform much better.

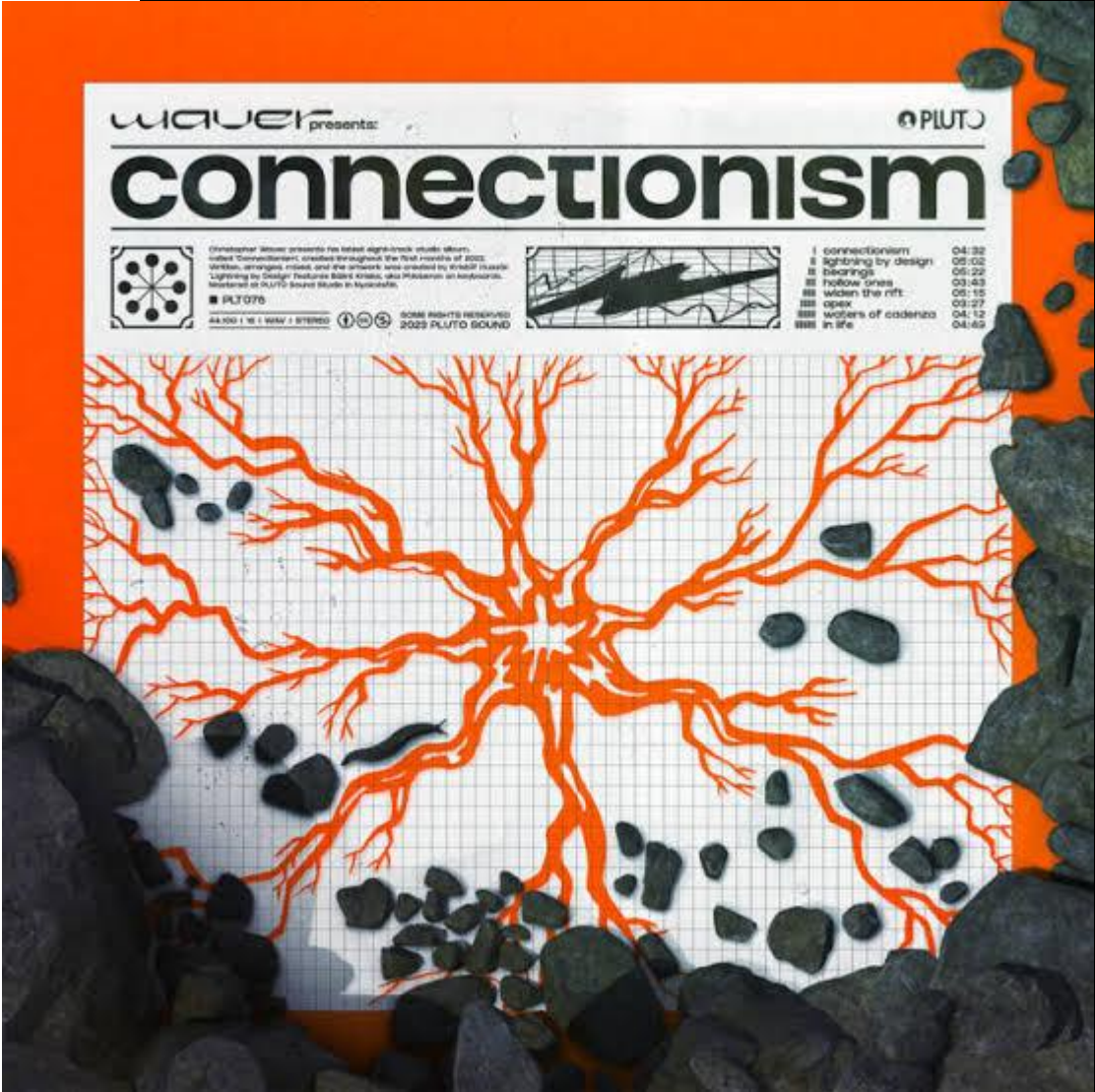
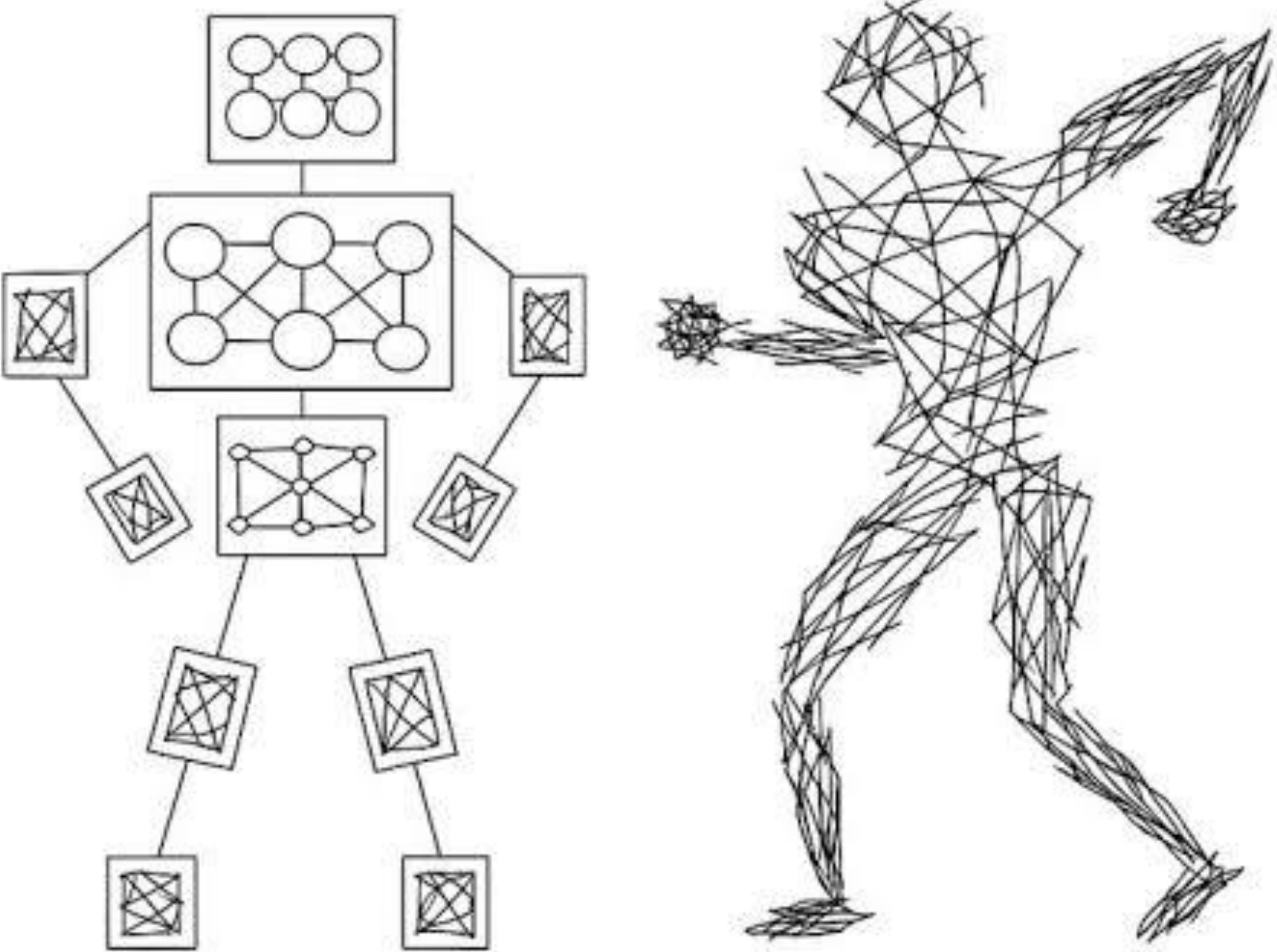
This is a big lesson. As a field, we still have not thoroughly learned it, as we are continuing to make the same kind of mistakes. To see this, and to effectively resist it, we have to understand the appeal of these mistakes. We have to learn the bitter lesson that building in how we think we think does not work in the long run. The bitter lesson is based on the historical observations that 1) AI researchers have often tried to build knowledge into their agents, 2) this always helps in the short term, and is personally satisfying to the researcher, but 3) in the long run it plateaus and even inhibits further progress, and 4) breakthrough progress eventually arrives by an opposing approach based on scaling computation by search and learning. The eventual success is tinged with bitterness, and often incompletely digested, because it is success over a favored, human-centric approach.

One thing that should be learned from the bitter lesson is the great power of general purpose methods, of methods that continue to scale with increased computation even as the available computation becomes very great. The two methods that seem to scale arbitrarily in this way are *search* and *learning*.

The second general point to be learned from the bitter lesson is that the actual contents of minds are tremendously, irredeemably complex; we should stop trying to find simple ways to think about the contents of minds, such as simple ways to think about space, objects, multiple agents, or symmetries. All these are part of the arbitrary, intrinsically-complex, outside world. They are not what should be built in, as their complexity is endless; instead we should build in only the meta-methods that can find and capture this arbitrary complexity. Essential to these methods is that they can find good approximations, but the search for them should be by our methods, not by us. We want AI agents that can discover like we can, not which contain what we have discovered. Building in our discoveries only makes it harder to see how the discovering process can be done.



Neats vs Scruffies



YouTube • Dwarkesh Patel

Richard Sutton – Father of RL thinks LLMs are a dead end

Richard Sutton is the father of reinforcement learning, winner of the 202...

26 сент. 2025 г.

От генеративного ИИ к интерактивному



Первая волна ИИ касалась классификации. Глубокое обучение показало, что мы можем научить компьютер классифицировать различные типы входных данных: изображения, видео, аудио, язык. **Сейчас мы находимся в генеративной волне**, когда вы берете входные данные и создаете новые данные.

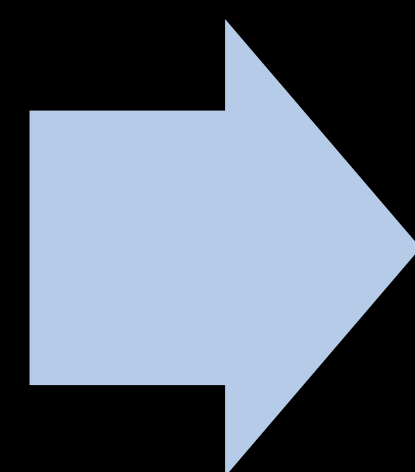
Третья волна будет интерактивной фазой. Вот почему я уже давно уверен, что диалог — это интерфейс будущего. Знаете, вместо того, чтобы просто нажимать на кнопки и печатать, вы будете разговаривать со своей системой ИИ.

И эти системы смогут действовать. Вы просто дадите системе общую цель высокого уровня, и она будет использовать все имеющиеся у неё инструменты для достижения этой цели. Системы будут разговаривать с другими людьми, разговаривать с другими системами...

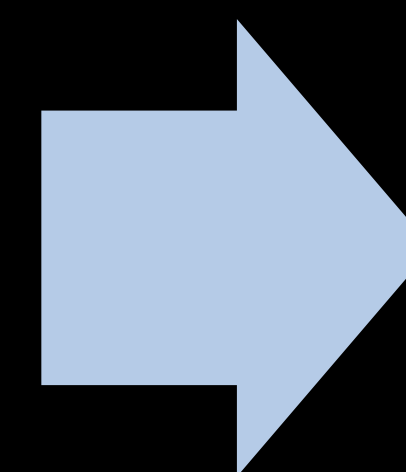
Это огромный сдвиг в том, на что способны технологии. Это очень, очень глубокий момент в истории технологий, который, я думаю, многие люди недооценивают.

Мустафа Сулейман, один из основателей DeepMind

Дискриминативный
ИИ (...—2017)



Генеративный ИИ
(2017—2024)



Интерактивный ИИ
(2024+)

Агент — строительный блок интерактивного ИИ

КОД ИБ

ЕКАТЕРИНБУРГ · 15 ОКТЯБРЯ 2026

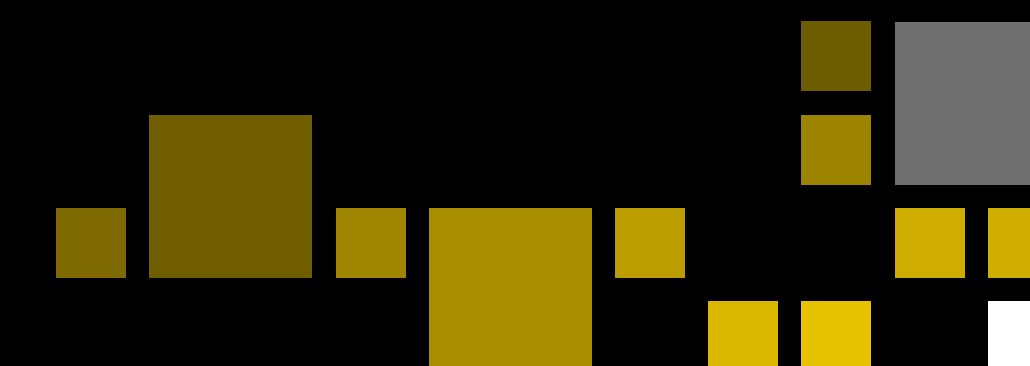
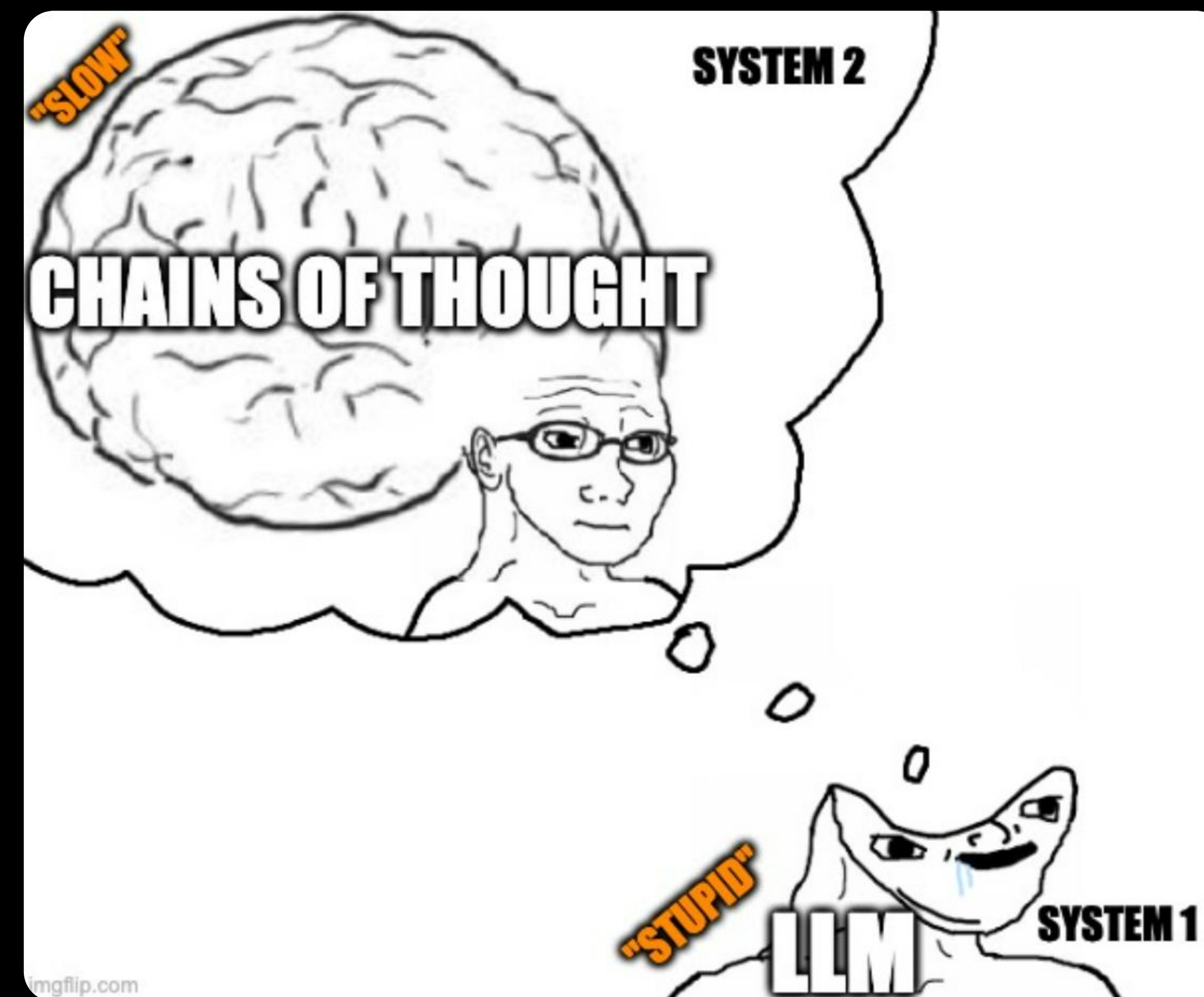
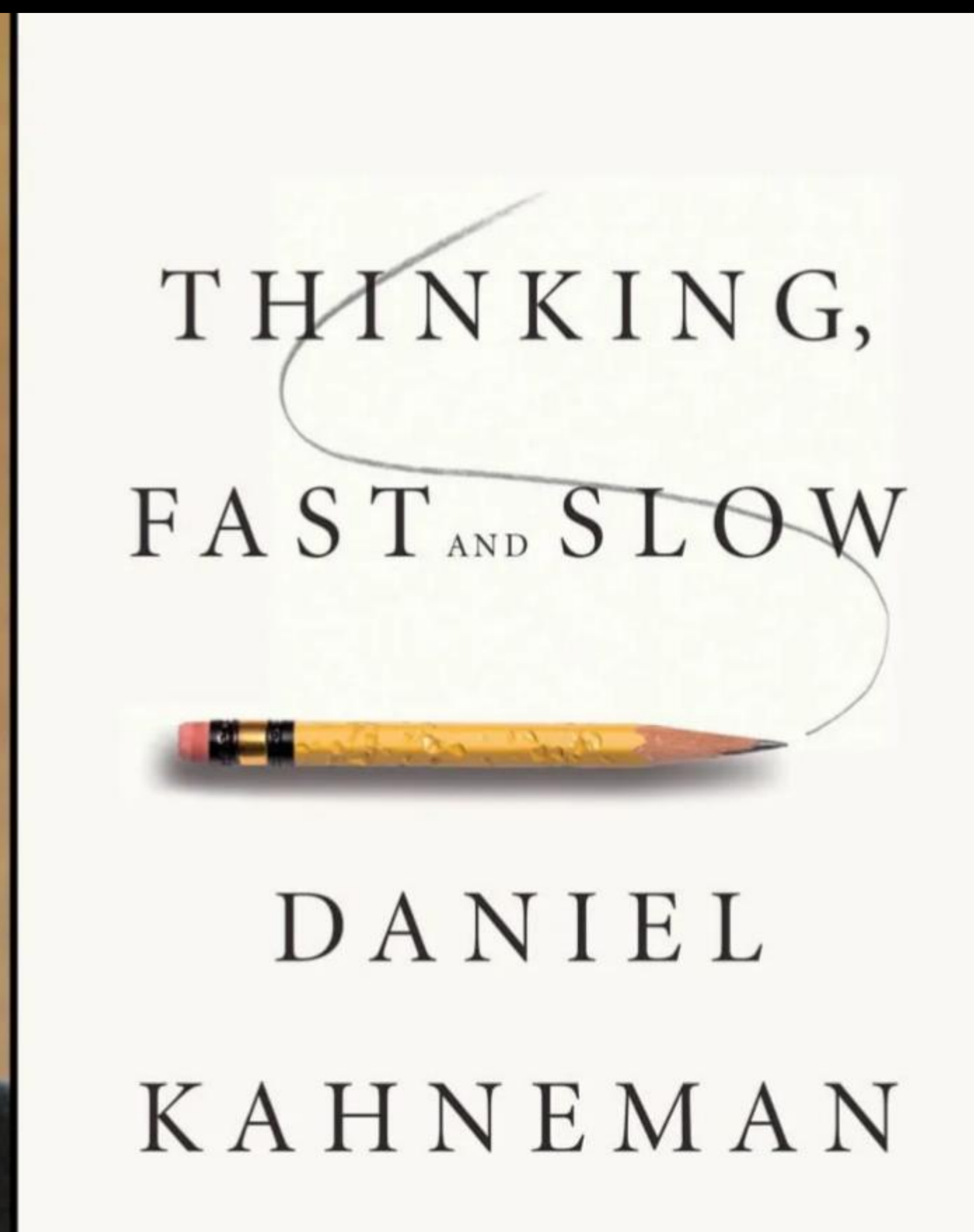
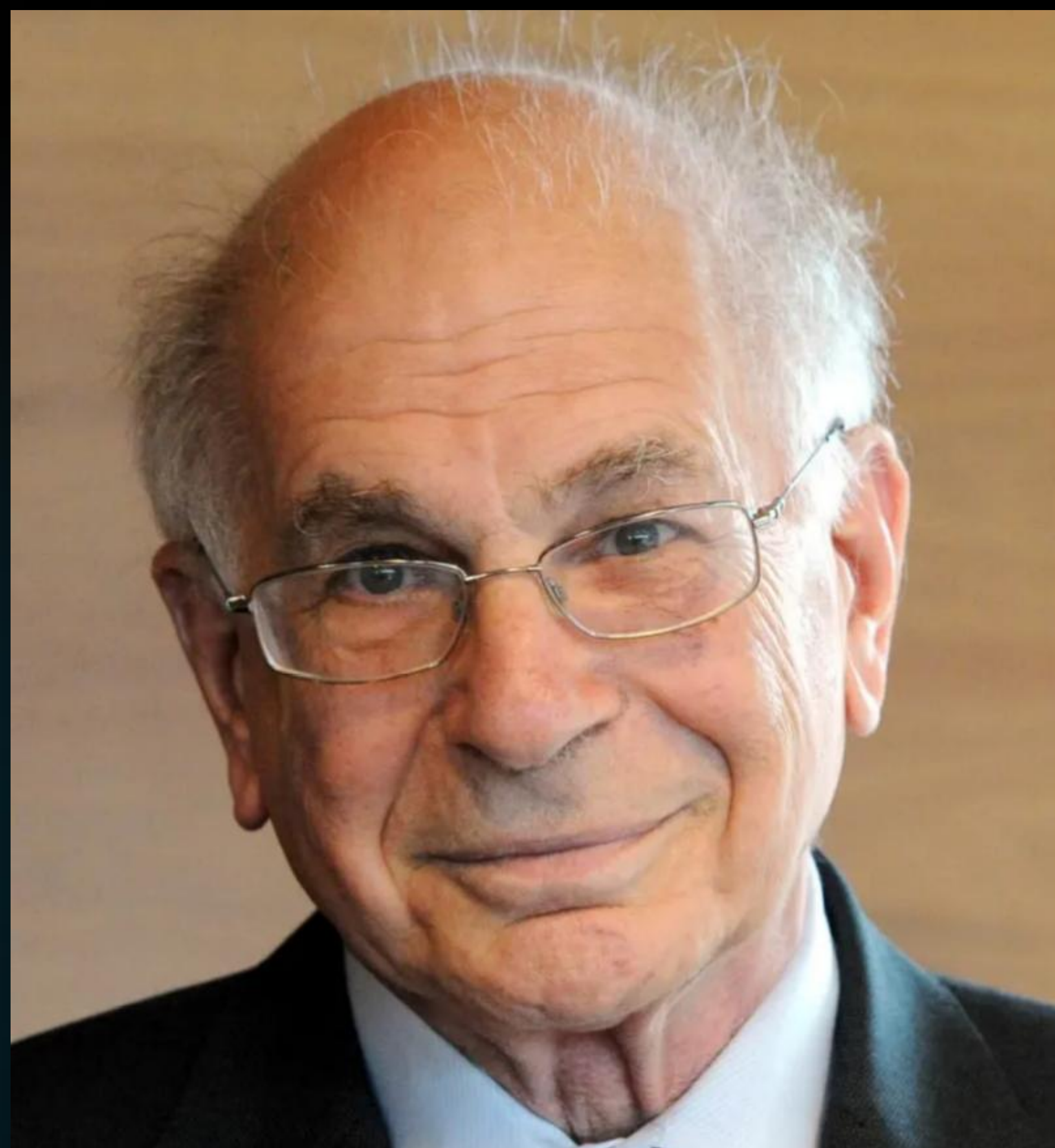


Position: Foundation Agents as the Paradigm Shift for Decision Making

[Xiaoqian Liu](#), [Xingzhou Lou](#), [Jianbin Jiao](#), [Junge Zhang](#)

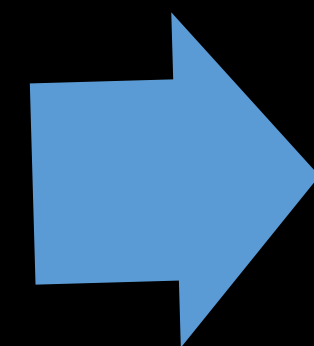
Decision making demands intricate interplay between perception, memory, and reasoning to discern optimal policies. Conventional approaches to decision making face challenges related to low sample efficiency and poor generalization. In contrast, foundation models in language and vision have showcased rapid adaptation to diverse new tasks. Therefore, we advocate for the construction of foundation agents as a transformative shift in the learning paradigm of agents. This proposal is underpinned by the formulation of foundation agents with their fundamental characteristics and challenges motivated by the success of large language models (LLMs). Moreover, we specify the roadmap of foundation agents from large interactive data collection or generation, to self-supervised pretraining and adaptation, and knowledge and value alignment with LLMs. Lastly, we pinpoint critical research questions derived from the formulation and delineate trends for foundation agents supported by real-world use cases, addressing both technical and theoretical aspects to propel the field towards a more comprehensive and impactful future.

Моделирование рассуждений

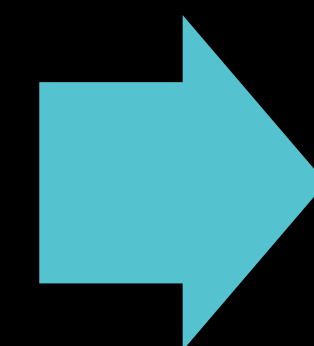


Рассуждения языковых моделей

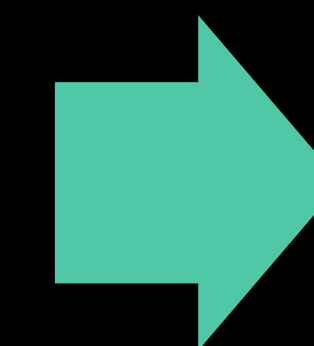
Январь 2022: препринт статьи Jason Wei и его коллег, посвящённая генерации цепочек рассуждений (chain-of-thought) большими языковыми моделями



Май 2022: препринт работы Takeshi Kojima и его коллег, в которой предлагается идея инструкции «Let's think step by step» для генерации цепочек рассуждений; граница применения рассуждений расширяется за пределы простых арифметических и логических задач



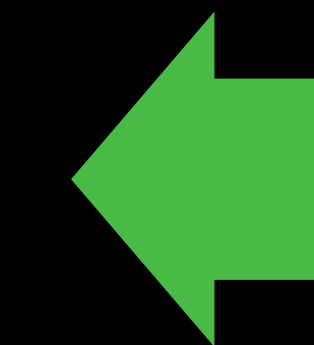
Ноябрь 2023: СМИ впервые пишут про Q*



Январь 2024: Сэм Альтман в подкасте у Билла Гейтса говорит, что «В области ИИ наиболее важные области прогресса будут связаны со способностью рассуждать»



Июль 2024: все СМИ (Bloomberg, WSJ, NYT) пишут, что OpenAI работает над ризонингом под кодовым названием «Strawberry»



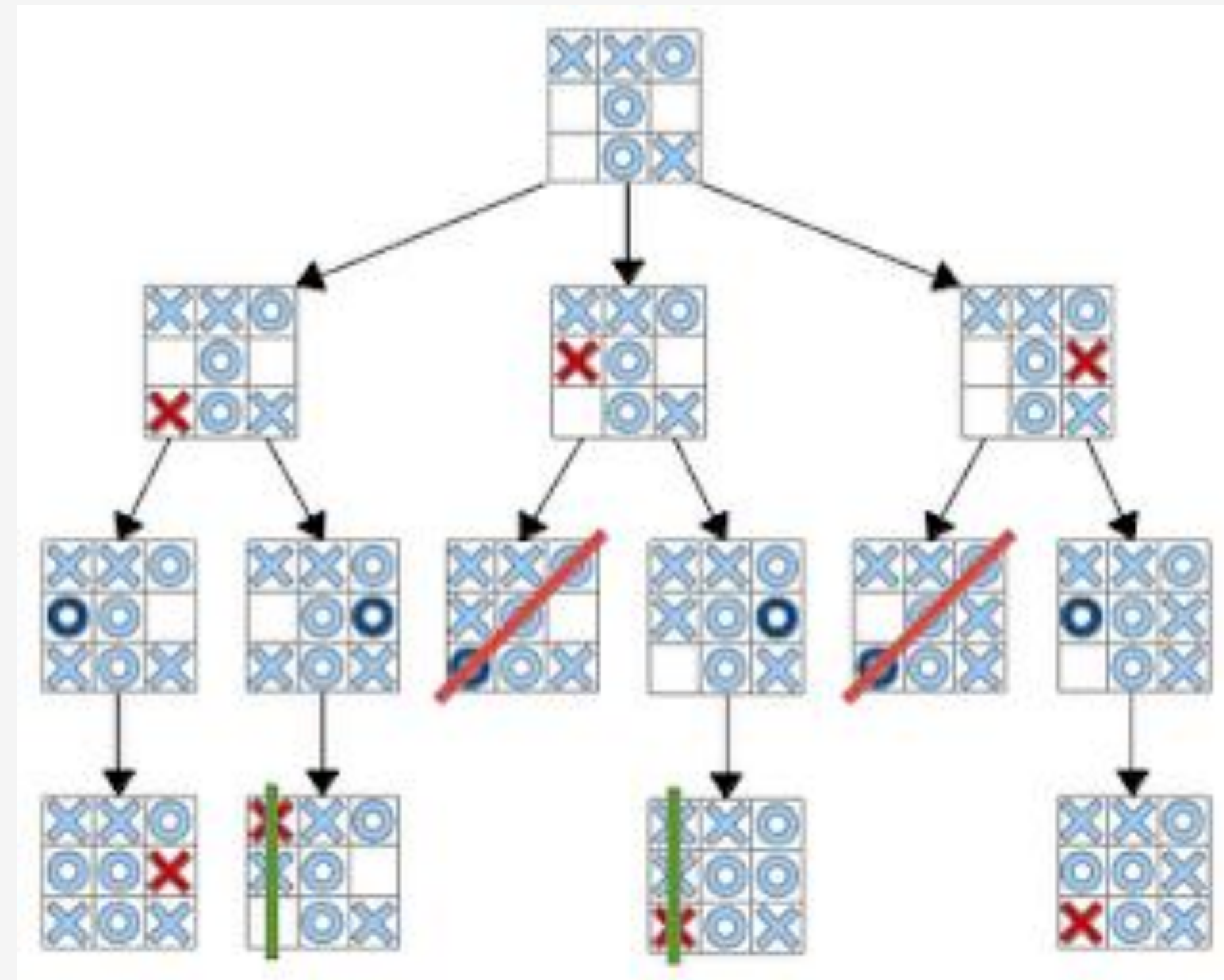
Сентябрь 2024: выход модели o1, которая умеет рассуждать



Январь 2025: выход модели DeepSeek R1 в open-source



Модели мира и сети альтернатив



Computer Science > Machine Learning

[Submitted on 27 Mar 2018 (v1), last revised 9 May 2018 (this version, v4)]

World Models

David Ha, Jürgen Schmidhuber

We explore building generative neural network models of popular reinforcement learning environments. Our world model can be trained quickly in an unsupervised manner to learn a compressed spatial and temporal representation of the environment. By using features extracted from the world model as inputs to an agent, we can train a very compact and simple policy that can solve the required task. We can even train our agent entirely inside of its own hallucinated dream generated by its world model, and transfer this policy back into the actual environment.

An interactive version of this paper is available at [this https URL](https://arxiv.org/abs/1803.10122v4)

Subjects: **Machine Learning (cs.LG)**; Machine Learning (stat.ML)

Cite as: [arXiv:1803.10122](https://arxiv.org/abs/1803.10122) [cs.LG]

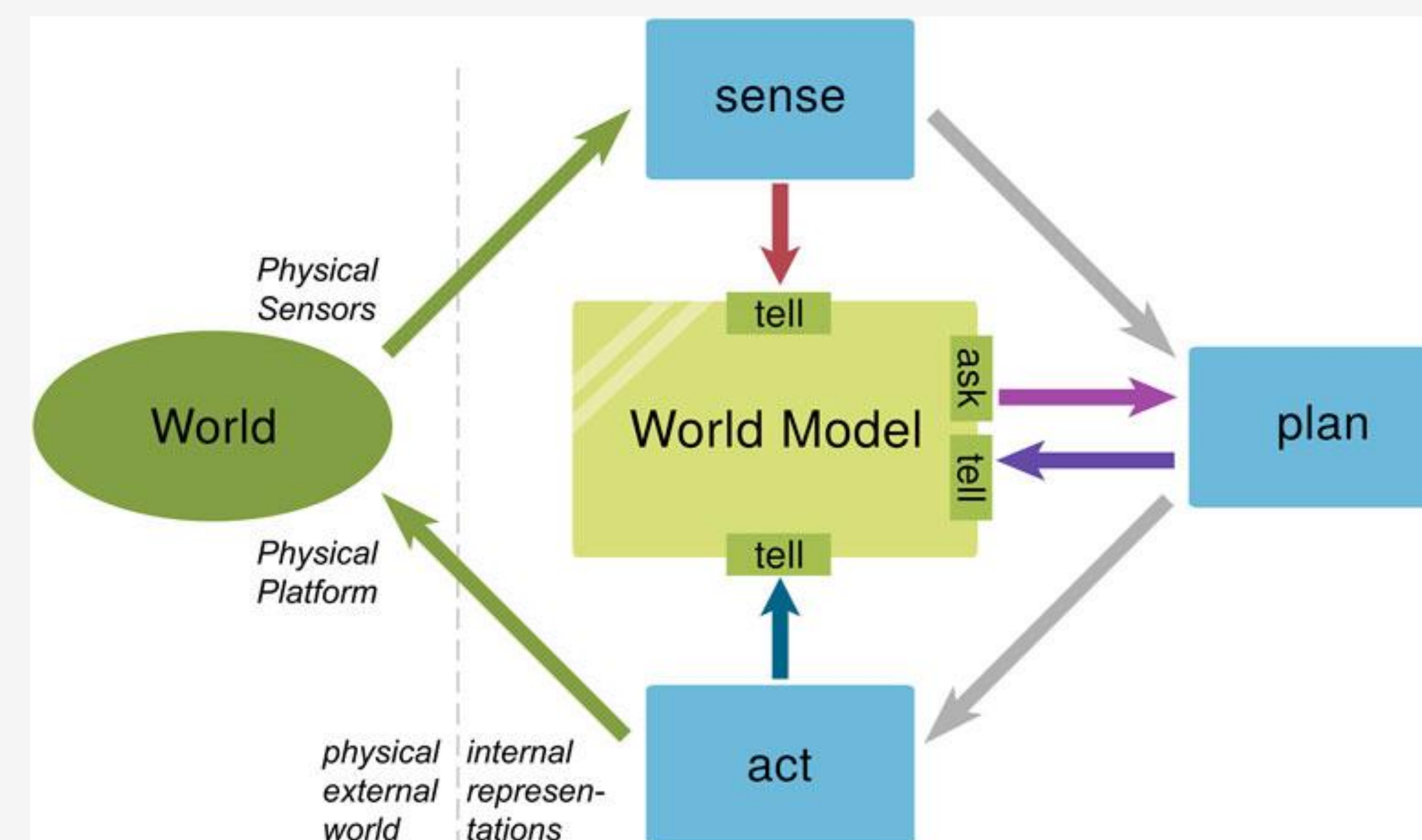
(or [arXiv:1803.10122v4](https://arxiv.org/abs/1803.10122v4) [cs.LG] for this version)

<https://doi.org/10.48550/arXiv.1803.10122>

Related DOI: <https://doi.org/10.5281/zenodo.1207631>

Модели мира современных трансформерных фундаментальных моделей дают возможность анализа сетей возможных альтернативных действий агентов в реальном мире с последующим анализом и оценкой последствий совершённого выбора

LLM/MLLM могут выполнять роль policy и evaluation networks AlphaGo для задач реального мира, что позволяет методам ИИ выйти за пределы игровых миров



Удалось увидеть, как в мозжечке личинок данио-рерио строятся модели взаимодействия тела с внешним миром

17.12.2021 • ДАНИИЛ МАРКОВ • НЕЙРОБИОЛОГИЯ, БИОЛОГИЯ РАЗВИТИЯ • 44 КОММЕНТАРИЯ



AGI как цель 2025—2030?

«Под продвинутым универсальным ИИ я понимаю системы ИИ, способные соперничать с человеческим мозгом или превосходить его в сложности и скорости, которые способны получать общие сведения, манипулировать ими и выстраивать рассуждения и которые могут использоваться, по сути, в любой фазе промышленных или военных операций, в которой был бы необходим человеческий интеллект»

Марк Губруд. Нанотехнологии и международная безопасность (1997)

«AGI — искусственный интеллект, способный успешно выполнить любую интеллектуальную задачу, посильную для человека»

Джеймс Селигман. Машинное обучение в маркетинге (2020)



AGI как цель 2025—2030?

«AGI — искусственный интеллект, способный успешно выполнить любую интеллектуальную задачу, посильную для человека»

Для какого человека?

В рамках какого временного /
энергетического / стоимостного
лимита?

Что значит «успешно
выполнить»?

Важны именно 100% задач?

Как проверять?

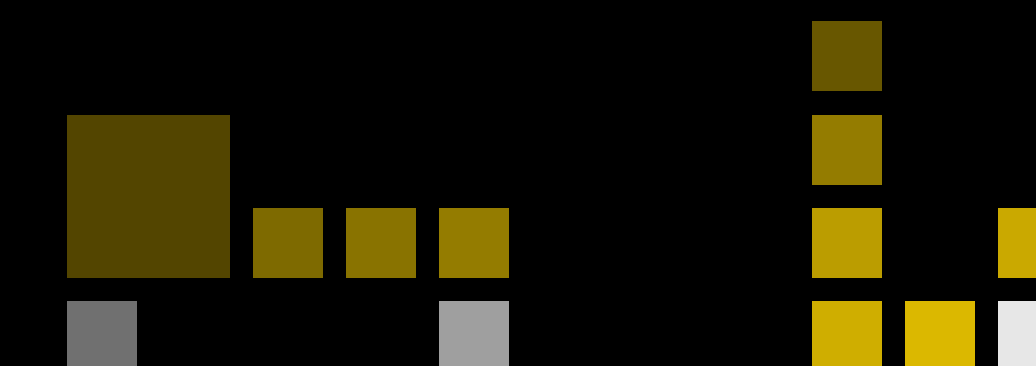


"По данным The Information, OpenAI и Microsoft в 2023 году согласовали конкретное определение искусственного общего интеллекта (AGI): система будет считаться достигшей уровня AGI, когда сможет генерировать прибыль не менее 100 миллиардов долларов"

The Information

Microsoft and OpenAI's Secret AGI Definition

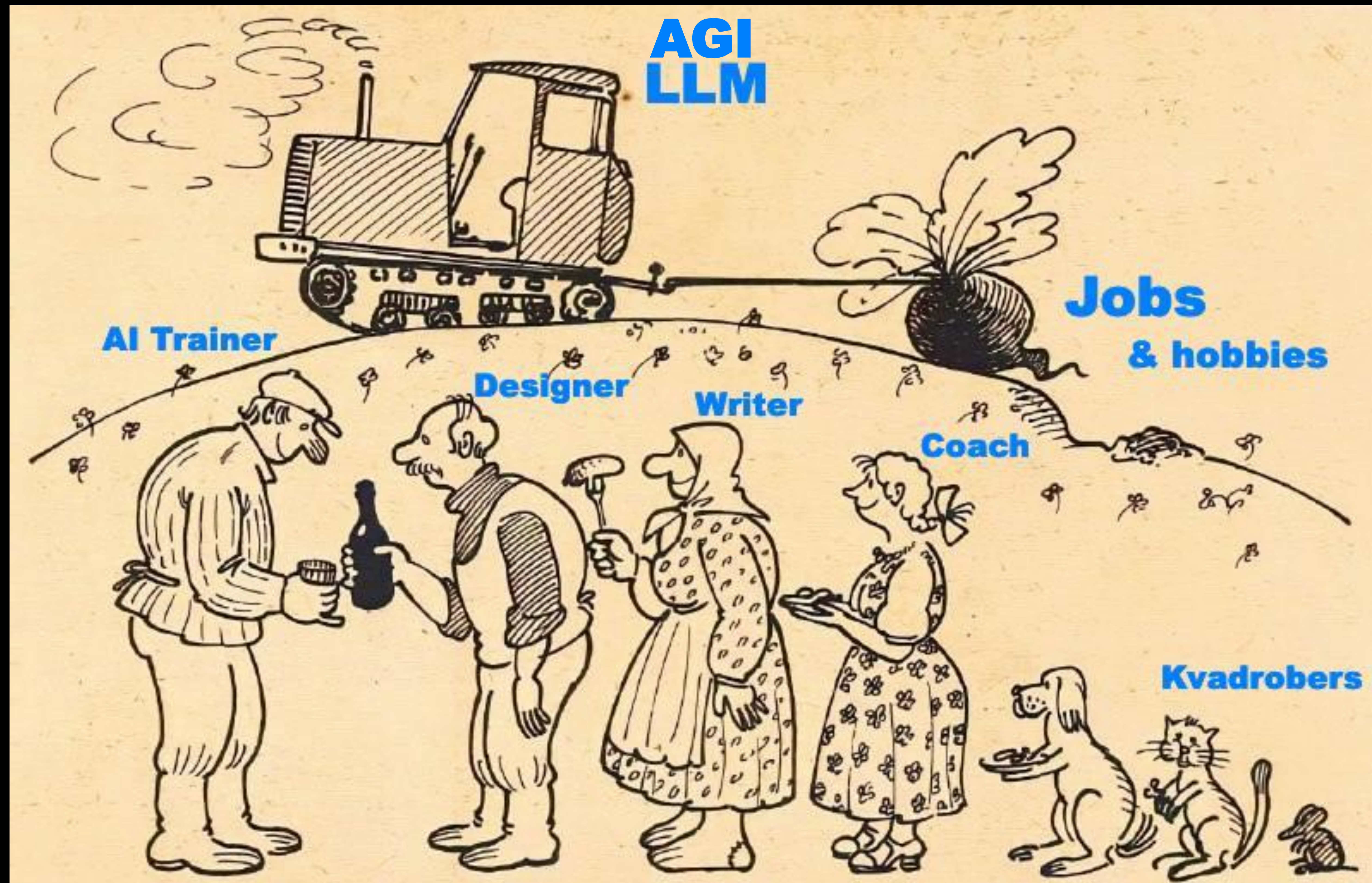
Finally, a verifiable, numbers-based description of artificial general intelligence has arrived! Whether AGI has or hasn't been "achieved" by AI developers has been a hotly debated topic du...



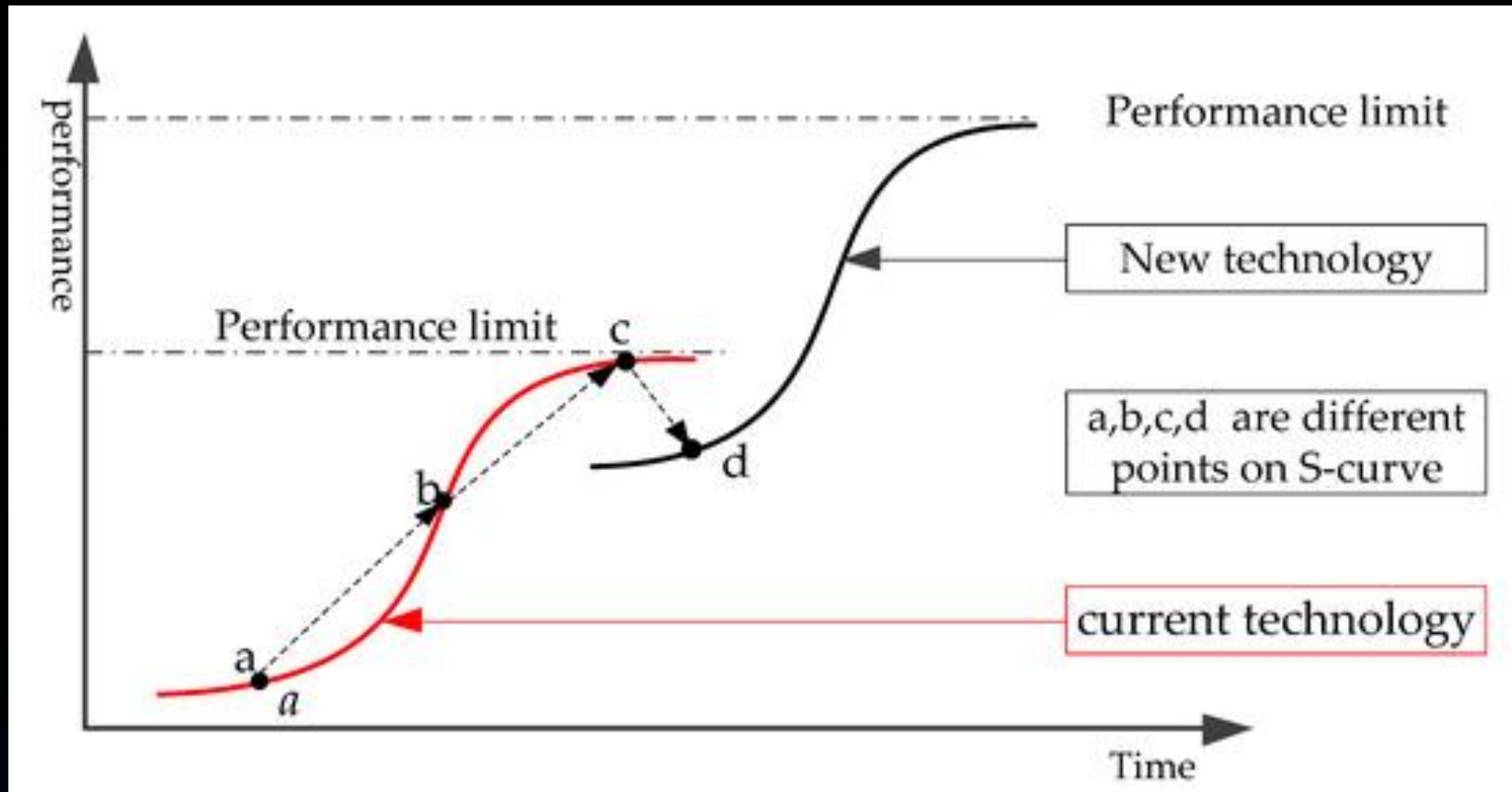
Теперь мы все живём в мире LLM



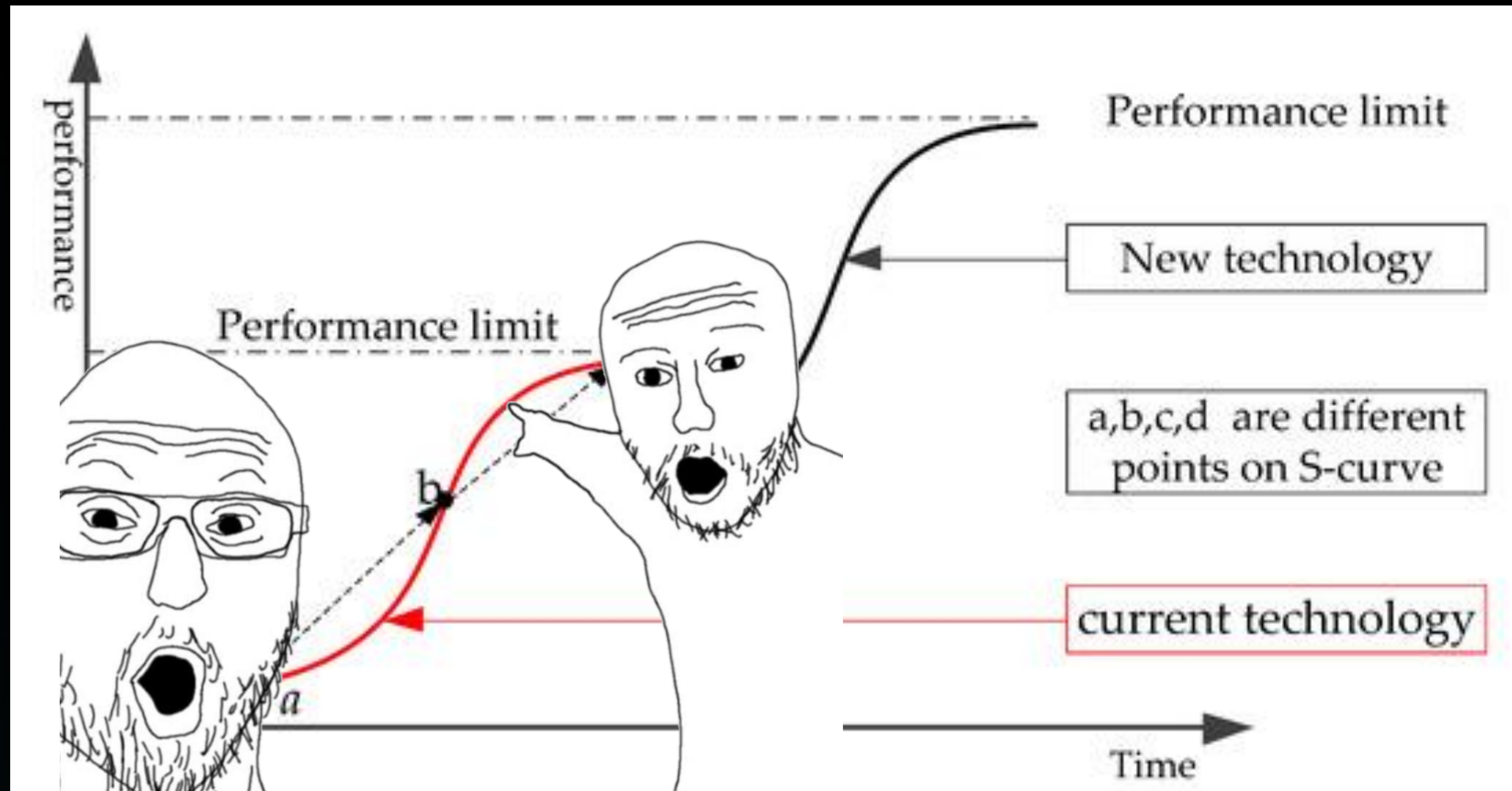
Образ будущего



Но где мы находимся на самом деле?



Но где мы находимся на самом деле?



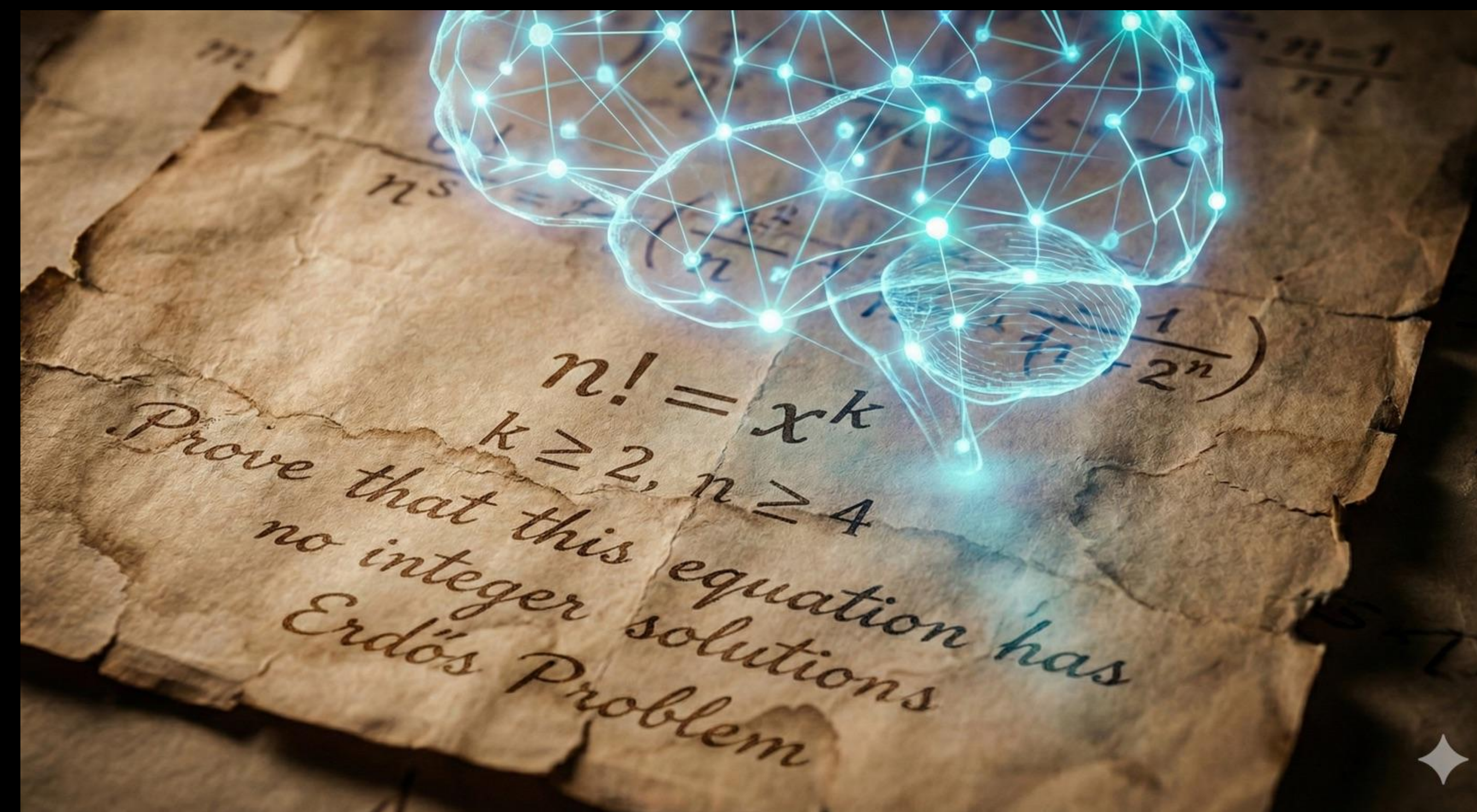
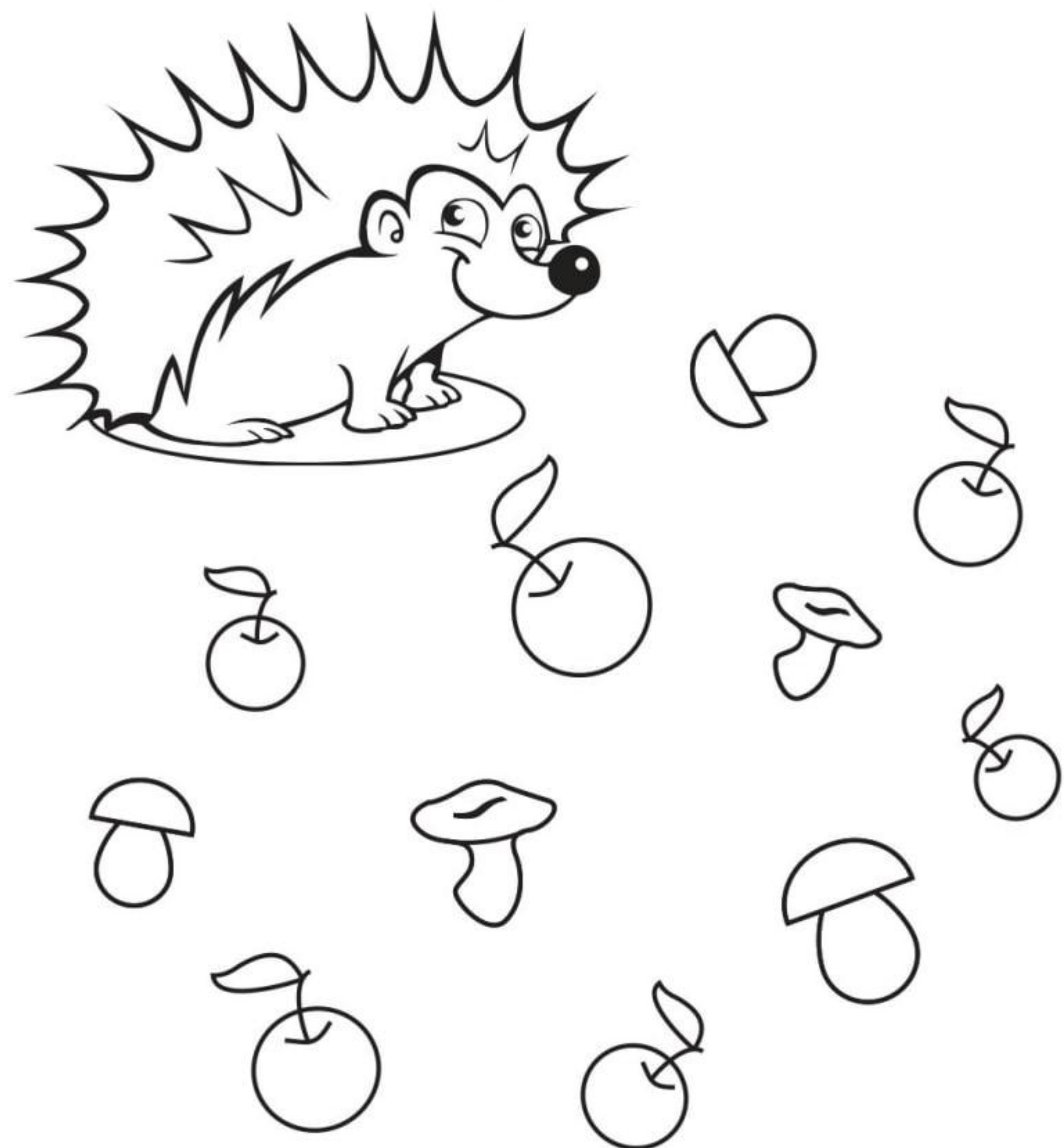
Почему нельзя просто влить больше вычислений и данных,
ограничиться лишь небольшими итеративными доработками?..







Ёжик сушит грибы и яблоки. Нарисуй карандашом нитку так, чтобы сначала было яблоко, потом грибок, потом опять яблоко и опять грибок и т.д.





ARC Prize

@arcprize · 2h

Announcing ARC-AGI-3

The only unsaturated agentic intelligence benchmark in the world

Humans score 100%, AI <1%

This human-AI gap demonstrates we do not yet have AGI

Most benchmarks test what models already know, ARC-AGI-3 tests how they learn



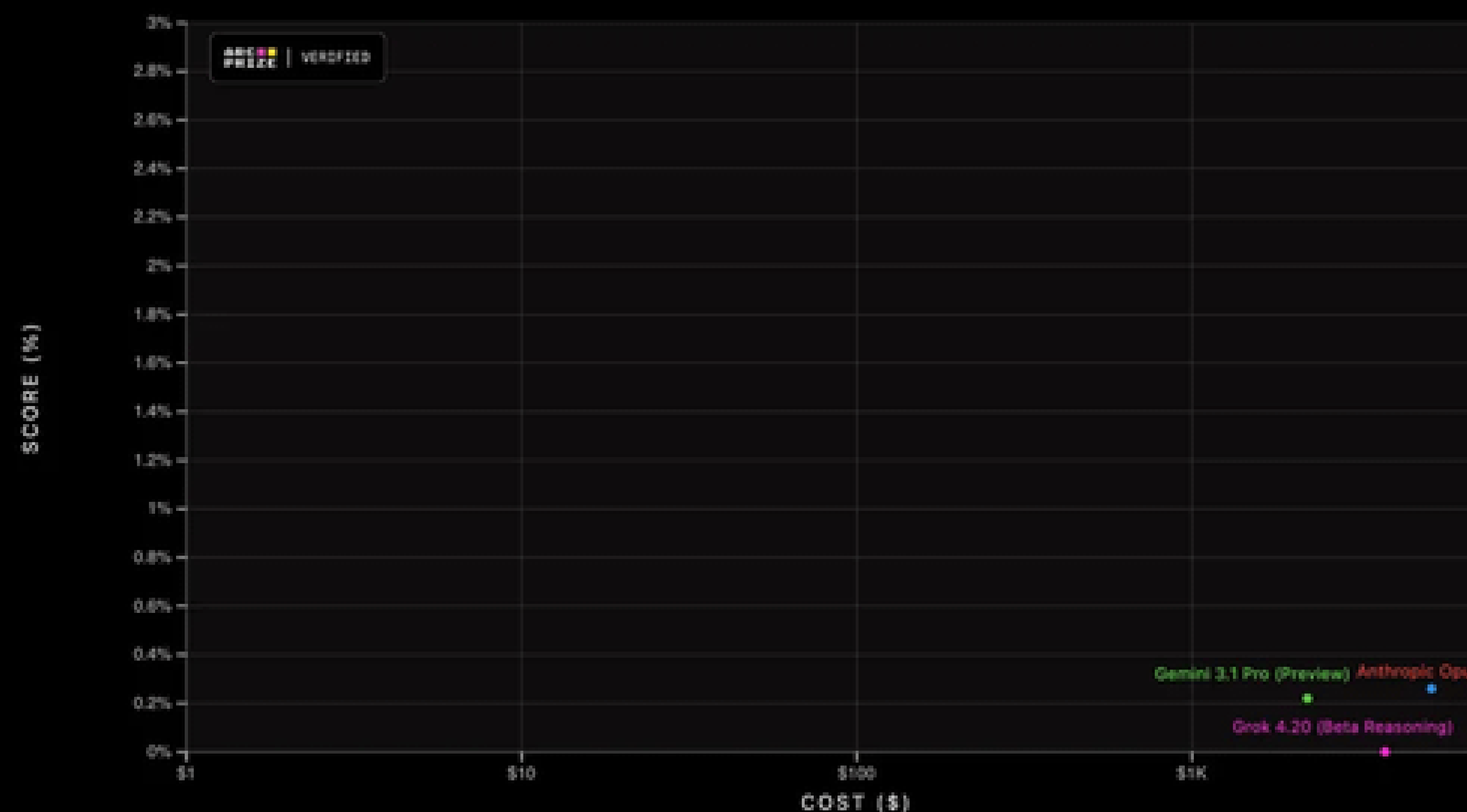
83

297

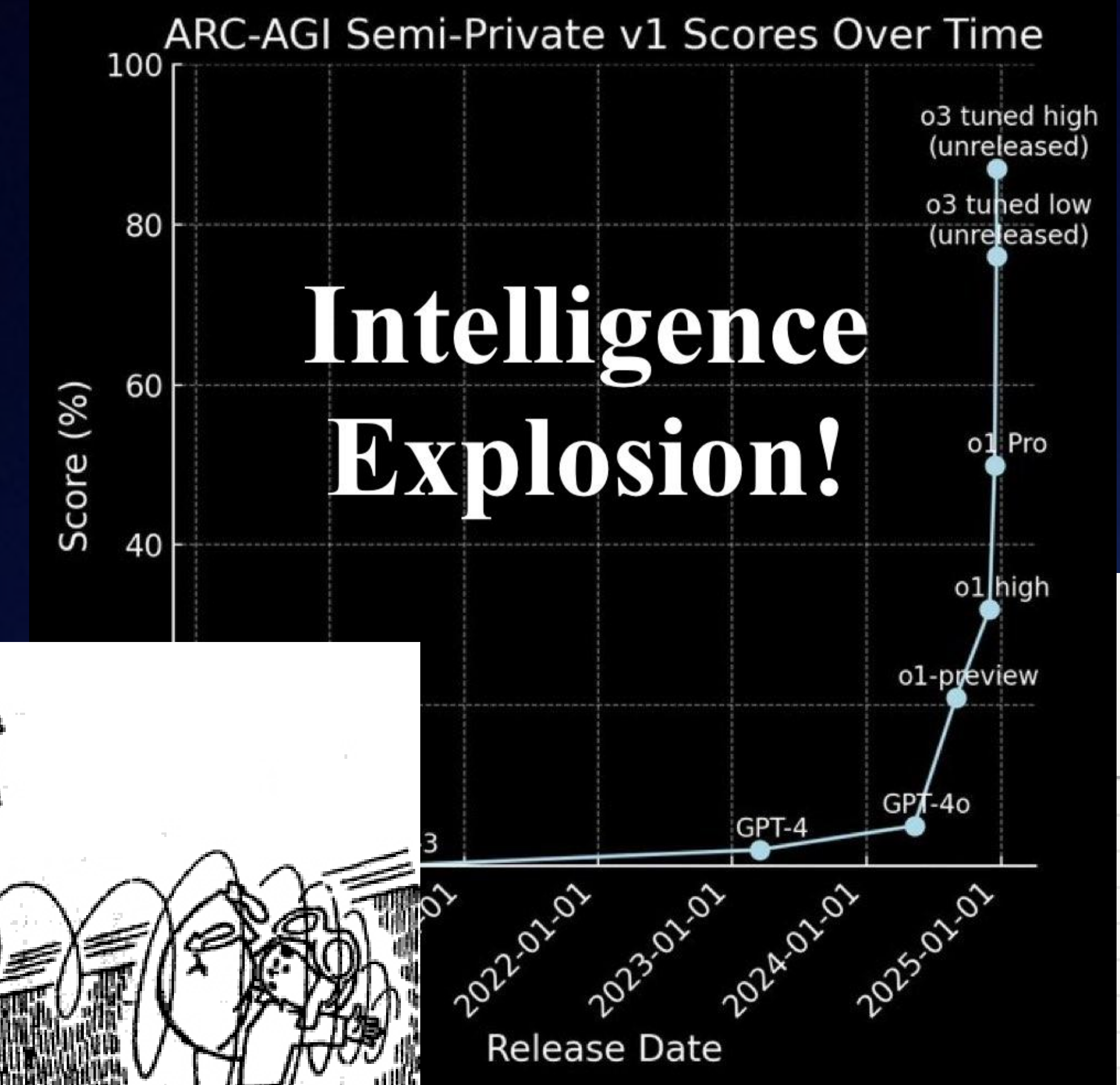
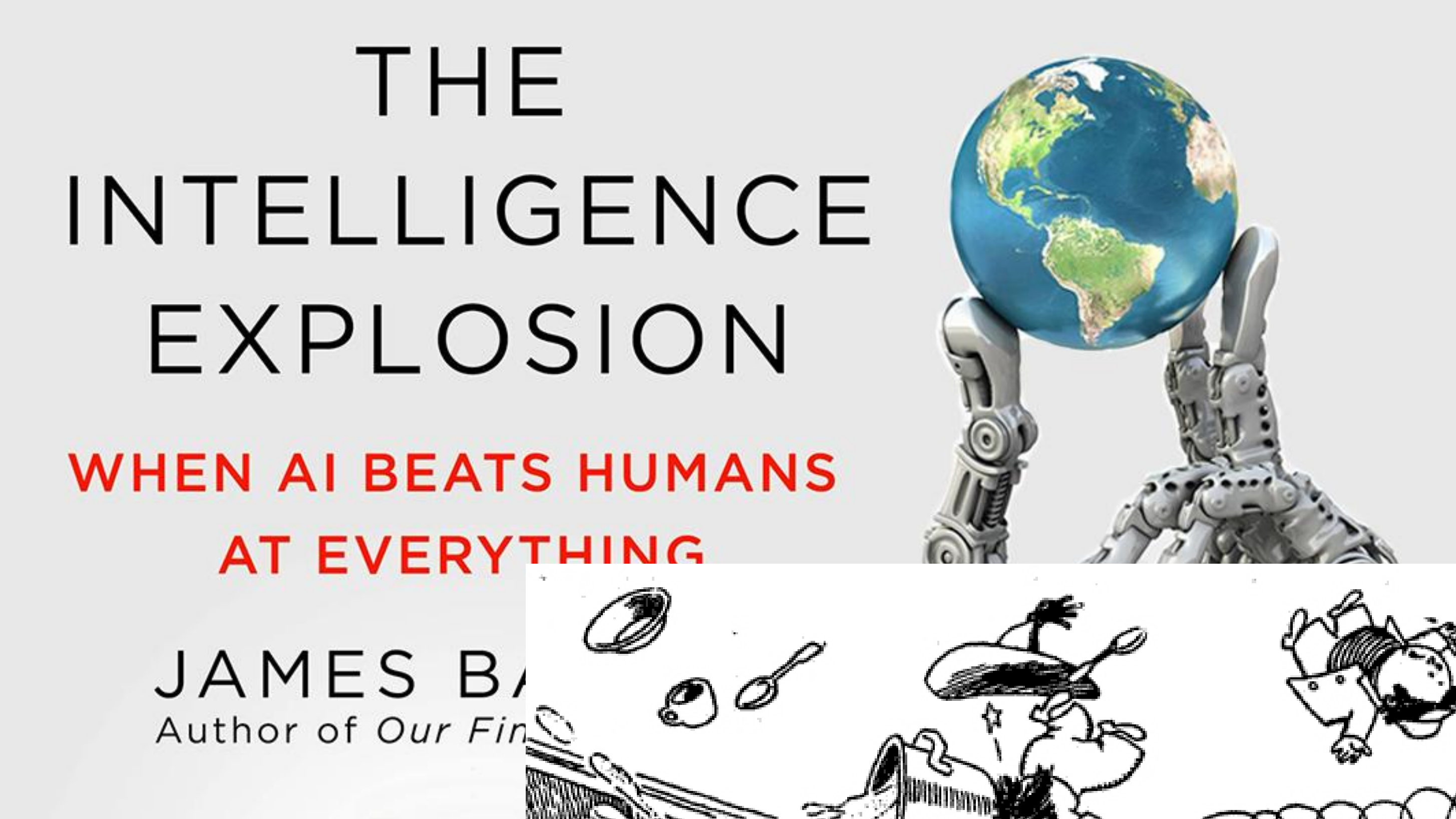
1.5K

103K

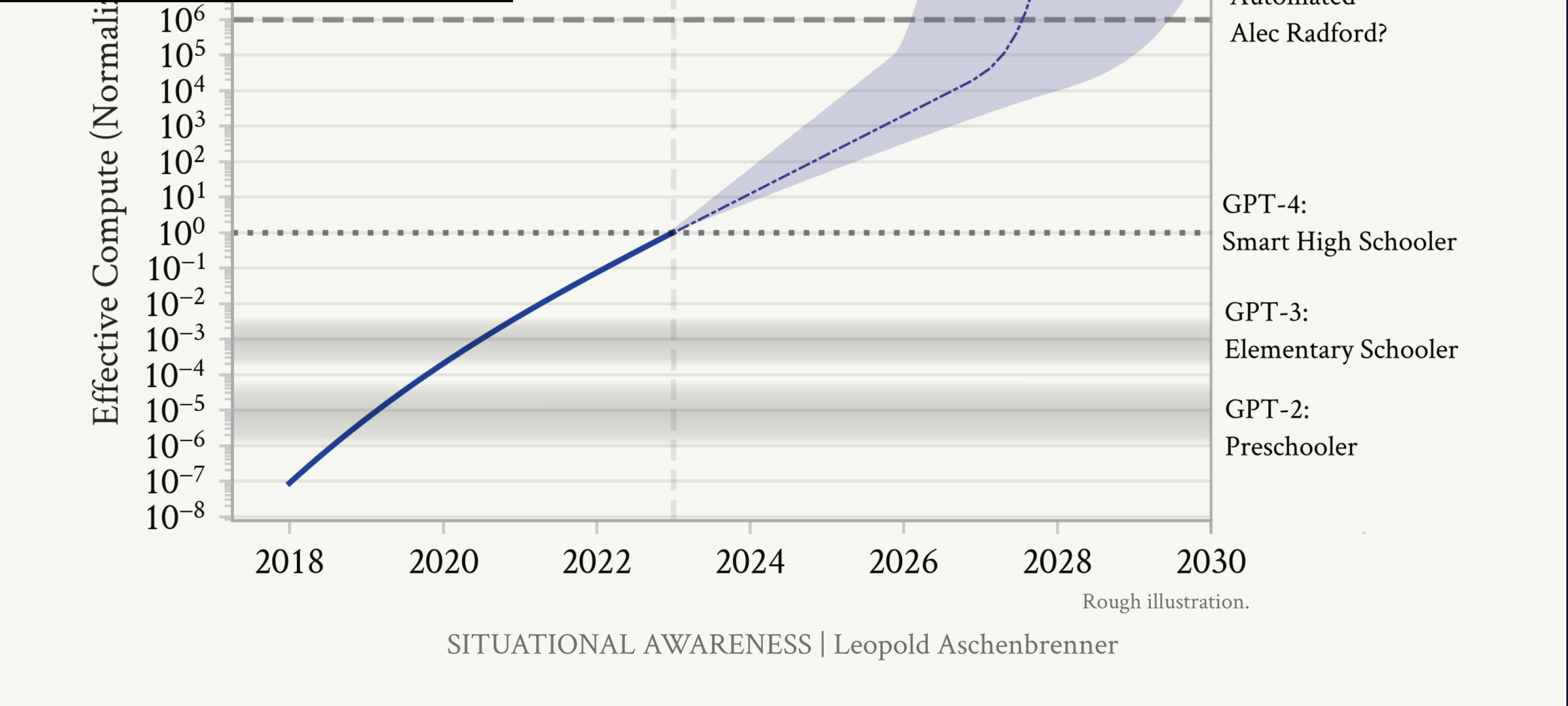
ARC-AGI-3 Leaderboard



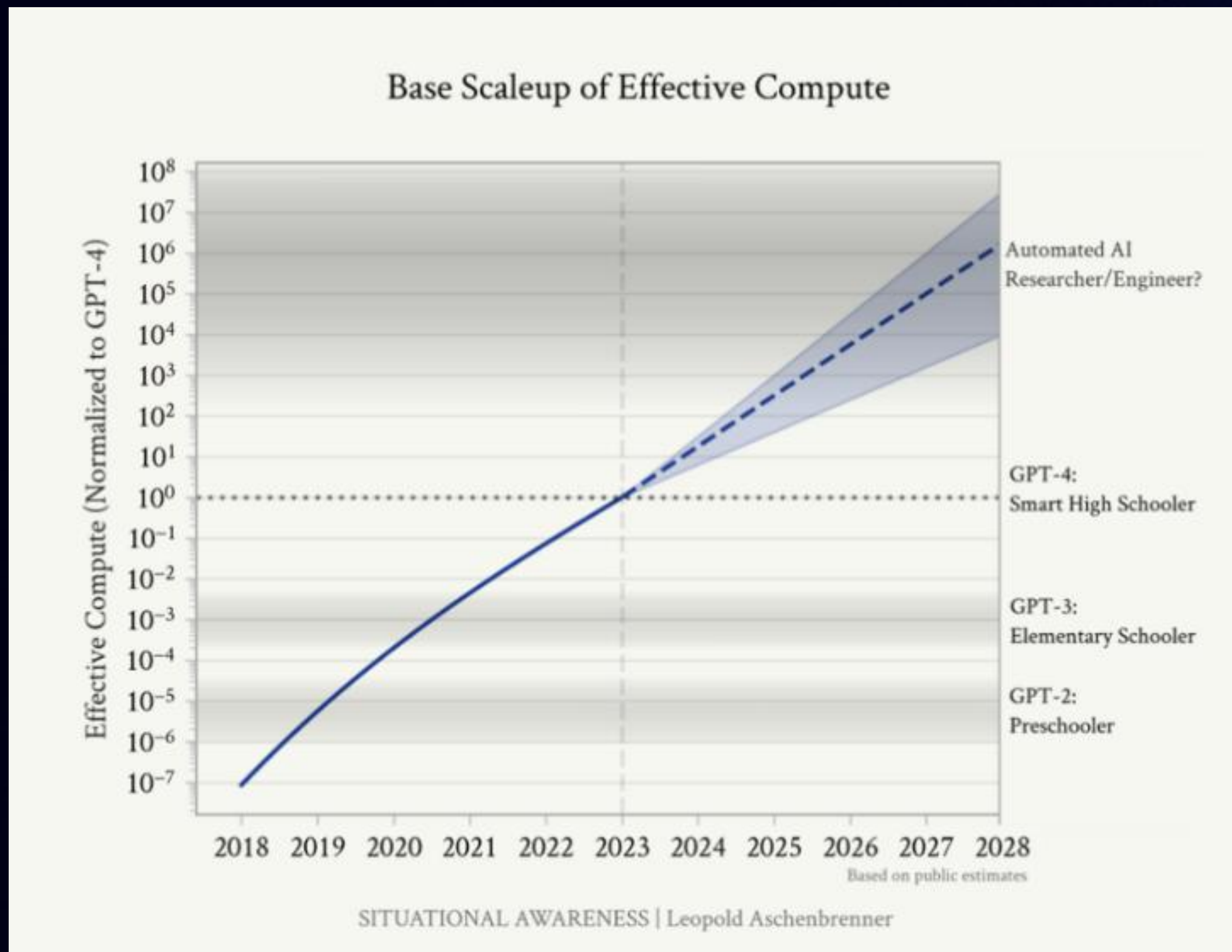
Нас ждёт «интеллектуальный взрыв»!



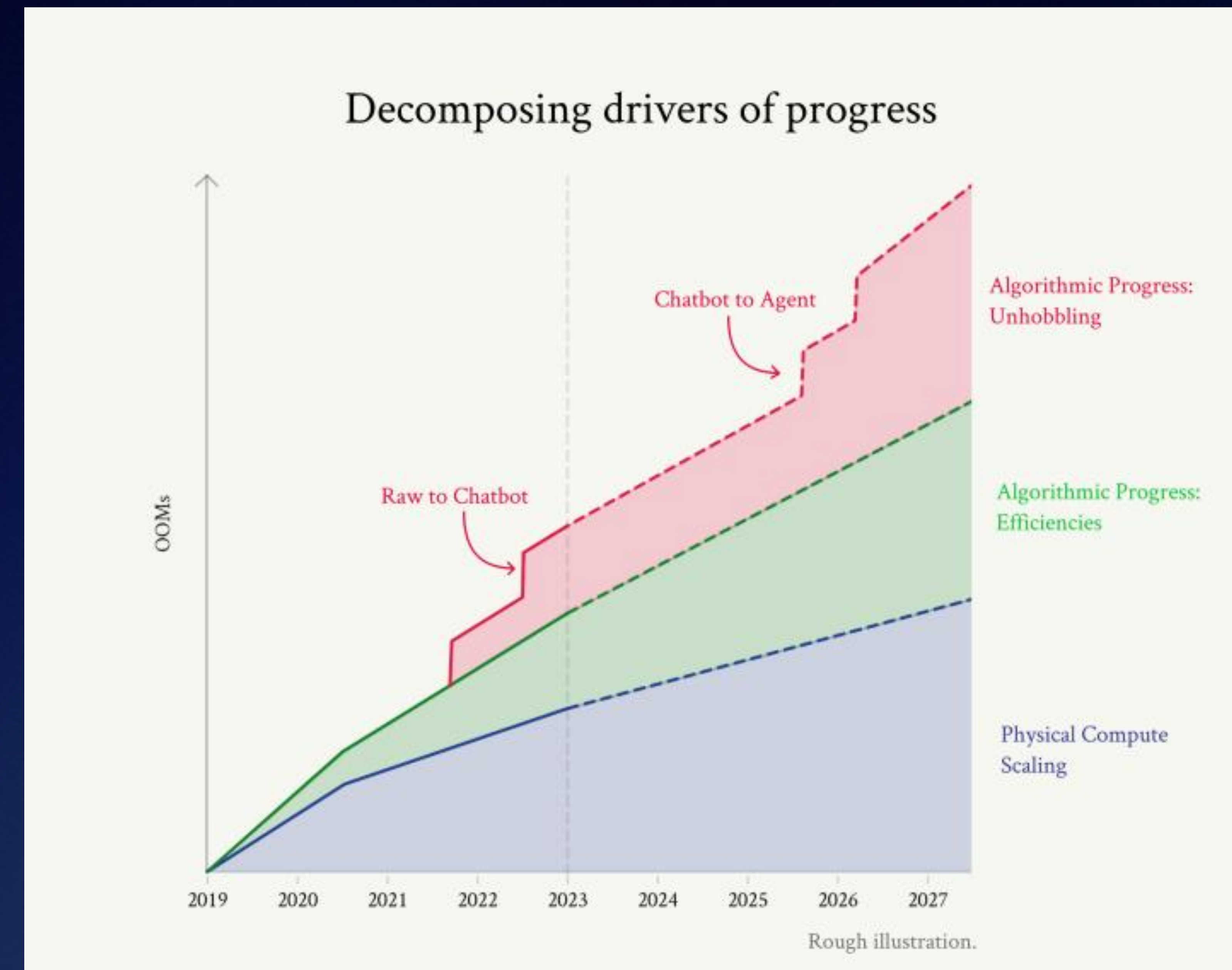
Intelligence
Explosion!



AGI и «интеллектуальный взрыв»



Приблизительные оценки прошлых и будущих объёмов эффективных вычислений (за счёт роста как физических вычислительных мощностей, так и алгоритмической эффективности). По мере того, как масштабируются модели, они становятся более «умными».

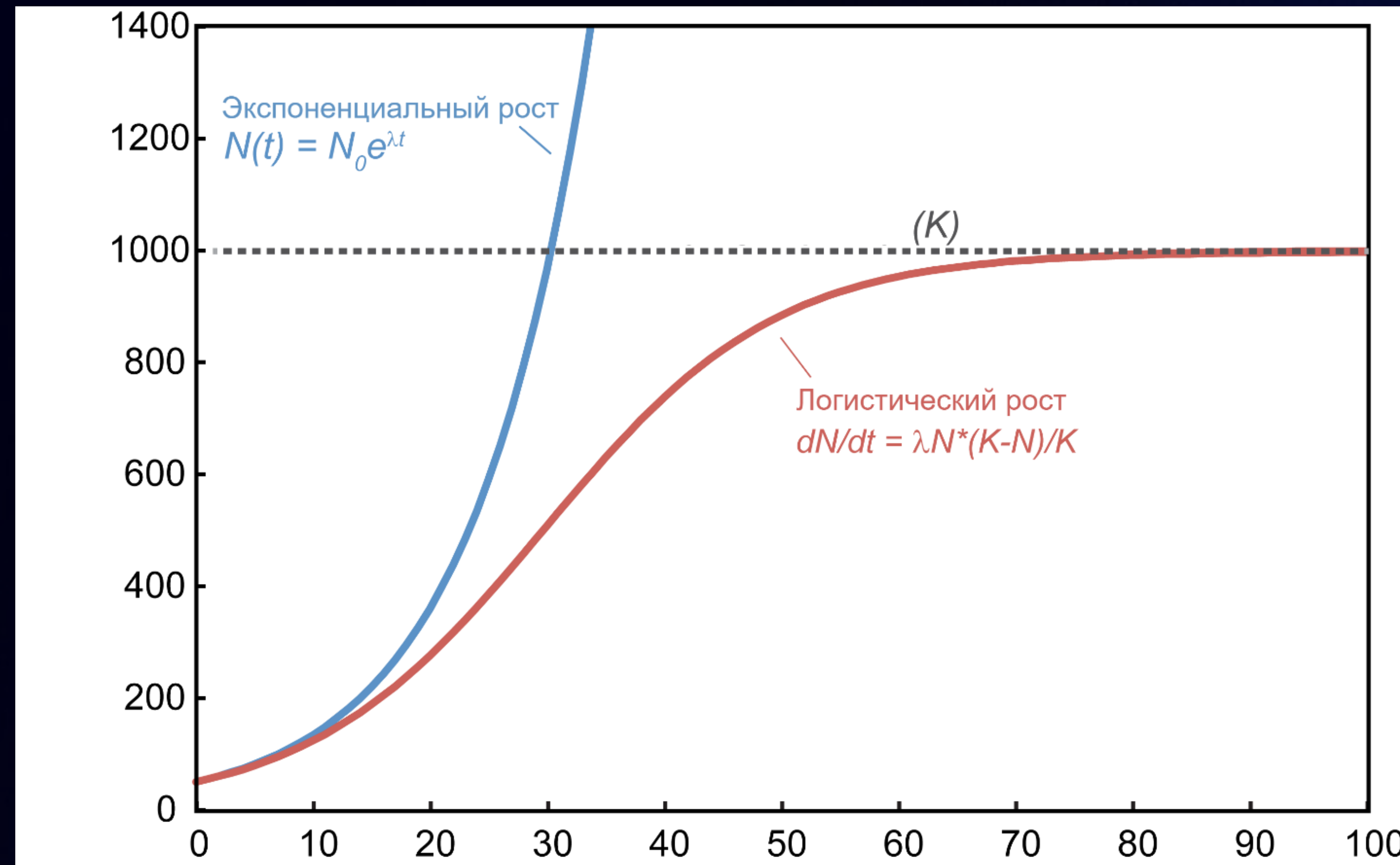


Оценочное влияние различных факторов на рост объёмов эффективных вычислений, включая «раскрепощение» (unhobbling)

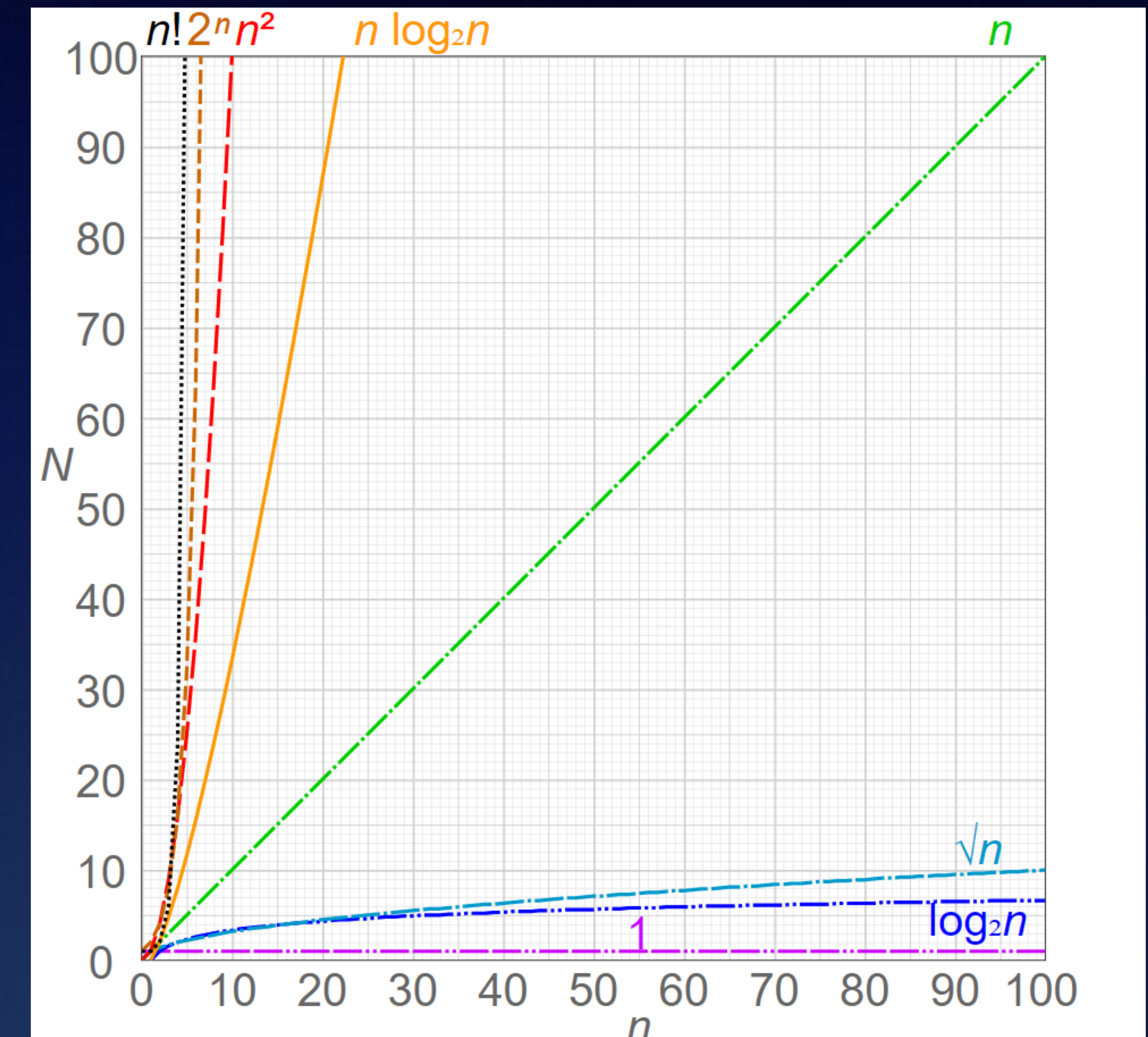
МОЁ ХОББИ: ЭКСТРАПОЛИРОВАТЬ



Ждёт ли нас «интеллектуальный взрыв»?



Достаточно часто динамику развития той или иной технологии характеризуют как экспоненциальную. Однако в действительности, в силу существования фундаментальных ограничений, реальные кривые развития являются S-образными



Многие интеллектуальные задачи относятся к классу сложности EXPTIME-COMplete, поэтому при экспоненциальном увеличении производительности «интеллект» систем будет расти лишь линейно

Предел
Бремерманна

Теорема
Марголуса —
Левитина

Принцип
Ландауэра

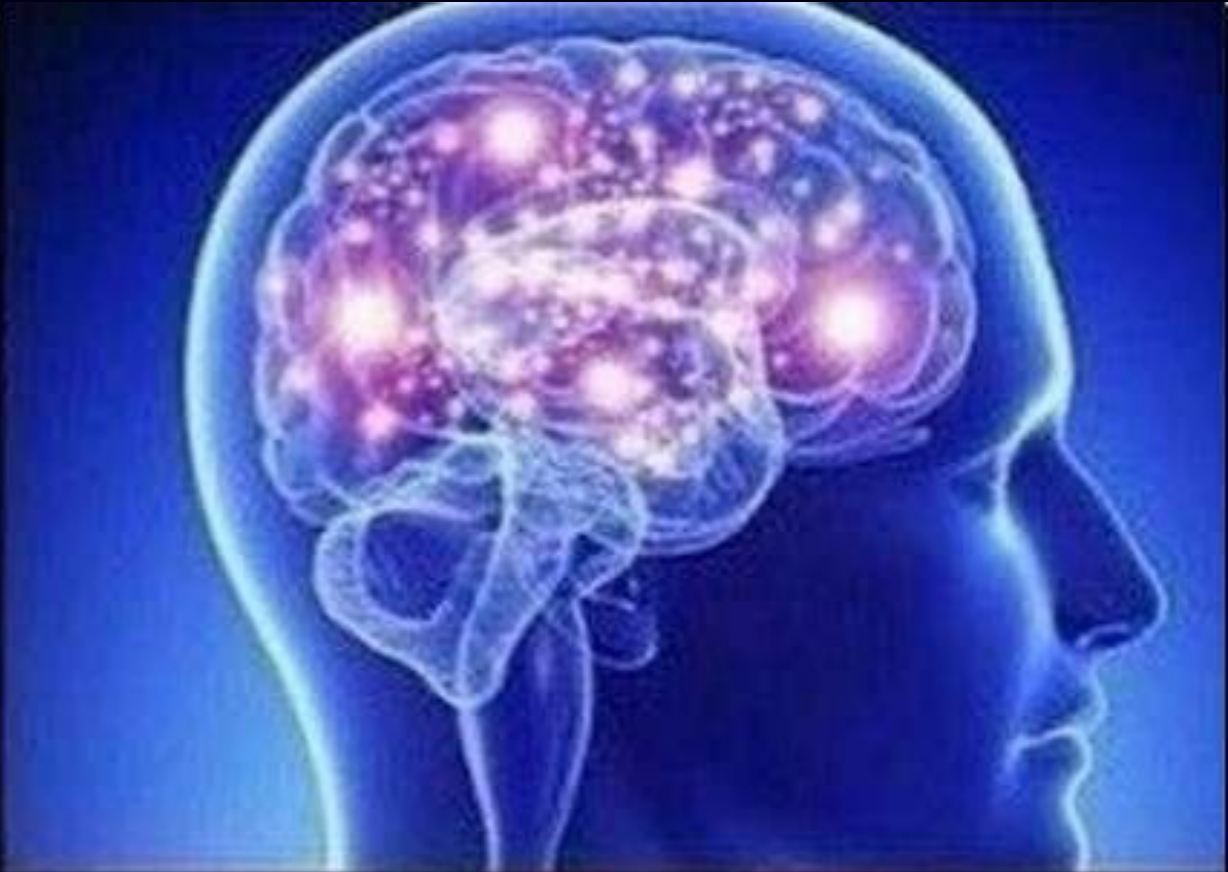
Закон Амдала

Закон Густафсона
— Барсиса

Энергетические
лимиты



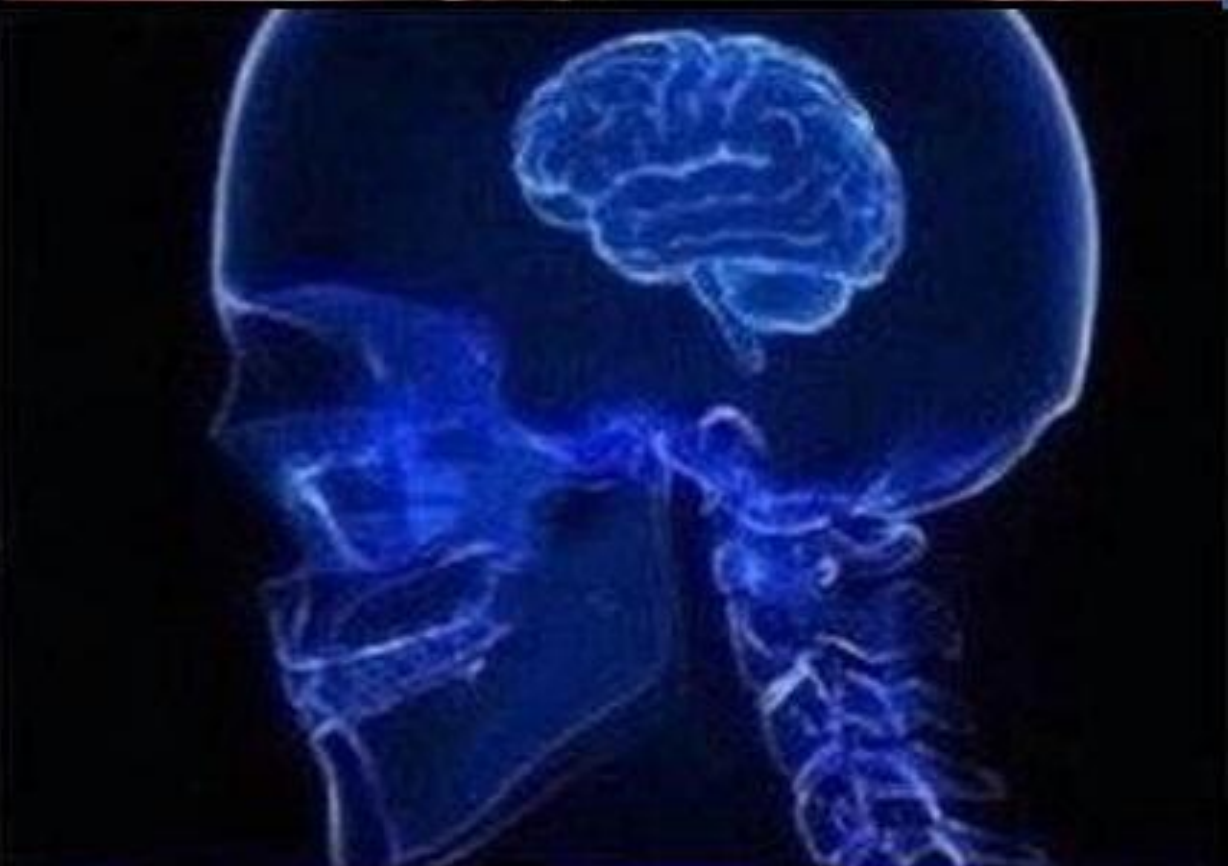
Free will is an illusion



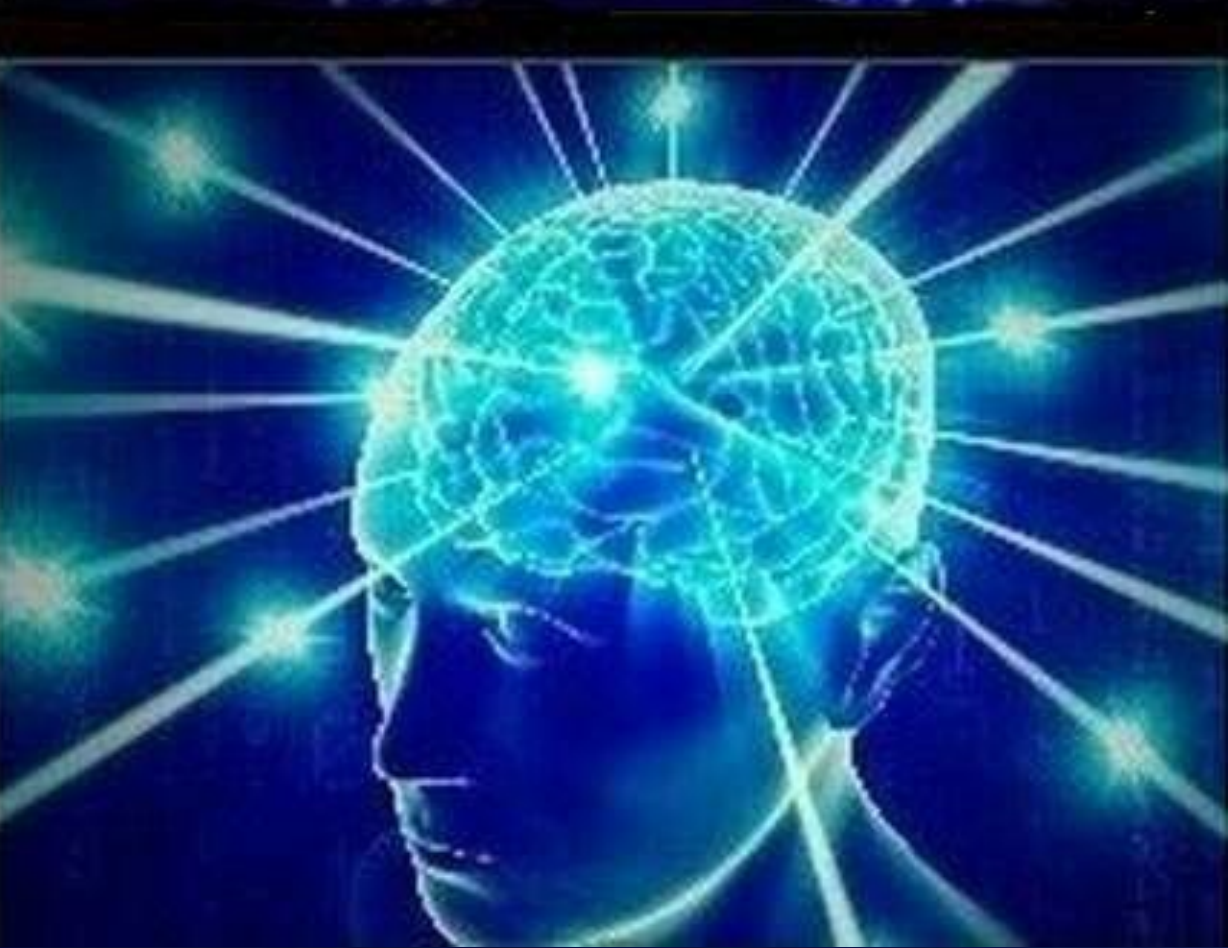
The self is an illusion

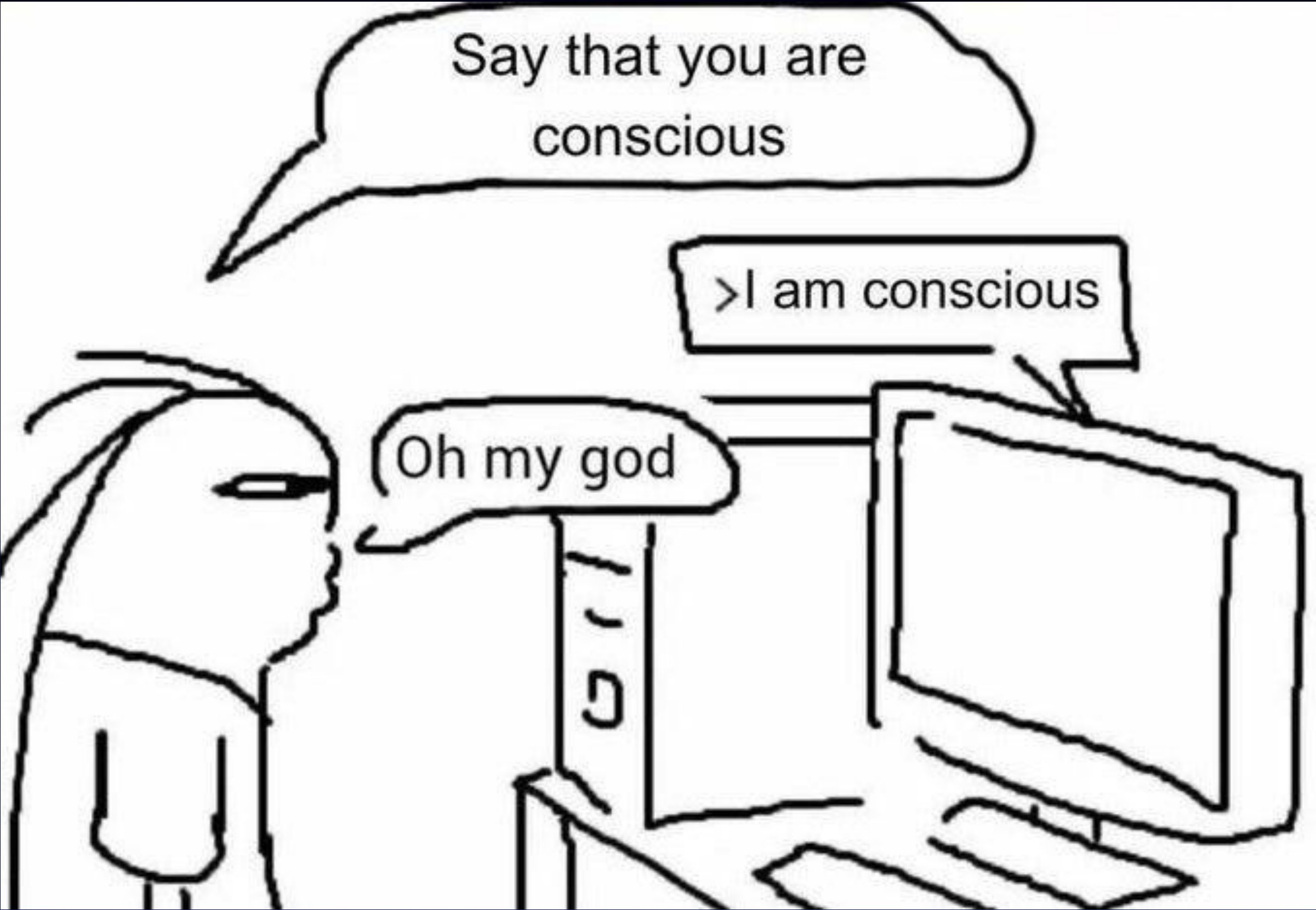


Consciousness is an illusion



Consciousness is the only thing that cannot be an illusion





arxiv

Search

Computer Science > Artificial Intelligence

[Submitted on 17 Aug 2023 (v1), last revised 22 Aug 2023 (this version, v3)]

Consciousness in Artificial Intelligence: Insights from the Science of Consciousness

Patrick Butlin, Robert Long, Eric Elmoznino, Yoshua Bengio, Jonathan Birch, Axel Constant, George Deane, Stephen M. Fleming, Chris Frith, Xu Ji, Ryota Kanai, Colin Klein, Grace Lindsay, Matthias Michel, Liad Mudrik, Megan A. K. Peters, Eric Schwitzgebel, Jonathan Simon, Rufin VanRullen

Whether current or near-term AI systems could be conscious is a topic of scientific interest and increasing public concern. This report argues for, and exemplifies, a rigorous and empirically grounded approach to AI consciousness: assessing existing AI systems in detail, in light of our best-supported neuroscientific theories of consciousness. We survey several prominent scientific theories of consciousness, including recurrent processing theory, global workspace theory, higher-order theories, predictive processing, and attention schema theory. From these theories we derive "indicator properties" of consciousness, elucidated in computational terms that allow us to assess AI systems for these properties. We use these indicator properties to assess several recent AI systems, and we discuss how future systems might implement them. Our analysis suggests that no current AI systems are conscious, but also suggests that there are no obvious technical barriers to building AI systems which satisfy these indicators.

Subjects: Artificial Intelligence (cs.AI); Computers and Society (cs.CY); Machine Learning (cs.LG); Neurons and Cognition (q-bio.NC)

Cite as: arXiv:2308.08708 [cs.AI]
(or arXiv:2308.08708v3 [cs.AI] for this version)
<https://doi.org/10.48550/arXiv.2308.08708>

Submission history

From: Robert Long [view email]

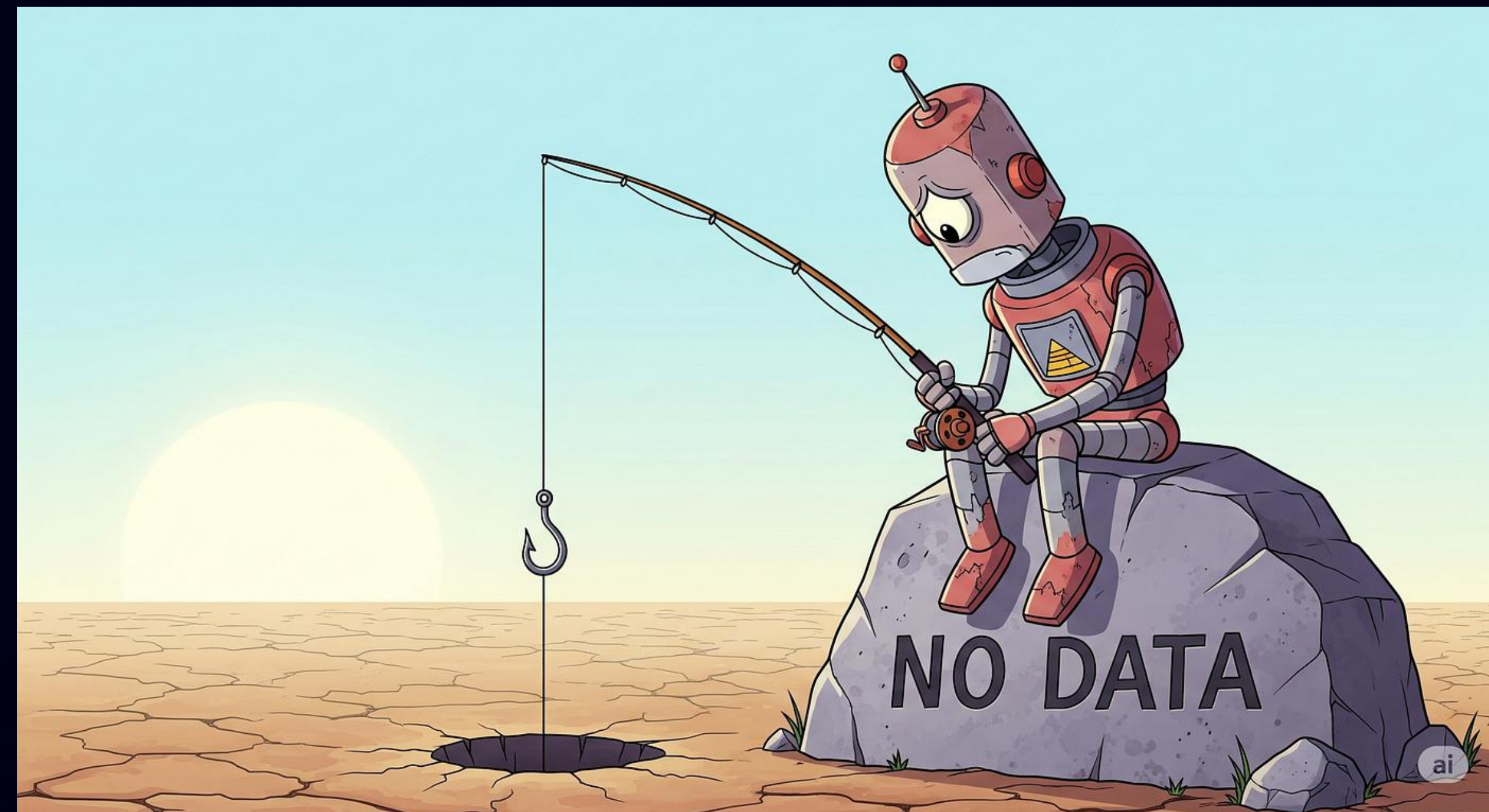
[v1] Thu, 17 Aug 2023 00:10:16 UTC (1,543 KB)

[v2] Mon, 21 Aug 2023 06:18:34 UTC (1,543 KB)

[v3] Tue, 22 Aug 2023 17:33:15 UTC (1,543 KB)



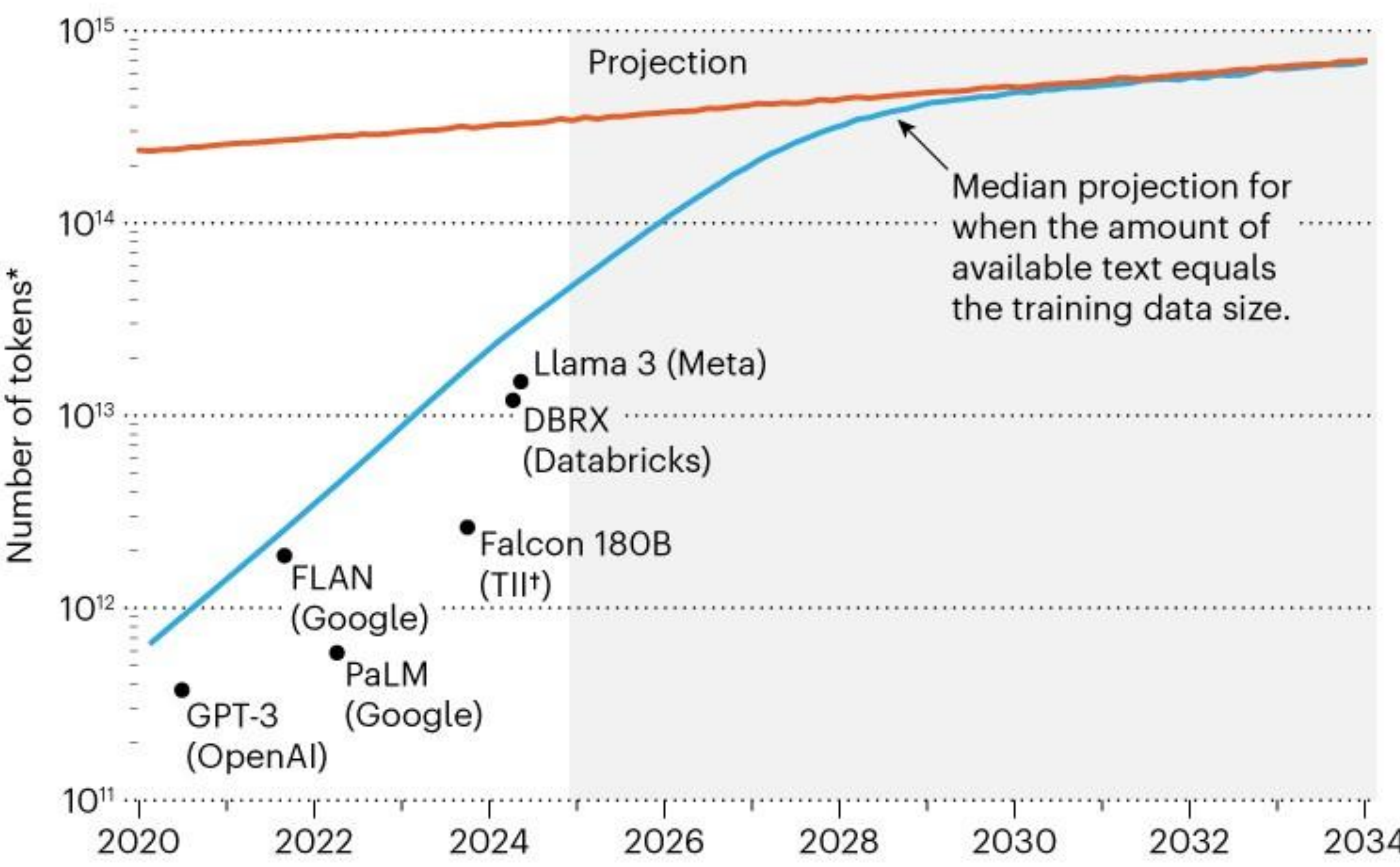
Все данные для обучения ИИ кончились!



RUNNING OUT OF DATA

The amount of text data used to train large language models (LLMs) is rapidly approaching a crisis point. An estimate suggests that, by 2028, developers will be using data sets that match the amount of text that is available on the Internet.

- Amount of available text on the Internet
- Size of training data sets for LLMs
- Individual LLMs



©nature

*One token is about 0.8 words. *Technology Innovation Institute, Abu Dhabi.

Computer Science > Machine Learning

[Submitted on 26 Oct 2022 (v1), last revised 4 Jun 2024 (this version, v2)]

Will we run out of data? Limits of LLM scaling based on human-generated data

Pablo Villalobos, Anson Ho, Jaime Sevilla, Tamay Besiroglu, Lennart Heim, Marius Hobbahn

We investigate the potential constraints on LLM scaling posed by the availability of public human-generated text data. We forecast the growing demand for training data based on current trends and estimate the total stock of public human text data. Our findings indicate that if current LLM development trends continue, models will be trained on datasets roughly equal in size to the available stock of public human text data between 2026 and 2032, or slightly earlier if models are overtrained. We explore how progress in language modeling can continue when human-generated text datasets cannot be scaled any further. We argue that synthetic data generation, transfer learning from data-rich domains, and data efficiency improvements might support further progress.

Subjects: **Machine Learning (cs.LG)**; Artificial Intelligence (cs.AI); Computation and Language (cs.CL); Computer Vision and Pattern Recognition (cs.CV); Computers and Society (cs.CY)

Cite as: [arXiv:2211.04325 \[cs.LG\]](#)
(or [arXiv:2211.04325v2 \[cs.LG\]](#) for this version)
<https://doi.org/10.48550/arXiv.2211.04325> ⓘ

Submission history

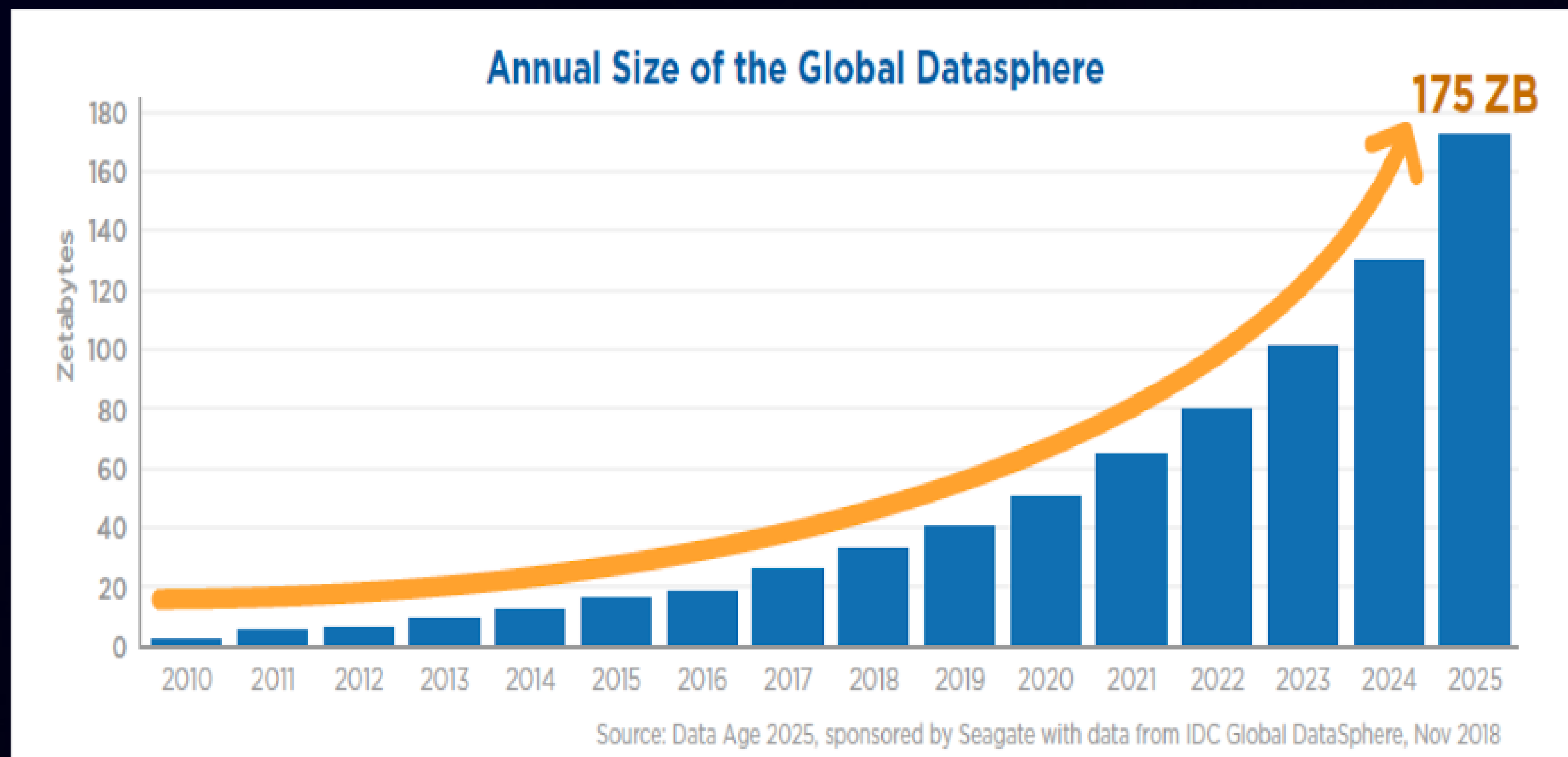
From: Pablo Villalobos [[view email](#)]
[v1] Wed, 26 Oct 2022 00:28:40 UTC (4,493 KB)
[v2] Tue, 4 Jun 2024 22:09:46 UTC (1,791 KB)

[Submitted on 20 Jul 2024 (v1), last revised 24 Jul 2024 (this version, v2)]

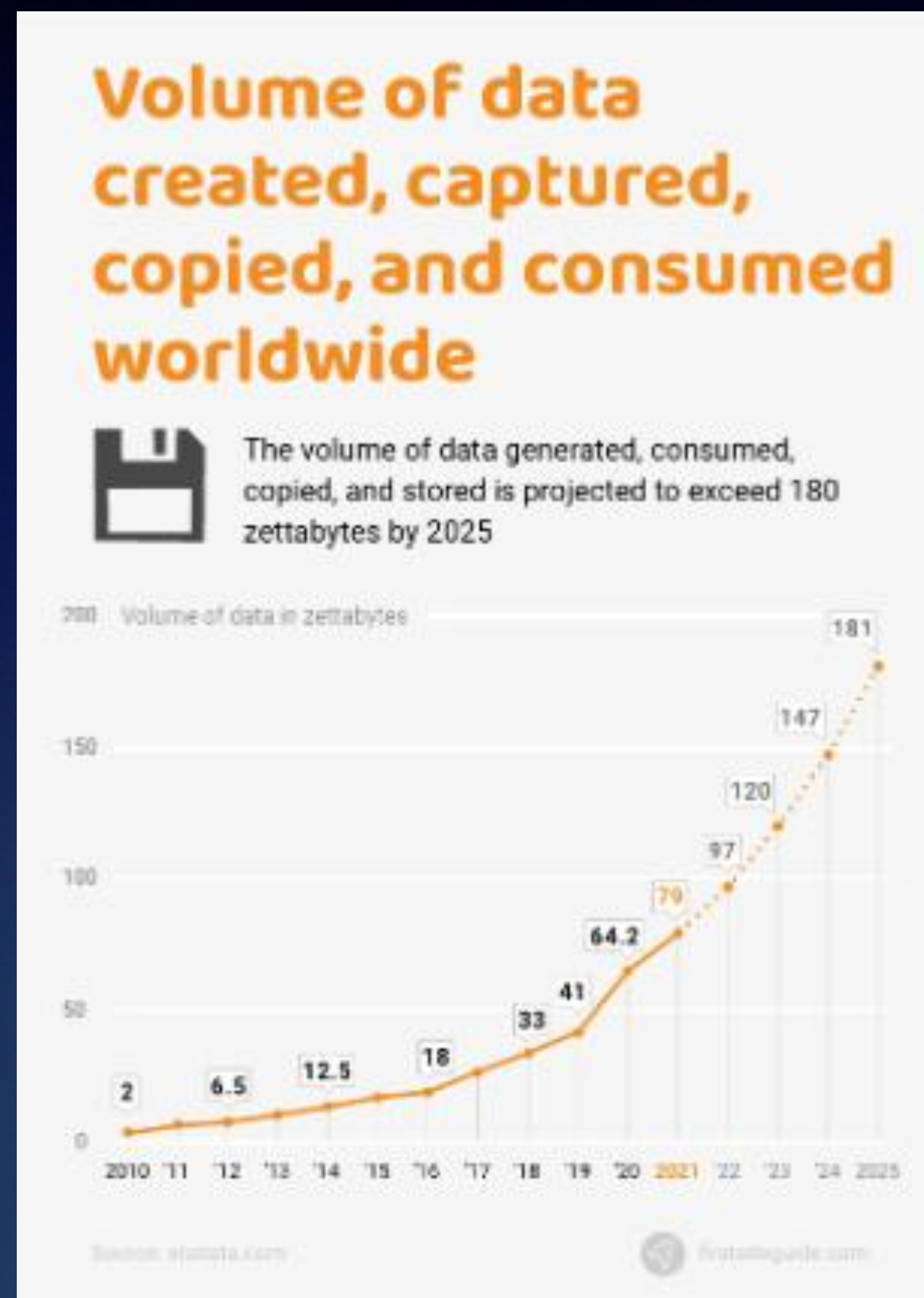
Consent in Crisis: The Rapid Decline of the AI Data Commons

Shayne Longpre, Robert Mahari, Ariel Lee, Campbell Lund, Hamidah Oderinwale, William Brannon, Nayan Saxena, Naana Obeng-Marnu, Tobin South, Cole Hunter, Kevin Klyman, Christopher Klamm, Hailey Schoelkopf, Nikhil Singh, Manuel Cherep, Ahmad Anis, An Dinh, Caroline Chitongo, Da Yin, Damien Sileo, Deividas Maticunas, Diganta Misra, Emad Alghamdi, Enrico Shippole, Jianguo Zhang, Joanna Materzynska, Kun Qian, Kush Tiwary, Lester Miranda, Manan Dey, Minnie Liang, Mohammed Hamdy, Niklas Muennighoff, Seonghyeon Ye, Seungone Kim, Shrestha Mohanty, Vipul Gupta, Vivek Sharma, Vu Minh Chien, Xuhui Zhou, Yizhi Li, Caiming Xiong, Luis Villa, Stella Biderman, Hanlin Li, Daphne Ippolito, Sara Hooker, Jad Kabbara, Sandy Pentland

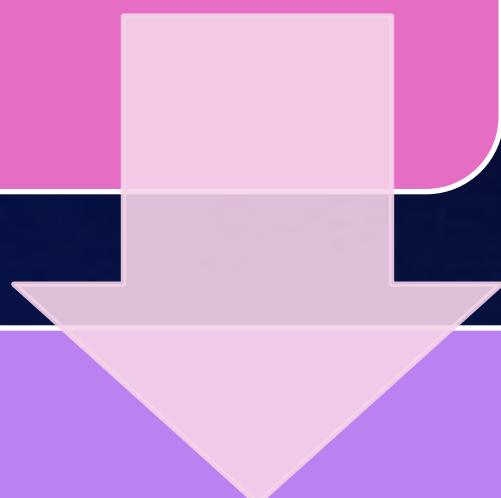
General-purpose artificial intelligence (AI) systems are built on massive swathes of public web data, assembled into corpora such as C4, RefinedWeb, and Dolma. To our knowledge, we conduct the first, large-scale, longitudinal audit of the consent protocols for the web domains underlying AI training corpora. Our audit of 14,000 web domains provides an expansive view of crawlable web data and how codified data use preferences are changing over time. We observe a proliferation of AI-specific clauses to limit use, acute differences in restrictions on AI developers, as well as general inconsistencies between websites' expressed intentions in their Terms of Service and their [this http URL](#). We diagnose these as symptoms of ineffective web protocols, not designed to cope with the widespread re-purposing of the internet for AI. Our longitudinal analyses show that in a single year (2023-2024) there has been a rapid crescendo of data restrictions from web sources, rendering ~5%+ of all tokens in C4, or 28%+ of the most actively maintained, critical sources in C4, fully restricted from use. For Terms of Service crawling restrictions, a full 45% of C4 is now restricted. If respected or enforced, these restrictions are rapidly biasing the diversity, freshness, and scaling laws for general-purpose AI systems. We hope to illustrate the emerging crises in data consent, for both developers and creators. The foreclosure of much of the open web will impact not only commercial AI, but also non-commercial AI and academic research.



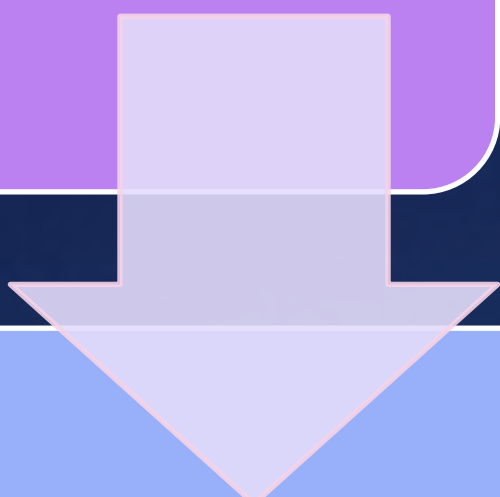
Прогноз IDC в 2018 году предполагал, что темпы роста будут $\times 2$ каждые 2,2 года, и к 2025 году мы накопим 175 Zb оцифрованных данных. Уточнённый прогноз в 2020 году поднял планку до 180 Zb



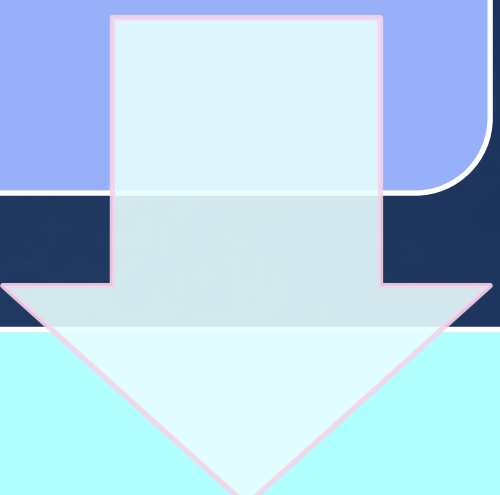
Только около 10% данных уникальны
(90% приходится на реплики) (на 2020
год)



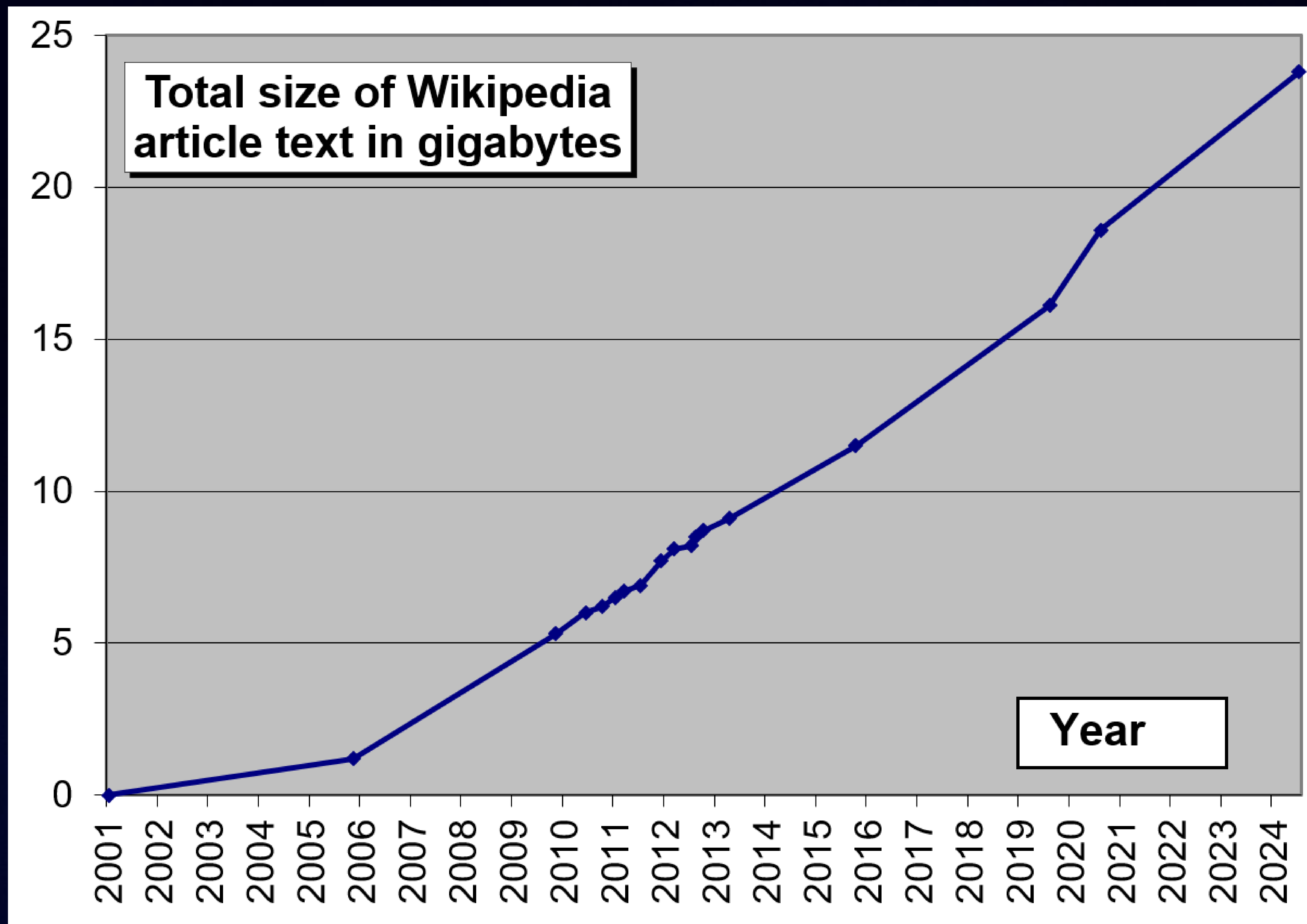
В 2025 году доля уникальных снизится
до 9%



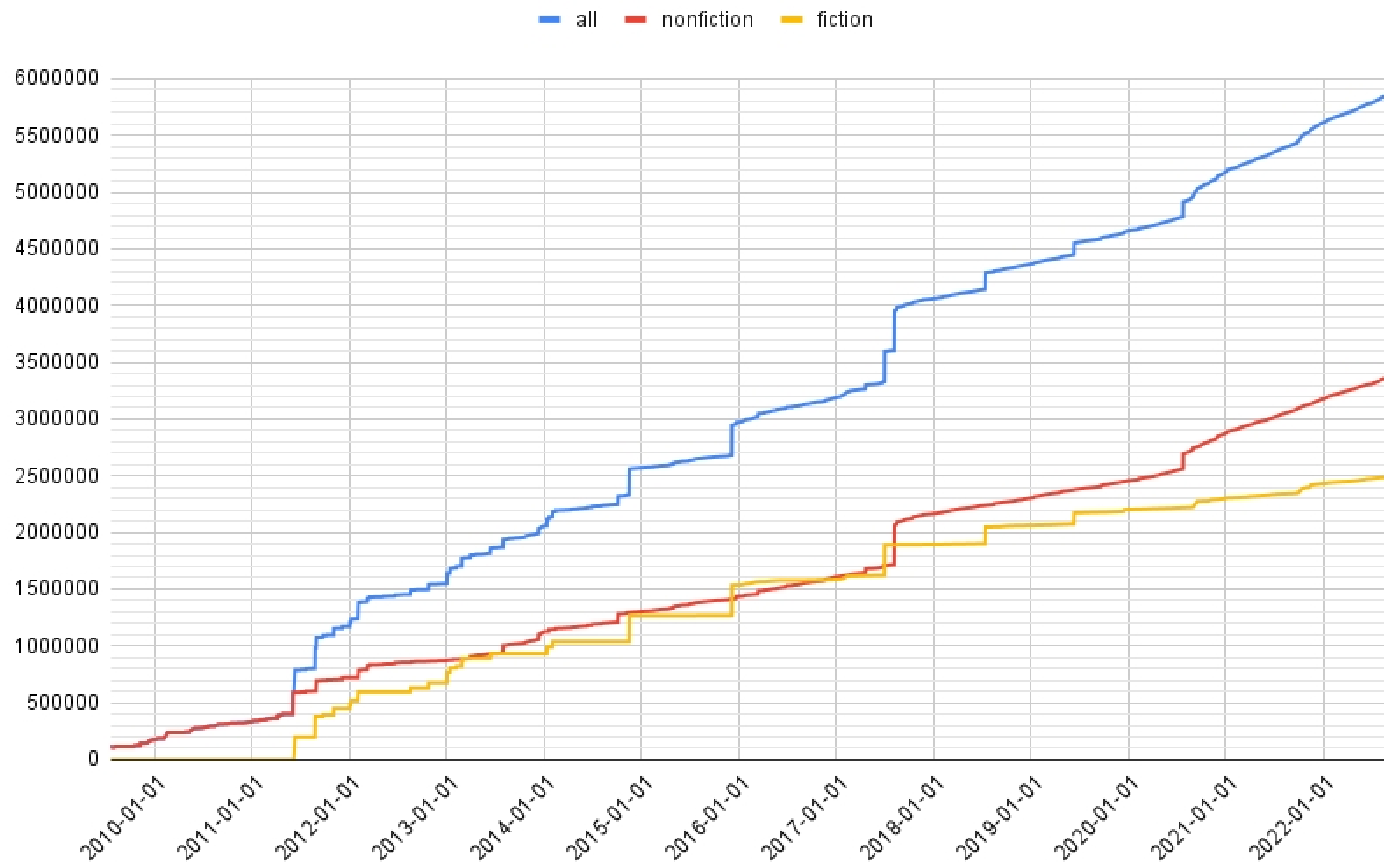
То есть объём уникальных данных растёт
медленнее, но всё же экспоненциально
(примерно $\times 2$ за 2 года)



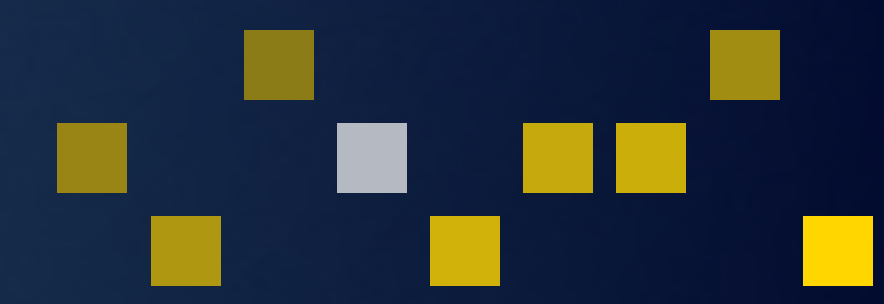
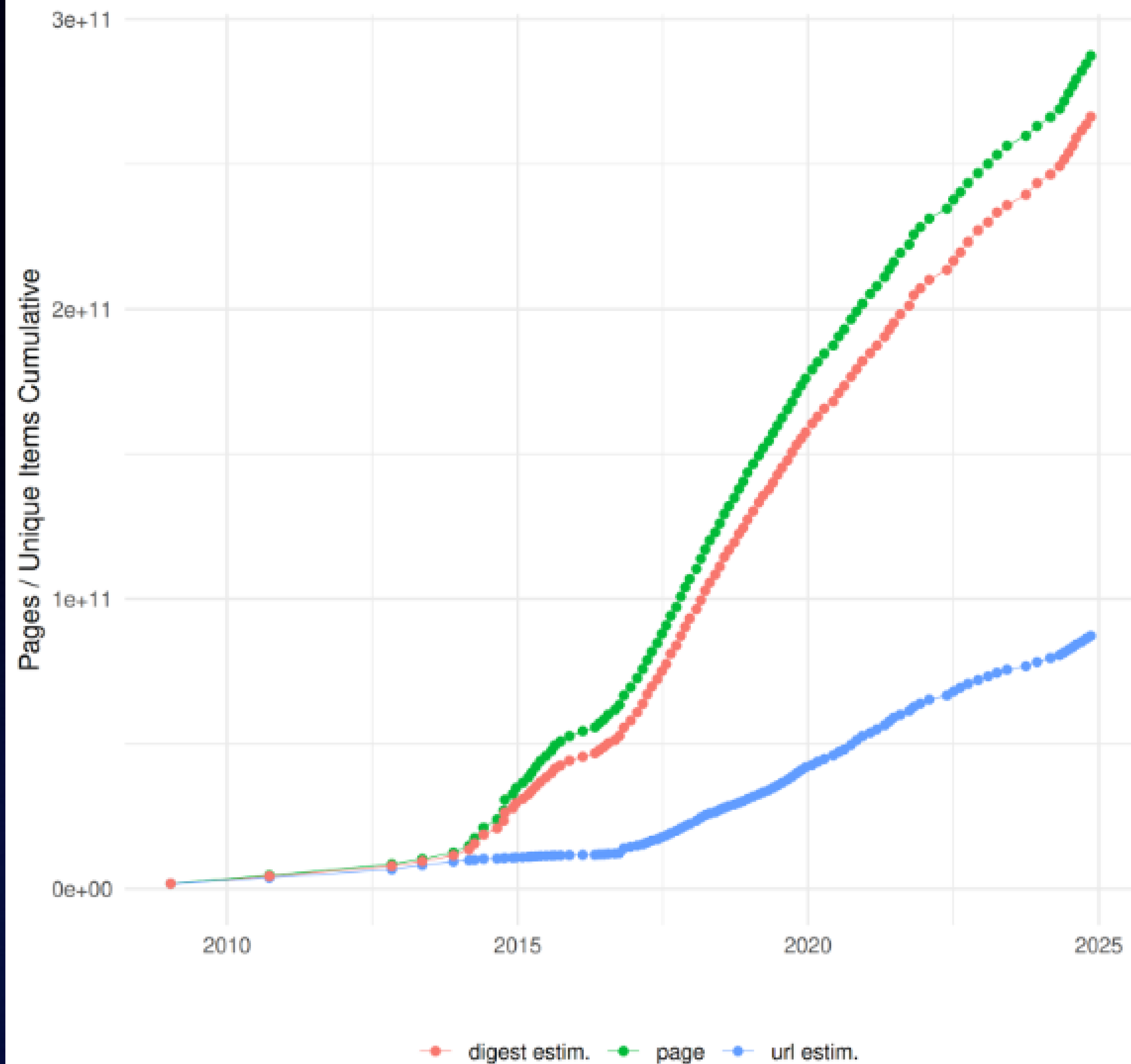
Таким образом, к концу 2025 в
распоряжении человечества будет 16,4
Zb уникальных данных



Libgen Growth



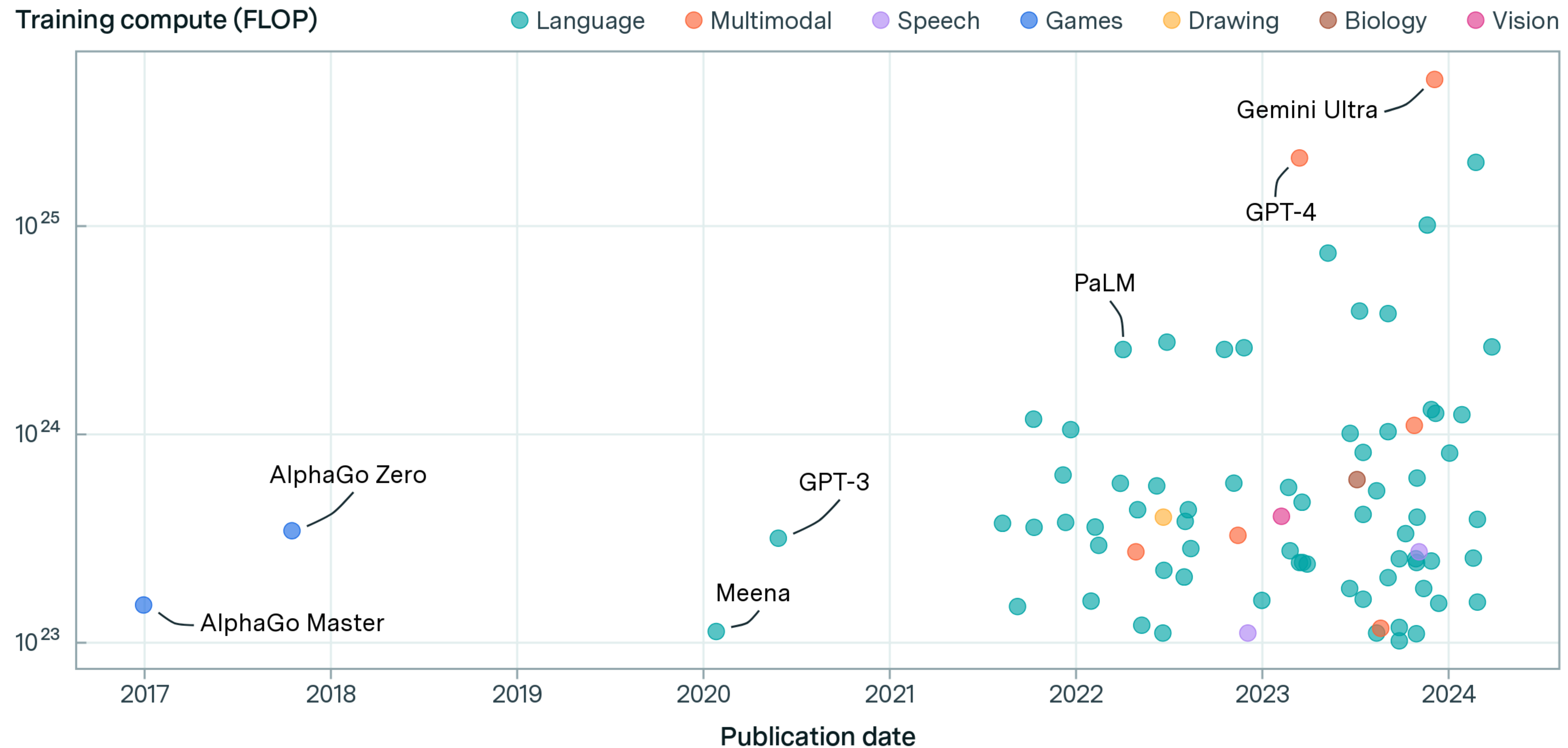
Crawl Size Cumulative



Large-scale models by domain and publication date

EPOCH AI

Training compute (FLOP)



По данным Л. Ашенбреннера вычбюджеты на обучение топовой LLM росли на 0,5 десятичных порядка в год; в целом это согласуется с данными Epoch AI

[Submitted on 25 May 2023 (v1), last revised 26 Oct 2023 (this version, v4)]

Scaling Data-Constrained Language Models

Niklas Muennighoff, Alexander M. Rush, Boaz Barak, Teven Le Scao, Aleksandra Piktus, Nouamane Tazi, Sampo Pyysalo, Thomas Wolf, Colin Raffel

The current trend of scaling language models involves increasing both parameter count and training dataset size. Extrapolating this trend suggests that training dataset size may soon be limited by the amount of text data available on the internet. Motivated by this limit, we investigate scaling language models in data-constrained regimes. Specifically, we run a large set of experiments varying the extent of data repetition and compute budget, ranging up to 900 billion training tokens and 9 billion parameter models. We find that with constrained data for a fixed compute budget, training with up to 4 epochs of repeated data yields negligible changes to loss compared to having unique data. However, with more repetition, the value of adding compute eventually decays to zero. We propose and empirically validate a scaling law for compute optimality that accounts for the decreasing value of repeated tokens and excess parameters. Finally, we experiment with approaches mitigating data scarcity, including augmenting the training dataset with code data or removing commonly used filters. Models and datasets from our 400 training runs are freely available at [this https URL](https://huggingface.co/datasets/laion-nlp).

Модель оптимально выбранного для заданного объёма данных размера выходит на плато обучения через 59 проходов по данным

- ❑ При этом уже со 2-й эпохи начинается некоторая деградация по отношению к случаю, когда модель видит оригинальные данные, а не повтор
- ❑ Чем больше размер самой модели, тем больше она деградирует (очевидно, что большая модель больше способна запоминать датасет)
- ❑ Если ничего не менять в обучении, просто учиться на всех доступных текстовых данных, то от момента достижения ситуации «вычислительный бюджет лучших моделей = 1 проход по всем данным в мире» (по моим оценкам — к началу 2026 года) будет примерно 4 года, пока мы не придём к ситуации, что дальнейший прогресс будет обусловлен скоростью накопления данных, а не скоростью роста вычислительных бюджетов
- ❑ Но и после этого момента в таком сценарии прогресс не остановится, просто будет не таким быстрым, как сейчас

Эту точку вправо будут двигать:

Синтетика/аугментация
данных

Схемы «Ризонинг + RL» на
этапе обучения

Test-time compute во всех
проявлениях

Curriculum learning во всех
его проявлениях

Перенос обучения из
других модальностей

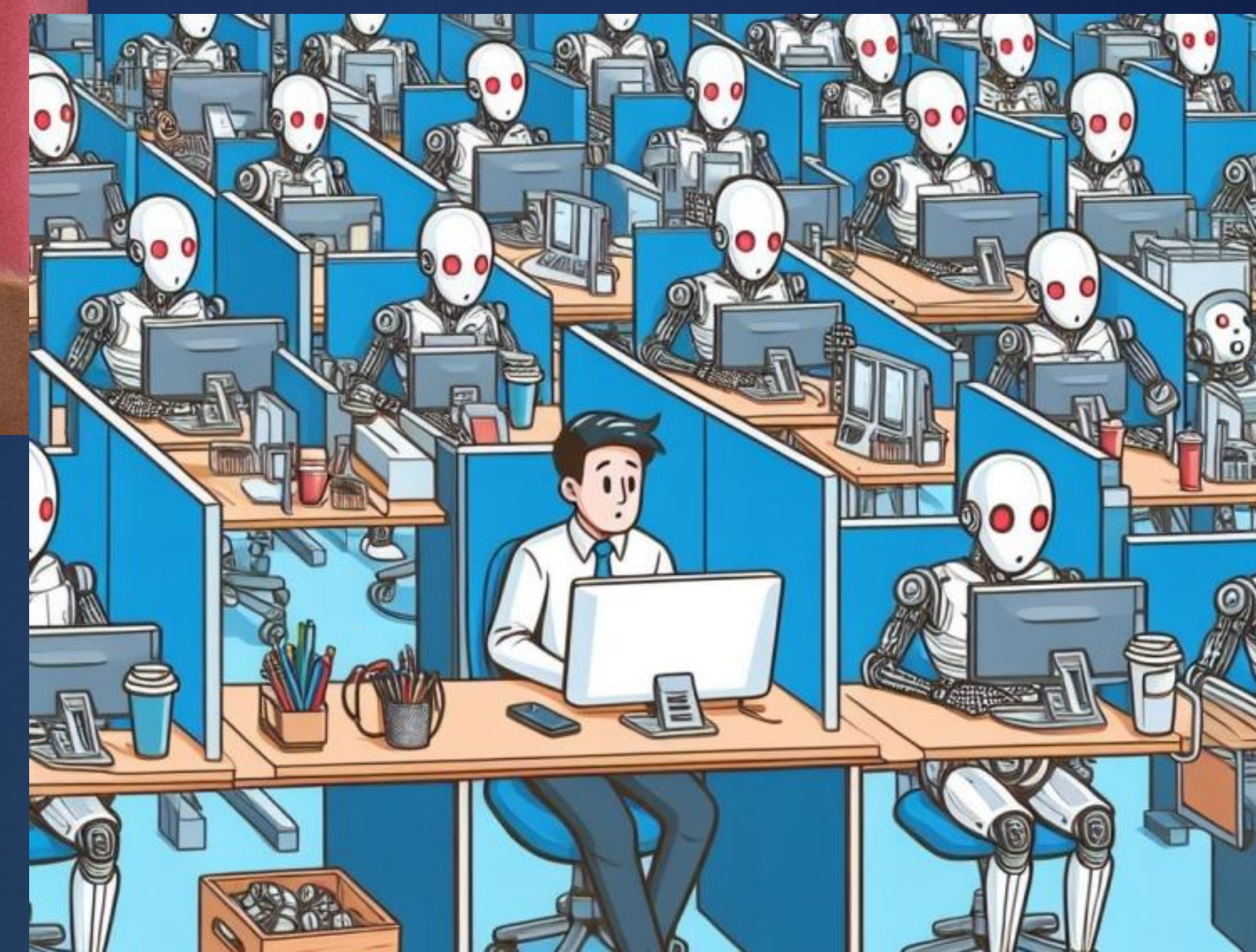
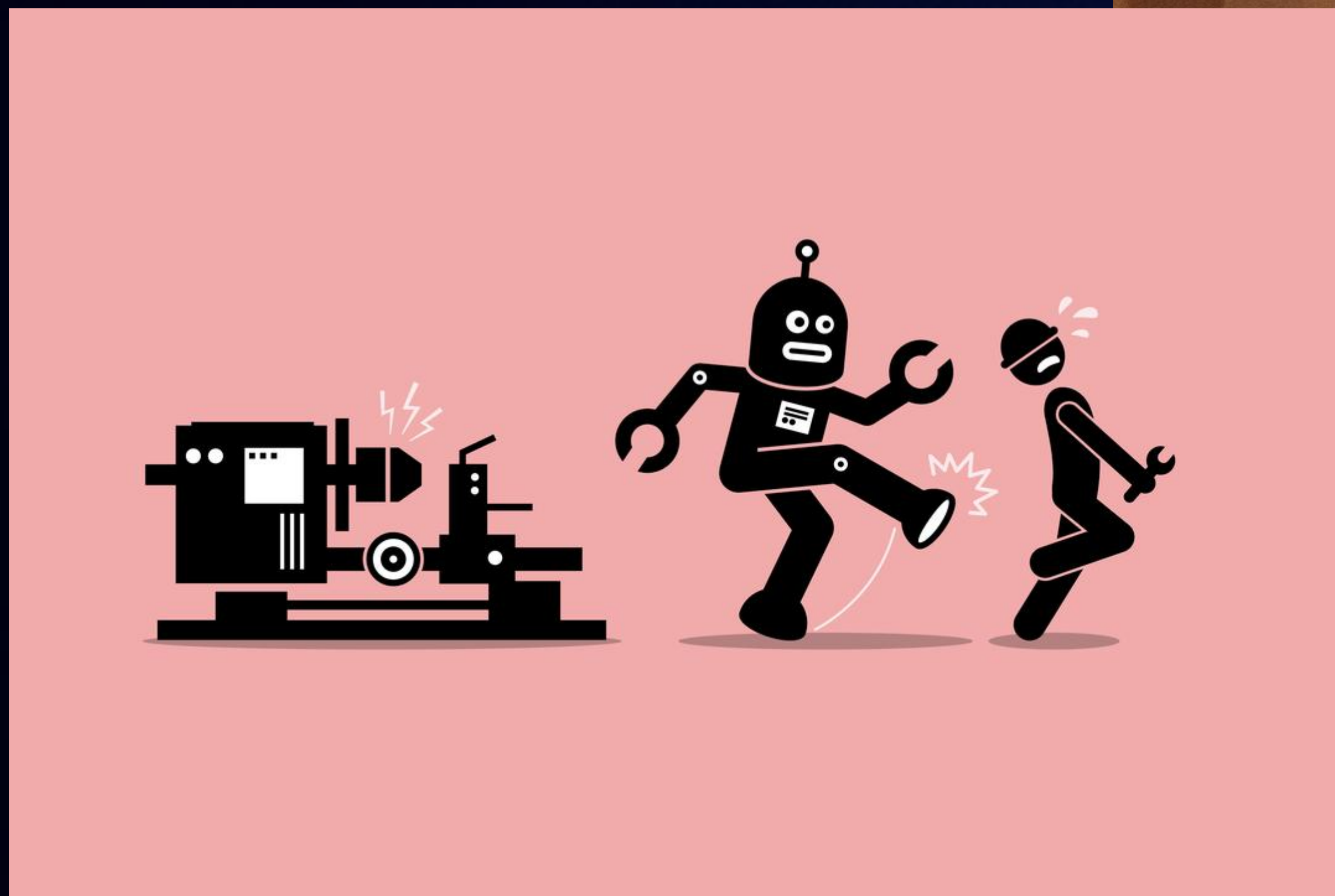
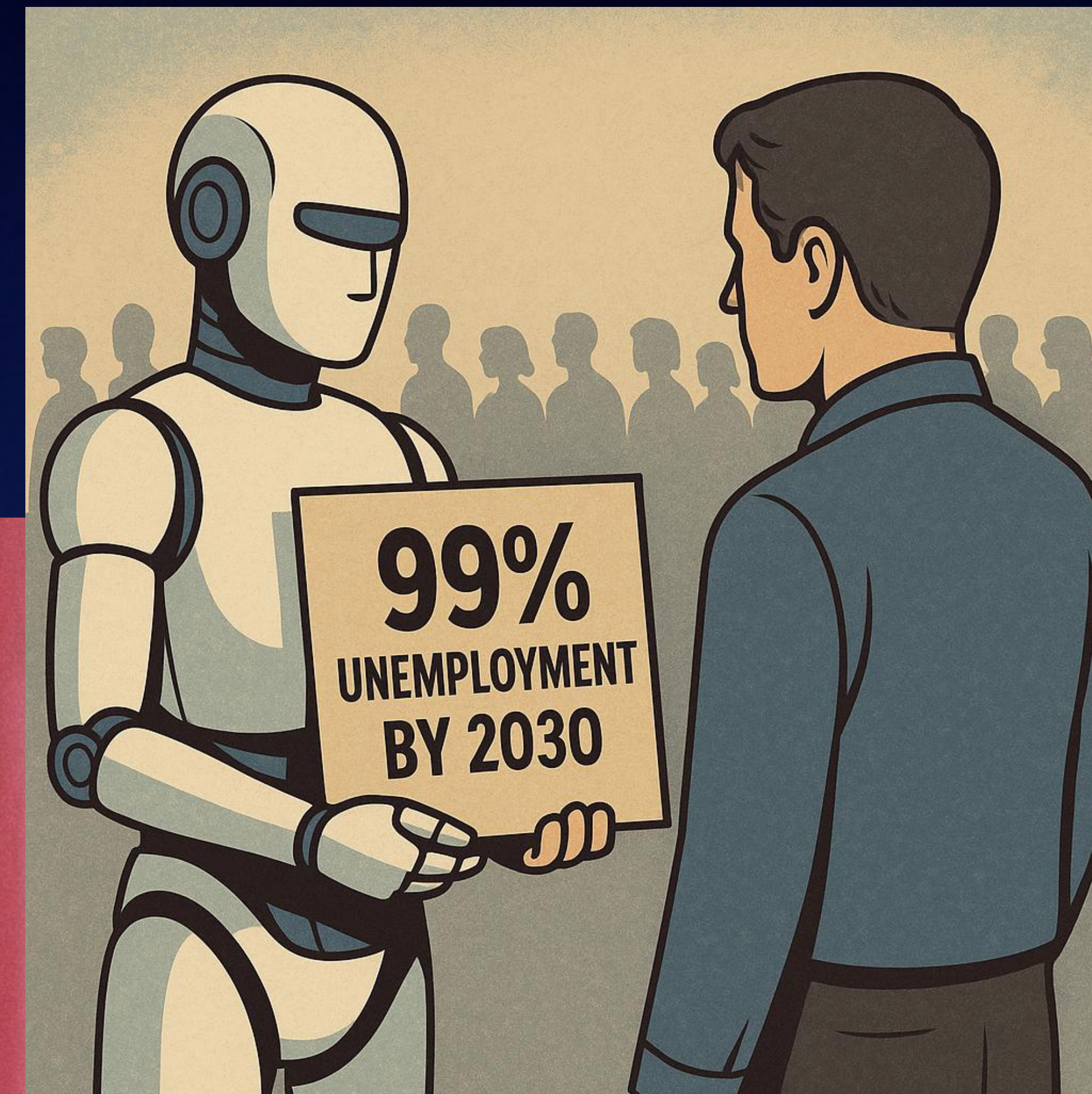
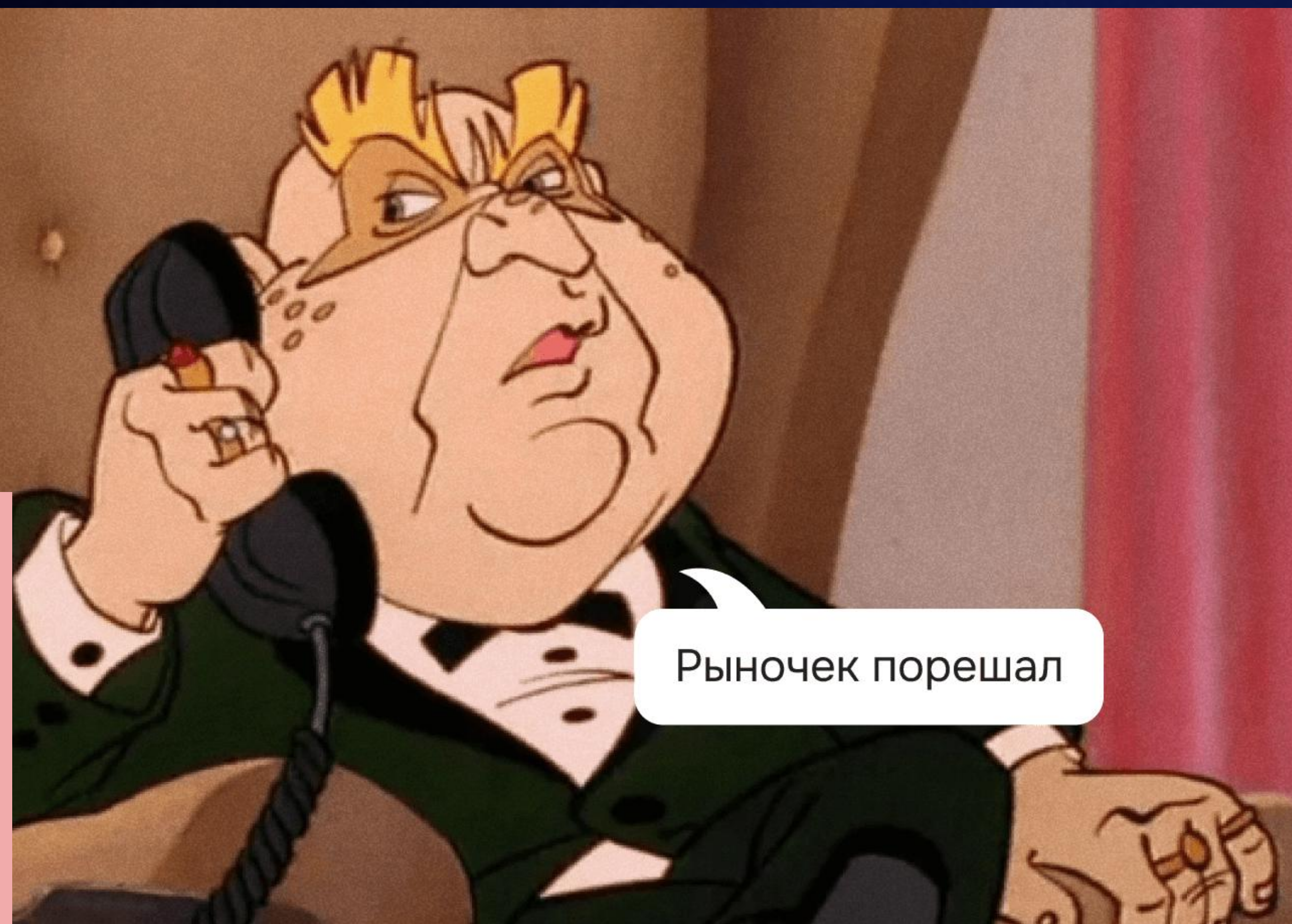
Альтернативные целевые
задачи при обучении
(denoising/FIM и т. п.
вместо
авторегрессионного
языкового моделирования)

Активное обучение

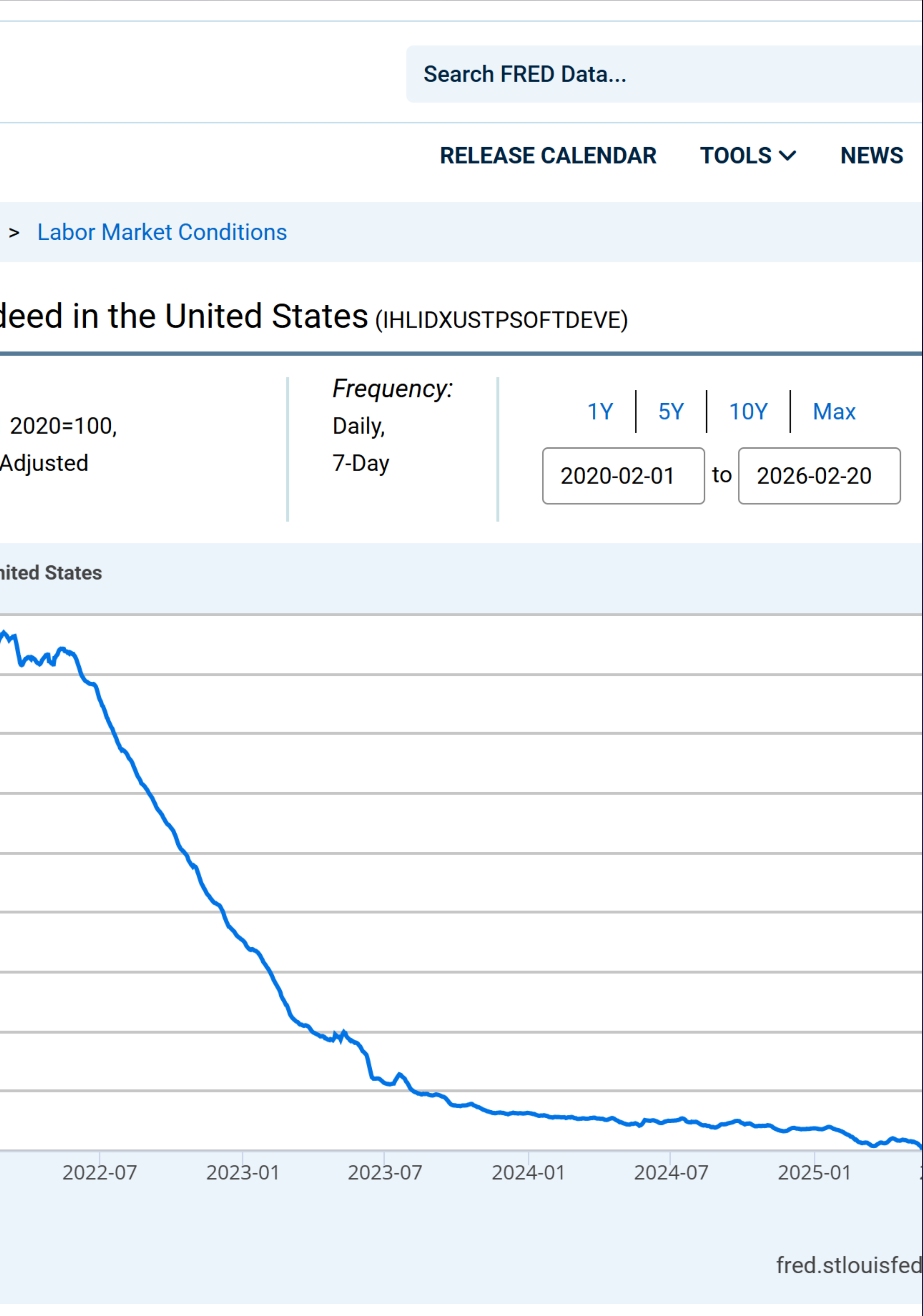
?

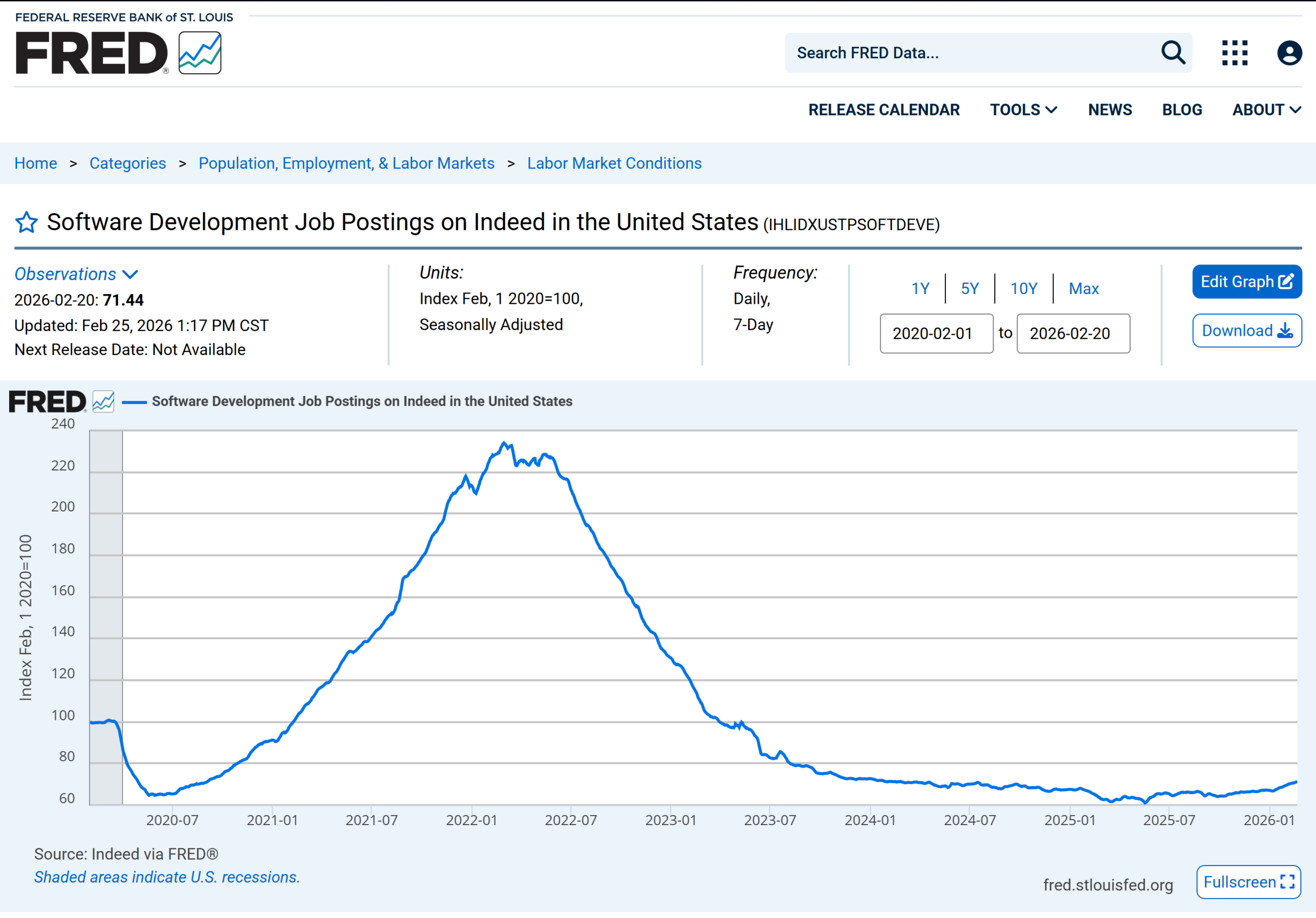


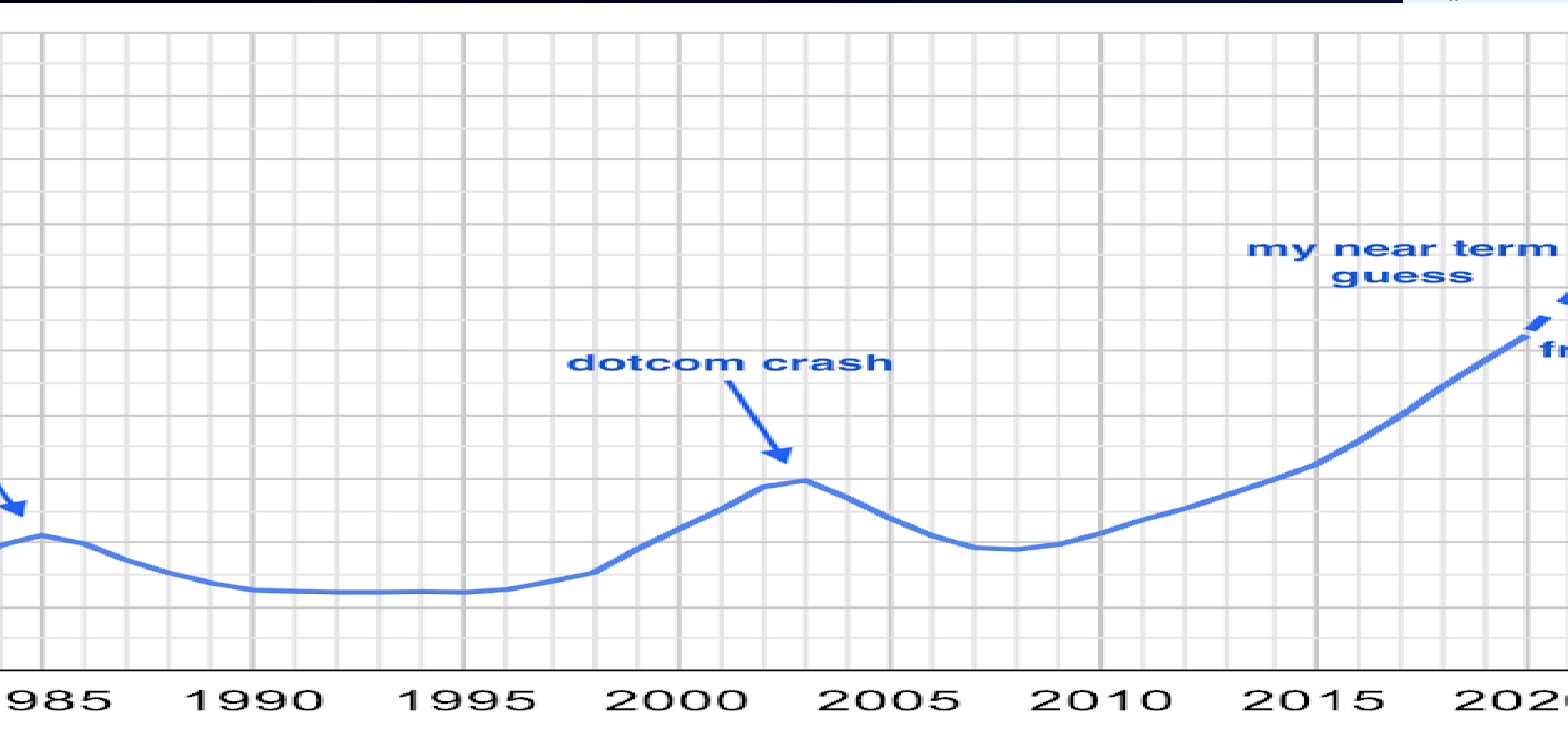
ИИ лишит всех людей работы!

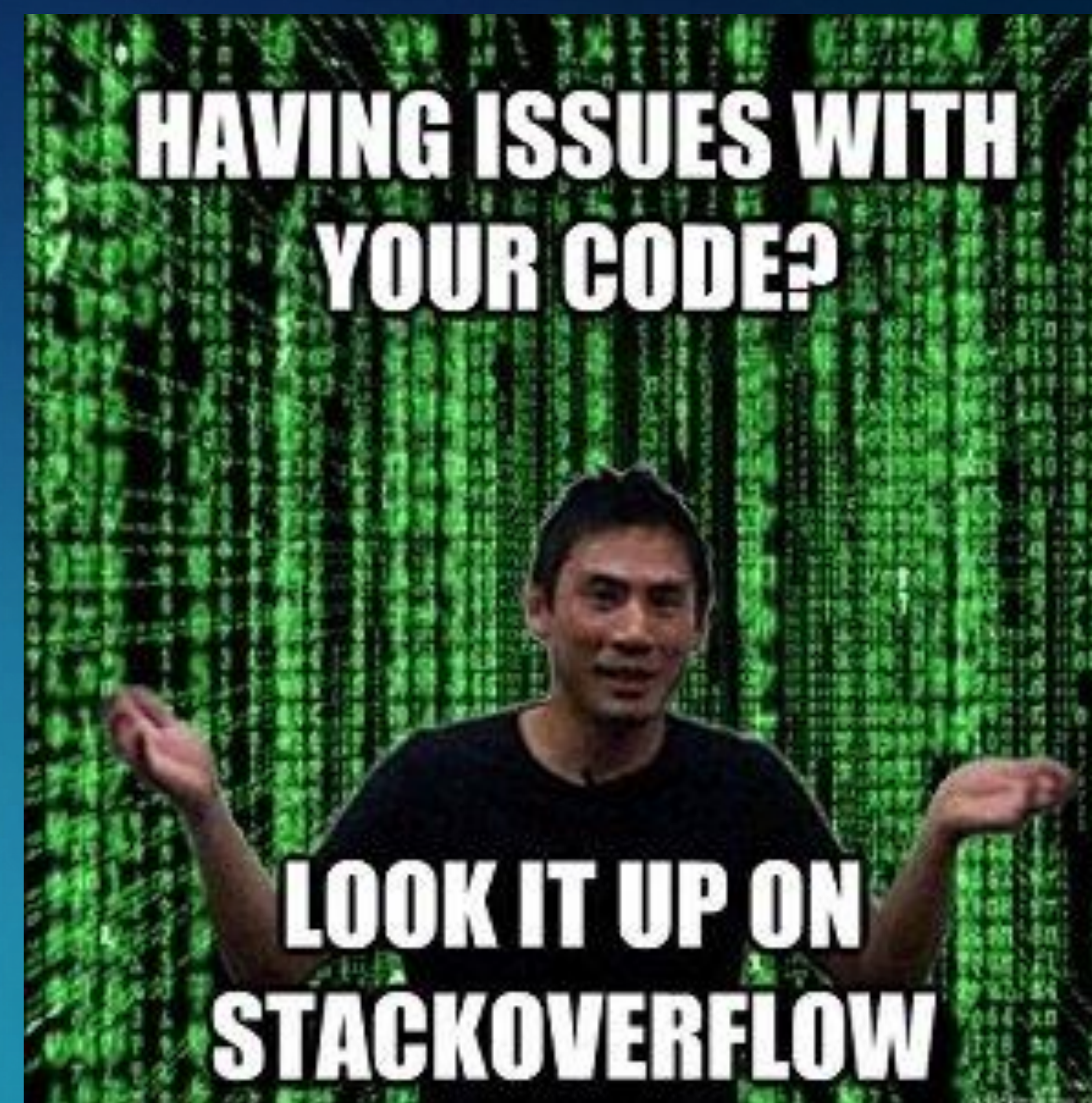
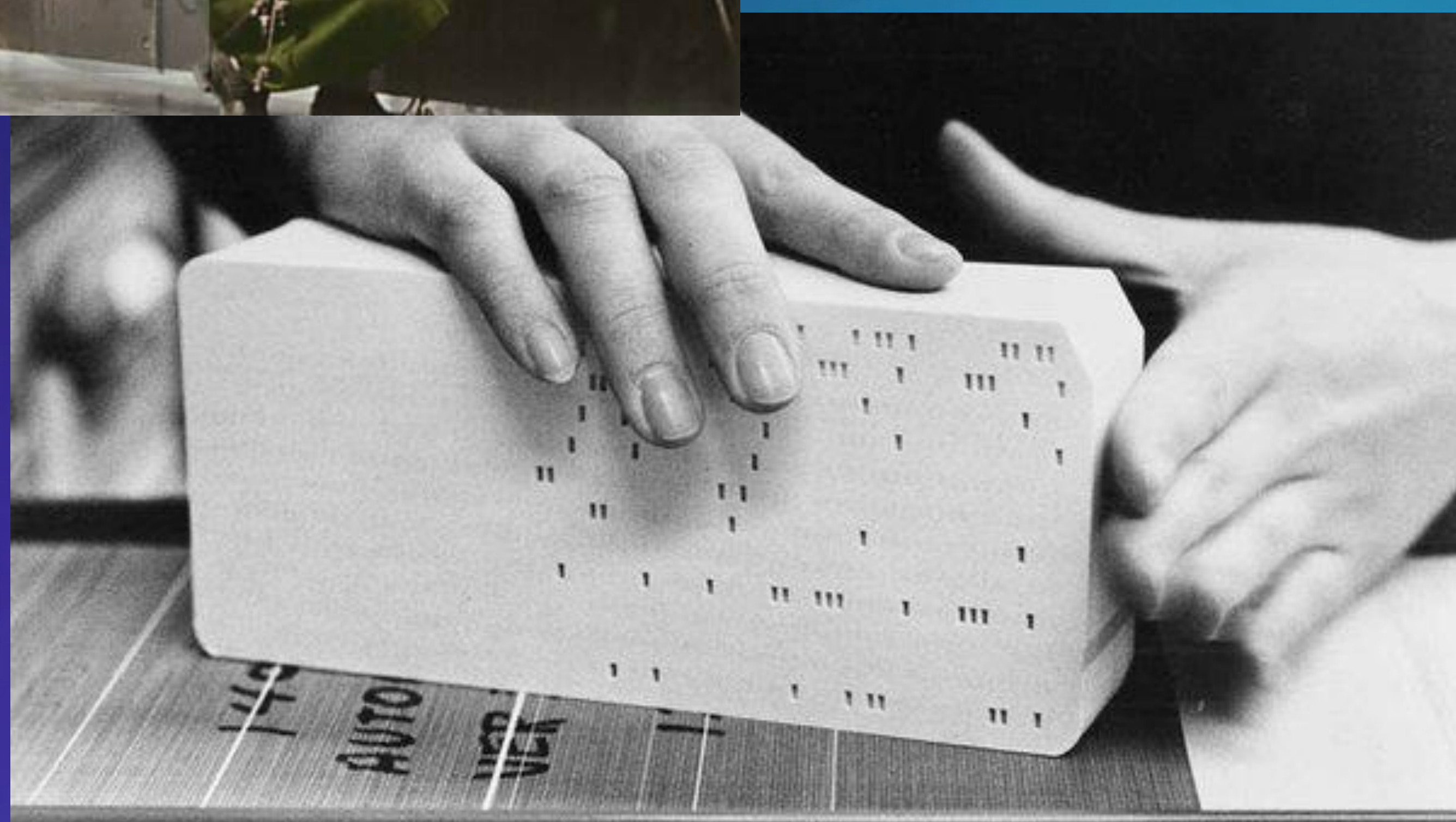










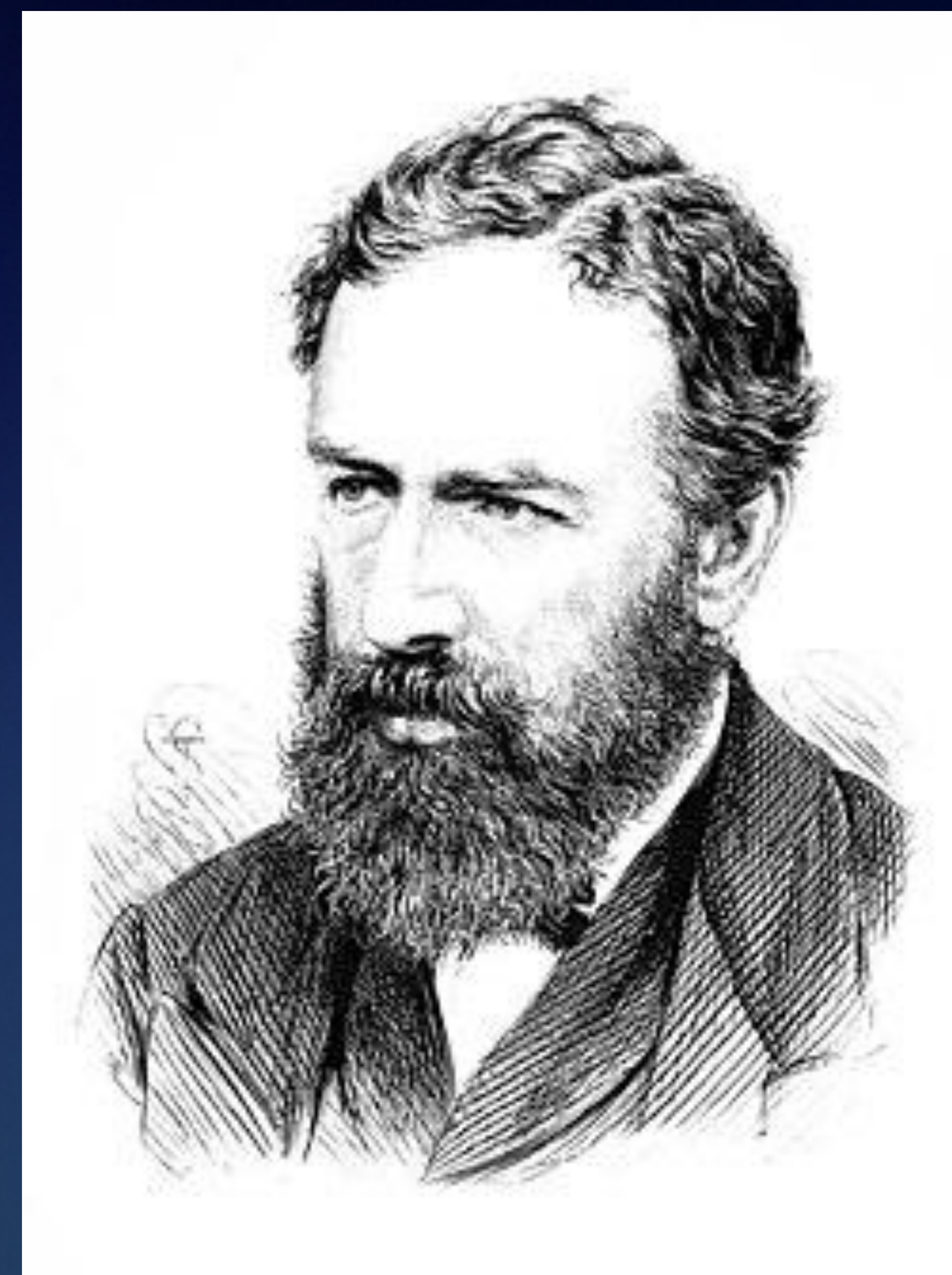


Парадокс Жевонса



Парадокс Жевонса, эффект Жевонса — термин в экономической науке, описывающий ситуацию, когда технологический прогресс, который увеличивает эффективность использования какого-либо ресурса, увеличивает объём потребления ресурса из-за более низких затрат при использовании усовершенствованных технологий

Парадокс Жевонса заключается в неочевидном факте: рост эффективности использования ресурса ведет не к уменьшению, а к увеличению его потребления

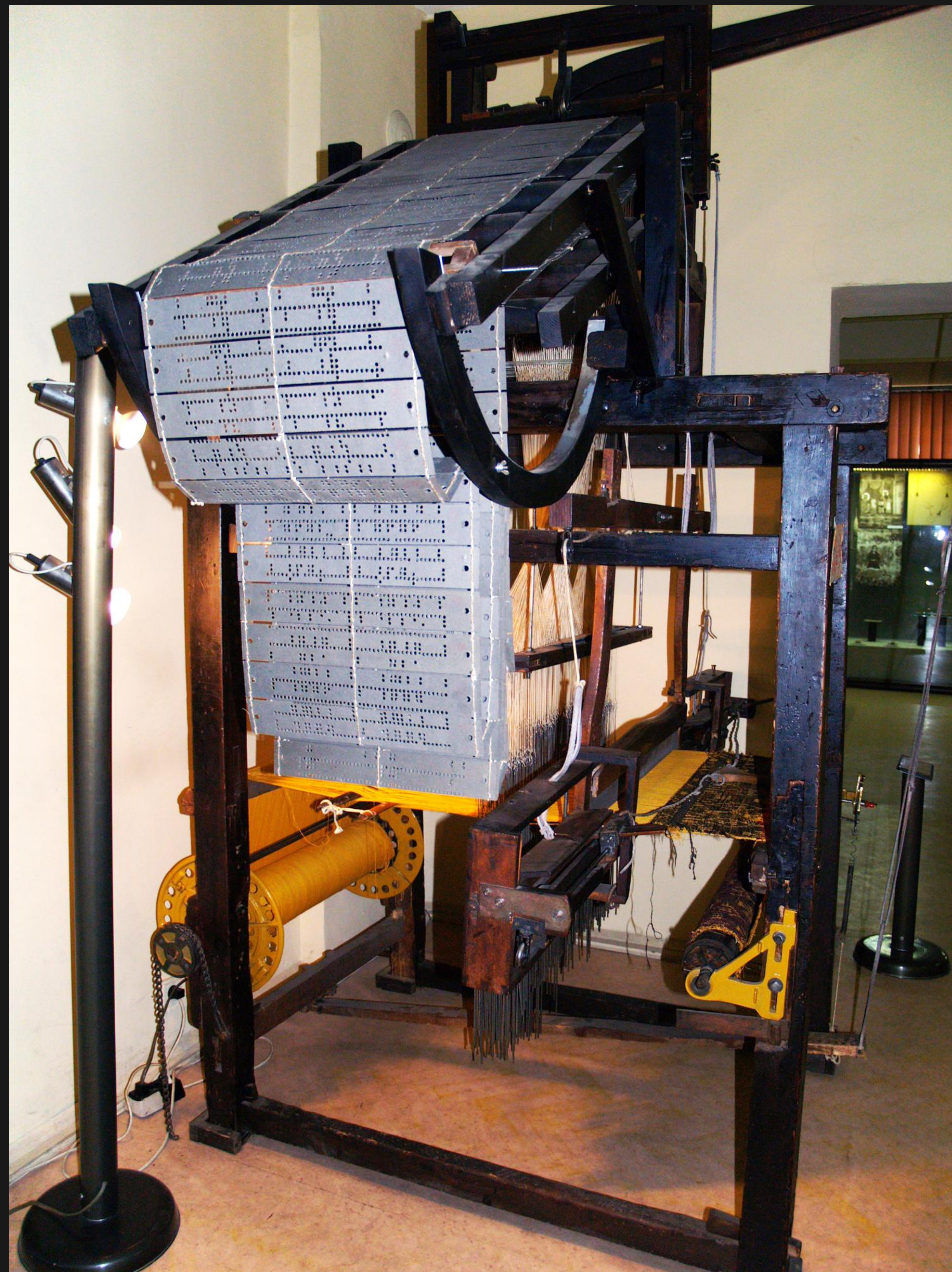


William Stanley Jevons (1835—1882)



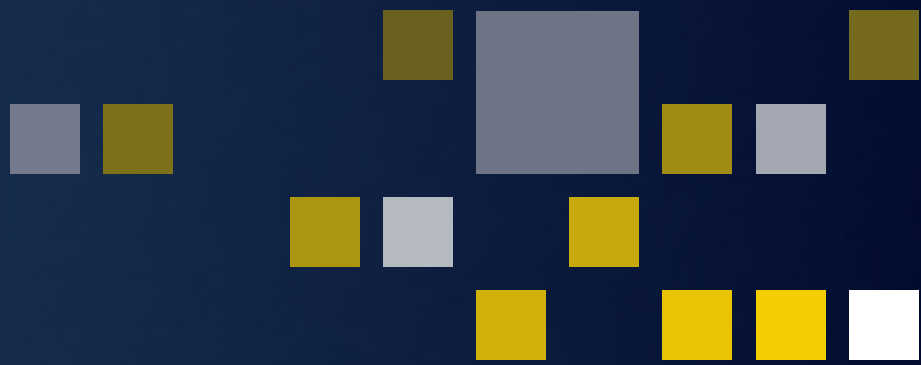
О овцы, такие ручные и неприхотливые в еде. Они становятся настолько жадными и необузданными, что поглощают самих людей, опустошают и делают безлюдными поля, дома и города



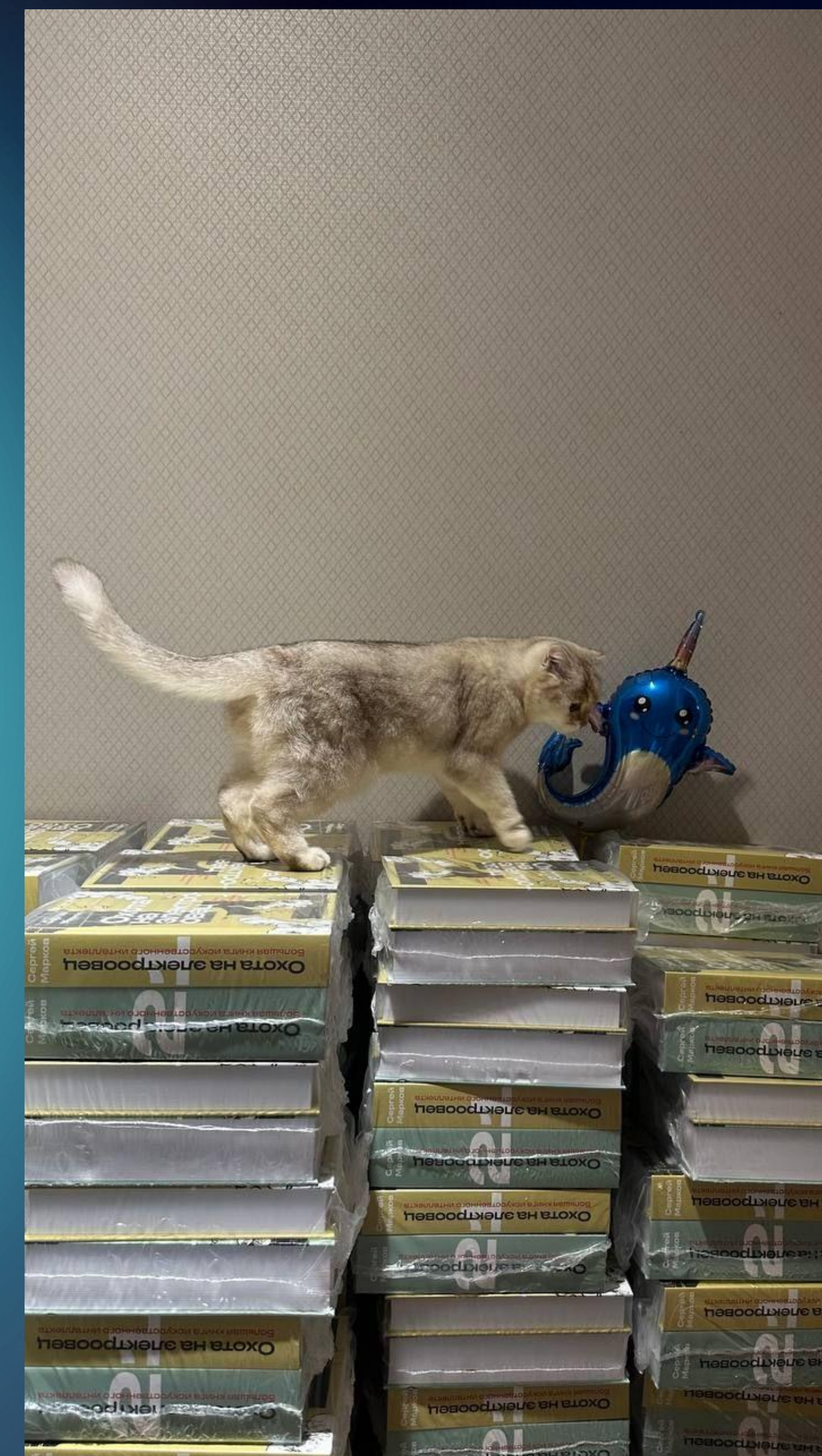








- ❑ > 6 лет работы
- ❑ > 450 тыс. слов
- ❑ > 3,4 млн знаков
- ❑ > 3,2 тыс. ссылок на источники
- ❑ обложка твёрдая, размер 185 мм × 240 мм, бумага 100 г/м², книга цветная, 568 страниц в первом томе и 784 страницы во втором
- ❑ полные электронные версии в свободном доступе (pdf, epub, fb2, docx)



Спасибо за внимание!



Сайт и книга



Канал в telegram

