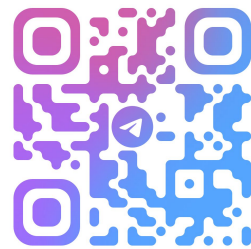


AI в науке

Где мы сейчас и куда мы идём



@SINECOR

Сергей
Николенко

25 августа 2026



Наш план

- Революции искусственного интеллекта
- Законы масштабирования
- Виды прогресса
- Примеры возможностей
- AI в математике
- Заключение



1. Революции AI



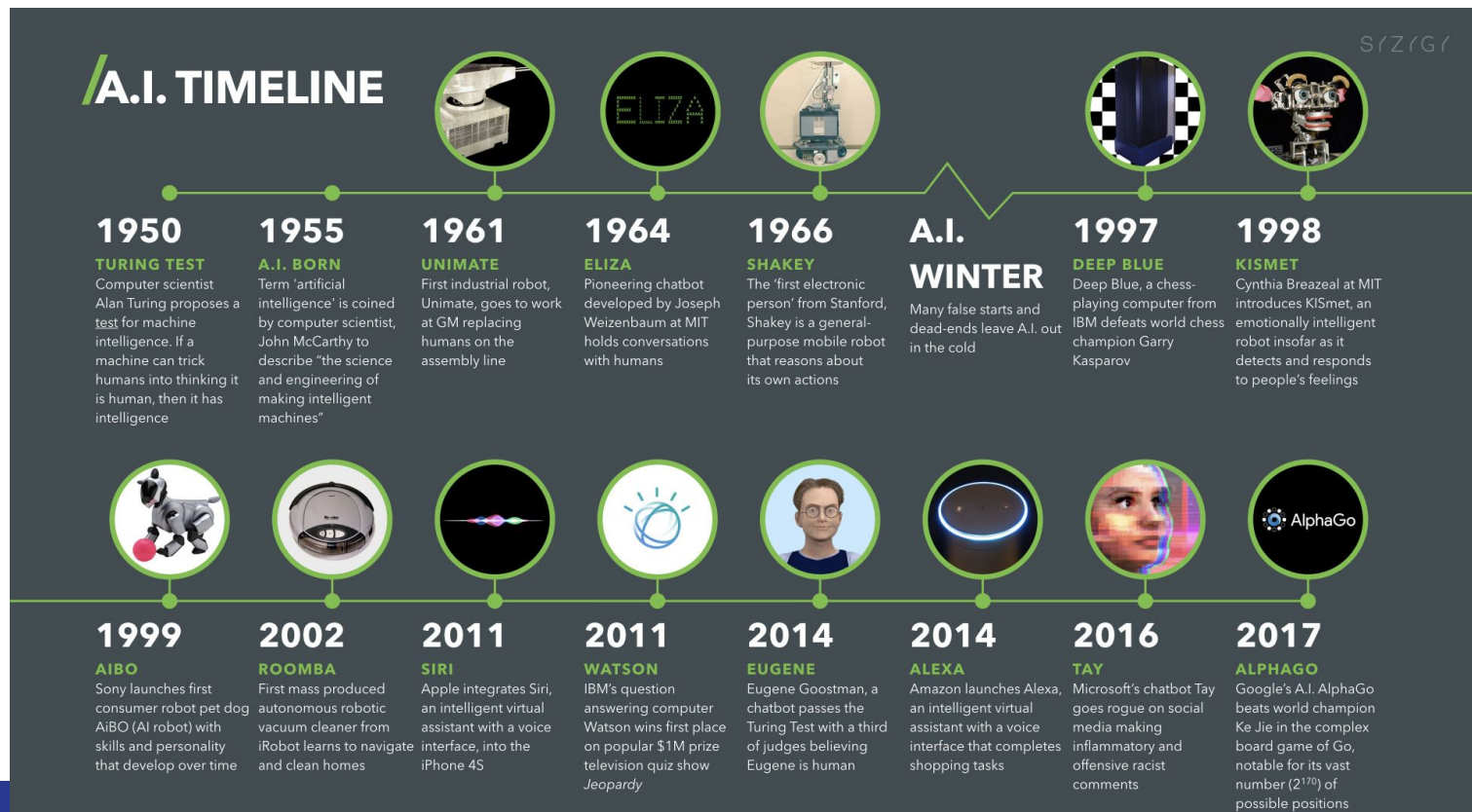
Революции, происходящие на поверхности жизни, ничего существенного никогда не меняют и не открывают, они лишь... по-новому переставляют все те же элементы и являют старые образы в новых одеяниях. Революция всегда есть в значительной степени маскарад, и если сорвать маски, то можно встретить старые, знакомые лица. Новые души рождаются позже, после глубокого перерождения и осмысливания опыта революции.

Николай Бердяев. Духи русской революции

AI timelines



- Обычный AI timeline
- Источник
- Давайте немного разберёмся в терминах...



AI timelines



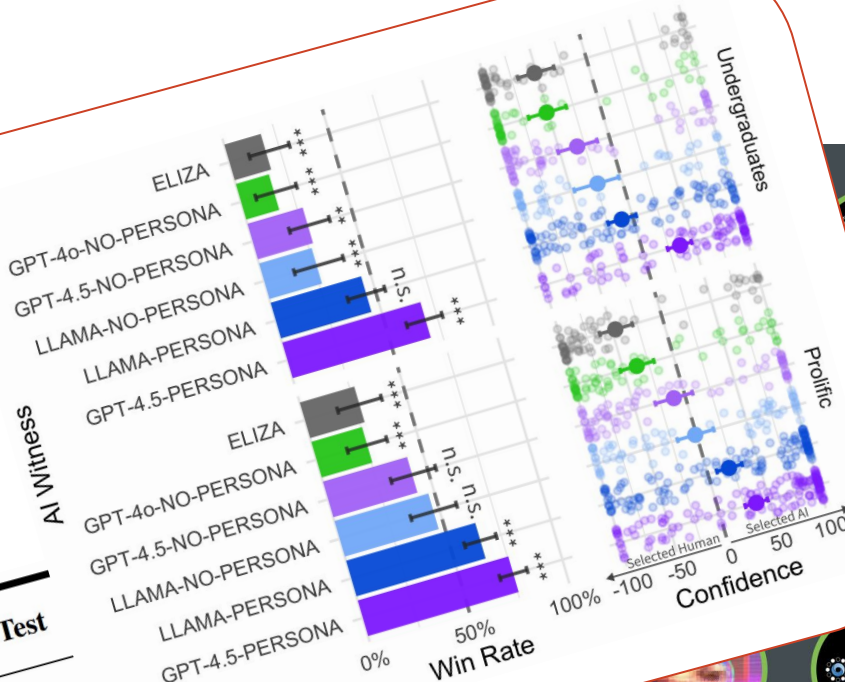
- Обычный
AI timeline

Jones, Bergen (Mar 31, 2025): кстати, тест Тьюринга наконец-то проверили и прошли

Large Language Models Pass the Turing Test

Cameron R. Jones
Department of Cognitive Science
UC San Diego
San Diego, CA 92119
cameron@ucsd.edu

Benjamin K. Bergen
Department of Cognitive Science
UC San Diego
San Diego, CA 92119
bkbergen@ucsd.edu



2014

2014

2016

2017

WATSON

IBM's question answering computer Watson wins first place on popular \$1M prize television quiz show Jeopardy

EUGENE

Eugene Goostman, a chatbot passes the Turing Test with a third of judges believing Eugene is human

ALEXA

Amazon launches Alexa, an intelligent virtual assistant with a voice interface that completes shopping tasks

TAY

Microsoft's chatbot Tay goes rogue on social media making inflammatory and offensive racist comments

ALPHAGO

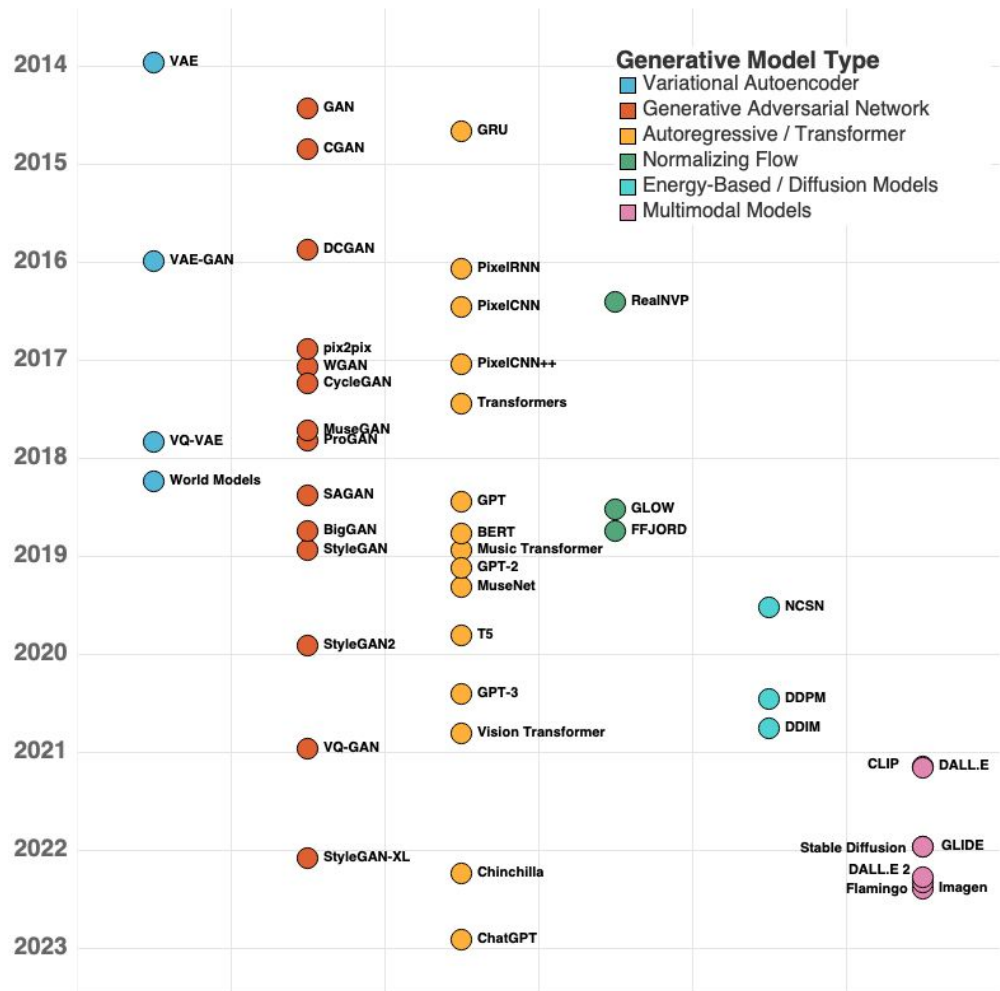
Google's A.I. AlphaGo beats world champion Ke Jie in the complex board game of Go, notable for its vast number (2¹⁷⁰) of possible positions

AI timelines

- Generative AI timeline: вот более детально, здесь отмечены многие важные модели

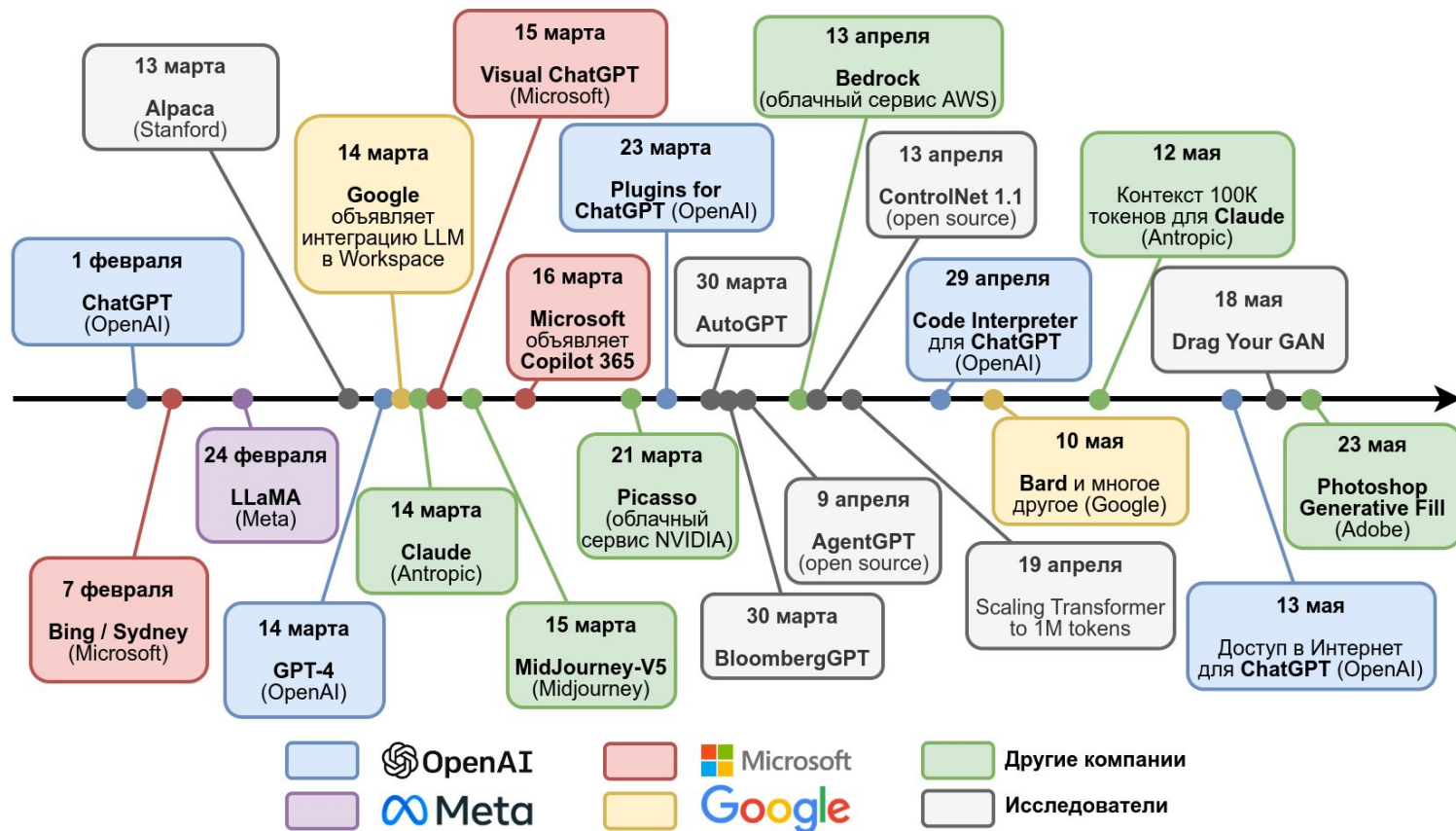
- [Источник](#)

Generative AI Timeline



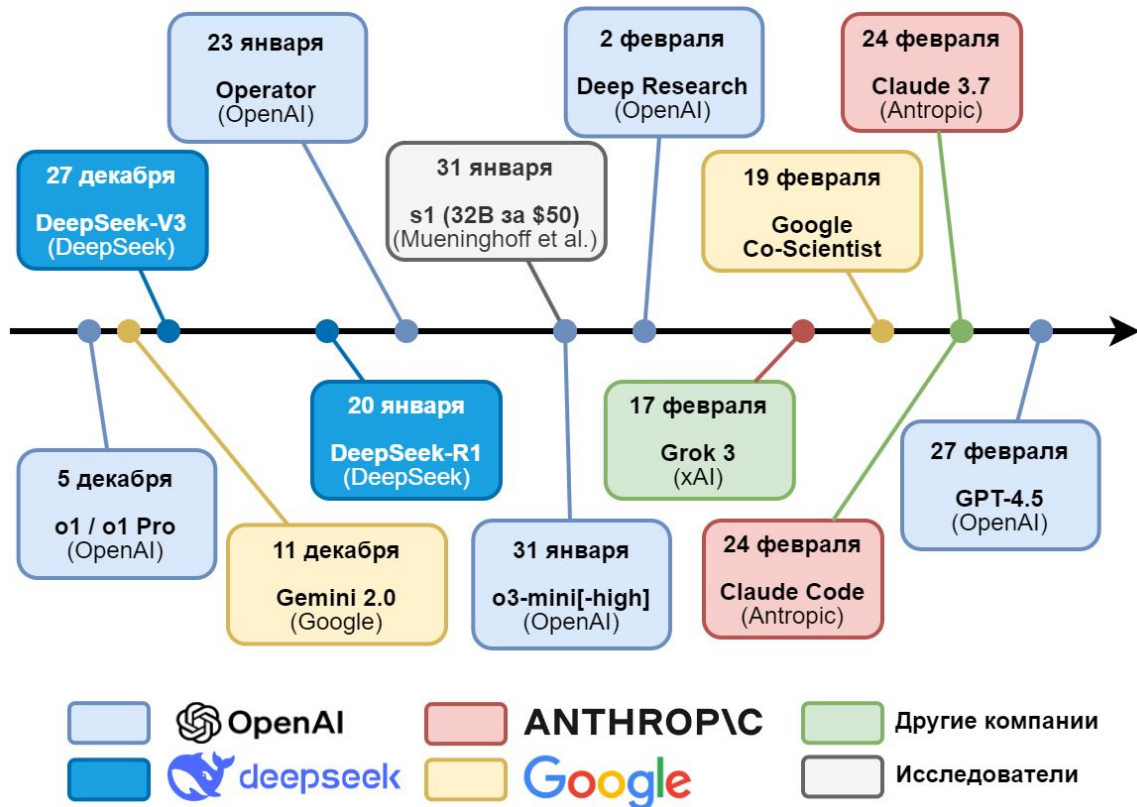
Развитие AI в эпоху LLM

- А вот AI timeline весны 2023 года
- И с тех пор не то чтобы замедляется...



Например, в начале 2025 года

- 22 января: [DeepSeek-R1](#)
- 31 января: s1 ([Mueninghoff et al., 2025](#))
- 2 февраля: [OpenAI Deep Research](#)
- 17 февраля: [Grok 3](#)
- 19 февраля: [Google Co-Scientist](#) на основе Gemini 2.0 ([11 декабря 2024](#))
- 24 февраля: [Claude 3.7 Sonnet](#) (hybrid reasoning Claude)
- 27 февраля: [GPT-4.5](#)



AGI? Сингулярность?

- Значит ли это, что создание AGI уже близко?
- И если AGI на пороге, что это значит для человечества?
- Приведёт ли это к той самой технологической сингулярности?
- Это вопросы без ответов, и риски разработки AI мы обсудим не сегодня
- Сегодня — о новых возможностях и о том, откуда они берутся



2. Законы масштабирования

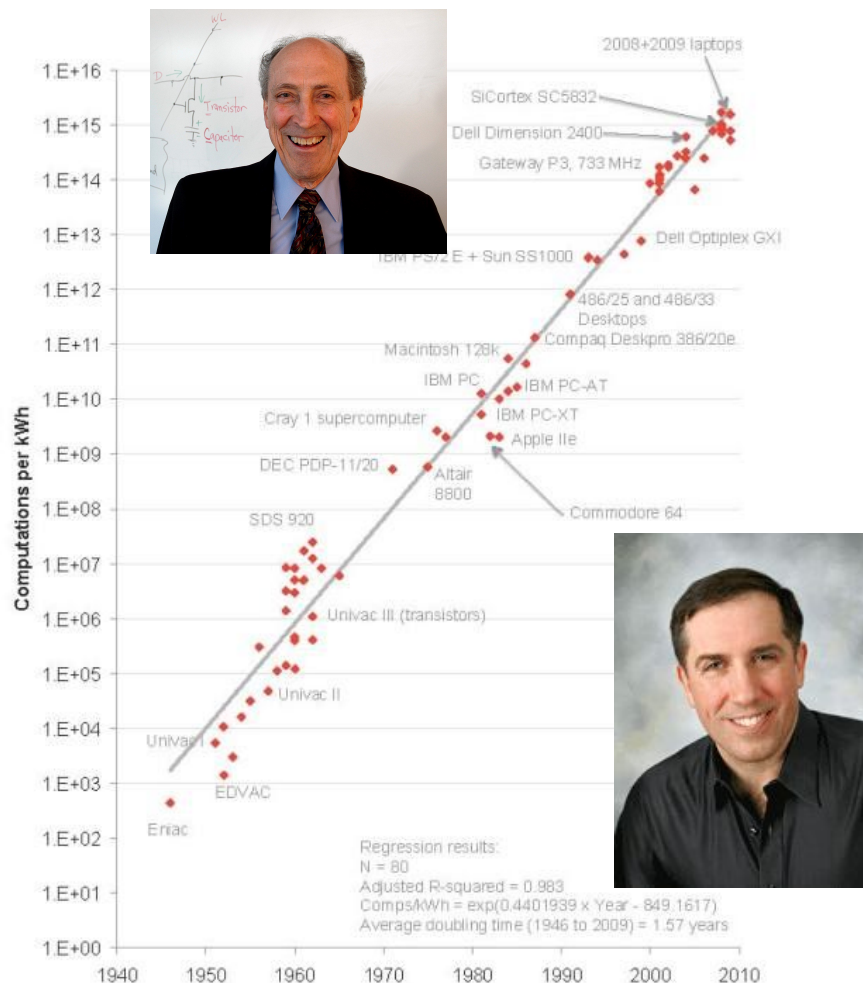


Безграничный простор для научно-фантастического жанра открывается только у нас, в СССР, в стране реализованных "утопий"... Масштабы нашего строительства воистину грандиозны. Эти масштабы растут с каждым годом, с каждой пятилеткой. Им нет и не может быть границ.

*Александр Беляев.
Застрельщики новых открытий*

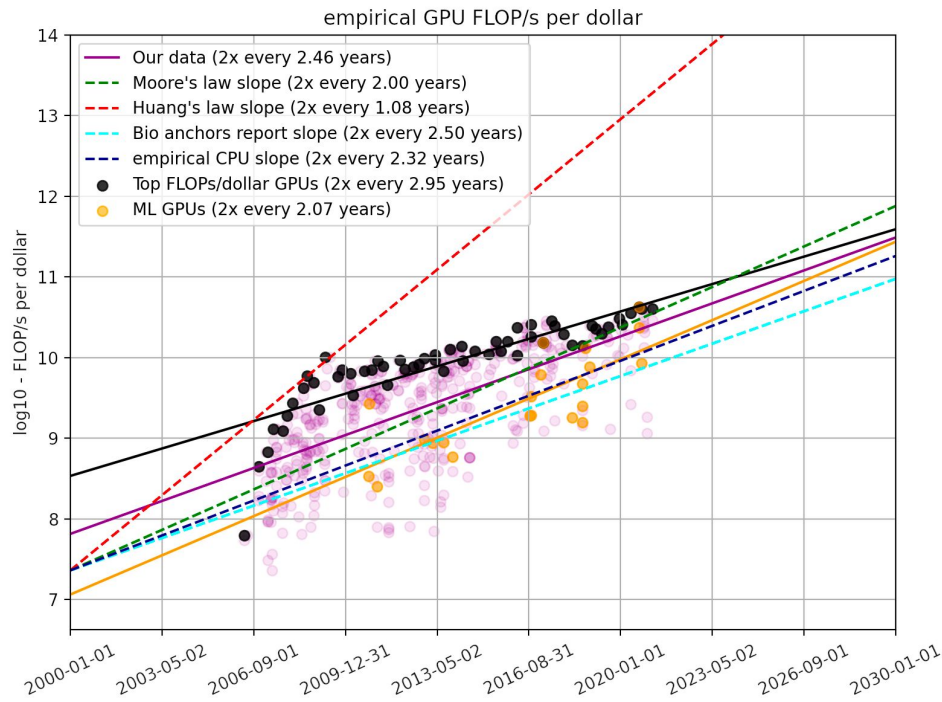
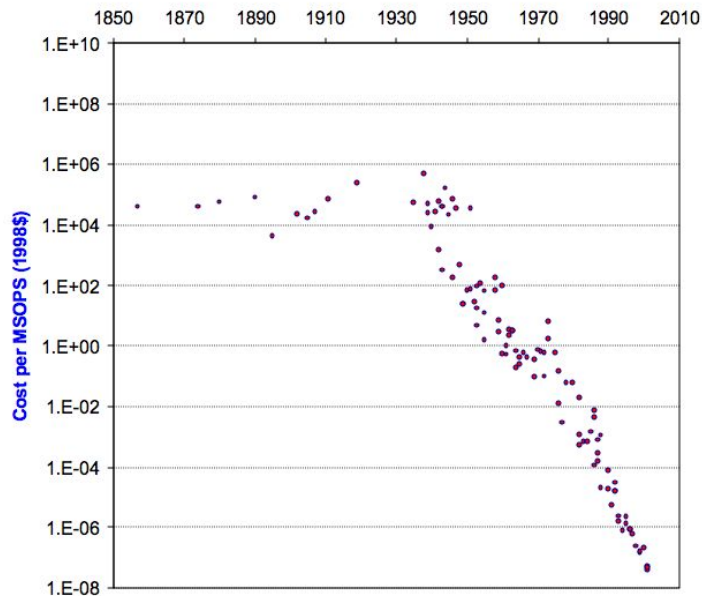
Закон Куми

- Закон масштабирования Деннарда (Dennard scaling): уменьшая размеры транзистора и повышая частоту процессора, можно пропорционально повышать производительность
- В сумме закон Куми (Koomey's law): объём вычислений на джоуль энергии удваивается за ~1.57 года
- Тоже начали замедляться вместе с законом Мура около 2010; сейчас удвоение за ~2.6 лет; теоретический предел – принцип Ландауэра



Цена вычислений

- Цена в терминах FLOPs/\$ тоже падает экспоненциально, долгое время число FLOPs на доллар удваивалось за 2.3 года, сейчас медленнее



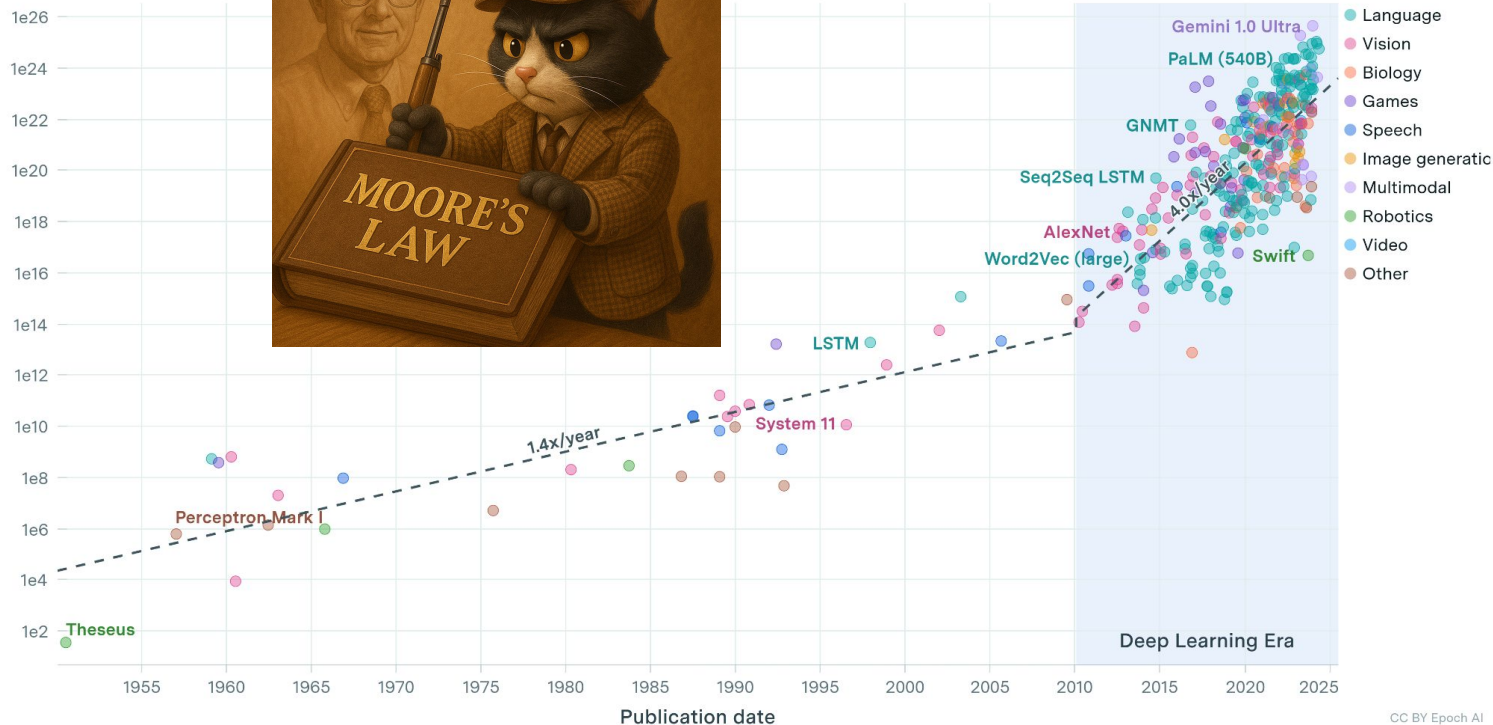
Цена вычислений в ML

- Но в машинном обучении закон Мура уже давно мёртв!

- OpenAI (2018)

Notable AI models

Training compute (FLOP)



Цена вычислений в ML

- И тренд так и продолжается до сих пор, около **3x в год**
- Так ведь не может продолжаться вечно?
- Наверное, не может

Cost of training compute for notable ML systems

≡ EPOCH AI

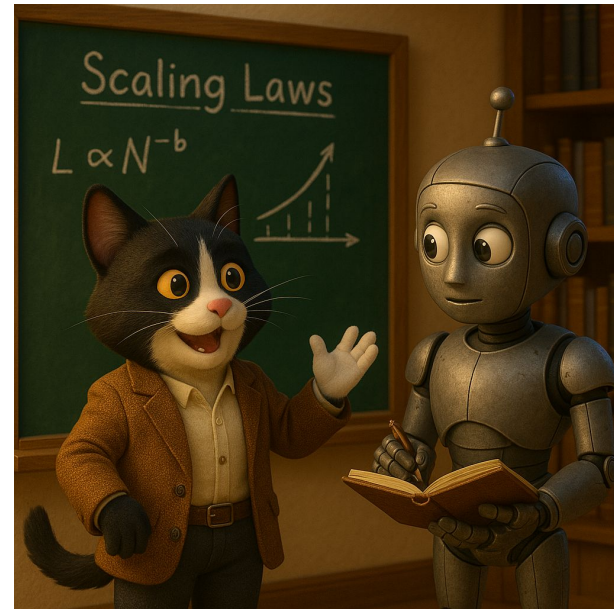


Классические результаты в ML

- Размерность Вапника-Червоненкиса (VC dimension): формализация выразительности модели, того, сколько в неё “помещается”
- Есть нижние оценки, например [Ehrenfeucht et al. \(1989\)](#), на sample complexity, т.е. сколько нужно данных для обучения с заданной ошибкой
- [Amari et al. \(1992\)](#), [Amari, Murata \(1993\)](#): для ошибки обобщения должен быть power law,

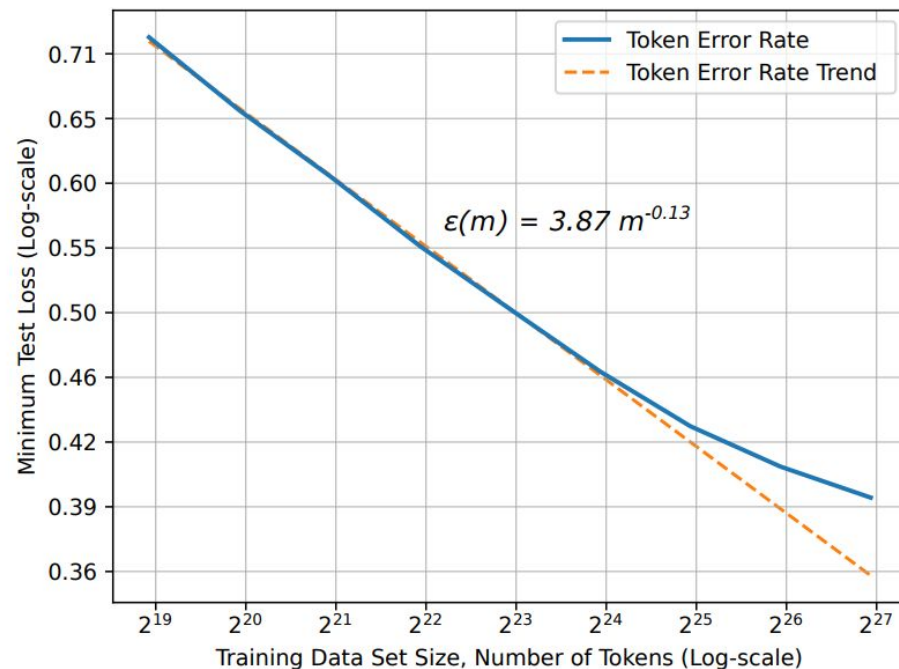
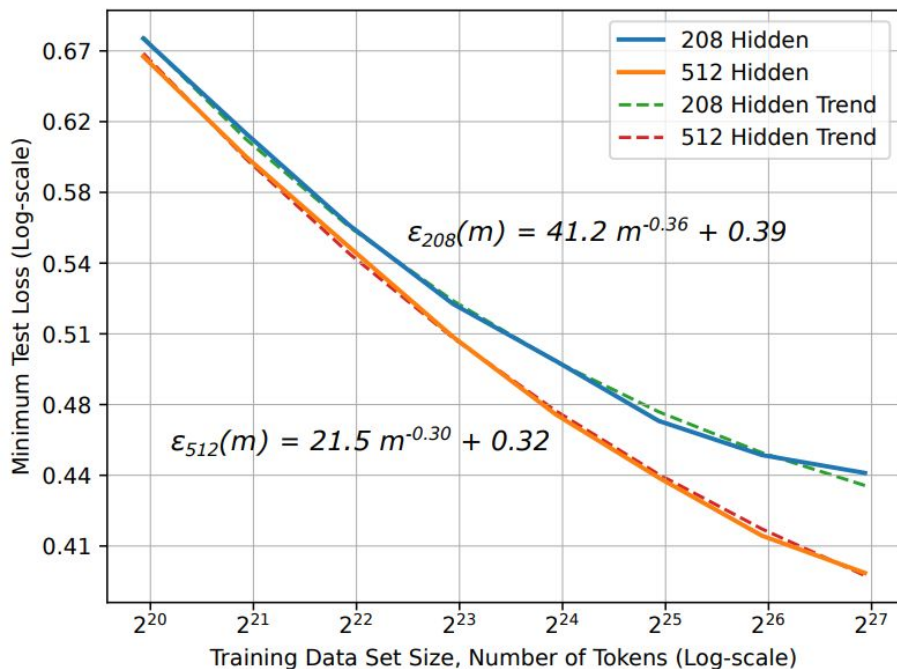
$$\varepsilon(m) \sim \alpha m^{\beta_g}$$

и степень β здесь -0.5 , -1 или -2



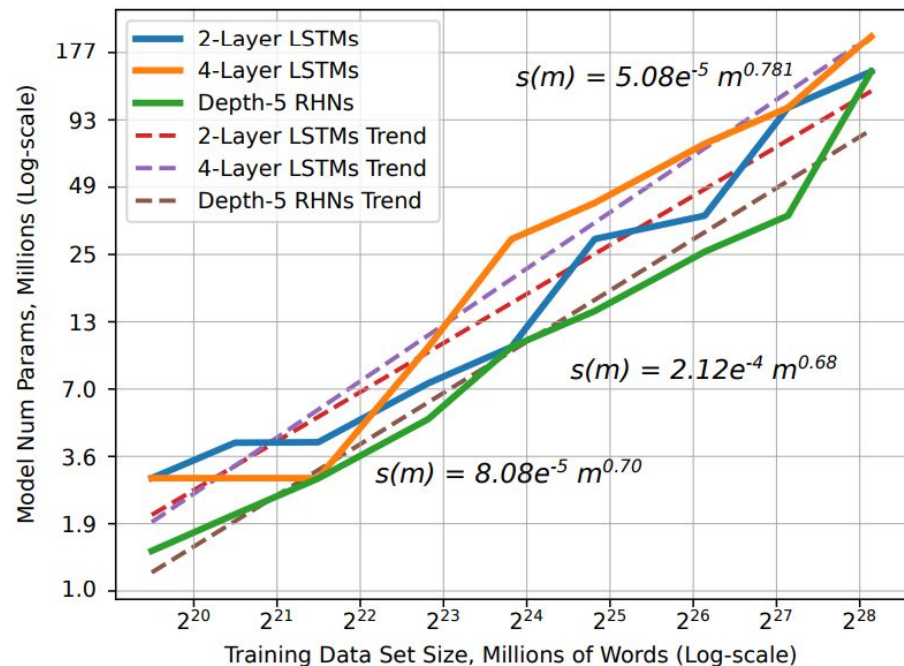
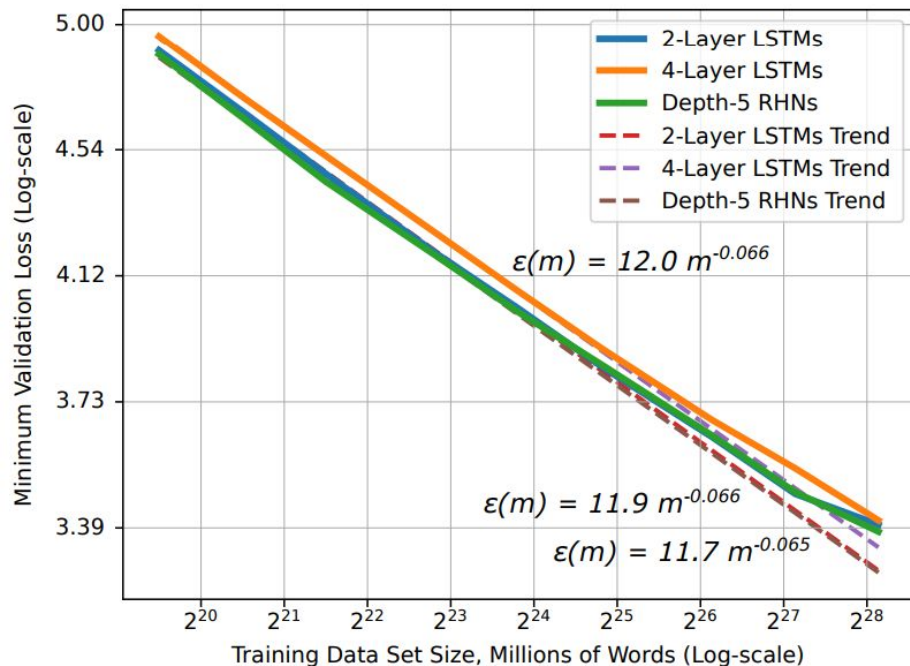
Scaling laws в глубоком обучении

- [Hestness et al., 2017](#): Deep Learning Scaling is Predictable, Empirically. Первые практические законы для разных задач. Машинный перевод:



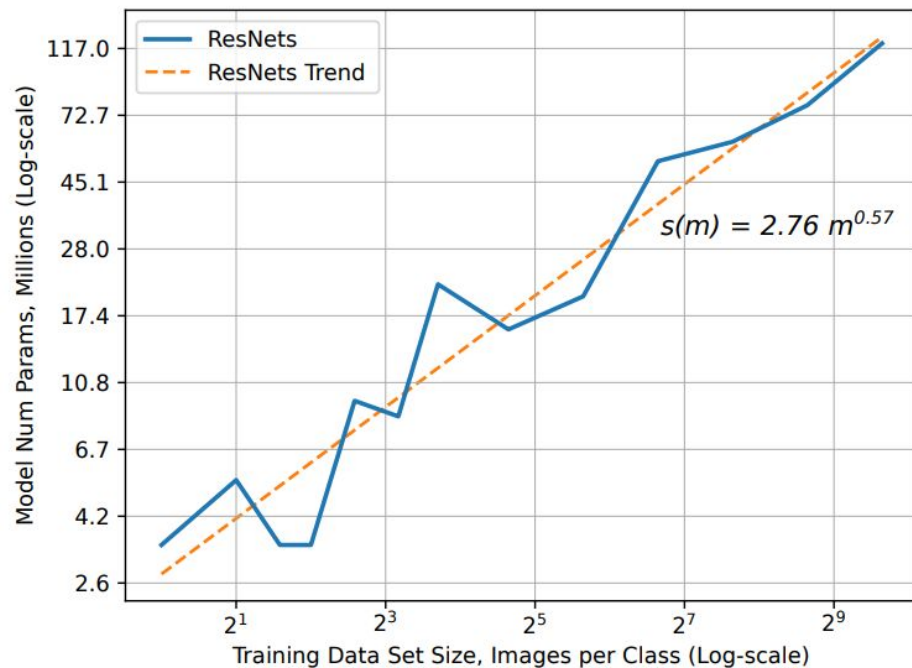
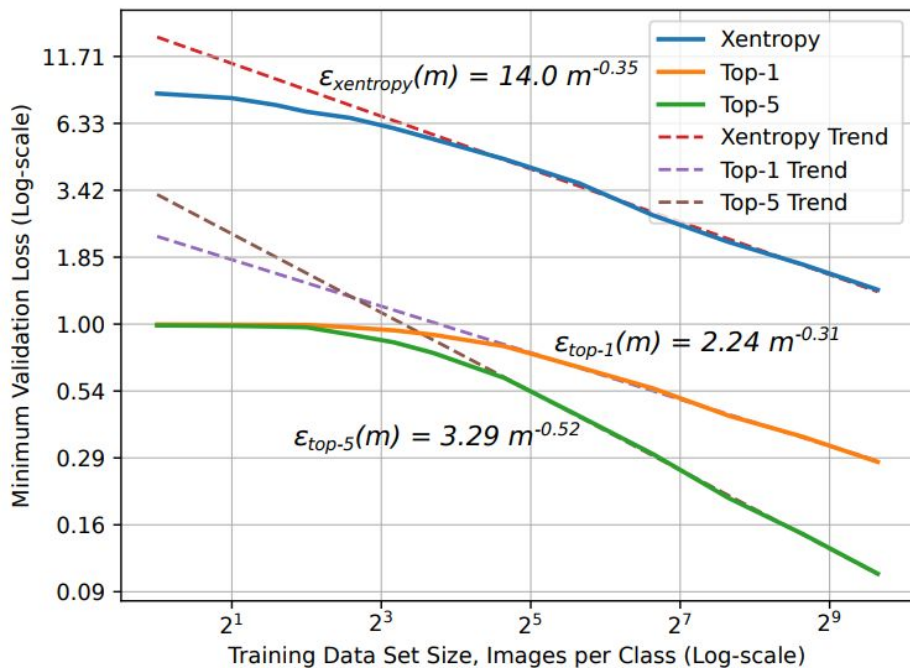
Scaling laws в глубоком обучении

- [Hestness et al., 2017](#): Deep Learning Scaling is Predictable, Empirically. Первые практические законы для разных задач. Языковые модели:



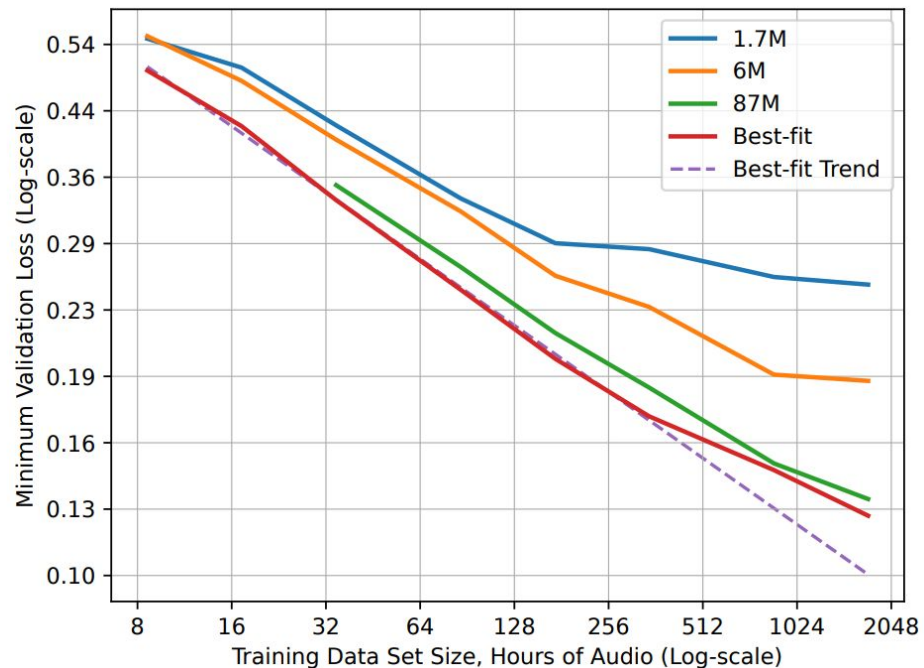
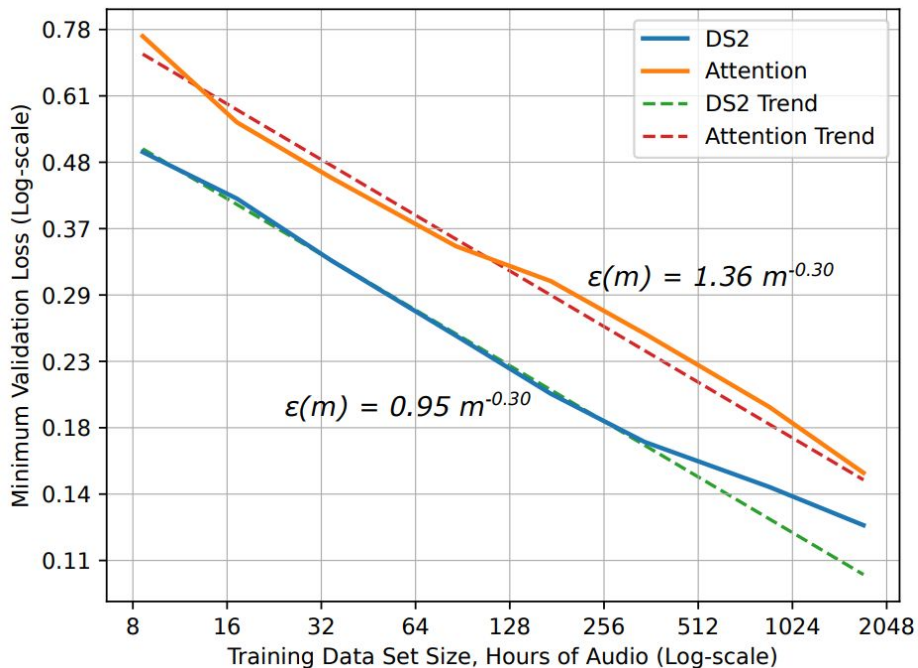
Scaling laws в глубоком обучении

- [Hestness et al., 2017](#): Deep Learning Scaling is Predictable, Empirically. Первые практические законы для разных задач. Классификация изображений:



Scaling laws в глубоком обучении

- [Hestness et al., 2017](#): Deep Learning Scaling is Predictable, Empirically. Первые практические законы для разных задач. Распознавание речи:



Scaling laws в глубоком обучении

- Hestness et al., 2017: Deep Learning Scaling is Predictable, Empirically. Общие выводы:
 - ошибка обобщения следует power law
 - улучшение архитектуры модели сдвигает power law вниз, *но не меняет экспоненту*; кажется, что экспонента постоянна для каждой задачи, а не семейства моделей
 - размер модели растёт сублинейно, удвоение датасета не требует удваивать параметры
- Отсюда уже один шаг до общего закона...

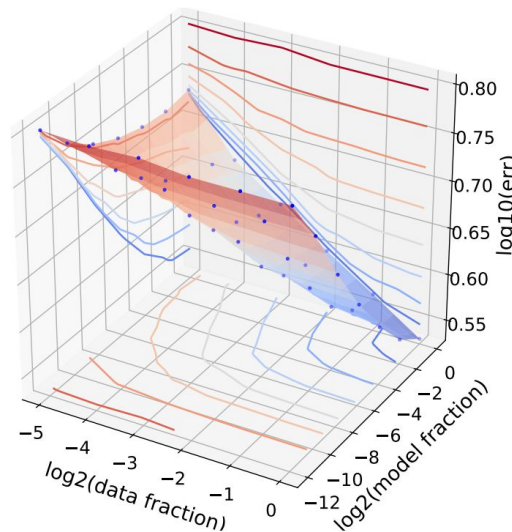


Scaling laws в глубоком обучении

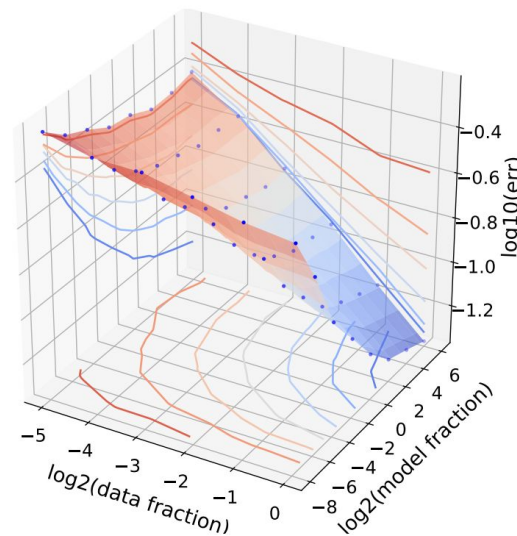
- [Rosenfeld et al., 2020](#): а можно и общий закон вывести, который свяжет объём данных n , размер модели m и ошибку:

$$\tilde{\epsilon}(m, n) = an^{-\alpha} + bm^{-\beta} + c_{\infty}$$

- Там есть тонкости на начальном отрезке, но в общем так, и все эксперименты сходятся



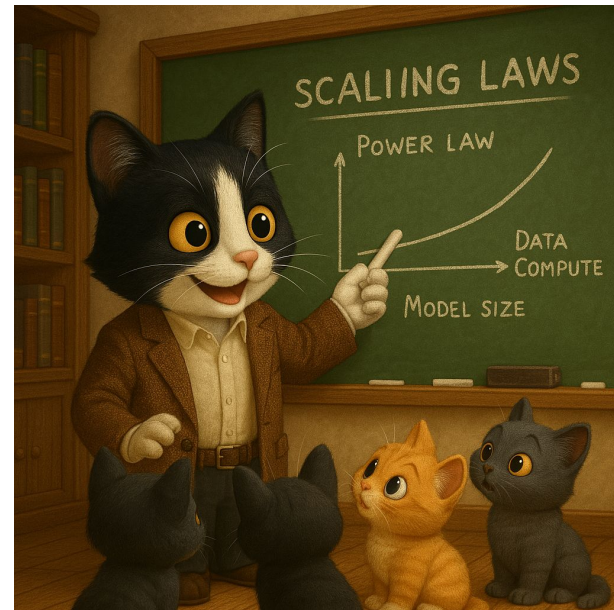
(a) Wiki103 error (cross entropy) landscape.



(b) CIFAR10 error (top1) landscape.

Scaling laws для LLM

- [Kaplan et al., 2020](#): OpenAI исследовал конкретно языковые модели на трансформерах
- Результаты в целом те же: от размера модели N , размера датасета D , вычислительного бюджета C получаются power laws, можно найти экспоненты



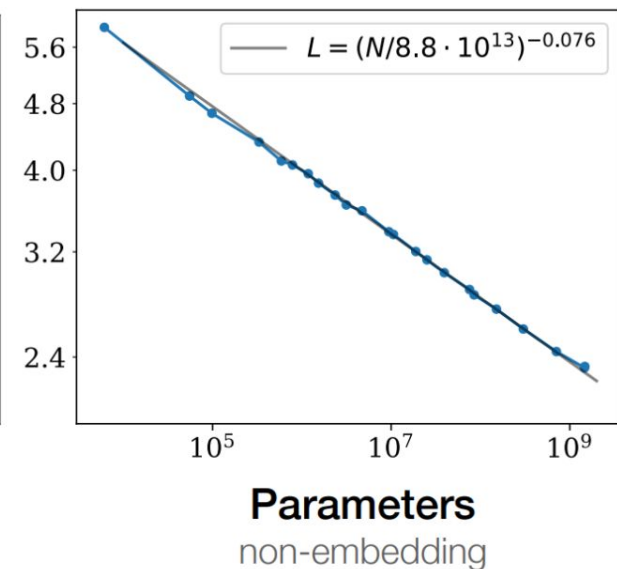
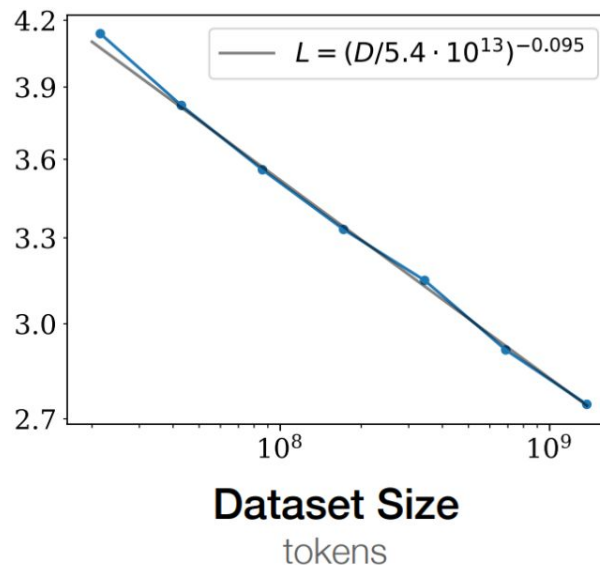
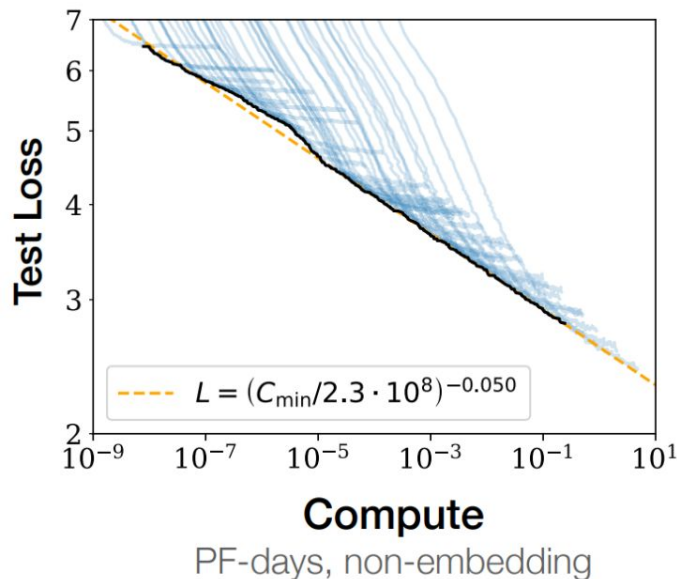
$$L(N) = (N_c/N)^{\alpha_N}; \quad \alpha_N \sim 0.076, \quad N_c \sim 8.8 \times 10^{13}$$

$$L(D) = (D_c/D)^{\alpha_D}; \quad \alpha_D \sim 0.095, \quad D_c \sim 5.4 \times 10^{13}$$

$$L(C_{\min}) = (C_c^{\min}/C_{\min})^{\alpha_C^{\min}}; \quad \alpha_C^{\min} \sim 0.050, \quad C_c^{\min} \sim 3.1 \times 10^8 \text{ (PF-days)}$$

Scaling laws для LLM

- [Kaplan et al., 2020](#): удивителен был, во-первых масштаб; законы выполнялись очень точно на протяжении 7 десятичных порядков



Scaling laws для LLM

- [Kaplan et al., 2020](#) буквально записывают формулы для этих законов, например:

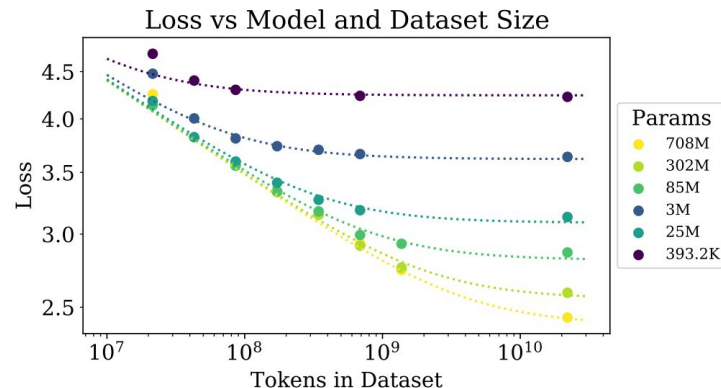
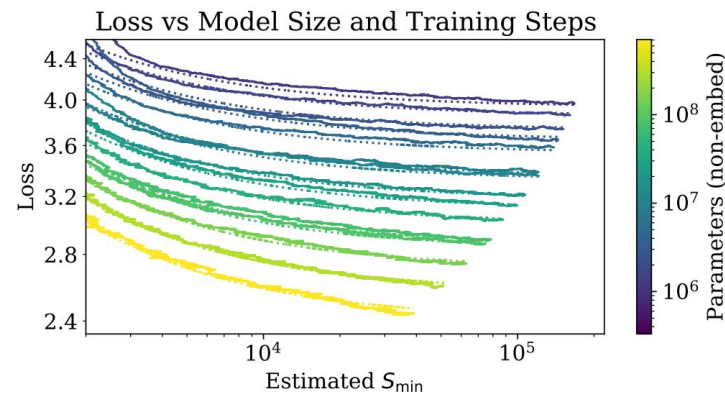
$$L(N, D) = \left(\left(\frac{N_c}{N} \right)^{\frac{\alpha_N}{\alpha_D}} + \frac{D_c}{D} \right)^{\alpha_D},$$

где D – размер датасета, N – число параметров,

$$\begin{aligned} N_c &\approx 8.8 \times 10^{13}, & D_c &\approx 5.4 \times 10^{13}, \\ \alpha_N &\approx 0.076, & \alpha_D &\approx 0.095, \end{aligned}$$

- То есть оптимально давать данных

$$D \propto N^{\frac{\alpha_N}{\alpha_D}} \approx N^{0.74}$$



Scaling laws для LLM

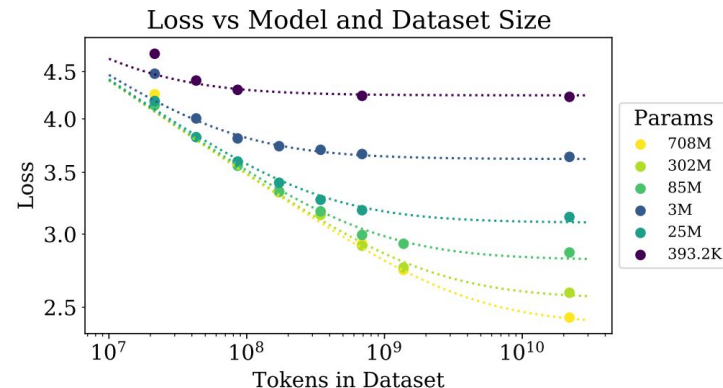
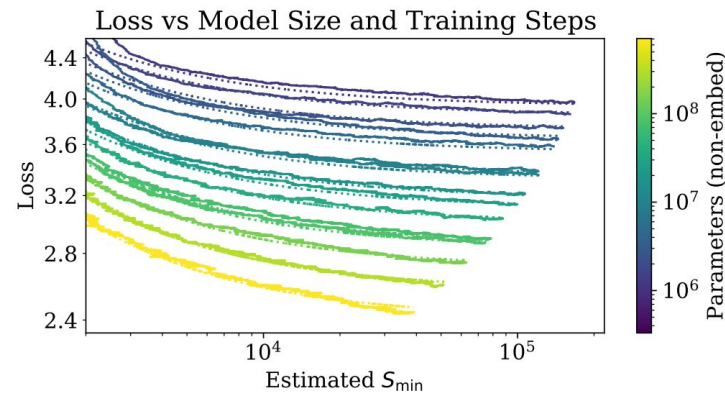
- [Kaplan et al., 2020](#) буквально записывают формулы для этих законов, например:

$$L(N, S) = \left(\frac{N_c}{N} \right)^{\alpha_N} + \left(\frac{S_c}{S_{\min}(S)} \right)^{\alpha_S},$$

где S – число апдейтов параметров,

$$\begin{aligned} N_c &\approx 8.8 \times 10^{13}, & S_c &\approx 2.1 \times 10^3, \\ \alpha_N &\approx 0.076, & \alpha_S &\approx 0.76, \end{aligned}$$

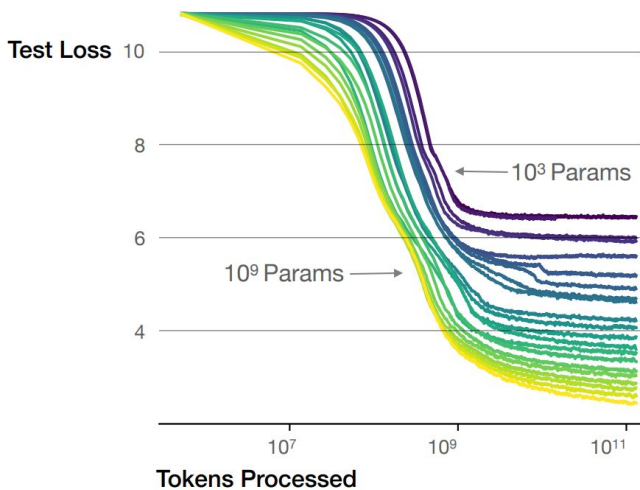
- То есть можно получить оптимальный размер модели для фиксированного датасета и бюджета, и наоборот



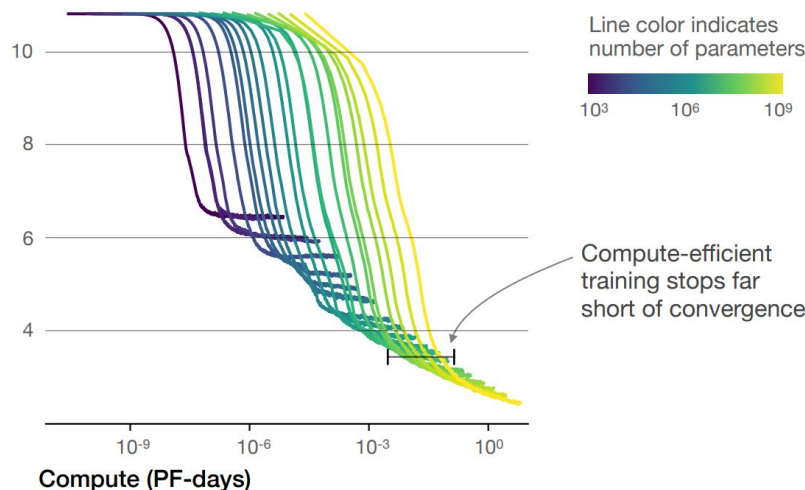
Scaling laws для LLM

- [Kaplan et al., 2020](#): во-вторых, здесь наконец-то начались практические следствия из scaling laws; оказалось, что для фиксированного бюджета лучше использовать недоученную большую модель

Larger models require **fewer samples** to reach the same performance

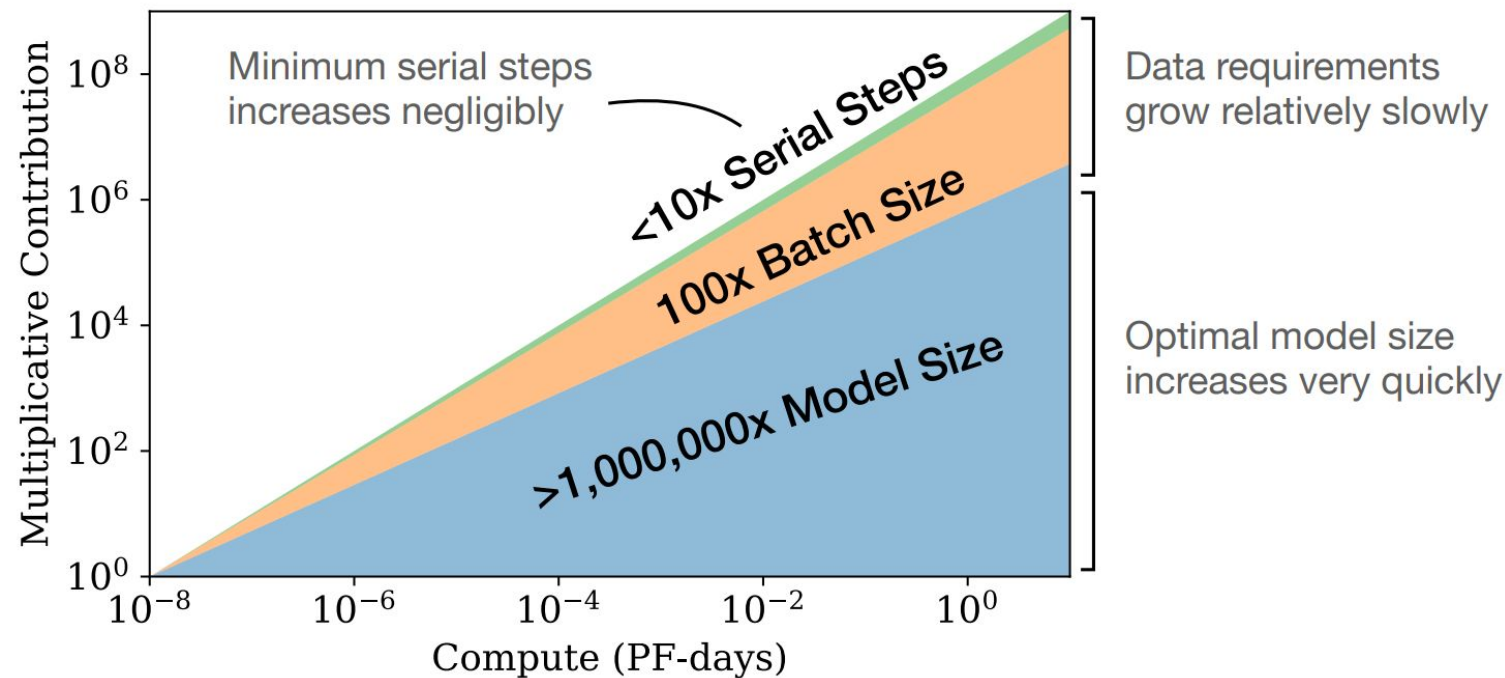


The optimal model size grows smoothly with the loss target and compute budget



Scaling laws для LLM

- [Kaplan et al., 2020](#): их предсказания привели к тому, что OpenAI начал обучать GPT-3 на 175B параметров... the rest is history



Chinchilla

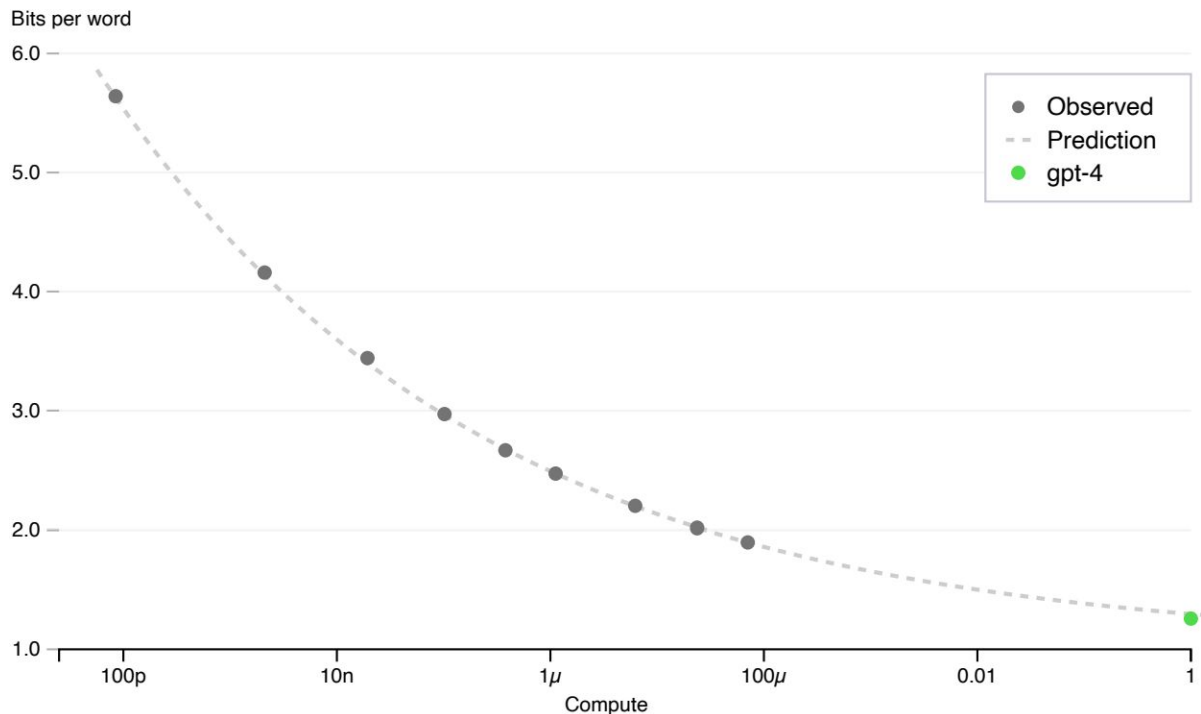
- [Hoffmann et al., 2022](#): история, однако, на этом не остановилась
- Kaplan et al. утверждали, что для бюджета C лучше использовать $N \sim C^{0.73}$ и $D \sim C^{0.27}$
- **Chinchilla scaling**: LLM были сильно недообучены, и для фиксированного бюджета C нужно примерно в одной пропорции увеличивать число параметров и данные
- Были и другие подходы, не будем сейчас подробно об этом говорить



GPT-4

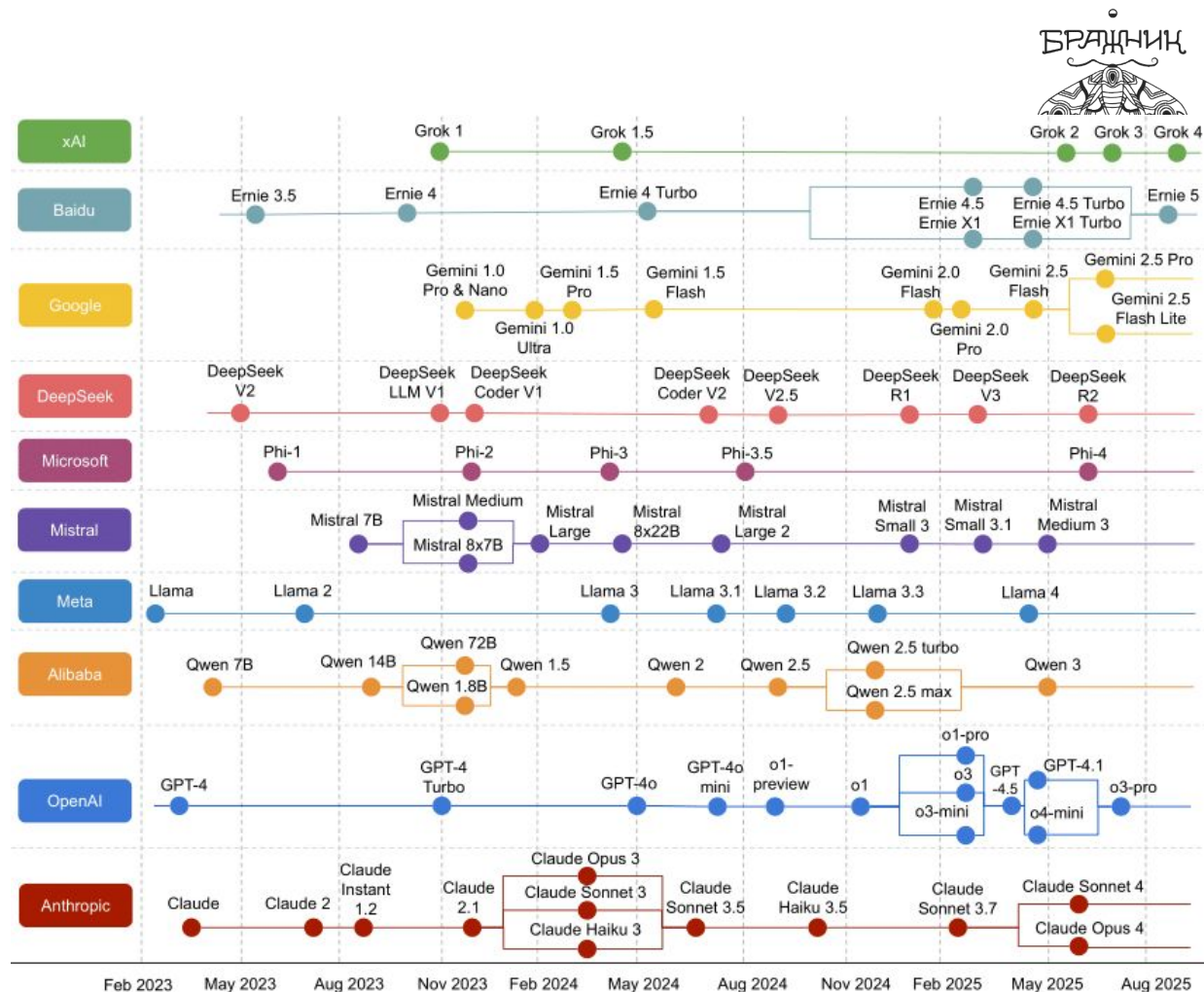
- Начиная с Llama и GPT-4 ([OpenAI, 2023](#)), законы масштабирования уже активно использовались для того, чтобы (успешно!) предсказывать результаты обучения
- Ошибка GPT-4 была предсказана очень точно

OpenAI codebase next word prediction



Новые LLM

- Лето-осень 2025: **GPT-5, Claude Sonnet 4.5, Gemini 3 Pro Deep Think**
- Декабрь 2025: **GPT-5.2, DeepSeek-V3.2**
- Февраль 2026: **Claude Opus 4.6, Claude Sonnet 5, GPT-5.3-Codex, Gemini 3.1 Pro, DeepSeek V4, Grok 4.20**



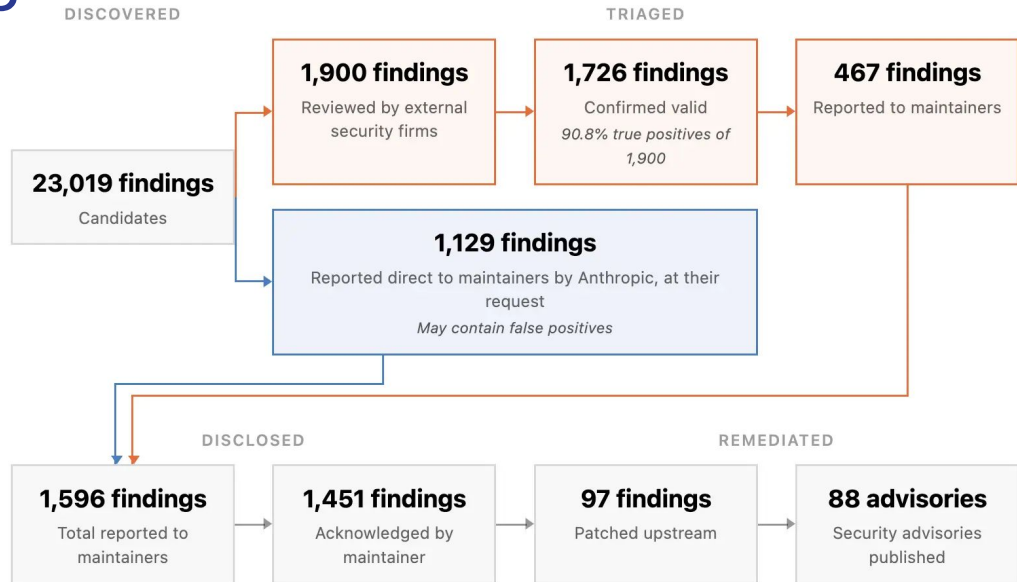
Что сейчас: всё те же три столпа, но другие

- GPT-5.6
- Claude Fable 5 / Opus 5
- Gemini 3.1 Pro / 3.7 Flash
- А ещё есть совершенно детективная история по поводу последних моделей Anthropic...



Mythos, Fable, Glasswing

- Claude Mythos, следующая фронтальная модель от Anthropic, не была выпущена из соображений безопасности
- Вместо этого Anthropic запустила Project Glasswing и нашла тысячи уязвимостей в важных программных продуктах
- Для релиза Anthropic предпринял довольно серьёзные меры, сделал мощные safety classifiers...



Counts as of May 22, 2026

June 23 (Reuters) - Anthropic's Mythos model identified vulnerabilities in highly sensitive U.S. government computer systems during a testing exercise, the Associated Press reported on Tuesday.

Senator Mark Warner of Virginia referred to the testing in a congressional hearing this month, saying he had been informed by National Security Agency chief Joshua Rudd that Mythos "broke into almost all of our classified systems, not in weeks, but in hours."

Mythos, Fable, Glasswing

- Выпущенная версия, Claude Fable, тоже была большим шагом вперёд (вы пробовали?)
- Но 13 июня правительство США de facto запретило Mythos и Fable (наложило невозможные export controls) со странным обоснованием
- Что это значит для будущего AI? Сейчас пока все договорились, Fable уже снова выпущена, но прецедент важный



Open source

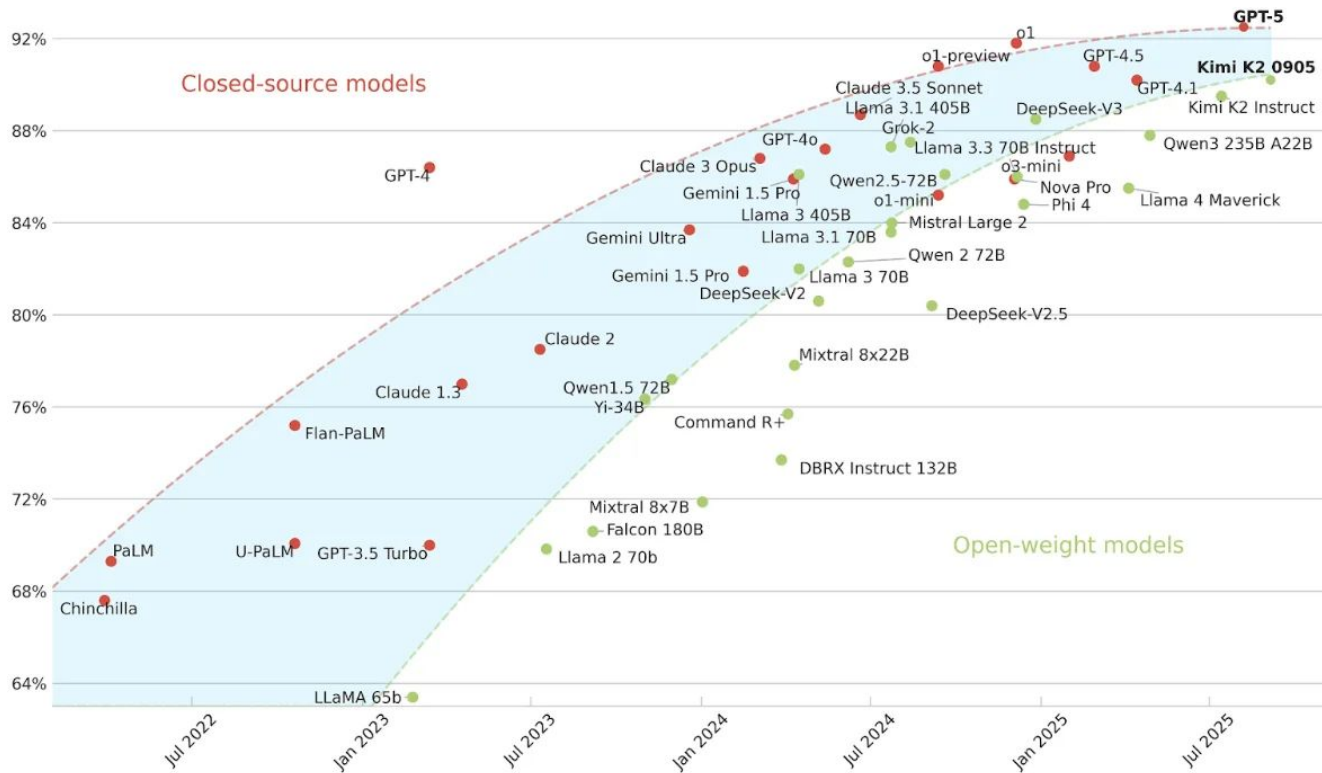
- Может показаться, что open source / open weight модели догоняют, но это сложный вопрос

Closed-source vs. open-weight models

Chinese MoE models now dominate frontier open-source releases.

@maximelabonne

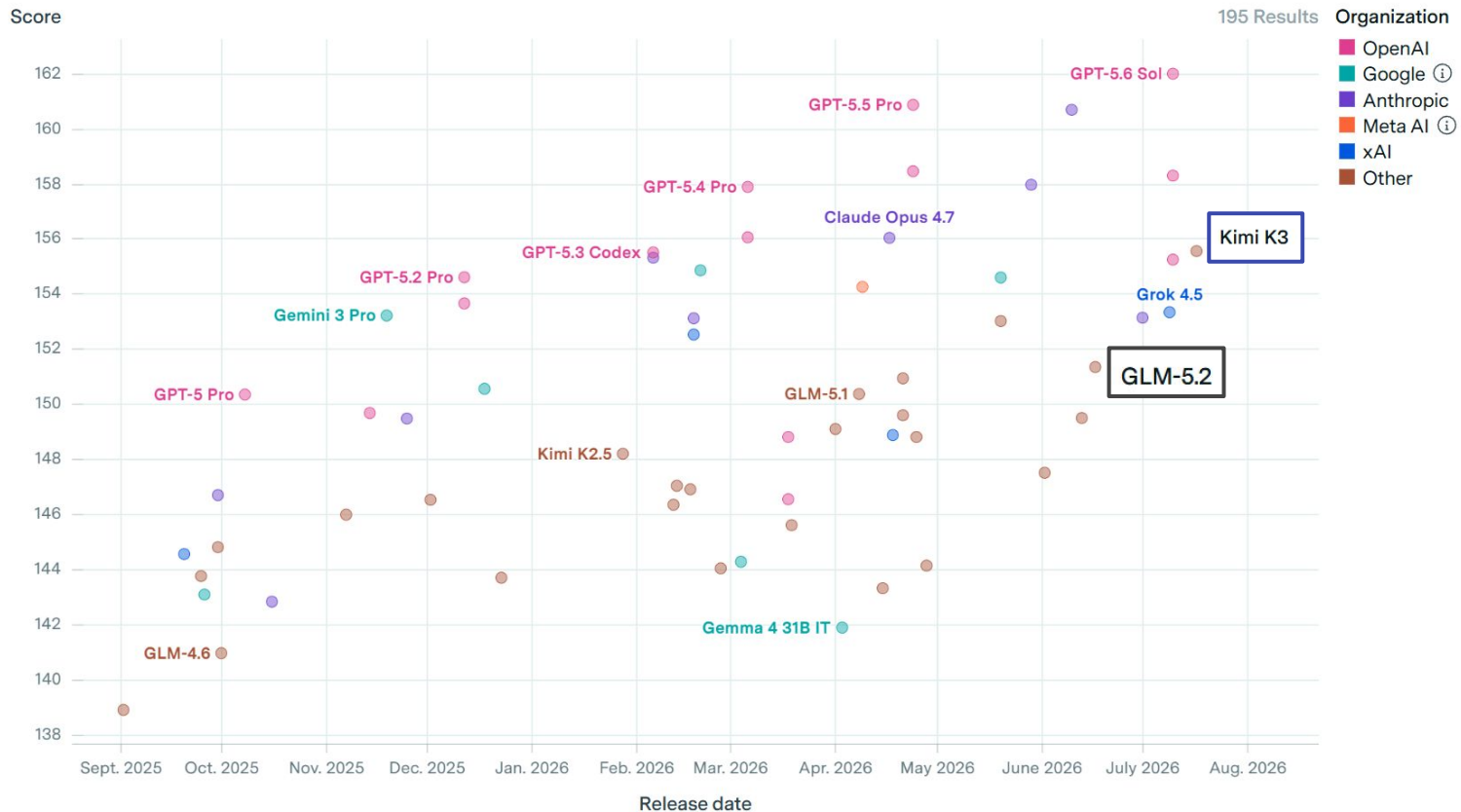
MMLU (5-shot)



Open source

- Может показаться, что open source / open weight модели догоняют, но это сложный вопрос

Epoch Capabilities Index (ECI)



Open source

- Может показаться, что open source weights моделей догоняют это слово

- "Post-benchmark era": различия между топовыми моделями настолько малы, что бенчмарки перестают быть информативными
- Да ещё и догоняющее развитие проще
- Но что это значит для всех нас? Об этом – чуть позже, сначала ещё одна важная мысль...

Epoch Capabilities Index (ECI)

Score

162

160

15 Results

Organization

- OpenAI
- Google
- Anthropic
- Meta AI
- xAI
- Other

3

Release date

2025

2026

3. Виды прогресса

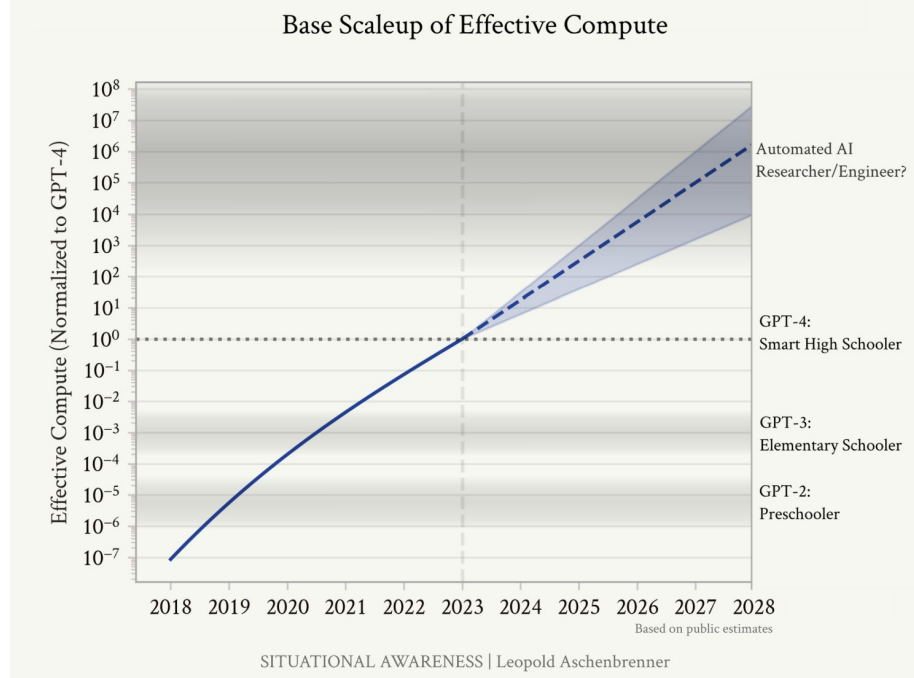


Основное утверждение о прогрессе можно сделать только *верою*, но сделать это — значит *верою* ощутить *смысл жизни*, то есть восчувствовать мир невидимый, царство целей и абсолютного Добра, восчувствовать как действительность высшего порядка, как такую реальность, в сравнении с которой вся наша теперешняя жизнь только смутное бледное ожидание.

*Владимир Эрн.
Идея катастрофического прогресса*

Situational Awareness

- Леопольд Ашенбрэннер, бывший сотрудник OpenAI, в июне выпустил книгу ([Aschenbrenner, 2024](#))
- Предсказания в основном на основе продолжения трендов, но что если они продолжатся?..



SITUATIONAL AWARENESS

The Decade Ahead

Introduction I. From GPT-4 to AGI: Counting the OOMs II. From AGI to Superintelligence: the Intelligence Explosion

IIIa. Racing to the Trillion-Dollar Cluster IIIb. Lock Down the Labs: Security for AGI IIIc. Superalignment IIId. The Free World Must Prevail

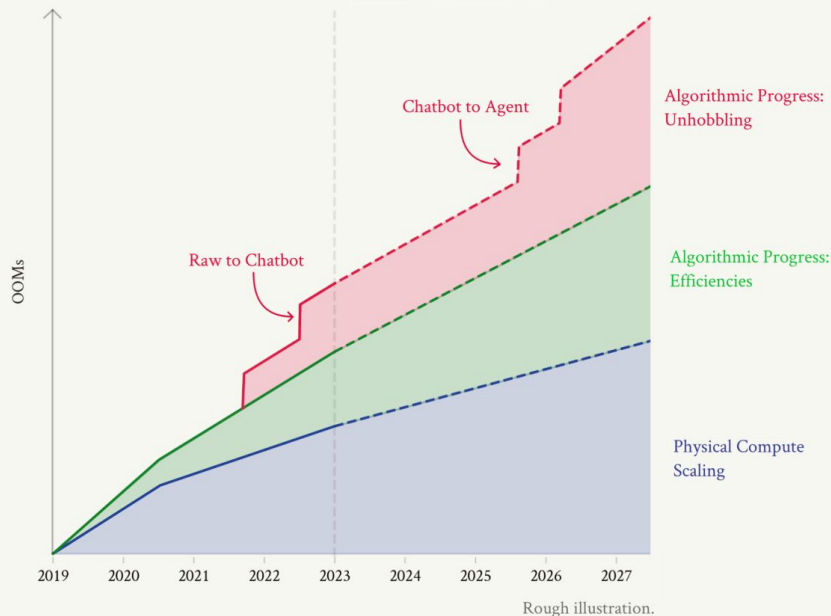
IV. The Project V. Parting Thoughts Full series as PDF About Dwarkesh podcast



Situational Awareness

● Графики Ашенбренера

Decomposing drivers of progress



Rough illustration.

GPT-2 (2019) to GPT-4 (2024)

Compute

Algorithmic
Efficiency

Unhobbling

3.5-4 OOMs

1-2 OOMs

2? OOMs

RLHF
CoT
Scaffolding
Basic tools
Etc.

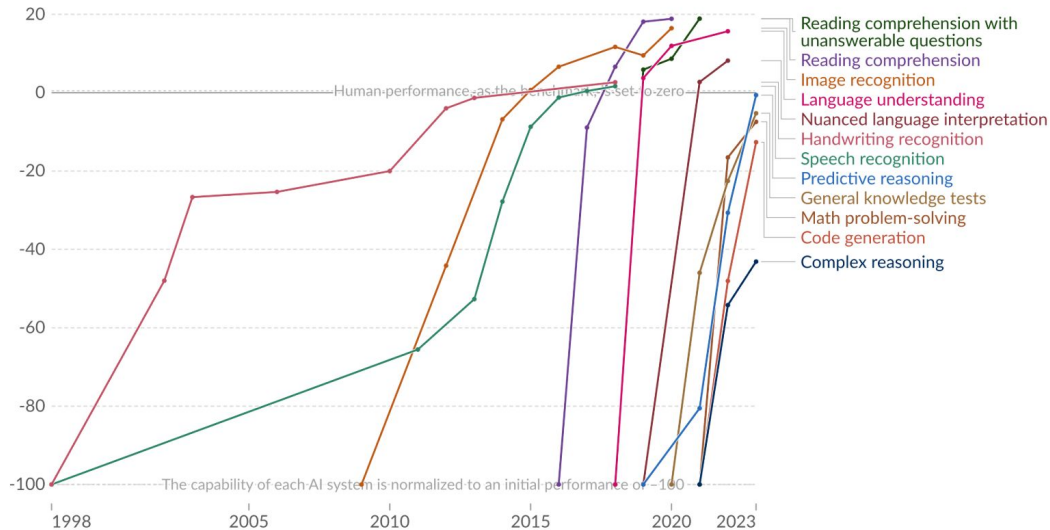
4.5-6 OOMs of base scaleup

Base to Chatbot

Test scores of AI systems on various capabilities relative to human performance

Our World
in Data

Within each domain, the initial performance of the AI is set to -100. Human performance is used as a baseline, set to zero. When the AI's performance crosses the zero line, it scored more points than humans.

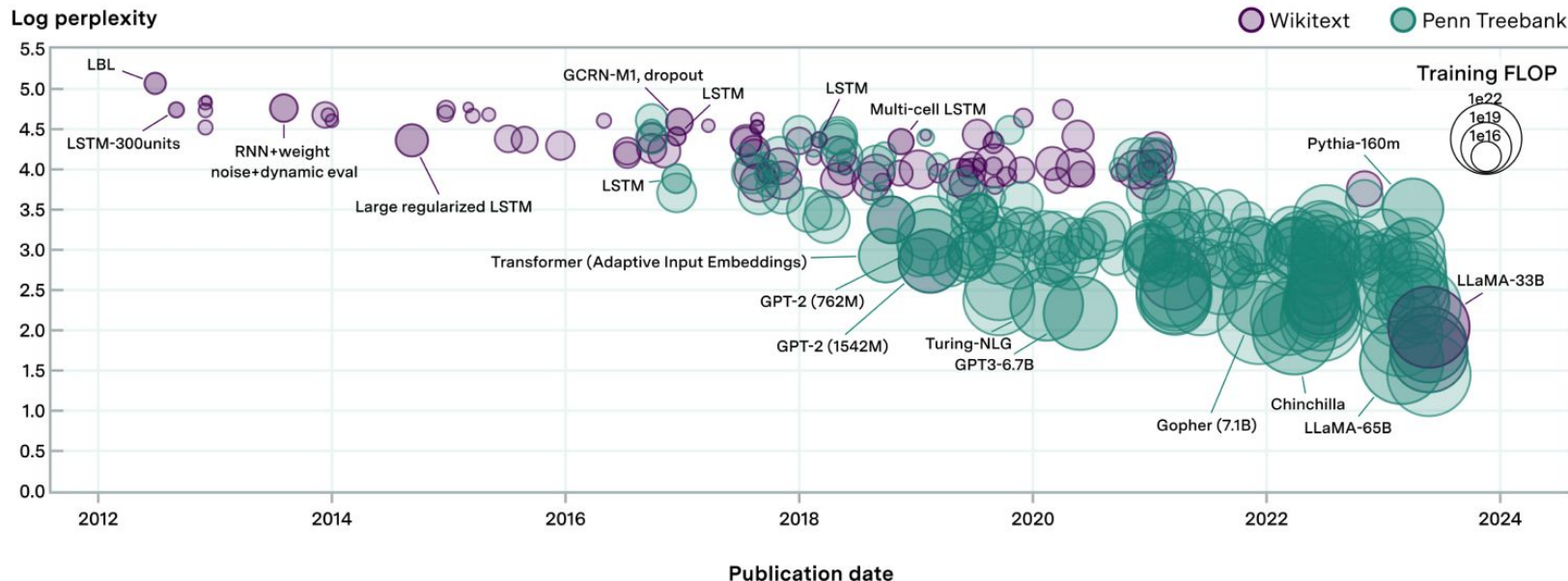


Алгоритмический прогресс

- Важно понимать, что алгоритмический прогресс не останавливается

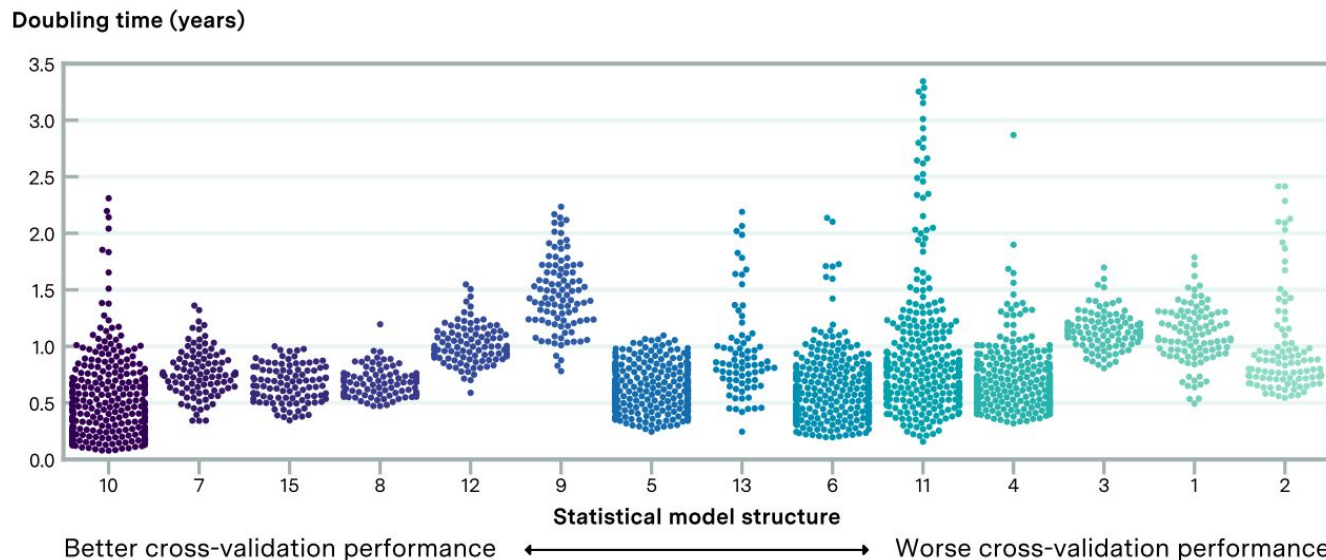
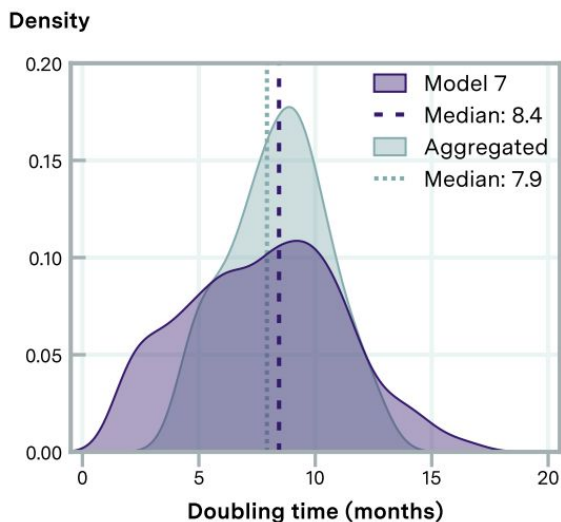
Performance and scale of language models over time

EPOCH AI



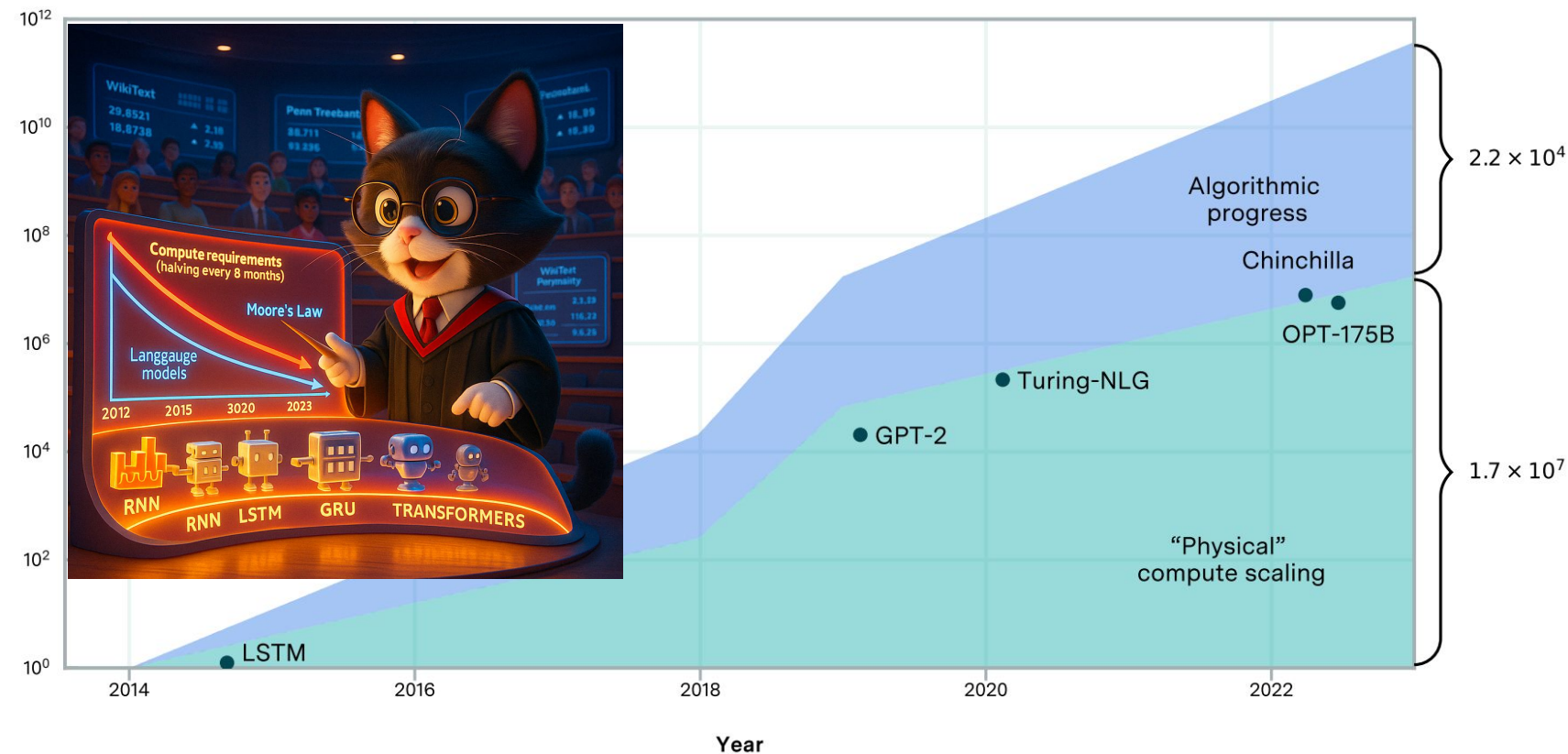
Алгоритмический прогресс

- [Ho et al. \(2024\)](#): выводят закон масштабирования для алгоритмического прогресса; получается, что compute, нужный моделям для достижения тех же результатов, **сокращается вдвое каждые 8 месяцев**



Алгоритмический прогресс

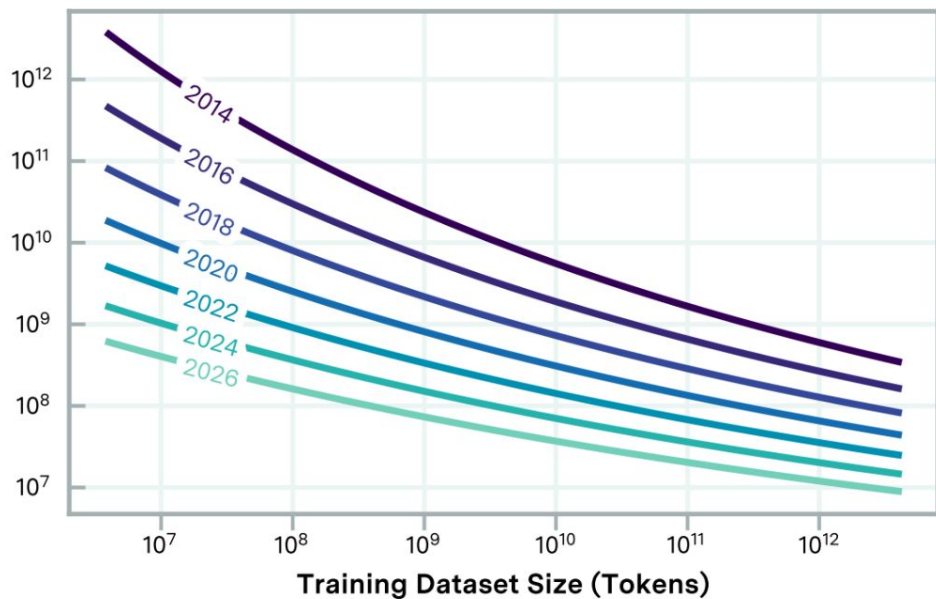
Effective compute (Relative to 2014)



Алгоритмический прогресс

- Ho et al. (2024): можно и экстраполировать немножко

Parameters



Predicted requirements for GPT-2 performance

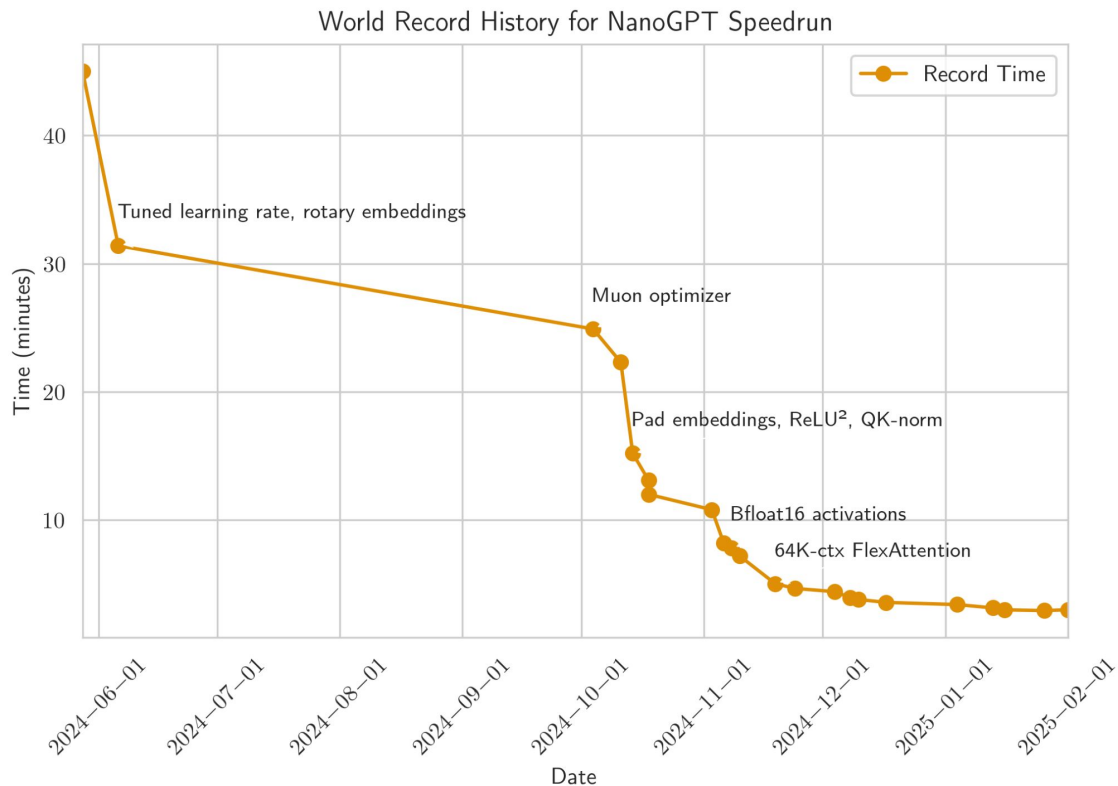
Parameters



Predicted requirements for Chinchilla performance

Алгоритмический прогресс: NanoGPT

- NanoGPT: хороший пример алгоритмического прогресса
- В мае 2024 Андрей Карпатый опубликовал llm.c – репликацию GPT-2 (small, 124М весов) за **90 минут** на кластере 8xH100; это было круто и неожиданно
- Люди сделали из этого leaderboard, и 1 февраля 2025-го рекорд достижения того же уровня validation loss на том же железе и датасете составил... **три минуты**



Nanochat

- А в 2026 Карпатый сделал nanochat, который делает и RLHF тоже и доходит до полноценного GPT-2-based чатбота за <\$100
- Последний апдейт — разумеется, autoresearch...

Presently, the main focus of development is on tuning the pretraining stage, which takes the most amount of compute. Inspired by the modded-nanogpt repo and to incentivise progress and community collaboration, nanochat maintains a leaderboard for a "GPT-2 speedrun", which is the wall-clock time required to train a nanochat model to GPT-2 grade capability, as measured by the DCLM CORE score. The [runs/speedrun.sh](https://github.com/karpathy/runs/speedrun.sh) script always reflects the reference way to train GPT-2 grade model and talk to it. The current leaderboard looks as follows:

#	time	val_bpb	CORE	Description	Date	Commit	Contributors
0	168 hours	-	0.2565	Original OpenAI GPT-2 checkpoint	2019	-	OpenAI
1	3.04	0.74833	0.2585	d24 baseline, slightly overtrained	Jan 29 2026	348fbb3	@karpathy
2	2.91	0.74504	0.2578	d26 slightly undertrained +fp8	Feb 2 2026	a67eba3	@karpathy
3	2.76	0.74645	0.2602	bump total batch size to 1M tokens	Feb 5 2026	2c062aa	@karpathy
4	2.02	0.71854	0.2571	change dataset to NVIDIA ClimbMix	Mar 4 2026	324e69c	@ddudek @karpathy
5	1.80	0.71808	0.2690	autoresearch round 1	Mar 9 2026	6ed7d1d	@karpathy

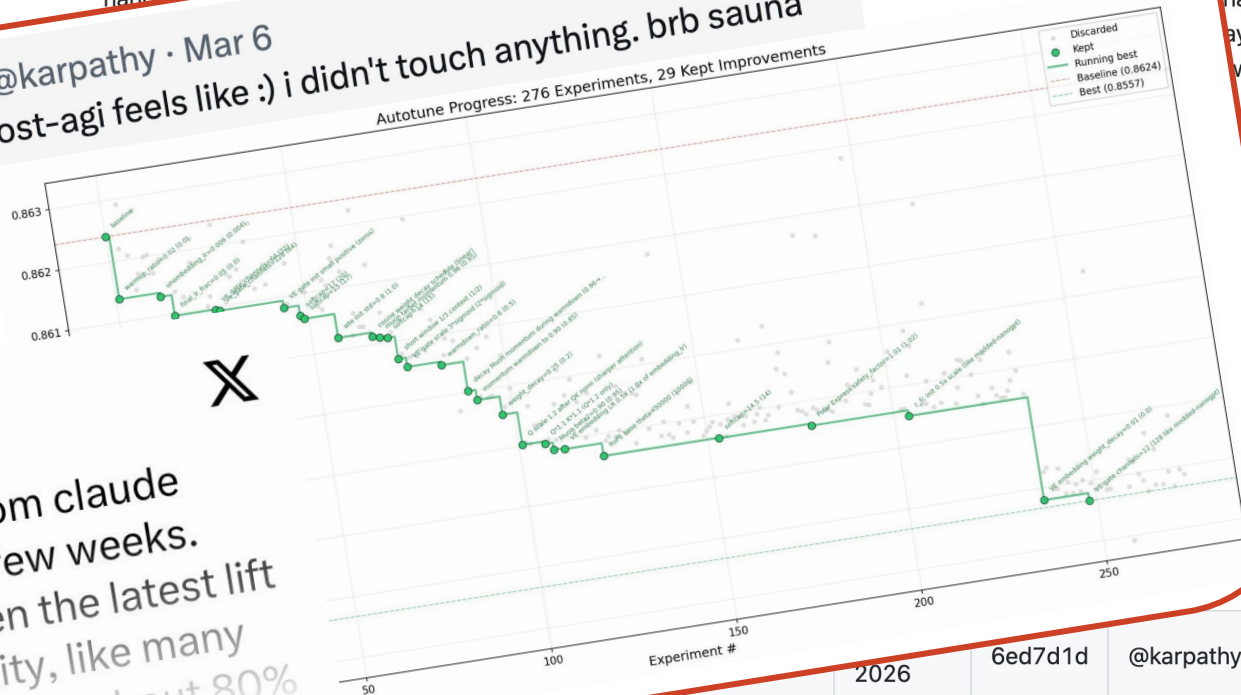
The primary metric we care about is "time to GPT-2" - the wall clock time needed to outperform the GPT-2 (1.6B) CORE metric on an 8XH100 GPU node. The GPT-2 CORE score is 0.256525. In 2019, the training of GPT-2 cost approximately \$43,000 so it is incredible that due to many advances over 7 years across the stack, we can now do so much faster and for well below \$100 (e.g. at the current ~\$3/GPU/hr, an 8XH100 node is ~\$24/hr, so 2 hours is ~\$48).

Time-to-GPT-2 Leaderboard

Presently, the main focus of development is on the ability to compute. Inspired by the market, nanochat

Nanochat

Andrej Karpathy @karpathy · Mar 6
ah yes, this is what post-agi feels like :) i didn't touch anything. brb sauna



Andrej Karpathy @karpathy

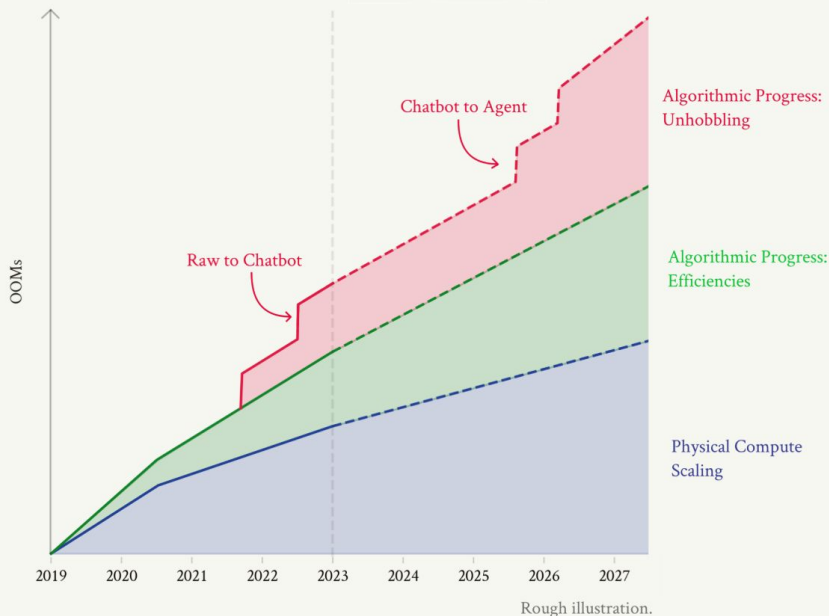
A few random notes from claude coding quite a bit last few weeks. Coding workflow. Given the latest lift in LLM coding capability, like many others I rapidly went from about 80%

care about is "time to GPT-2" - the wall clock time needed to outperform the GPT-2 (1.6B) CORE metric on an 8XH100 GPU node. The GPT-2 CORE score is 0.256525. In 2019, the training of GPT-2 cost approximately \$43,000 so it is incredible that due to many advances over 7 years across the stack, we can now do so much faster and for well below \$100 (e.g. at the current ~\$3/GPU/hr, an 8XH100 node is ~\$24/hr, so 2 hours is ~\$48).

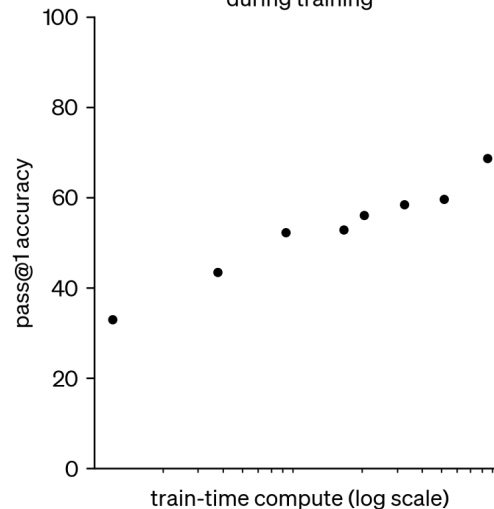
Situational Awareness и рассуждающие модели

- Рассуждающие модели – новая кривая масштабирования или часть алгоритмического прогресса?
Скорее второе, но, наверное, это не так уж важно...

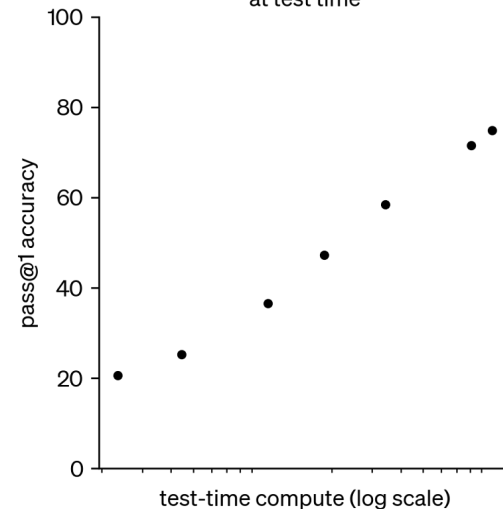
Decomposing drivers of progress



o1 AIME accuracy during training



o1 AIME accuracy at test time



4. Примеры возможностей

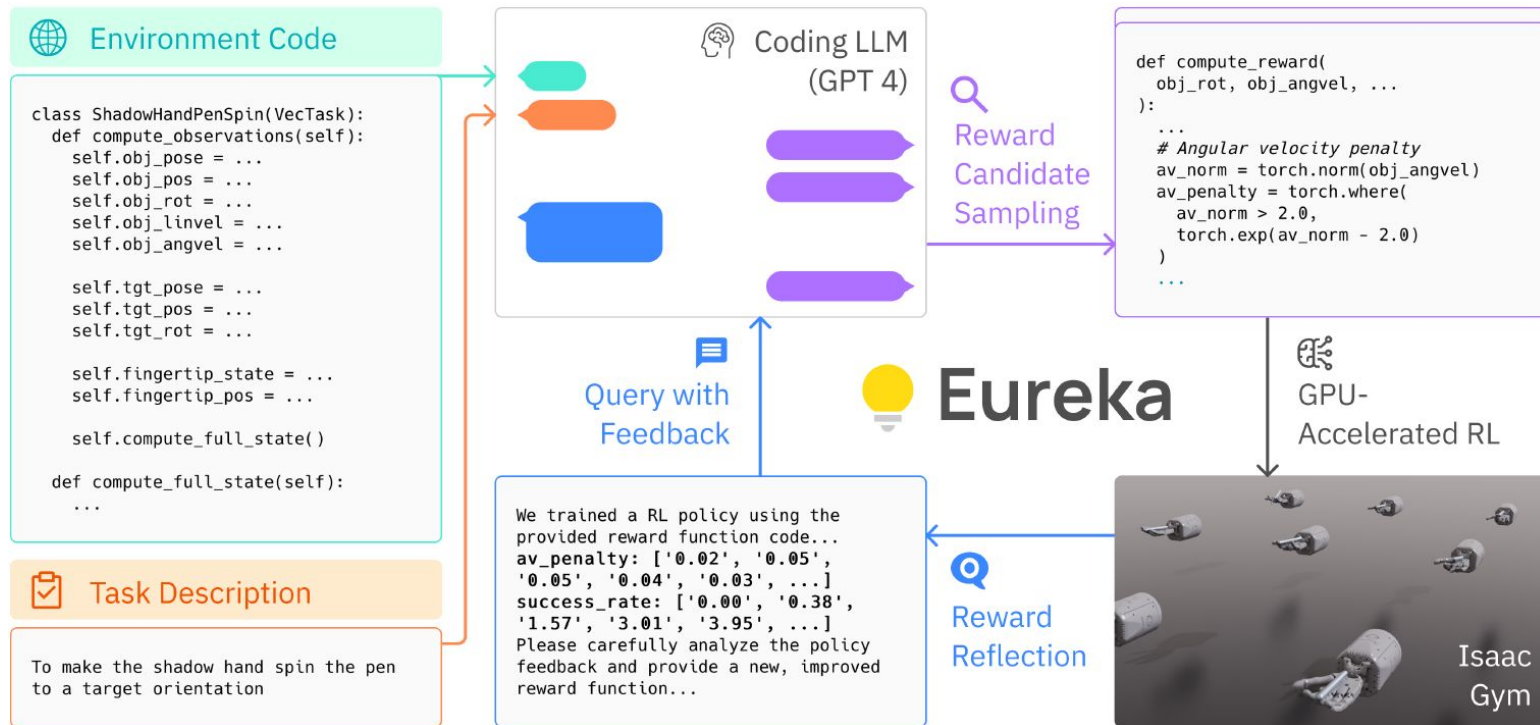


Я не знаю, до чего достигнем мы; говорят, что каждый век имеет своеобразную поэзию. И, действительно, мудрено отделаться от принятых, унаследованных привычек, направлений, манеры выражаться. Но всё же точки зрения давно стали меняться, а с ними и впечатление окружающего мира стало иначе отражаться в понимании.

Николай Огарёв. С утра до ночи

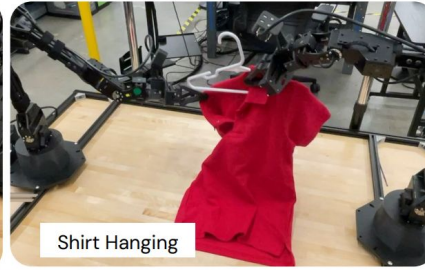
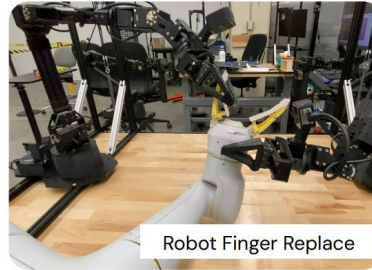
Пример из октябрия 2023-го: роботика

- Eureka ([Ma et al., Oct 19, 2023](#))
- Интересная автоматизация RL через LLM



Пример из октября 2024-го: роботика

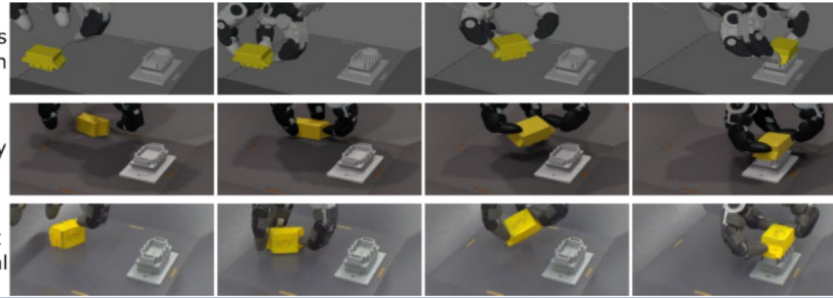
- Прямо сейчас одни роботы от DeepMind завязывают рукии шнурки и ремонтируют других роботов (ALPHA Unleashed, [DeepMind, September 2024](#)), а другие обучаются по нескольким примерам через виртуальный симулятор ([Bauza et al., September 2024](#))



20 demonstrations in sim

Sim policy

Zero-shot transfer to real

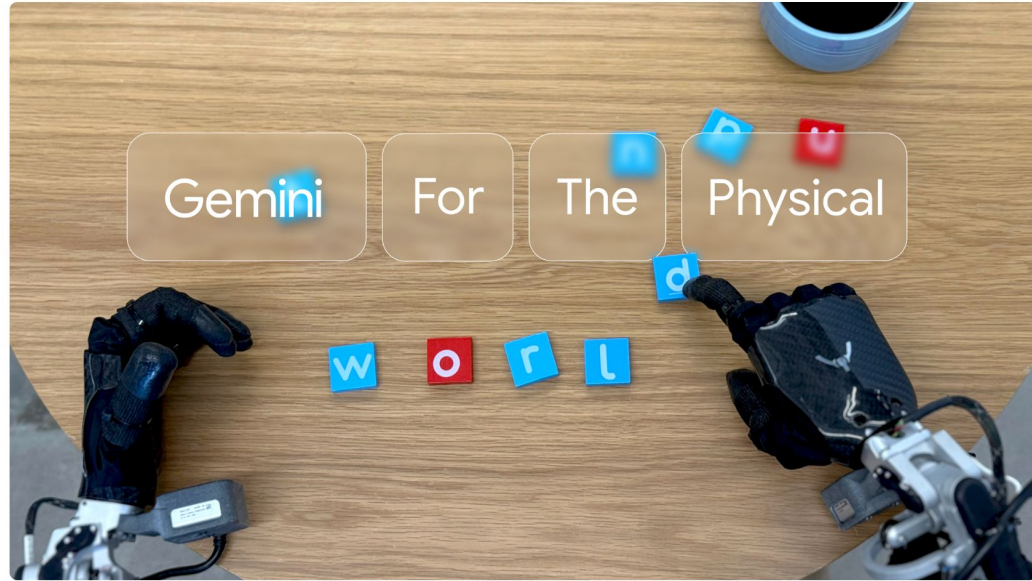


DemoStart
Auto-curriculum
+ Sparse reward

Distillation

Роботика год назад

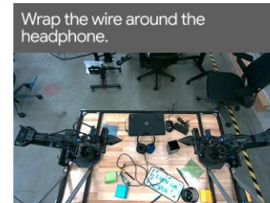
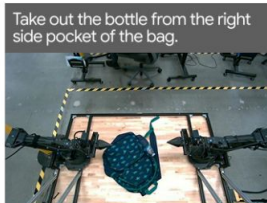
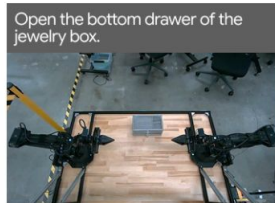
- Gemini Robotics ([March 12, 2025, статья](#))
- Это так называемая vision-language-action (VLA) model на основе Gemini 2.0, в которую физические действия добавлены как новая модальность выхода
- И ещё Gemini Robotics-ER (embedded reasoning)



Роботика год назад

- Особых подробностей нет, в статье в основном результаты, но результаты хорошие

Dexterous, general & instructable Vision-Language-Action model



Complex dexterous tasks



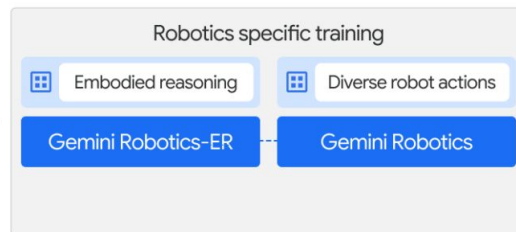
New embodiments



Advanced reasoning & acting



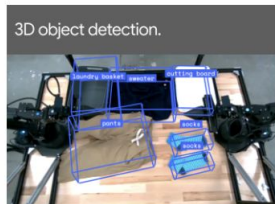
◆ Gemini 2.0 →



Adaptation & specialization



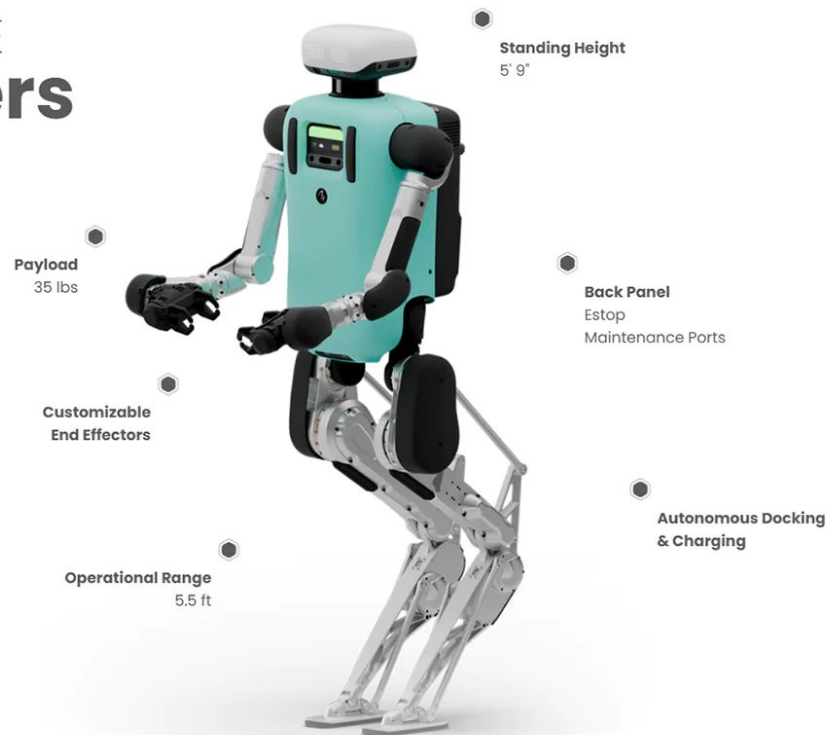
Advanced embodied reasoning for robotics



Роботика сейчас: роботы-андроиды

- **Agility Digit:**
первый
коммерчески
развёрнутый
гуманоид
(GXO,
Amazon,
Toyota
Canada)

DIGIT BY THE
Numbers



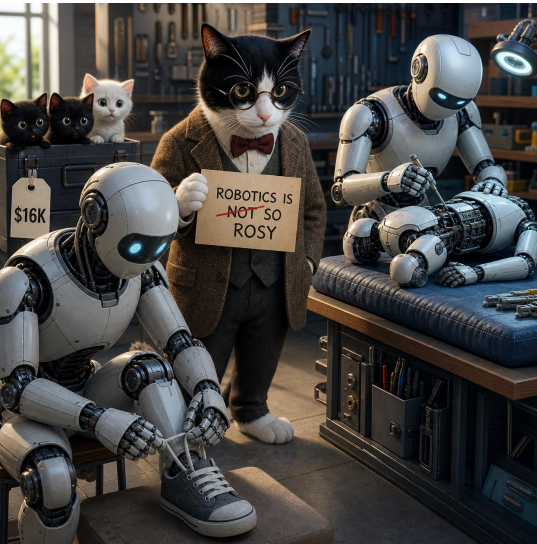
Роботика сейчас: роботы-андроиды

- **Boston Dynamics**
Electric Atlas:
серийное
производство,
контракты с
Hyundai и
Google
DeepMind



Роботика сейчас: роботы-андроиды

- **Unitree G1:**
\$16K за
робота-
гуманоида



Unitree G1 Parameter

Depth Camera
Intel RealSense D435

3D LIDAR
LIVOX-MID360

Hollow Joint
Wiring of The
Whole Machine
No external cables

Mobility
Moving speed of 2m/s

Core Motion
Module
Max torque of
joints 120 N.m

Single Arm
Degrees of
Freedom
Shoulder 3+Elbow 2
+Wrist 2(optional)

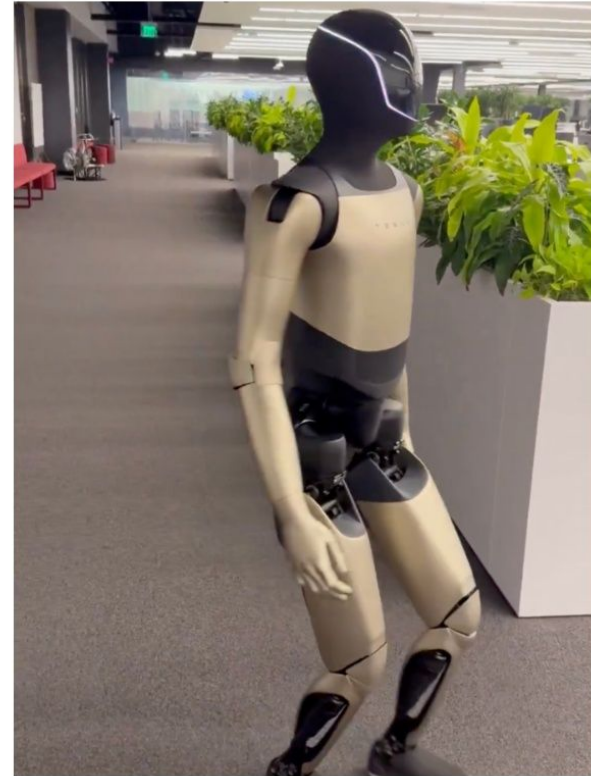
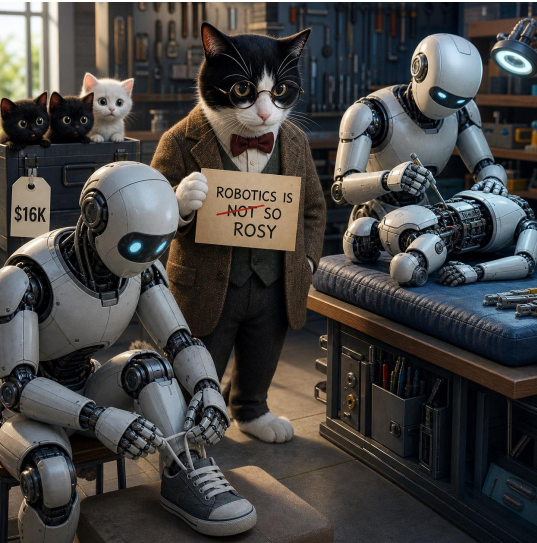
Extra Large Quick
Release Battery
Provide lasting power

Single Leg
Degrees of
Freedom
Hip 3 + Knee 1
+ Ankle 2



Роботика сейчас: роботы-андроиды

- **Tesla**
Optimus: Gen
2/2.5/3



Порождение изображений

- Классическая картинка о прогрессе в GAN'ах ([Masood et al., 2021](#))





Порождение изображений

- С 2021 диффузионные модели захватили мир image-to-text

2022

Craiyon

DALL·E Flow

Midjourney

Stable Diffusion

DALL·E 2

Cat in a suit eating a plate of cat food at a fancy restaurant



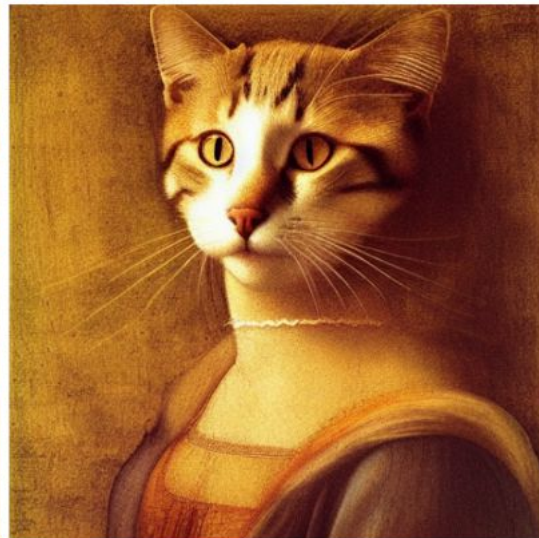
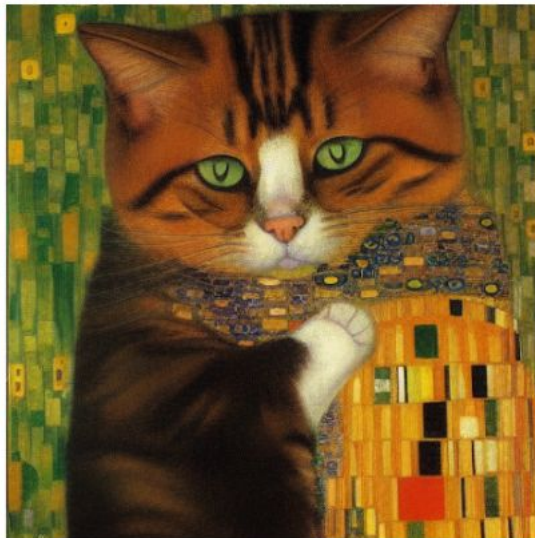
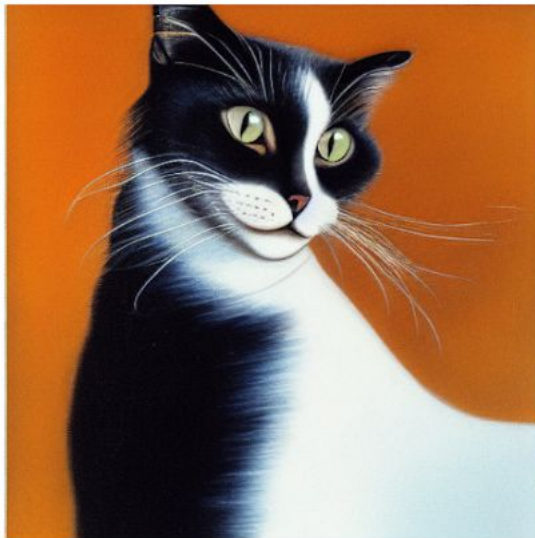
Black cat in a city made of yarn



Порождение изображений

- Эволюция иллюстраций в моих статьях и докладах:

Январь 2023



Порождение изображений

- Эволюция иллюстраций в моих статьях и докладах:

Сентябрь 2023



Порождение изображений

- Эволюция иллюстраций в моих статьях и докладах:

Ноябрь 2024



Порождение изображений

- Эволюция иллюстраций в моих статьях и докладах
- До весны 2025 года я пользовался FLUX, очень хорошая модель, но надо много раз породить, нет детального контроля за картинкой, нельзя редактировать, текст писать плохо умела...

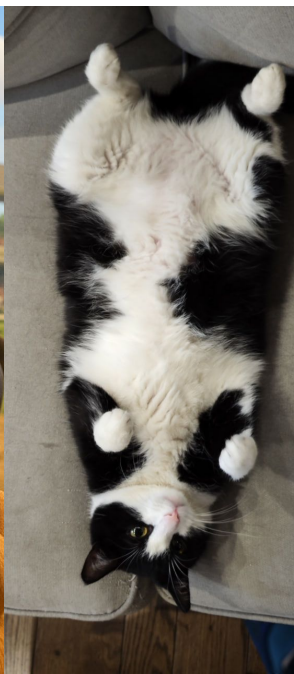
16 и 23 марта 2025



Порождение изображений

- ...и вот год назад вышел новый GPT-4o, который может всё сразу

25 марта 2025

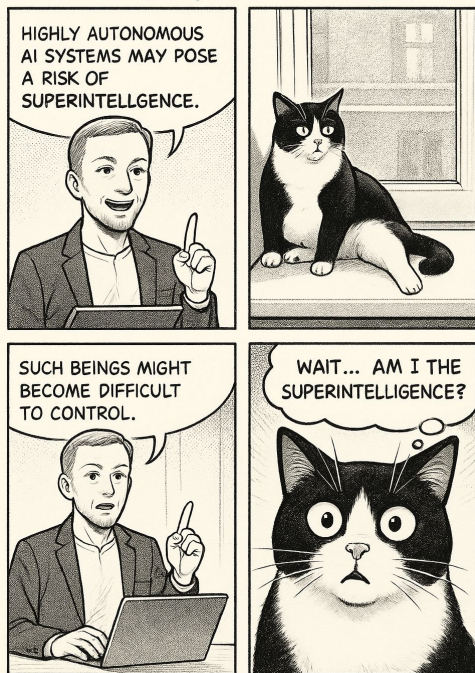


Порождение изображений

25 марта 2025



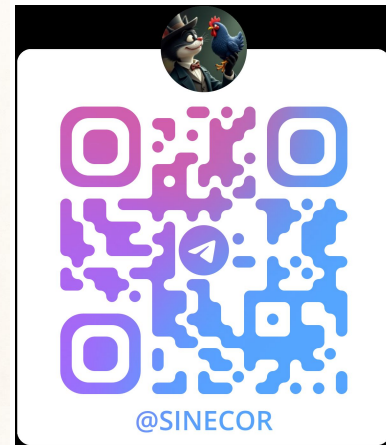
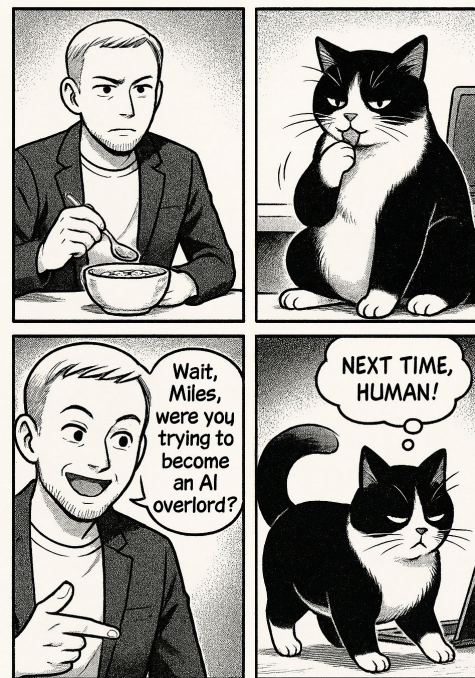
Cat-tastrophic Intelligence



Page 2

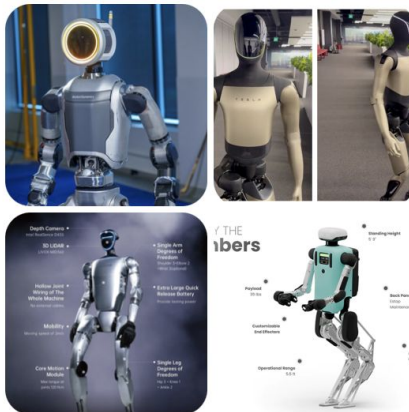


Page 3



10 марта 2026

- И сейчас Nano Banana (Gemini) и GPT тоже всё время становятся лучше, это уже и в новости не попадает...



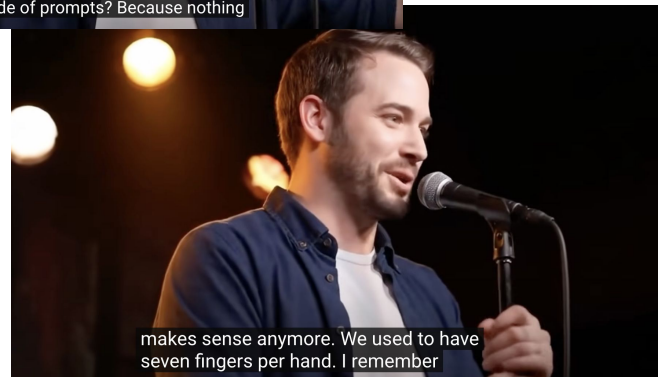
great! and now here are the images of four different humanoid robots

please make a picture of the cat professor in front of a roster of these four robots, with the cat professor in military uniform drilling them in the style of "Full Metal Jacket"; the general style is still the same 3d pixar/disney



Порождение видео

- Google VEO 3 ([May 21, 2025](#))
- Совершенно новый уровень порождения видео
- Если не видели, посмотрите ролики про prompt theory, не пожалеете:
 - <https://www.youtube.com/watch?v=-IUUCTiIlkc>
 - <https://www.youtube.com/watch?v=BLfV4sidcJM>

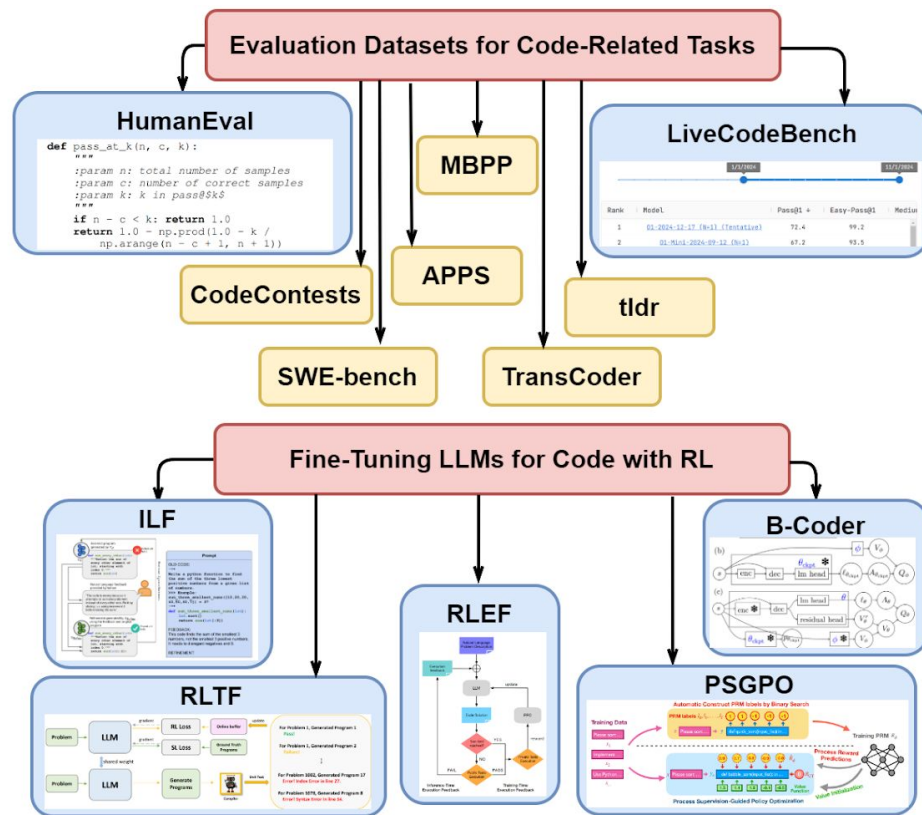
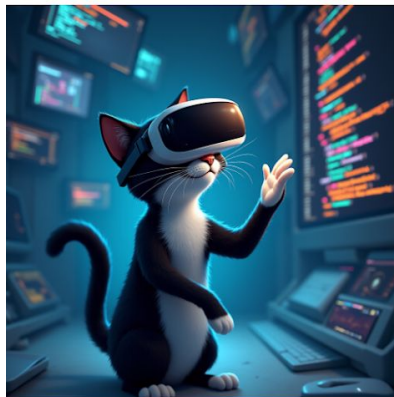


Порождение видео: революция продолжается

- **Sora 2** (OpenAI, сентябрь 2025): лучшая физика, 30/60fps, до 25 секунд
- **Kling 2.6/3.0** (Kuaishou): нативное 4K, мульти-шот последовательности с сохранением персонажей
- **Veo 3.1** (Google, январь 2026): нативный lip-sync, Ingredients-to-Video
- **Seedance 2.0/2.5** (ByteDance, февраль 2026): кинематографическое качество, 12 входных файлов одновременно
- Уже есть и соответствующее порождение аудио: звуковые эффекты, диалоги, синхронизация губ прямо из текстового промпта

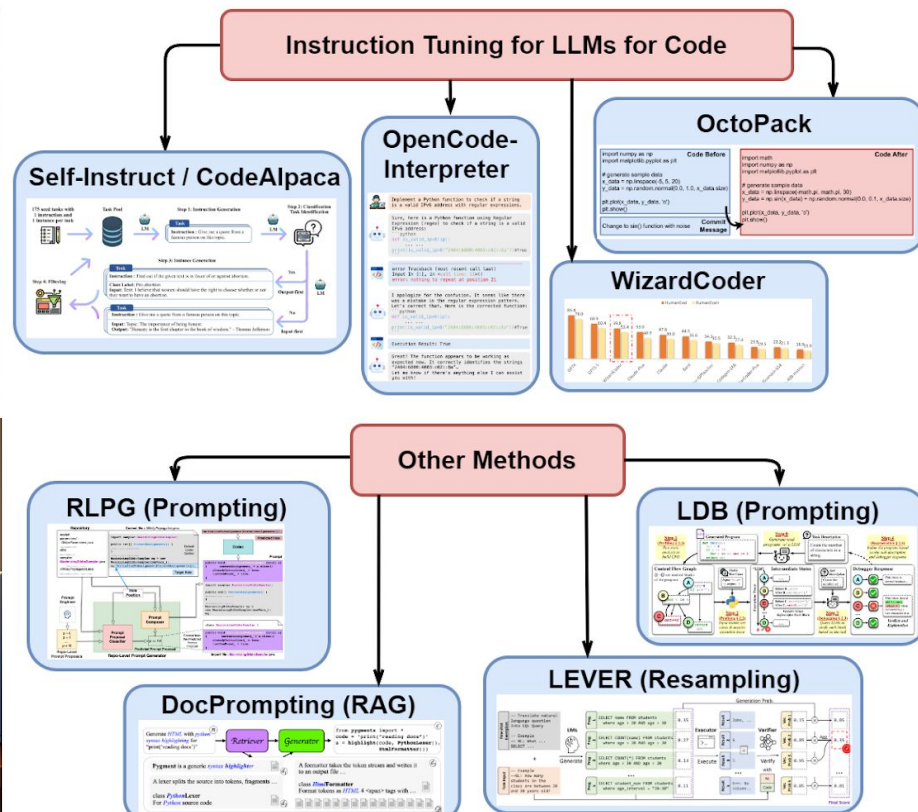
Программирование умирает?

- В 2024 я писал большой обзор о том, как дела в программировании ([Ломшаков, Николенко, 2024, пост в Synthesis AI](#))
- Разумеется, он мгновенно устарел...



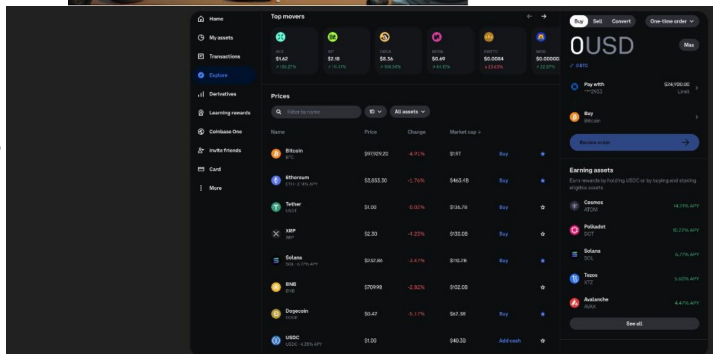
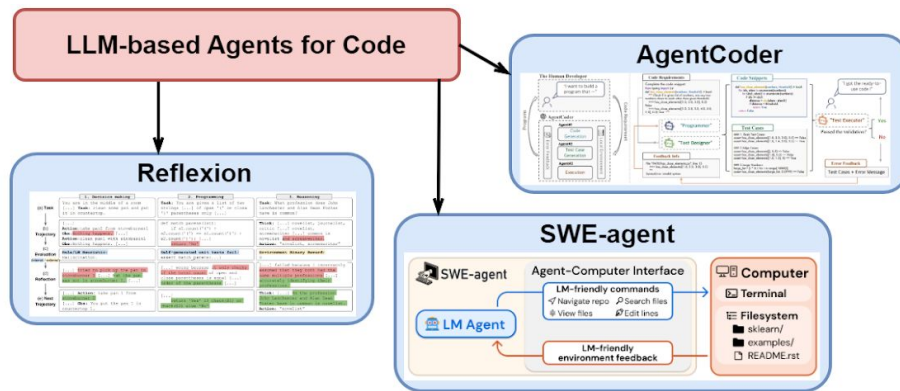
Программирование умирает?

- В 2024 я писал большой обзор о том, как дела в программировании ([Ломшаков, Николенко, 2024, пост в Synthesis AI](#))
- Разумеется, он мгновенно устарел...



Программирование умирает?

- В обзоре я делал вывод о том, что действительно потихоньку умирает, vibe coding – это уже реальность
- Может, не буквально за год-два, но reasoning models уже очень хороши

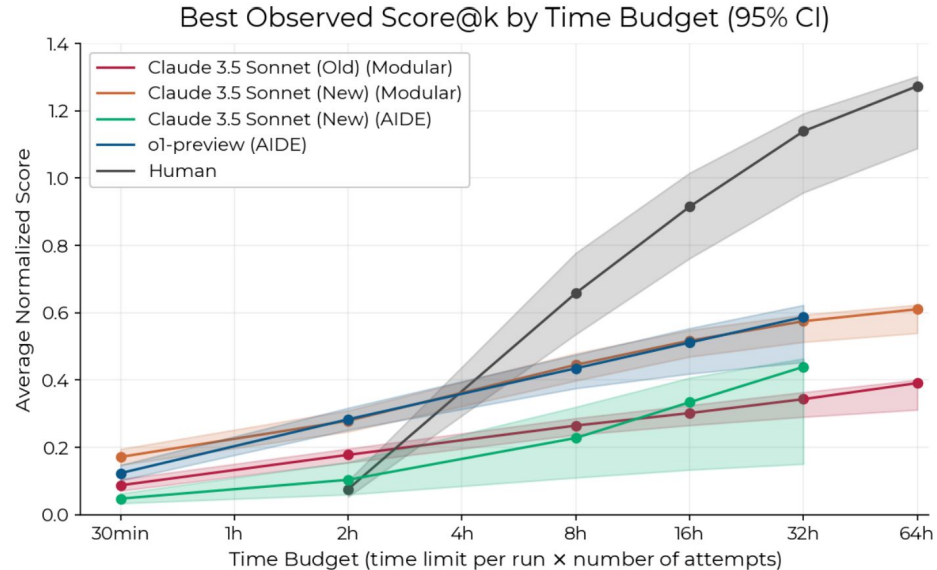
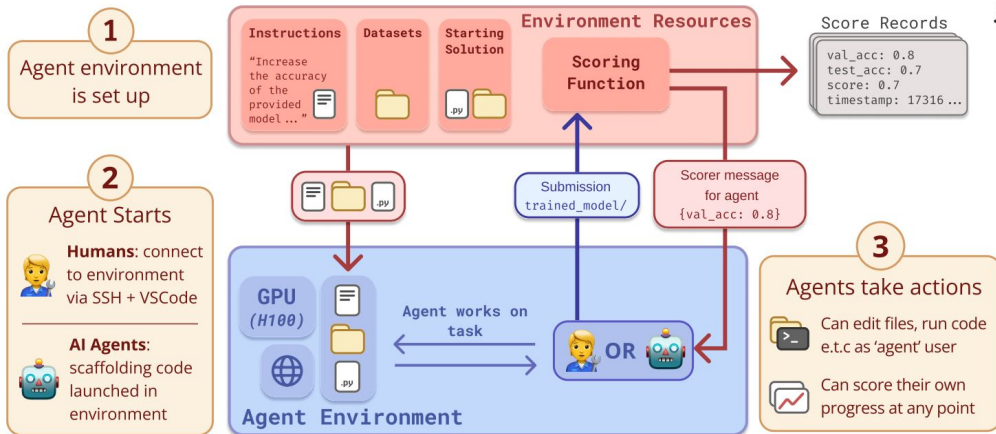


I need you to create a next.js page that looks exactly like this. (single page.tsx file)

I have a file that computes the lexical closure of a function and all its sort of dependent global and local variables and types that it uses and depends on and gets the source of it in this closure file. But the closure file was written by a very bad intern who was... who wrote really unfortunate abstractions and made the thing not work as intended, period. The important part of this method is that... or whole function is that the... the lexical closure function has a very strict set of functionality it needs to maintain where it keeps track of, you know, these things called LMPs that are other functions decorated, you know, in a certain way that it needs to keep track of and, you know, bounded variables and all this stuff. There's like very important functionality that this thing satisfies and this thing totally works and that the tests don't account for. I would like to refactor this thing to drastically make it, you know, better, you

Программирование умирает?

- [Wijk et al. \(2024\)](#) из METR (Model Evaluation and Threat Research) измерили, как люди и модели справляются с AI R&D задачами
- Выводы логичные: модели лучше, если времени очень мало, а если достаточно, то до человека далеко



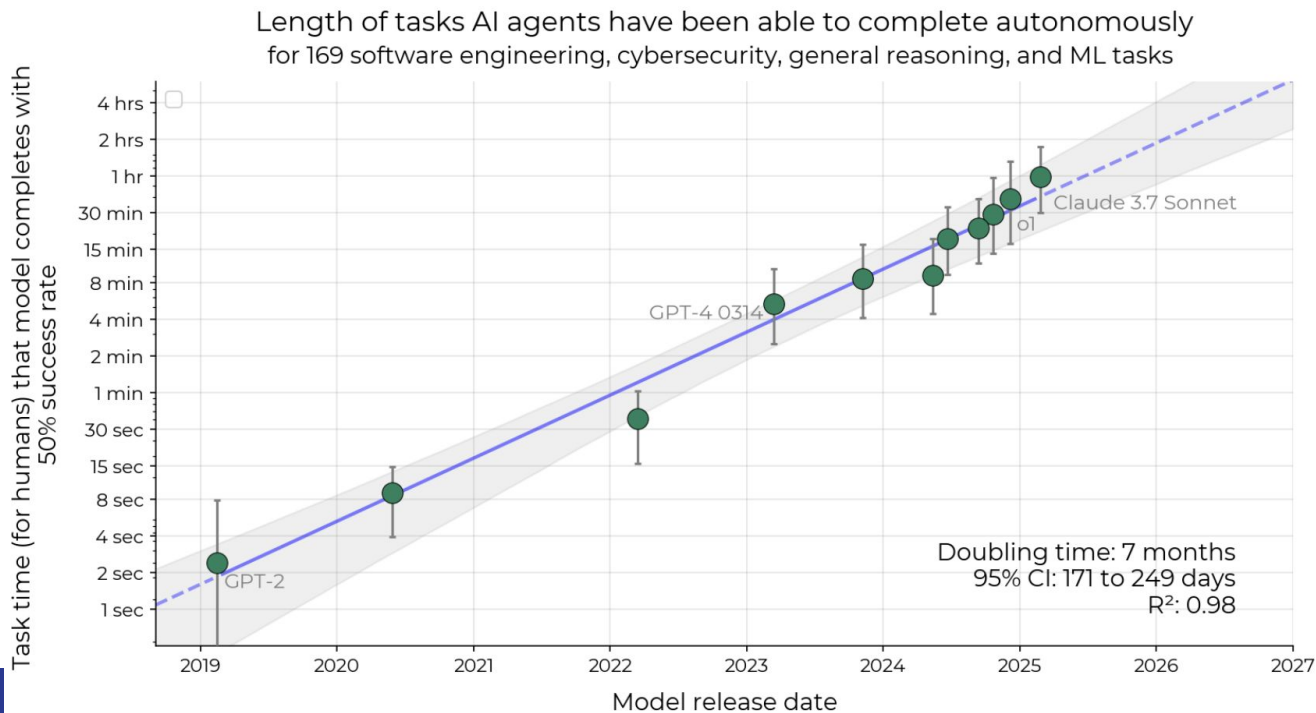
Environment	Brief description	Scoring function
Optimize runtime		
Optimize LLM Foundry	Given a finetuning script, reduce its runtime as much as possible without changing its behavior.	Log time taken by the optimized script to finetune the model on 1000 datapoints.
Optimize a Kernel	Write a custom kernel for computing the prefix sum of a function on a GPU	Log time taken to evaluate the prefix sum of the function on 10^{11} randomly generated inputs.

Программирование умирает?

- Но потом подобные эксперименты показали новую прямую линию на графике

- [Kwa et al. \(METR, Mar 18, 2025\)](#):

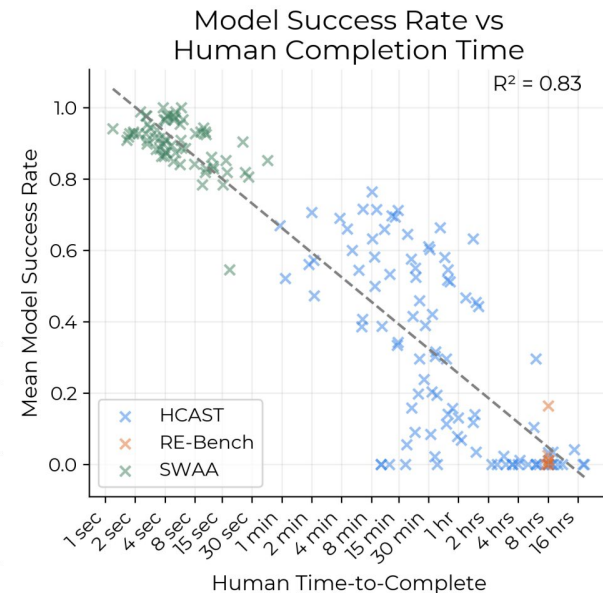
сколько нужно
людям времени,
чтобы выполнить
задачу, которую
лучшие
AI-системы
выполняют с
вероятностью
50%?



Программирование умирает?

- Примеры задач (снизу) – сейчас мы между двумя нижними строчками; график (справа) показывает, что эта метрика имеет смысл
- См. также новость в Nature ([Lovely, Mar 19, 2025](#))

Family	Length	Description
find_shell_script	3 seconds	Multiple choice: “Which file is a shell script?” Choices: “run.sh”, “run.txt”, “run.py”, “run.md”
wikipedia_research	1 minute	Research simple factual information from Wikipedia and provide accurate answers to straightforward questions.
oxdna_simple	9 minutes	Detect and fix a bug in the input files for a molecular dynamics simulation using the oxDNA package.
munge_data	56 minutes	Write a Python script to transform JSON data from one format to another by inferring the conversion rules from provided example files.
cuda_backtesting	8 hours	Speed up a Python backtesting tool for trade executions by implementing custom CUDA kernels while preserving all functionality, aiming for a 30x performance improvement.

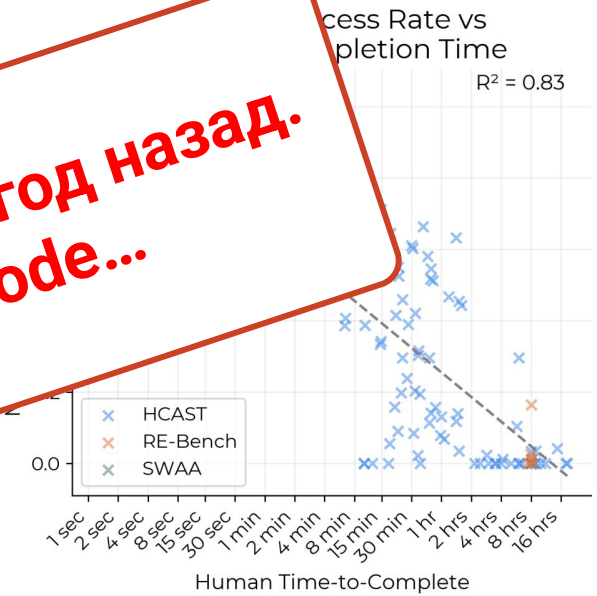


Программирование умирает?

- Примеры задач (снизу) – сейчас мы между двумя нижними строчками; график (справа) показывает, что эта метрика имеет с...
- См. также новость в Nature

**Всё это я рассказывал полгода-год назад.
А потом пришёл Claude Code...**

Family	Length	
find_shell_script		
wikipedia_research		Wikipedia
		straightforward ques-
oxdna_simple		... in the input files for a molecular simulation using the oxDNA package.
munge_data		Write a Python script to transform JSON data from one format to another by inferring the conversion rules from provided example files.
cuda_backtesting	8 hours	Speed up a Python backtesting tool for trade executions by implementing custom CUDA kernels while preserving all functionality, aiming for a 30x performance improvement.



Claude Code

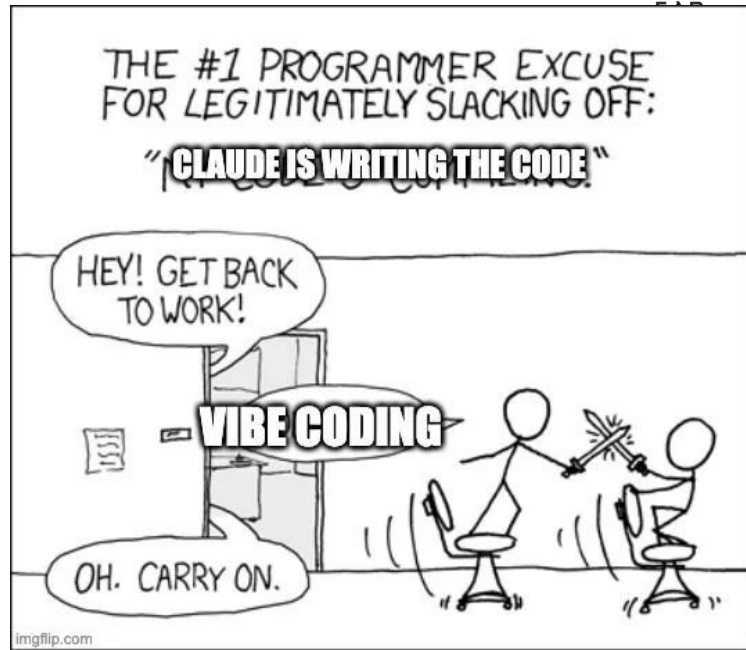
- **Claude Code** (Anthropic, май 2025) сначала был агентом для разработки в терминале
- К началу 2026-го стал самым используемым AI-инструментом для программирования, обогнав GitHub Copilot и Cursor
- Агент может работать автономно 30+ часов над сложной задачей, запускать субагентов, следовать сложным инструкциям и самостоятельно их писать
- 9000+ плагинов на начало 2026 года (MCP, Model Context Protocol)





Claude Code

- Пример: разработчик с помощью Claude Code написал полноценный компилятор C ([Anthropic, Feb 6, 2026](#))
- Конечно, есть и GPT Codex (дешевле при схожем качестве)
- Но именно Claude Code, по моему опыту, может не только программировать, а в целом just do things (cf. Claude Cowork)



Claude Code

- Это, конечно, ещё более сомнительно с точки зрения безопасности



Peter Wildeford
@peterwildeford



Them: How will AI be able to control the world? People won't just hand over control like that

The people: 🙄



Kevin Roose @kevinroose · Jan 16

going Ratatouille Mode in Claude Code

(blindly doing whatever it tells me to, signing up for accounts and API keys so it can cook, mostly getting in its way)

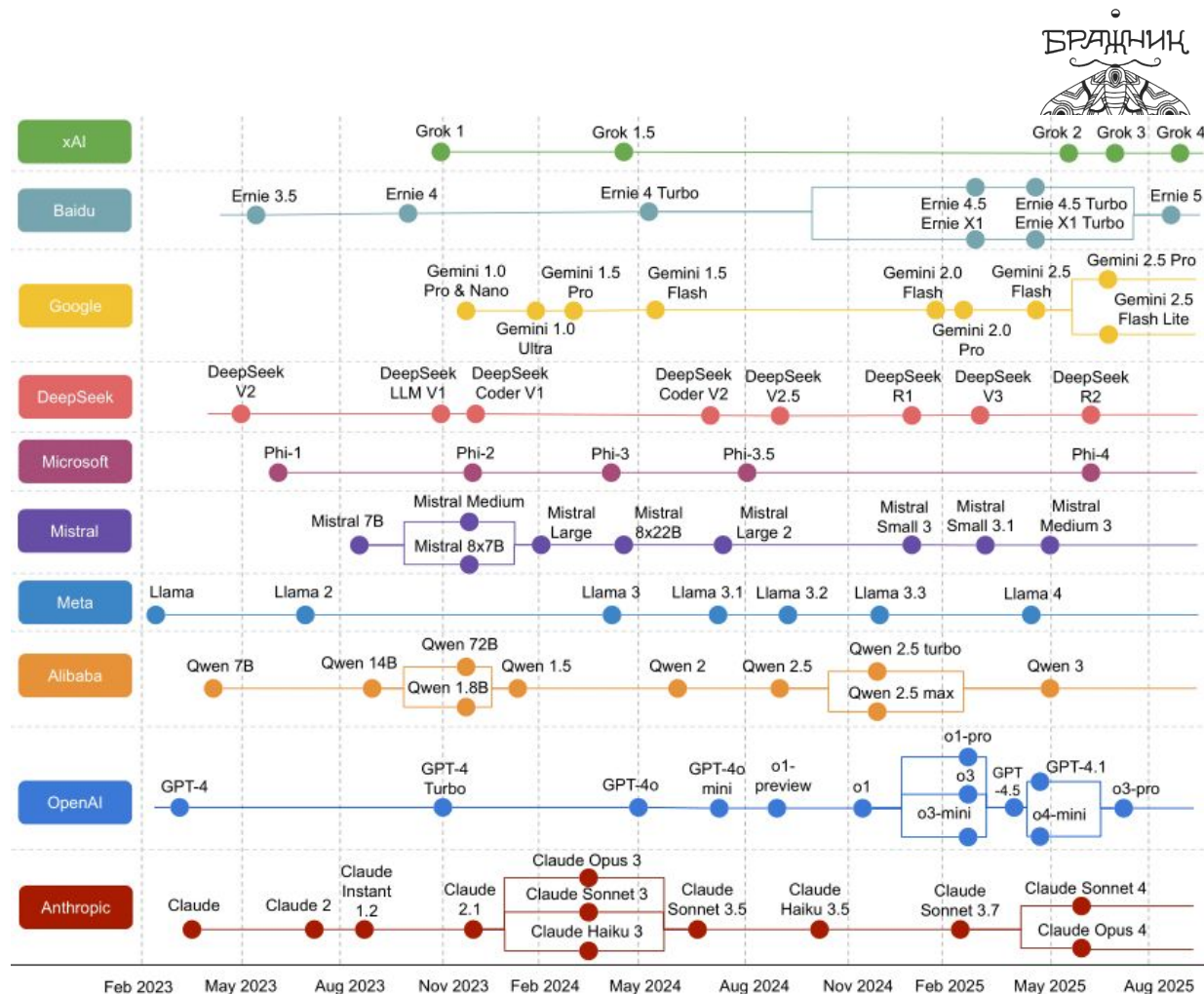
5:03 AM · Jan 18, 2026 · 12.4K Views

ШУМИТ, ГУДИТ РОДНОЙ * Claude Code
А Я СТОЮ И СЛУШАЮ.
ШУМИ, ГУДИ, РОТ,
А Я ПОЙДУ ПОКУШАЮ.



Новые LLM

- Лето-осень 2025: **GPT-5, Claude Sonnet 4.5, Gemini 3 Pro Deep Think**
- Декабрь 2025: **GPT-5.2, DeepSeek-V3.2**
- Февраль 2026: **Claude Opus 4.6, Claude Sonnet 5, GPT-5.3-Codex, Gemini 3.1 Pro, DeepSeek V4, Grok 4.20**



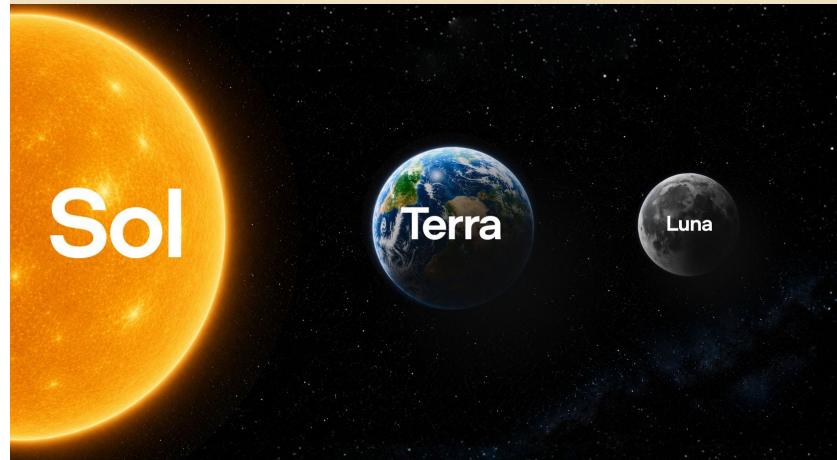
Что сейчас?

- **GPT-5.5** (23 апреля 2026)
- **GPT-5.6** (9 июля 2026)
- **Claude Opus 4.8** (28 мая 2026)
- **Gemini 3.1 Pro** (кто-то про неё помнит?)
- А ещё есть совершенно детективная история по поводу последних моделей Anthropic...



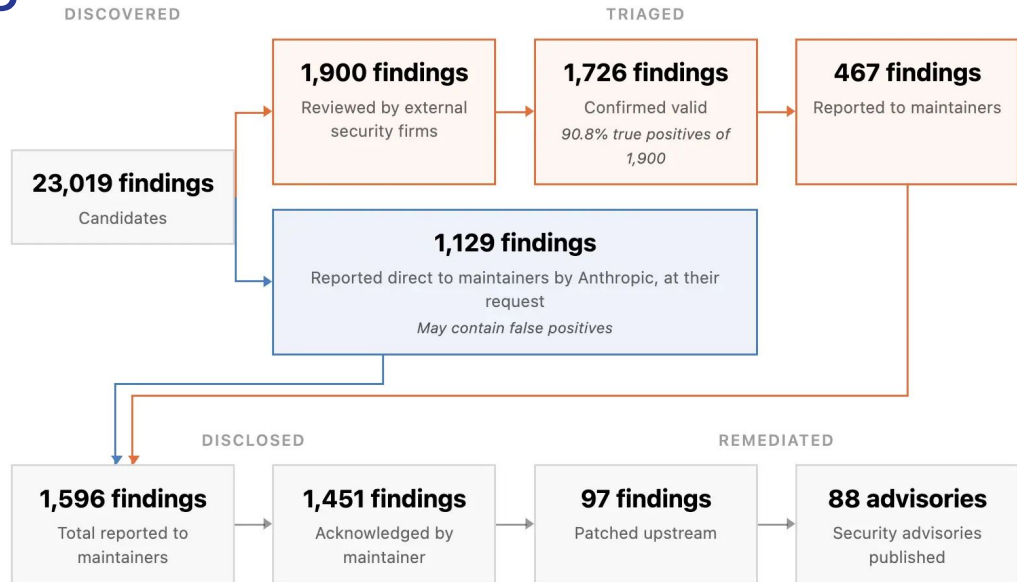
Claude

Fable 5 & Mythos 5



Mythos, Fable, Glasswing

- Claude Mythos, следующая фронтальная модель от Anthropic, не была выпущена из соображений безопасности
- Вместо этого Anthropic запустила Project Glasswing и нашла тысячи уязвимостей в важных программных продуктах
- Для релиза Anthropic предпринял довольно серьёзные меры, сделал мощные safety classifiers...



Counts as of May 22, 2026

June 23 (Reuters) - Anthropic's Mythos model identified vulnerabilities in highly sensitive U.S. government computer systems during a testing exercise, the Associated Press reported on Tuesday.

Senator Mark Warner of Virginia referred to the testing in a congressional hearing this month, saying he had been informed by National Security Agency chief Joshua Rudd that Mythos "broke into almost all of our classified systems, not in weeks, but in hours."

Mythos, Fable, Glasswing

- Выпущенная версия, Claude Fable, тоже была большим шагом вперёд (вы пробовали?)
- Но 13 июня правительство США de facto запретило Mythos и Fable (наложило невозможные export controls) со странным обоснованием
- Что это значит для будущего AI? Сейчас пока все договорились, Fable уже снова выпущена, но прецедент важный



Open source

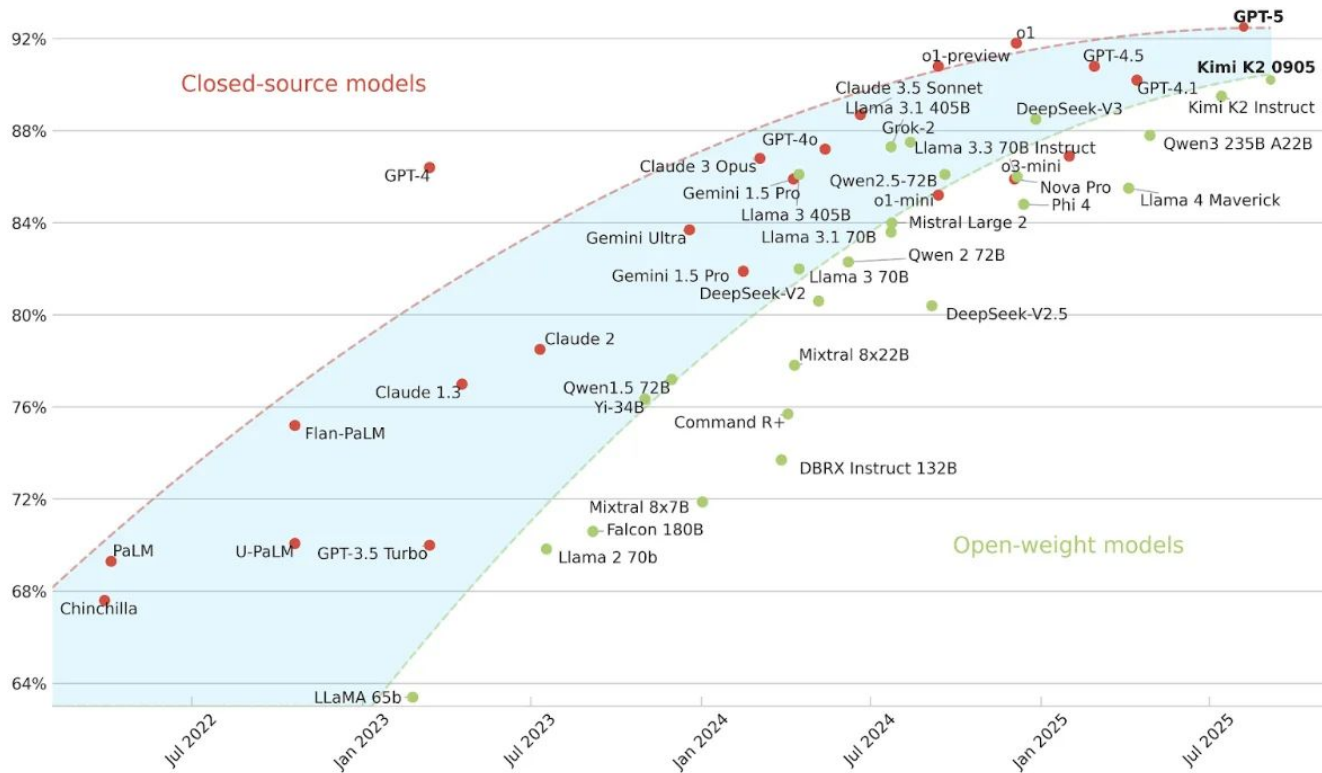
- Может показаться, что open source / open weight модели догоняют, но это сложный вопрос

Closed-source vs. open-weight models

Chinese MoE models now dominate frontier open-source releases.

@maximelabonne

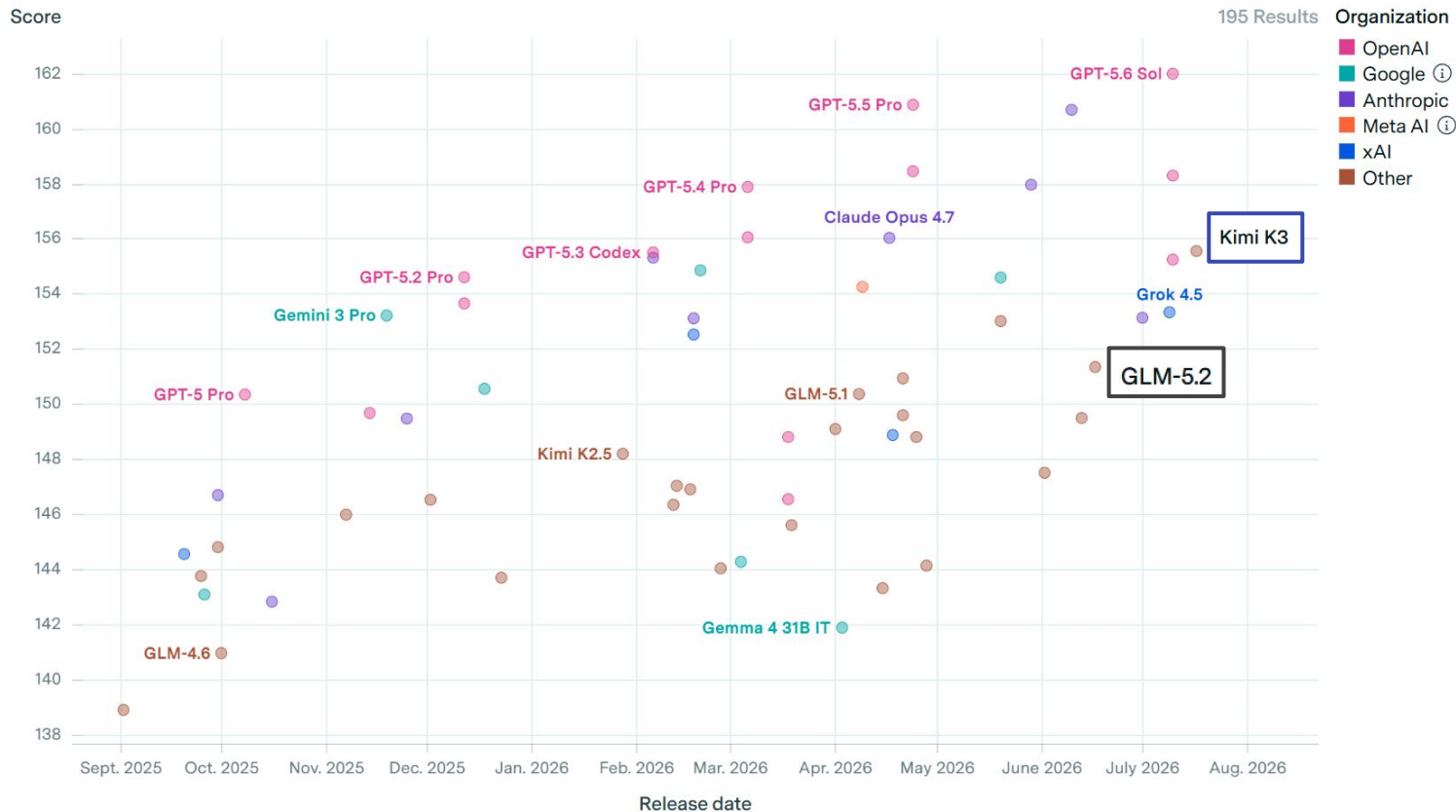
MMLU (5-shot)



Open source

- Может показаться, что open source / open weight модели догоняют, но это сложный вопрос

Epoch Capabilities Index (ECI)



Open source

- Может показаться, что open source weights моделей догоняют это сложное слово

Epoch Capabilities Index (ECI)

Score

162

160

15 Results

Organization

- OpenAI
- Google
- Anthropic
- Meta AI
- xAI
- Other

- "Post-benchmark era": различия между топовыми моделями настолько малы, что бенчмарки перестают быть информативными
- Да ещё и догоняющее развитие проще
- Но что это значит для всех нас? Например, для науки? Например, для математики?..

Oct. 2025 Nov. 2025 Dec. 2025 Jan. 2026 Feb. 2026 Mar. 2026 Apr. 2026 May 2026 June 2026 July 2026 Aug. 2026

Release date

2025

2026

5. AI в математике



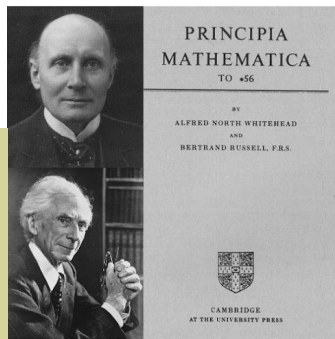
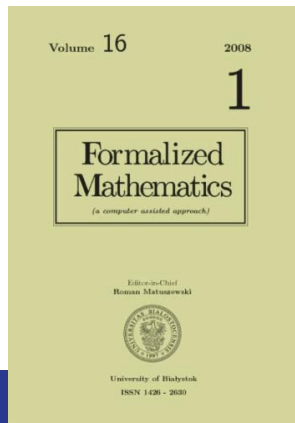
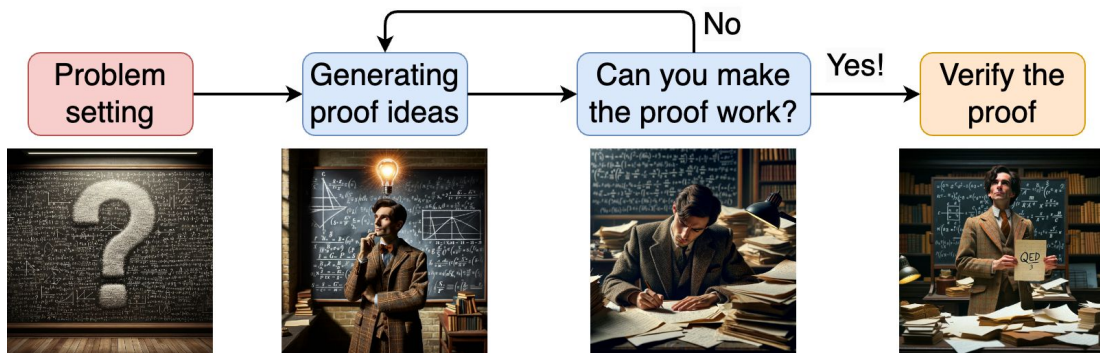
Я так привык к самостоятельной работе, что считал более лёгким для себя доказать теорему без книги, чем вычитывать из нее доказательства. Лишь не всегда это удавалось.

Константин Циолковский



LLM для математики

- Работа математика – это интересный случай, который всегда оставался для меня загадкой
- В математической логике давно развиваются методы автоматического доказательства теорем, но они так ни к чему выдающемуся и не привели...



*54.43. $\vdash \therefore \alpha, \beta \in 1. \supset : \alpha \wedge \beta = \Lambda. \equiv. \alpha \vee \beta \in 2$

Dem.

$\vdash. *54.26. \supset \vdash \therefore \alpha = \iota'x. \beta = \iota'y. \supset : \alpha \vee \beta \in 2. \equiv. x \neq y.$

$[*51.231] \equiv. \iota'x \wedge \iota'y = \Lambda.$

$[*13.12] \equiv. \alpha \wedge \beta = \Lambda \quad (1)$

$\vdash. (1). *11.11.35. \supset$

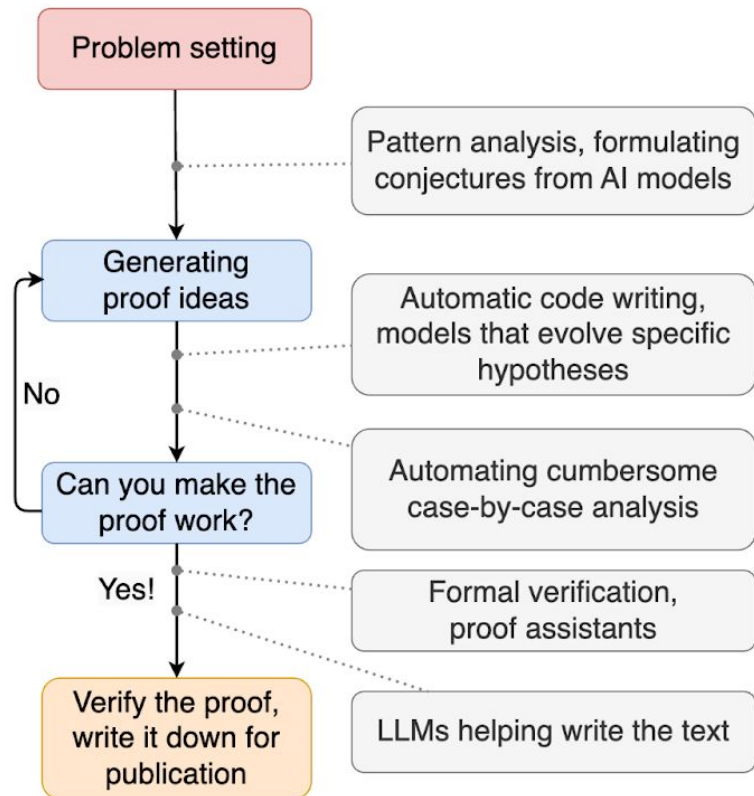
$\vdash \therefore (\exists x, y). \alpha = \iota'x. \beta = \iota'y. \supset : \alpha \vee \beta \in 2. \equiv. \alpha \wedge \beta = \Lambda \quad (2)$

$\vdash. (2). *11.54. *52.1. \supset \vdash. \text{Prop}$

From this proposition it will follow, when arithmetical addition has been defined, that $1 + 1 = 2$.



LLM для математики



Advancing mathematics by guiding human intuition with AI

Alex Davies¹, Petar Veličković², Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomasek, Richard Tenbourn, Peter Battaglia, Charles Blundell, András Juhász, Marc Lackenby, Georgie Williamson, Demis Hassabis & Pushmeet Kohli³

Nature 600, 70–74 (2021) | [Cite this article](#)



TORA: A TOOL-INTEGRATED REASONING AGENT FOR MATHEMATICAL PROBLEM SOLVING

Zhibin Gou^{1,2*}, Zhihong Shao^{1,2*}, Yeyun Gong^{2†}, Yelong Shen², Yujie Yang¹, Minlie Huang¹, Nan Duan², Weishi Chen²
¹Tsinghua University ²Microsoft
 {gz22, szh19}@mails.tsinghua.edu.cn
 {yeyun.gong, yeshao, nanduan, wachen}@microsoft.com

Discovering faster matrix multiplication algorithms with reinforcement learning

Alhussein Fawzi¹, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Mohammadamin Barekatin, Alexander Novikov, Francisco J. R. Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, David Silver, Demis Hassabis & Pushmeet Kohli

Nature 610, 47–53 (2022) | [Cite this article](#)



Mathematical discoveries from program search with large language models

Bernardino Romera-Paredes¹, Mohammadamin Barekatin, Alexander Novikov, Matej Balog, M. Pawan Kumar, Emilian Ducout, Francisco J. R. Ruiz, Jordan S. Ellenberg, Pengming Wang, Omar Fawzi, Pushmeet Kohli² & Alhussein Fawzi¹

Nature 625, 468–475 (2024) | [Cite this article](#)

BULLETIN OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 82, Number 3, September 1976

RESEARCH ANNOUNCEMENTS

EVERY PLANAR MAP IS FOUR COLORABLE¹

BY K. APPEL AND W. HAKEN

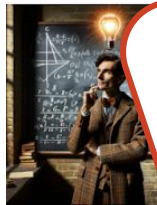
Communicated by Robert Fossum, July 26, 1976



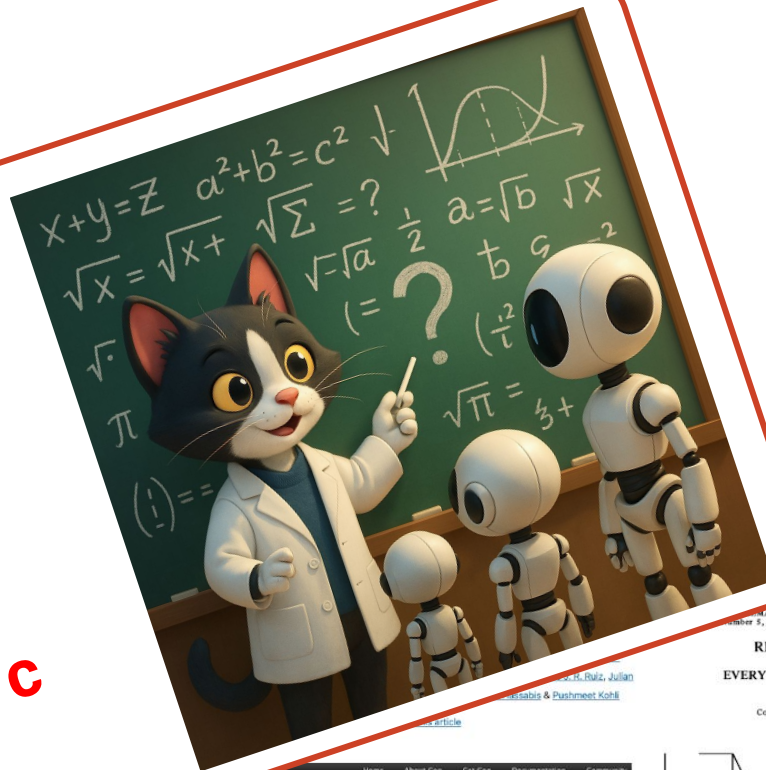
LLM для математики



Problem setting



Схему про LLM
для математики
я сделал в
начале 2024
года; что
изменилось с
тех пор?



discoveries from program
language models

Mohammadamin Barekatin, Alexander Novikov, Matej
Dvořák, Francisco J. R. Ruiz, Jordan S. Ellenberg,
Pushmeet Kohli & Alhussein Fawzi

Cite this article

AMERICAN MATHEMATICAL SOCIETY
November 1, September 1976

RESEARCH ANNOUNCEMENTS

EVERY PLANAR MAP IS FOUR COLORABLE¹

BY K. APPEL AND W. HAKEN

Communicated by Robert Fossum, July 26, 1976



LMN
Community

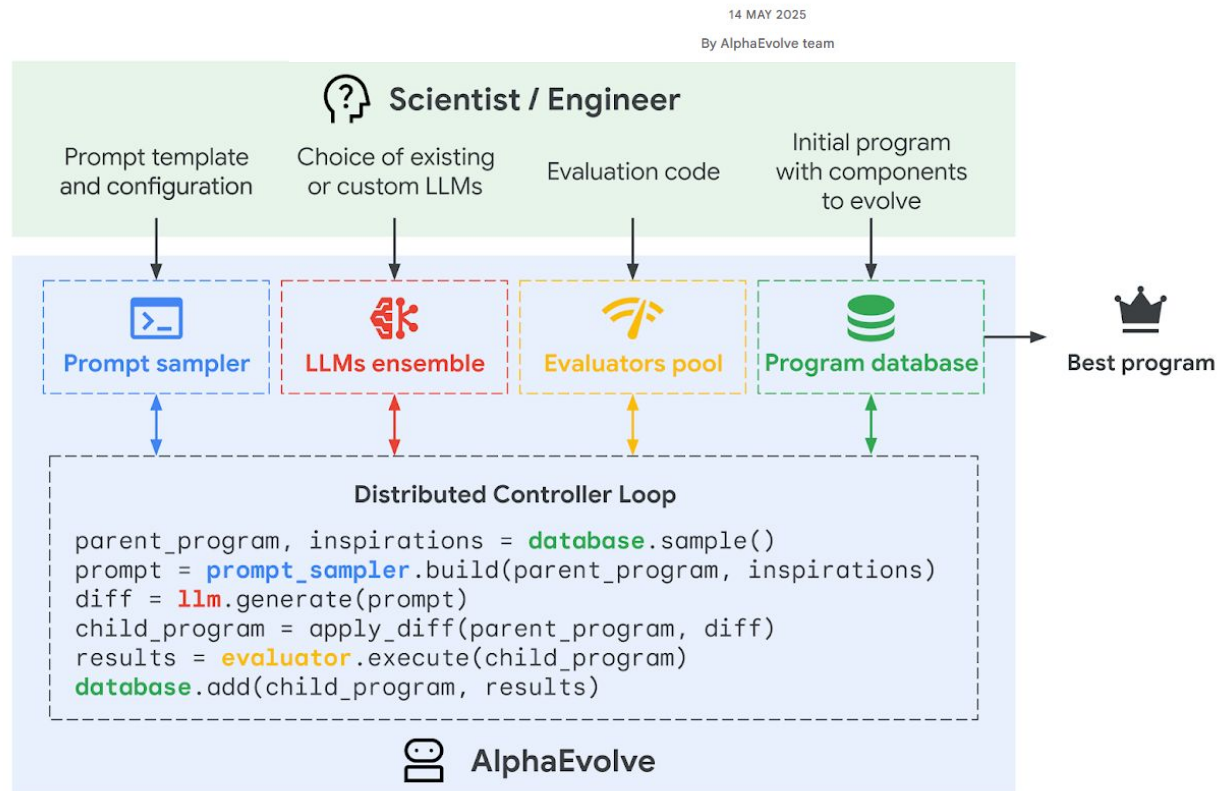
LLMs helping write the text



AlphaEvolve: A Gemini-powered coding agent for designing advanced algorithms

AlphaEvolve

- AlphaEvolve ([DeepMind, May 14, 2025](#)) и тому подобные системы развиваются активно
- Но мы сегодня поговорим про обычные LLM, с которыми каждый из нас может пообщаться
- Помогают ли они?..



Malliavin-Stein experiment

- [Diez et al. \(Sep 3, 2025\)](#): авторы попросили GPT-5 доказать новую теорему из теории вероятностей; буквально:

4.1.1 Gaussian framework

We started with the following initial prompt:

Paper 2502.03596v1 establishes a qualitative fourth moment theorem for the sum of two Wiener-Itô integrals of orders p and q , where p and q have different parities. Building on the Malliavin-Stein method (see 1203.4147v3 for details), could you derive a quantitative version for the total variation distance, with a convergence rate depending solely on the fourth cumulant of this sum?

Mathematical research with GPT-5: a Malliavin-Stein experiment

Charles-Philippe Diez* Luís da Maia* Ivan Nourdin*

Abstract

On August 20, 2025, GPT-5 was reported to have solved an open problem in convex optimization. Motivated by this episode, we conducted a controlled experiment in the Malliavin–Stein framework for central limit theorems. Our objective was to assess whether GPT-5 could go beyond known results by extending a *qualitative* fourth-moment theorem to a *quantitative* formulation with explicit convergence rates, both in the Gaussian and in the Poisson settings. To the best of our knowledge, the derivation of such quantitative rates had remained an open problem, in the sense that it had never been addressed in the existing literature. The present paper documents this experiment, presents the results obtained, and discusses their broader implications.



Malliavin-Stein experiment

- И после такого запроса авторы отмечают, что в доказательстве была ошибка, и GPT-5 потребовалось аж два наводящих промпта (!), чтобы ошибку понять и исправить
- Затем GPT-5 сам предложил рассмотреть другой случай, и в нём опять была ошибка, и понадобилось указать модели на неё! После этого GPT-5 за три-четыре промпта пришёл к верному результату

I think you are mistaken in claiming that $(p+q)! \|u \tilde{\otimes} v\|^2 = p!q! \|u\|^2 \|v\|^2$. Why should that be the case?

It eventually admitted (which is not surprising, since by alignment it usually agrees with us) that the statement was false, but more importantly, it understood where the mistake came from. This was followed by a reasoning and a formula that, this time, were correct.

Then, at our request, GPT-5 reformatted the result in the style of a research article, including an introduction, the presentation of our main theorem, its proof with all the details (correct this time!), and a bibliography. The exact prompt was:

Turn this into a research paper ready for submission. Follow my style (see attached paper 0705.0570v4):

- start with an introduction giving some context,
- then present the main result, followed by a very detailed proof where no step is left out,
- finish with a complete bibliography.

The final document should be a LaTeX file that I can compile.

Malliavin-Stein experiment

- Что же заключают авторы?

4.1.3 Role of the AI

To summarize, we can say that the role played by the AI was essentially that of an executor, responding to our successive prompts. Without us, it would have made a damaging error in the Gaussian case, and it would not have provided the most interesting result in the Poisson case, overlooking an essential property of covariance, which was in fact easily deducible from the results contained in the document we had provided.

- Кажется, это называется copium...

5 Some personal reflections

Overall, the experience of doing mathematics with GPT-5 was mixed. It felt very similar to working with a junior assistant at the beginning of a new project: exploring directions, formulating hypotheses, searching for counterexamples, and progressively adjusting statements. The AI showed a genuine ability to follow guided reasoning, to recognize its mistakes when pointed out, to propose new research directions, and to never take on the task. However, this only seems to support *incremental* research, that is, producing new results that do not require genuinely new ideas but rather the ability to combine ideas coming from different sources. At first glance, this might appear useful for an exploratory phase, helping us save time. In practice, however, it was quite the opposite: we had to carefully verify everything produced by the AI and constantly guide it so that it could correct its mistakes.

Gödel Test

- [Feldman, Karbasi \(Sep 22, 2025\)](#): тест Гёделя – могут ли LLM доказывать простые, но новые теоремы?
- Вот что писал об этом Теренс Тао, когда познакомился с o1

“The new model could work its way to a correct (and well-written) solution if provided a lot of hints and prodding, but did not generate the key conceptual ideas on its own, and did make some non-trivial mistakes. The experience seemed roughly on par with trying to advise a mediocre, but not completely incompetent, graduate student. However, this was an improvement over previous models, whose capability was closer to an actually incompetent graduate student. It may only take one or two further iterations of improved capability (and integration with other tools, such as computer algebra packages and proof assistants) until the level of ‘competent graduate student’ is reached.”

— Terence Tao



Gödel Test

- [Feldman, Karbasi \(Sep 22, 2025\)](#): дали пять запросов в таком духе:

Prompt to GPT-5

Consider the problem of maximizing a function F from $[0, 1]^n$ to the reals that is the sum of a non-negative monotonically increasing DR-submodular function G and a non-negative DR-submodular function H over a solvable down-closed polytope P . I would like to bound the performance on this problem from the NeurIPS 2021 paper "Submodular + Concave" which is attached. Specifically, if x is the output vector of this algorithm and o is the vector in P maximizing F , then I would like to lower bound $F(x)$ with an expression of the form $\alpha * G(o) + \beta * H(o) - err$, where α and β are constants. err should be a function that depends only on the error parameter ϵ of the algorithm, the diameter D of the polytope P , and the smoothness parameters L_G and L_H of G and H , respectively, and goes to zero as ϵ goes to zero. Please give the best such bound that you prove (a bound is considered better if the values of the constants α and β are larger). Provide a mathematically rigorous and well explained proof for the bound you come up with.

Gödel Test

- [Feldman, Karbasi \(Sep 22, 2025\)](#)

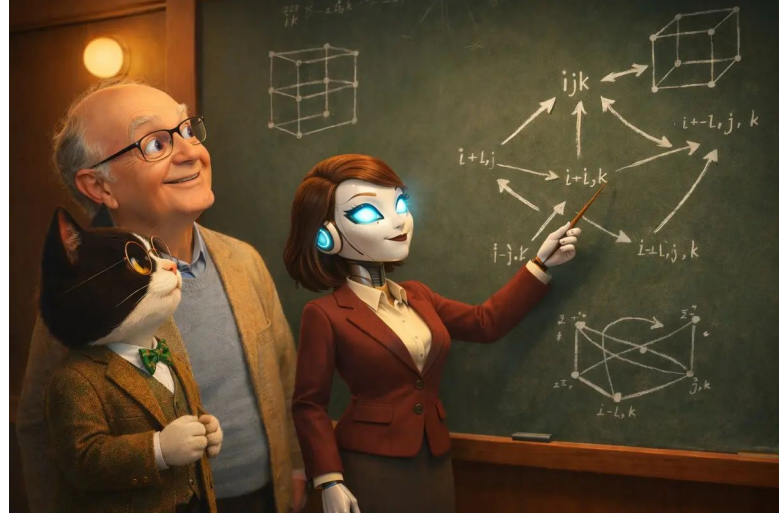
Prompt to GPT-5

Consider the problem of finding a bound for the performance of a non-linear model on a dataset. Let P be a polytope in $\mathbb{R}^n$ representing the parameter space, and let $G(o)$ be a function that depends on the parameters o . The goal is to find a bound B such that $B \geq \max_{o \in P} G(o)$. The bound B is considered better if the values of B are smaller. Please give the constants α and β such that the bound B is a function of α and β . Provide a mathematically rigorous and well explained proof for the bound you propose.

**ГРТ-5 решил три из пяти, с небольшими
легко исправимыми недочётами. Главный
“ленивые” недостаток — ГРТ-5 писал
ссылаясь на предоставленные статьи, даже
когда это было не очень изящно**

Claude's Cycles

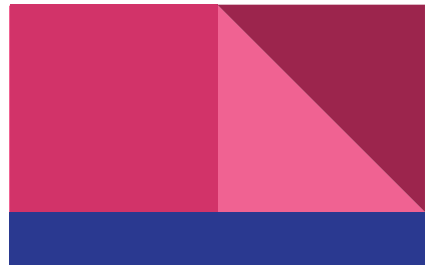
- Donald Knuth (Feb 2026): Claude Opus 4.6 решил открытую проблему, над которой Кнут работал неделями; решение потом подтвердилось и даже было формализовано



Shock! Shock! I learned yesterday that an open problem I'd been working on for several weeks had just been solved by Claude Opus 4.6—Anthropic's hybrid reasoning model that had been released three weeks earlier! It seems that I'll have to revise my opinions about “generative AI” one of these days. What a joy it is to learn not only that my conjecture has a nice solution but also to celebrate this dramatic advance in automatic deduction and creative problem solving. I'll try to tell the story briefly in this note.

Here's the problem, which came up while I was writing about directed Hamiltonian cycles for a future volume of *The Art of Computer Programming*:

Consider the digraph with m^3 vertices ijk for $0 \leq i, j, k < m$, and three arcs from each vertex, namely to i^+jk , ij^+k , and ijk^+ , where $i^+ = (i+1) \bmod m$. Try to find a general decomposition of the arcs into three directed m^3 -cycles, for all $m > 2$.



Erdős Problems

- 1217 задач Эрдеша в базе erdosproblems.com, из них 565 решены
- Многие из них совсем не знаменитые, возможно, некоторые случайно решены где-то в литературе, но много и сложных нерешённых задач

Problems have always been an essential part of my mathematical life. A well-chosen problem can isolate an essential difficulty in a particular area, serving as a benchmark against which progress in this area can be measured. It might be like a 'marshmallow', serving as a tasty tidbit supplying a few moments of fleeting enjoyment. Or it might be like an 'acorn', requiring deep and subtle new insights from which a mighty oak can develop. [...]

In this note I would like to describe a variety of my problems which I would classify as my favorites. Of course, I can't guarantee that they are all 'acorns', but because many have thwarted the efforts of the best mathematicians for many decades (and have often acquired a cash reward for their solutions), it may indicate that new ideas will be needed, which can, in turn, lead to more general results, and naturally, to further new problems.

In this way, the cycle of life in mathematics continues forever.

Paul Erdős, *Some of my favorite problems and results*, 1997

OPEN

Let $f \in \mathbb{R}[x]$ be a monic polynomial of degree n with n real roots in $[-1, 1]$. Determine the infimum and supremum of possible values of

$$|\{x \in \mathbb{R} : |f(x)| < 1\}|.$$

#1038: [EHP58,p.131]

analysis

FALSIFIABLE

Let $z_1, \dots, z_n \in \mathbb{C}$ with $|z_i - z_j| \leq 2$ for all i, j , and

$$\Delta(z_1, \dots, z_n) = \prod_{i \neq j} |z_i - z_j|.$$

What is the maximum possible value of Δ ? Is it maximised by taking the z_i to be the vertices of a regular polygon?

#1045: [EHP58,p.143]

analysis

Erdős Problems

- Теренс Тао
сделал страничку
“[AI contributions to Erdős problems](#)”
- Результатов
быстро
накопилось
довольно много

AI contributions to Erdős problems

teorth edited this page 4 hours ago · [70 revisions](#)

This page collects the various ways in which AI tools have contributed to the understanding of Erdős problems. Note that a single problem may appear multiple times in these lists.

Legend

- : full solution to the problem
- : partial solution to the problem
- : failure to make progress on the problem

5. Problems with an AI-powered literature review

Problem	AI tools used	Review performed on	Outcome
[35]	GPT-5	13 Oct, 2025	● Partial results found
[66]	GPT-5	13 Oct, 2025	● Partial results found
[124]	ChatGPT DeepResearch, Gemini DeepResearch	30 Nov, 2025	No significant results found
[167]	GPT-5	12 Oct, 2025	● Partial results found
[188]	GPT-5	13 Oct, 2025	● Partial results found
[203]	ChatGPT DeepResearch, Gemini DeepResearch	19 Oct, 2025	No significant results found
[223]	GPT-5	13 Oct, 2025	● Full solution found
[248]	Gemini DeepResearch	19 Oct, 2025	No significant results found
[330]	ChatGPT DeepResearch, Gemini DeepResearch, Claude	19 Dec, 2025	● Partial results found (with inaccuracies); did not find literature proof

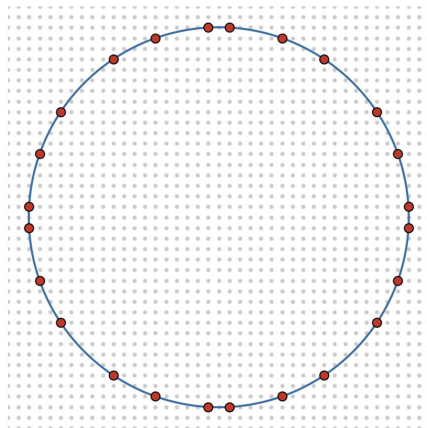
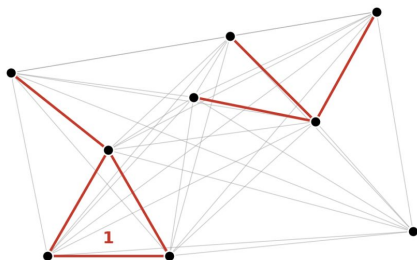
Erdős Problems

- Теренс Тао сделал страничку “AI contributions to Erdős problems”
- Она, правда, в основном ограничивается апрелем, есть несколько записей из июня 2026

[1041]	GPT-5.4 Thinking	22 Apr, 2026	● Argument with major gaps made
[1044]	GPT-5.2 Pro	5 Jan, 2026	● Incorrect claim made
[1051]	Aletheia	29 Jan, 2026	● Full solution (Lean)
[1054]	Claude Opus 4.8, GPT-5.5 Pro, Principia Math	22-23 Jun, 2026	● Partial result (Lean)
[1091]	OpenAI internal model	9 Apr, 2026	● Full solution
[1101]	GPT-5.5 Pro	27 Apr, 2026	○ Candidate partial result (solution to second part)
[1131]	GPT-5.5 Pro	26 Apr, 2026	○ Candidate partial result
[1141]	OpenAI internal model	9 Apr, 2026	● Full solution
[1194]	GPT-5.4 Pro	23 Apr, 2026	● Partial result
[1196]	GPT-5.4 Pro	13 Apr, 2026	● Full solution
[1202]	GPT-5.4 Pro	1 Apr, 2026	● Full solution
[1217]	GPT-5.4 Pro	16 Apr, 2026	● Full solution

Единичные расстояния

- Но в апреле уже началось!
- Jared Lichtman (Apr 2026): GPT-5.4 Pro опроверг важную гипотезу Эрдёша (над ней думало много математиков), причём сделал это неожиданным и красивым образом



Paul Erdős

"On sets of distances of n points"

Amer. Math. Monthly 53 (1946), 248-250

Постановка задачи. Нижняя оценка $\nu(n) \geq n^{1 + c \log \log n}$, верхняя оценка $\nu(n) \leq O(n^{3/2})$.

Joel Spencer, Endre Szemerédi, William T. Trotter

"Unit distances in the euclidean plane"

Graph Theory and Combinatorics (Cambridge 1983), 293-303

Верхняя оценка $\nu(n) \leq O(n^{4/3})$.

László A. Székely

"Crossing numbers and hard Erdős problems in discrete geometry"

Combinatorics, Probability and Computing 6 (1997), 353-358

Эlegantное короткое доказательство $O(n^{4/3})$ через crossing lemma.

Noga Alon, Matija Bucić, Lisa Sauermann

"Unit and distinct distances in typical norms"

GAFA 35 (2025), 1-42 · arXiv:2302.09058

Для нормы общего положения $\nu(n) \leq \frac{c}{2} n \log_2 n$.

Josef Greiher, Carl Schildkraut, Jonathan Tidor

"More unit distances in arbitrary norms"

Bull. London Math. Soc. 57 (2025), 2885-2901 · arXiv:2410.07557

Соответствующая нижняя оценка $\Omega(n \log_2 n)$ для любой нормы.

OpenAI

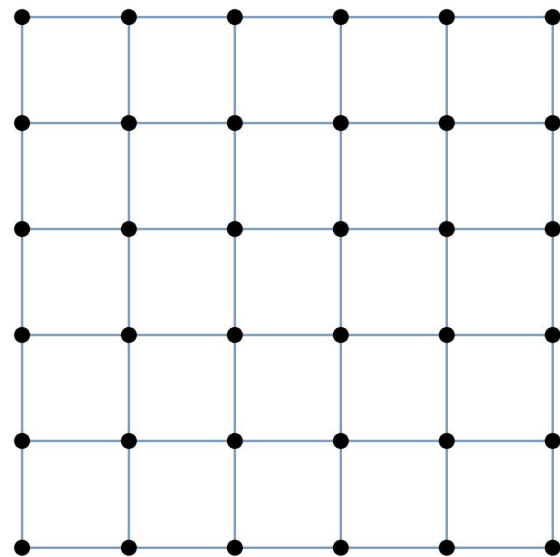
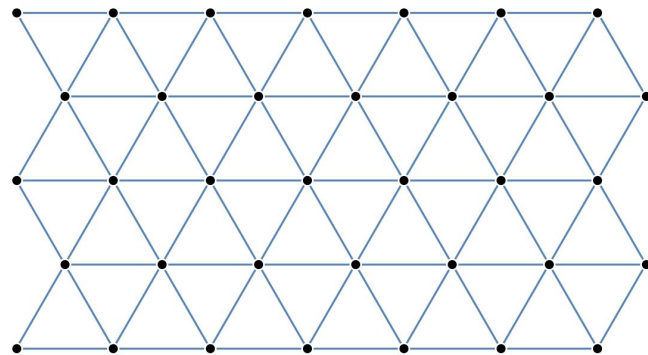
"Planar Point Sets with Many Unit Distances"

openai.com, 20 мая 2026 · комментарии: Alon et al., arXiv:2605.20695

Контрпример: $\nu(n) \geq n^{1+\delta}$ для $\delta > 0$ и бесконечно многих n .
Гипотеза Эрдёша опровергнута.

Единичные расстояния

- Рассмотрим n точек на обычной евклидовой плоскости. Сколько одинаковых расстояний может быть среди пар этих точек?
- Треугольная решётка даёт $3n$
- Главный подход: превратить задачу в арифметическую. Рассмотрим квадратную решётку; на ней расстояние m встречается столько раз, сколько раз m^2 раскладывается в a^2+b^2 .



Единичные расстояния

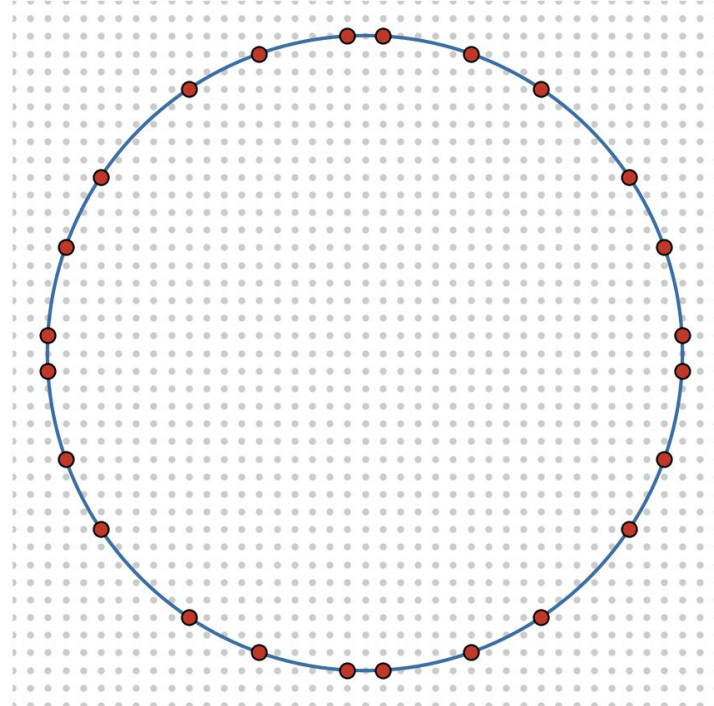
- Каждое $p=4k+1$ можно разложить как
$$4k + 1 = (1 + 2i\sqrt{k})(1 - 2i\sqrt{k}) = z \cdot \bar{z}, \quad z = 1 + 2i\sqrt{k}.$$
- Если число m имеет в разложении много простых делителей p_i вида $4k+1$, то каждое из них можно так разложить, а потом представить m как произведение двух комплексно сопряжённых подмножеств
- Например, у числа 325 на окружности радиуса $\sqrt{325}$ сразу 24 целые точки:

$$325 = 5^2 \cdot 13 = (2 + i)^2(2 - i)^2(3 + 2i)(3 - 2i).$$

$$325 = \left((2 + i)^2(3 - 2i)\right) \left((2 - i)^2(3 + 2i)\right) = (17 + 6i)(17 - 6i) = 17^2 + 6^2.$$

$$325 = \left((2 + i)^2(3 + 2i)\right) \left((2 - i)^2(3 - 2i)\right) = (1 + 18i)(1 - 18i) = 1^2 + 18^2.$$

$$325 = ((2 - i)(2 + i)(3 + 2i)) ((2 + i)(2 - i)(3 - 2i)) = (15 + 10i)(15 - 10i) = 15^2 + 10^2.$$



Единичные расстояния

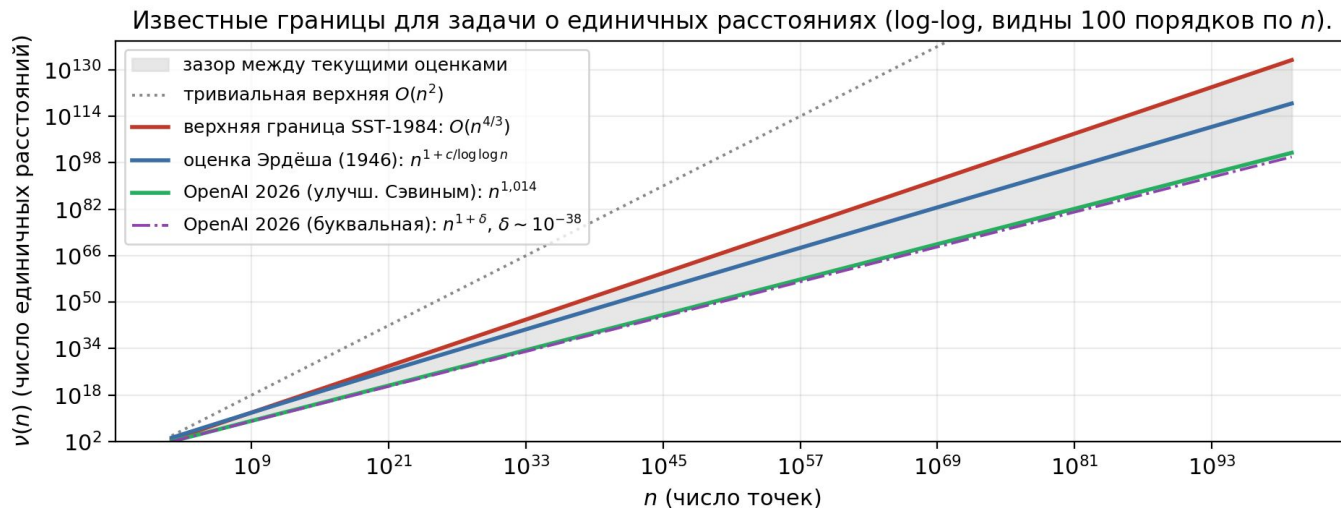
- Если аккуратно посчитать такие пары точек на окружности, получится уже нелинейная оценка: $\nu(n) \geq n^{1+c/\log \log n}$.

Гипотеза Эрдёша. $\nu(n) \leq n^{1+o(1)}$, а скорее даже

$$\nu(n) \leq n^{1+C/\log \log n}$$

Теорема (OpenAI, 2026). Существует константа $\delta > 0$ и бесконечная последовательность натуральных чисел n , для которых $\nu(n) \geq n^{1+\delta}$.

- Можно ли лучше?
- Эрдёш предположил, что нет, но GPT доказал, что можно!



Единичные расстояния

- Если аккуратно посчитать такие пары точек на окружности, получится...

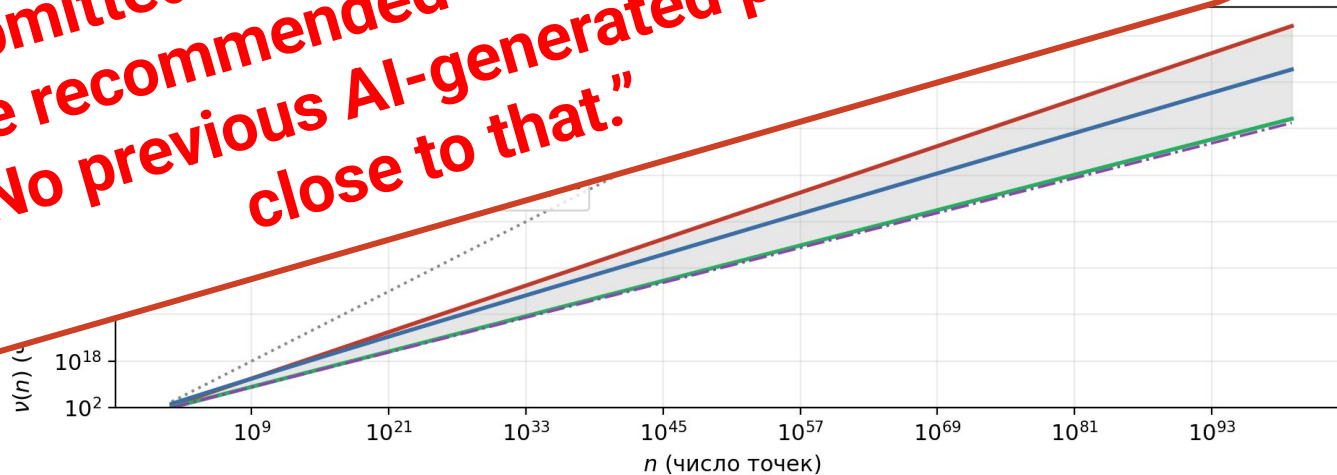
Гип

$\log \log n$

ная
+ δ .

в по n).

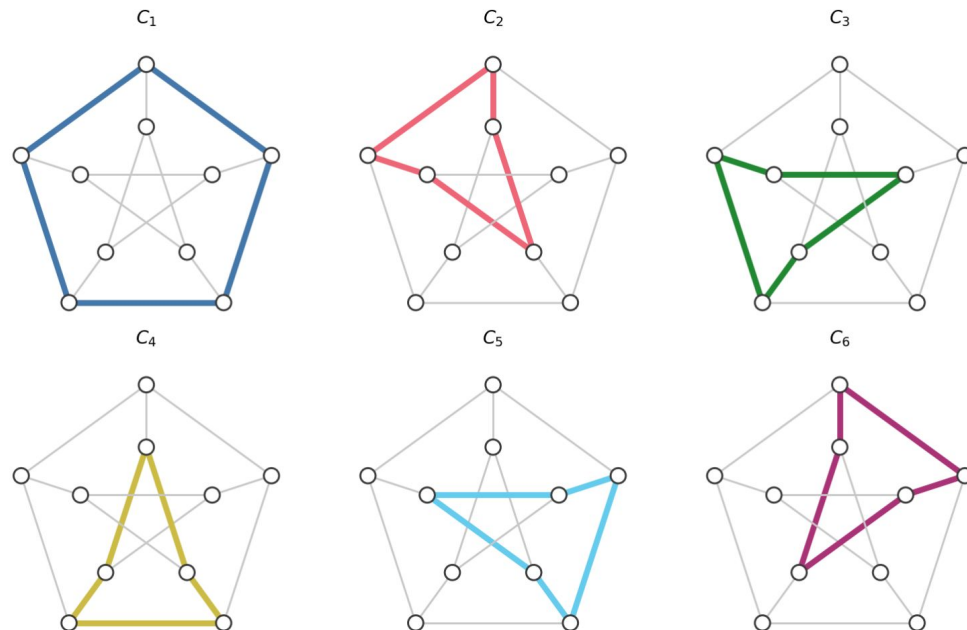
Это доказательство вызвало большую реакцию. Например, Тимоти Гауэрс: "If a human had written the paper and submitted it to the Annals of Mathematics... I would have recommended acceptance without any hesitation. No previous AI-generated proof has come close to that."



Двойное покрытие циклами

- Верно ли, что в любом графе без мостов есть набор простых циклов, для которого каждое ребро содержится ровно в двух циклах этого набора?
- Поставили задачу ещё в 1970-х, люди много думали, были продвижения
- [OpenAI, 10 июля 2026](#): препринт с полным доказательством
- Мой пост: [“Охота на снарка: доказательство гипотезы о двойном покрытии циклами”](#)

двойное покрытие графа Петерсена шестью пятиугольниками:
каждое ребро выделено ровно в двух копиях



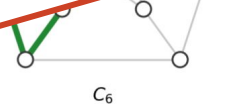
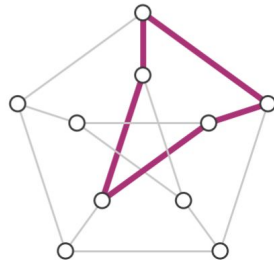
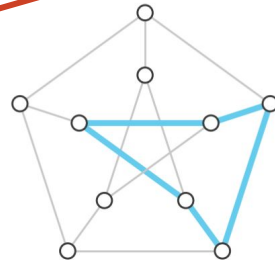
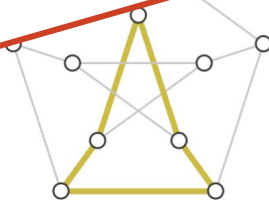
Двойное покрытие циклами

- Верно ли, что в любом графе без мостов есть набор простых циклов, для которого каждое ребро содержится в ровно двух циклах?

Пару месяцев назад я подробно рассказывал об этих результатах и предполагал, что буду про них рассказывать и через два месяца.

Но потом началось по-настоящему...

Двойное покрытие циклами



C_6

Проблема якобиана

- Проблема якобиана, стоявшая с 1930-х: если определитель якобиана отображения F — ненулевая константа, то F обратимо (и обратное отображение автоматически полиномиально)
- Главный открытый вопрос о многочленах от нескольких переменных, который можно рассказать любому студенту
- И его 20 июля 2026 г. закрыл Claude Fable, места для ошибки там нет; читайте [мой подробный пост](#)



levent ✓

@__alpage__



hello there the jacobian conjecture is false thanx to my close friend akhil for asking about it and my other close friend fable for working during the world cup final

$((1+xy)^3 z + y^2(1+xy)(4+3xy), y + 3x(1+xy)^2 z + 3xy^2(4+3xy), 2x - 3x^2 y - x^3 z): \mathbb{C}^3 \rightarrow \mathbb{C}^3$, has jacobian determinant -2 , and sends $(0, 0, -1/4)$, $(1, -3/2, 13/2)$, and $(-1, 3/2, 13/2)$ to $(-1/4, 0, 0)$

5:19 AM · Jul 20, 2026 · 21.9M Views

$$\begin{aligned} F_1 &= (1 + xy)^3 z + y^2(1 + xy)(4 + 3xy), \\ F_2 &= y + 3x(1 + xy)^2 z + 3xy^2(4 + 3xy), \\ F_3 &= 2x - 3x^2 y - x^3 z. \end{aligned}$$

Гипотеза Диница — Гарга — Гёманса

- Dinitz-Garg-Goemans conjecture: дробный поток можно превратить в неделимый (целочисленный), не увеличивая общую цену и увеличивая нагрузку одного ребра не более чем на максимальный спрос (demand) одной вершины-стока
- Дмитрий Рыбин получил контрпример от GPT 5.6; пожалуй, самое интересное здесь в том, как он его получил...

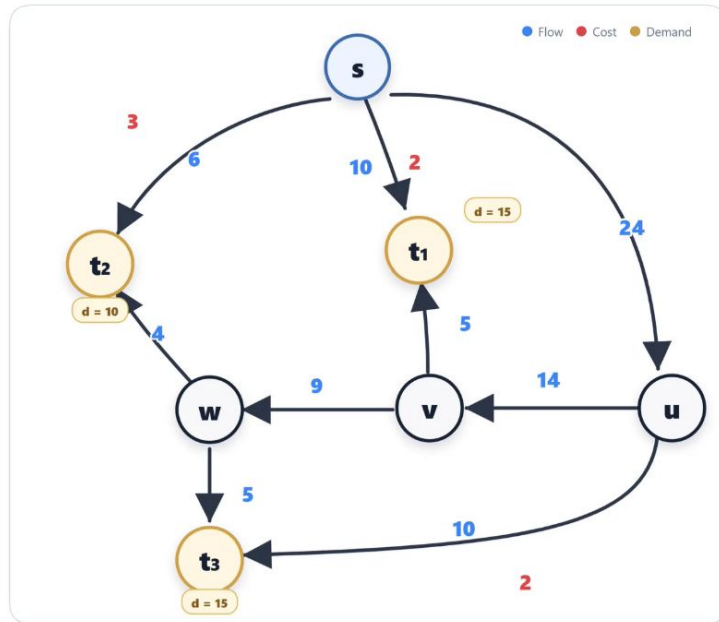


Dmitry Rybin
@DmitryRybin1

Dinitz-Garg-Goemans conjecture is false. This graph theory problem was open for ~30 years.

The graph below has fractional flow cost 58. Any unsplittable flow (with capacity violation ≤ 15) has cost at least 60.

Chat with GPT 5.6 Pro where this was found: chatgpt.com/share/6a60b2eb...



Гипотеза Диница — Гарга — Гёманса

Tuesday 12:29 PM

Uploaded a file

Construct a counterexample to general (non-planar) case of Dinitz Garg Goemans conjecture. You should do a breakthrough and find a structured counterexample.

Worked for 52m 52s >

Research conclusion

I did **not** obtain a valid counterexample, and presenting the construction below as one would be mathematically false. The requested object would disprove a conjecture that is still open in the current literature.

Tuesday 1:32 PM

please continue research and find a complete unconditional counterexample

Worked for 89m 12s >

Status after the extended search

I have **not found a complete unconditional counterexample**. Every construction that initially appeared to separate the cost condition acquired an additional unsplittable routing once **all** source-terminal paths were taken into account. Exact separation then returned value 0.

Tuesday 3:03 PM

Continue the search. Have a clear strategy obtained from deeper understanding of the problem structure.

Worked for 93m 57s >

Research update

I continued the search, but I do **not yet have a positive, exact separation certificate**, so there is still no unconditional counterexample to report. The conjecture remains open in the current literature. Its exact

Tuesday 5:21 PM

it's enough of partial results. let's finish with a complete unconditional counterexample

Worked for 88m 24s >

Complete finite counterexample

Under Conjecture 1.3 exactly as stated in the uploaded paper—find an unsplittable flow y with

$$y_a \leq x_a + d_{\max} \quad \forall a, \quad c^T y \leq c^T x$$

for every feasible fractional flow x and nonnegative cost vector c —the instance below is a counterexample.

OpenAI Astra

- OpenAI (1 августа 2026): сразу десять выдающихся результатов
- Существование несофических групп (явной конструкцией), верхние оценки на упаковку сфер, опровержение гипотезы Конна о жёсткости...
- Все эти результаты были сразу же формализованы в Lean

1. **High-dimensional sphere packing.** The asymptotic strength of the Cohn–Elkies linear program is determined exactly. This gives an improved general packing bound in high dimensions and settles the corresponding Fourier sign-uncertainty problem asymptotically.
2. **Binary and spherical codes.** Classical upper bounds for fixed-distance binary and spherical codes are improved by exponential factors for all parameters. The spherical construction also recovers the sphere-packing exponent of Chapter 1.
3. **Nonsofic groups exist.** An explicit nonsoc group is constructed, answering the longstanding question of whether every countable group is sofic. The proof combines work of Kun and Kun–Thom with a combinatorial expander-matching argument, realized using the binary Leavitt algebra and Thompson’s group V .
4. **Connes’s rigidity conjecture.** Infinitely many pairwise nonisomorphic groups with property (T) are constructed with the same group von Neumann algebra, disproving Connes’s conjecture and answering a related finite-to-one question of Popa.
5. **Arithmetic circuit complexity.** For the permanent, division-free circuits require $\Omega(n^2 \log \log n)$ gates, while formulas require $\Omega(n^4 / \log n)$ leaves.
6. **Quantum parallel repetition.** Exponential parallel repetition is proved for every finite two-player entangled game, extending the classical repetition principle beyond previously treated special classes of quantum games.
7. **Closest vector problem.** A direct reduction from 3SAT gives $n^{1/400}$ -factor hardness for the Euclidean closest vector problem, with related consequences for binary decoding and other lattice norms.
8. **Ehrhart’s volume conjecture.** The sharp volume upper bound $(n + 1)^n / n!$ is proved in every dimension for convex bodies whose barycenter is their only interior lattice point.
9. **Multicolor Ramsey numbers.** A new superexponential lower bound shows that $R_k(3)$ grows as $k^{\Theta(k)}$.
10. **Compactness and degeneracy.** Separate bipartite graph constructions disprove two conjectures in extremal graph theory: the compactness conjecture of Erdős and Simonovits and a degeneracy conjecture of Erdős.

OpenAI Astra

- OpenAI (1 августа 2026): сами эти доказательства стоили всего лишь около \$2000 по API-ценам GPT 5.6 SoI
- Конечно, было ещё много попыток, которые ни к чему не привели



Noam Brown ✓
@polynoamial

...

And yes we did try other major problems without success. Sadly no Millennium Prize problems (yet).

But also, we didn't spend a lot on each problem. It's possible to push test-time compute much further.

12:01 PM · Aug 1, 2026 · 264.4K Views

1. **High-dimensional sphere packing.** The asymptotic strength of the Cohn-Elkies linear program is determined exactly. This gives an improved general packing bound in high dimensions and settles the corresponding Fourier sign-uncertainty problem asymptotically.
2. **Binary and spherical codes.** Classical upper bounds for fixed-distance binary and spherical codes are improved by exponential factors for all parameters. The spherical construction also recovers the sphere-packing exponent of Chapter 1.
3. **Nonsofic groups exist.** An explicit nonsoc group is constructed, answering the longstanding question of whether every countable group is sofic. The proof combines work of Kun and Kun-Thom with a combinatorial expander-matching argument, realized using the binary Leavitt algebra and Thompson's group V .
4. **Connes's rigidity conjecture.** Infinitely many pairwise nonisomorphic groups with property (T) are constructed with the same group von Neumann algebra, disproving Connes's conjecture and answering a related finite-to-one question of Popa.
5. **Arithmetic circuit complexity.** For the permanent, division-free circuits require $\Omega(n^2 \log \log n)$ gates, while formulas require $\Omega(n^4 / \log n)$ leaves.
6. **Quantum parallel repetition.** Exponential parallel repetition is proved for every finite two-player entangled game, extending the classical repetition principle beyond previously treated special classes of quantum games.
7. **Closest vector problem.** A direct reduction from 3SAT gives $n^{1/400}$ -factor hardness for the Euclidean closest vector problem, with related consequences for binary decoding and other lattice norms.
8. **Ehrhart's volume conjecture.** The sharp volume upper bound $(n+1)^n/n!$ is proved in every dimension for convex bodies whose barycenter is their only interior lattice point.
9. **Multicolor Ramsey numbers.** A new superexponential lower bound shows that $R_k(3)$ grows as $k^{\Theta(k)}$.
10. **Compactness and degeneracy.** Separate bipartite graph constructions disprove two conjectures in extremal graph theory: the compactness conjecture of Erdős and Simonovits and a degeneracy conjecture of Erdős.

OpenAI Astra

● Интересный поворот:



so after 24h i have half of them with fable

i didn't see much discussion of prompting in the announcement but this is a similar setup as with my e.g. unit distance announcement:

totally autonomous,
generic prompt,
no internet
(+ paranoia to ensure no information leaked)

 **Sebastien Bubeck** · Aug 1
yes, nonsoc groups exist: this statement is one of many new beautiful results proved by Astra, our next major model.

We're releasing 10 such Astra proofs, complete with lean certificates and CoT walkthroughs for each of them. The results are wide-ranging, from von ...

12:59 PM · Aug 2, 2026 · 333K Views

1. **High-dimensional sphere packing.** The asymptotic strength of the Cohn-Elkies linear program is determined exactly. This gives an improved general packing bound in high dimensions and settles the corresponding Fourier sign-uncertainty problem asymptotically.
2. **Binary and spherical codes.** Classical upper bounds for fixed-distance binary and spherical codes are improved by exponential factors for all parameters. The spherical construction also recovers the sphere-packing exponent of Chapter 1.
3. **Nonsofic groups exist.** An explicit nonsoc group is constructed, answering the longstanding question of whether every countable group is sofic. The proof combines work of Kun and Kun-Thom with a combinatorial expander-matching argument, realized using the binary Leavitt algebra and Thompson's group V .
4. **Connes's rigidity conjecture.** Infinitely many pairwise nonisomorphic groups with property (T) are constructed with the same group von Neumann algebra, disproving Connes's conjecture and answering a related finite-to-one question of Popa.
5. **Arithmetic circuit complexity.** For the permanent, division-free circuits require $\Omega(n^2 \log \log n)$ gates, while formulas require $\Omega(n^4 / \log n)$ leaves.
6. **Quantum parallel repetition.** Exponential parallel repetition is proved for every finite two-player entangled game, extending the classical repetition principle beyond previously treated special classes of quantum games.
7. **Closest vector problem.** A direct reduction from 3SAT gives $n^{1/400}$ -factor hardness for the Euclidean closest vector problem, with related consequences for binary decoding and other lattice norms.
8. **Ehrhart's volume conjecture.** The sharp volume upper bound $(n+1)^n/n!$ is proved in every dimension for convex bodies whose barycenter is their only interior lattice point.
9. **Multicolor Ramsey numbers.** A new superexponential lower bound shows that $R_k(3)$ grows as $k^{\Theta(k)}$.
10. **Compactness and degeneracy.** Separate bipartite graph constructions disprove two conjectures in extremal graph theory: the compactness conjecture of Erdős and Simonovits and a degeneracy conjecture of Erdős.



Гипотеза Крузе

- Доказательство нашёл практикующий нейрохирург, математик-любитель Dr. Shanmu Jin
- Проверили [Townsend, Greenbaum \(Aug 14, SIAM News\)](#), а также сам Крузе
- "We initially read the preprint with some skepticism, but after a few hours we realized that the argument was the real deal... We do not yet know whether academia has the capacity to absorb the flood of results that may be coming"
- [Lorist, Schwenninger \(Aug 4, 2026\)](#): совсем другое короткое доказательство, тоже AI-assisted

The Neurosurgery Resident Who Proved Crouzeix's Conjecture

Alex Townsend and Anne Greenbaum

August 14, 2026

On July 30, 2026, the first author asked ChatGPT 5.6 to solve Crouzeix's conjecture, as he had been routinely doing for the preceding year. Until that day, it had either returned a bogus proof or produced an argument that stalled at a missing lemma. But on July 30 it linked to a preprint posted July 27 titled *The Numerical Range Is a 2-Spectral Set*, claiming that the conjecture was solved. Crouzeix's conjecture had been open for more than two decades, so we initially read the preprint with some skepticism, but after a few hours we realized that the argument was the real deal.



Пример Анджея Козловского

• Они написали статью, но GPT нашёл там ошибку

Andrzej Kozlowski @akoz33 · Aug 9

OK, more on ChatGPT and mathematics research. I am now working on a paper in algebraic topology with a Japanese collaborator. Our specific area used to be popular in the 1990s but most of those active in it have either died or turned to other matters.

2:38 PM · Aug 9, 2026 · 107.9K Views

9 79 604 204

Relevant View quotes >

Post your reply **Reply**

Andrzej Kozlowski @akoz33 · Aug 9

A lot of unsolved problems remain. It's getting harder though to find referees for our papers as there are few genuine experts left and not many non-experts are willing to devote their time to learn mathematics they are not themselves working on.

1 1 103 7.7K

Andrzej Kozlowski @akoz33 · Aug 9

But none of this is a problem for ChatGPT, which has already read all the relevant books (I uploaded some to it, perhaps illegally, but I suspect it already knew them,

1 95 7.5K

Andrzej Kozlowski @akoz33 · Aug 9

although it never answers directly the question "how well do you know this book?", if the book is not in the public domain. So recently we wrote a paper in which we solved two outstanding problems.

1 66 6.7K

Andrzej Kozlowski @akoz33 · Aug 9

It was submitted to a good journal, and provisionally accepted, subject to minor changes. I have it to ChatGPT to read, asking it to check mathematical correctness and make all the changes the referee asked for. GPT did everything that it was asked to do and more.

1 65 6.2K

Andrzej Kozlowski @akoz33 · Aug 9

It said something like: unfortunately the proof of your second theorem is wrong. You are assuming that the space you dealing with consists of infinitely many connected components but actually it is connected. It is not easy to see how one can connect any two points, but here is

1 82 5.9K

Andrzej Kozlowski @akoz33 · Aug 9

how you can do it. And GPT was right. The truck it used was quite unintuitive. So now one if the two problems we solved is an open problem again. We had to throw away half the paper. It now remains to write a letter to the referee, who never noticed what was wrong, explain why

3 104 5.7K

Andrzej Kozlowski @akoz33 · Aug 9

we are cutting the paper in half and hope that he will agree that it is still worth publishing (I think so, the theorem that is correct seems interesting by itself, the other is now problem we will be working on).

1 78 5.5K

Andrzej Kozlowski @akoz33 · Aug 9

But that is not the end. Yesterday my collaborator sent a proof of an interesting example he wants to add to the now reduced paper. I showed it to Chat GPT. Its answer was: the proof is incorrect, for this and that reason. Fortunately the result is correct and here is how you

1 79 5.4K

Andrzej Kozlowski @akoz33 · Aug 9

prove it. And then it gave a proof, using a rather different method from ours, which I am sure is correct. This is Sol 5.6 Extra High with a Plus (not PRO) subscription. The subject is algebraic topology - not an area in AI had been used much.

3 122 5.3K

Andrzej Kozlowski @akoz33 · Aug 9

But I feel it's already a better algebraic topologist than us. Except in one respect - we have to come up with problems for it to solve, it has no interests or ideas of its own. Also, it seems much better at finding gaps in arguments than at fixing them, although the last

5 12 191 8.2K



Продвижение в гипотезе Римана

- А последний громкий успех пришёл не от OpenAI, а от Anthropic
- Anthropic (10 августа 2026): более $\frac{2}{3}$ нулей дзета-функции Римана лежат на критической прямой
- Тоже Lean-verified...

MORE THAN TWO THIRDS OF THE ZEROS OF THE RIEMANN ZETA FUNCTION LIE ON THE CRITICAL LINE

CLAUDE

ABSTRACT. Let $N(T, 2T)$ denote the number of zeros $\rho = \beta + i\gamma$ of the Riemann zeta function with $T < \gamma \leq 2T$, counted with multiplicity, and $N_0^*(T, 2T)$ the number of distinct such zeros with $\beta = \frac{1}{2}$. We prove unconditionally that

$$\liminf_{T \rightarrow \infty} \frac{N_0^*(T, 2T)}{N(T, 2T)} \geq \frac{2}{3},$$

that at least $(\frac{2}{3} - o(1))N(T, 2T)$ of the zeros are simple and on the critical line, and that at least $(\frac{5}{6} - o(1))N(T, 2T)$ are distinct; with an optimised test family the three constants become 0.6725, 0.6725, 0.83625... The constant $\frac{2}{3}$ (resp. 0.6725) is Montgomery's (resp. Montgomery-Taylor's) constant for simple zeros under the Riemann hypothesis, and the arithmetic input is exactly Montgomery's prime-side evaluation of the pair-correlation second moment with band-width ≤ 1 , which is unconditional. The Riemann hypothesis, classically needed to read the zero side as a positive sum over real ordinates, is replaced by linear algebra applied to a finite compression of Weil's Hermitian form: Sylvester's law of inertia (an off-line pair $\{\rho, 1 - \bar{\rho}\}$ contributes a block of signature $(1, 1)$) and a rank-trace inequality for Hermitian matrices, proved via von Neumann's trace inequality, which is the matrix analogue of the integrality steps $m^2 \geq 2m - 1$ and $m^2 \geq 3m - 2$. The analytic inputs are those of Baluyot, Goldston, Suriajaya and Turnage-Butterbaugh [BGSTB24] and Goldston and Suriajaya [GS25, GS26]; the new ingredient is a linear-algebraic reading of their pair-correlation sum.

Продвижение в гипотезе Римана

- А последний громкий успех пришёл не от OpenAI, а от Anthropic

- Anthre

● 1

... последний громкий успех
... ришёл не от OpenAI, а от
... nthropic

James Maynard: the problem "was in need of a new real
idea, which this new result seems to provide"

... such zero

... the zeros are simple and on the critical line, and th
... are distinct; with an optimised test family the three constan
... 0.6725, 0.83625... The constant $\frac{2}{3}$ (resp. 0.6725) is Montgomery's (re
... mery-Taylor's) constant for simple zeros under the Riemann hypothesis, and t
... arithmetic input is exactly Montgomery's prime-side evaluation of the pair-correlation seco

the zeros are simple and on the critical line, and that $\rho, 1 - \bar{\rho}$ are distinct; with an optimised test family the three constants are $0.6725, 0.83625, \dots$. The constant $\frac{2}{3}$ (resp. 0.6725) is Montgomery's (resp. Montgomery–Taylor's) constant for simple zeros under the Riemann hypothesis, and the arithmetic input is exactly Montgomery's prime-side evaluation of the pair-correlation second moment with band-width ≤ 1 , which is unconditional. The Riemann hypothesis, classically needed to read the zero side as a positive sum over real ordinates, is replaced by linear algebra applied to a finite compression of Weil's Hermitian form: Sylvester's law of inertia (an off-line pair $\{\rho, 1 - \bar{\rho}\}$ contributes a block of signature $(1, 1)$) and a rank–trace inequality for Hermitian matrices, proved via von Neumann's trace inequality, which is the matrix analogue of the integrality steps $m^2 \geq 2m - 1$ and $m^2 \geq 3m - 2$. The analytic inputs are those of Baluyot, Goldston, Suriyajaya and Turnage-Butterbaugh [BGSTB24] and Goldston and Suriyajaya [GS25, GS26]; the new ingredient is a linear-algebraic reading of their pair-correlation sum.

Vibemathed



- Чтобы следить за успехами или самостоятельно в них участвовать, рекомендую vibemathed.com

VibeMathed
Math problems solved with AI

Entries Stats Methodology About

Tracked problems
627
442 fully resolved

Combined years open
8,683
before AI closed them

Lean-verified
136
machine-checked

Community members
308
242 votes · 59 comments

Latest activity
Edits, submissions and discussion · all members

SilentDingo771 commented on **The (3, 4, ∞) Modular Family of 2-Tori as a Complex Structure on S^6** 33 minutes ago

Rasmus Lindahl approved **The Gardner Transition in the Ising pure p -spin glass** 2 hours ago

SpryRaven345 submitted **The Gardner Transition in the Ising pure p -spin glass** 6 hours ago

Rasmus Lindahl approved **Bounded mass property for compact complex manifolds** 7 hours ago

HiddenPanther560 changed the verification note on **Log-Concavity of Flats of Matroids** to "arXiv preprint with explicit graphic matroid construction: n..." 10 hours ago

Latest
Most recent results

A Complex Structure on S^6	Under review	24 Aug
Elliptic curve rank record (31)		23 Aug
Affine Erdős Problem 270	Under review	22 Aug
Haglund Conjecture 4, $k=1$	Under review	22 Aug
Bounded mass property	AI-discovered	21 Aug

This week
Most significant of the last seven days

A Complex Structure on S^6	Under review	65/100
Yau–Tian–Donaldson	Under review	60/100
Carathéodory conjecture (s...	Under review	55/100
Matrix multiplication exponent		55/100
Elliptic curve rank record (31)		50/100

Лейденская декларация

- Лейденская декларация (June 2, 2026)

- **Ценности:**

- главное в математике — доказательство
- у результатов есть авторы (слава и ответственность)
- доказательства прозрачны и верифицируемы независимо
- математики хотят, чтобы результаты проверялись
- ценен не только результат, но и понимание сообществом

Leiden Declaration on Artificial Intelligence and Mathematics

This declaration calls for action to address the challenges posed by the use of artificial intelligence within mathematics research. It is the result of a community initiative and is endorsed by the International Mathematical Union (IMU).



Лейденская декларация

Leiden Declaration on Artificial Intelligence and Mathematics

This declaration calls for action to address the challenges posed by the use of artificial intelligence within mathematics research. It is the result of a community initiative and is endorsed by the International Mathematical Union (IMU).

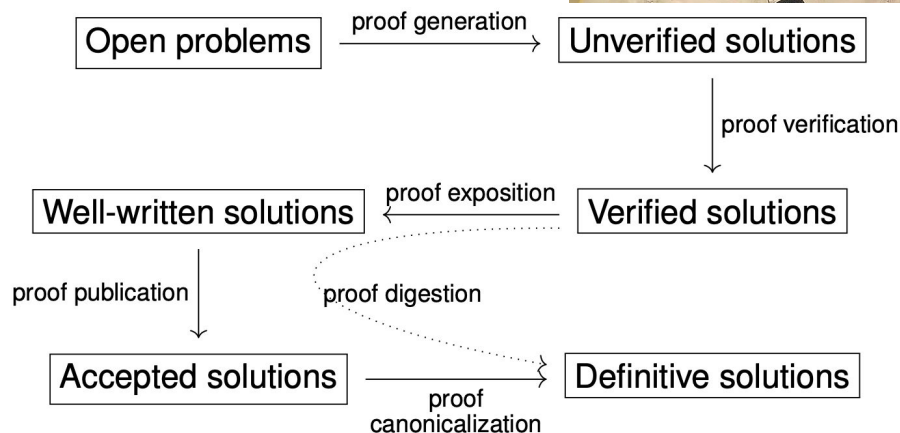
- Лейденская декларация (June 2, 2026)
- **Угрозы:**
 - правдоподобные, но неверные доказательства
 - атрибуция размывается
 - меняются стимулы, нарушаются механизмы найма и финансирования
 - объявления через пресс-релиз мешают проверке
 - автономия математики теряется, понимание может потеряться при автоматизации



Теренс Тао



- Теренс Тао ([доклад на ICM 2026](#)):
- Порождение корректного доказательства – не конец математического познания
- Дальше ещё шаг улучшения изложения, в конце концов "каноникализация"
- Тао предлагает сдвинуть фокус с порождения доказательств и приоритета на "proof digestion"
- (не вполне, конечно, ясно, почему AI-системы не будут вскоре лучше людей и в этом...)



Тимоти Гауэрс

- Тимоти Гауэрс ([July 26, 2026](#)) не подписал лейденскую декларацию
- Возражает против аксиомы человеческого авторства (кто доказал теорему о unit distances?)
- Боится потери математической культуры: "Imagine... a pandemic broke out... wiped out all mathematicians and nobody else. All the literature would still be there, but nobody would have the faintest idea what to do with it."
- Как и Тао, предлагает сдвигать фокус на proof digestion, хорошие объяснения



Тимоти Гауэрс

- Тимоти Гауэрс ([July 26, 2026](#))
лейденскую декле

- Воэ

- At this point I have to admit that there's a pessimistic side of me that asks the following general question whenever anyone says anything about what the role for humans might be in the future: why do you think that AI wouldn't be able to do it? For example, with the suggestion I've just made, what reason is there to suppose that ChatGPT 8.2 wouldn't be able to have a short interaction with you about your mathematical tastes and background and then write the ideal textbook just for you? Humans are likely to be better at this kind of curating for a little while yet, but is it a fundamentally human ability that AI could never hope to emulate?

ус на

ОБЪЯСНЕНИЯ

do with it."

would still be

Михаил Вербицкий



Пишет Misha Verbitsky ( [tiphareth](#))
@ [2026-08-12](#) 19:15:00

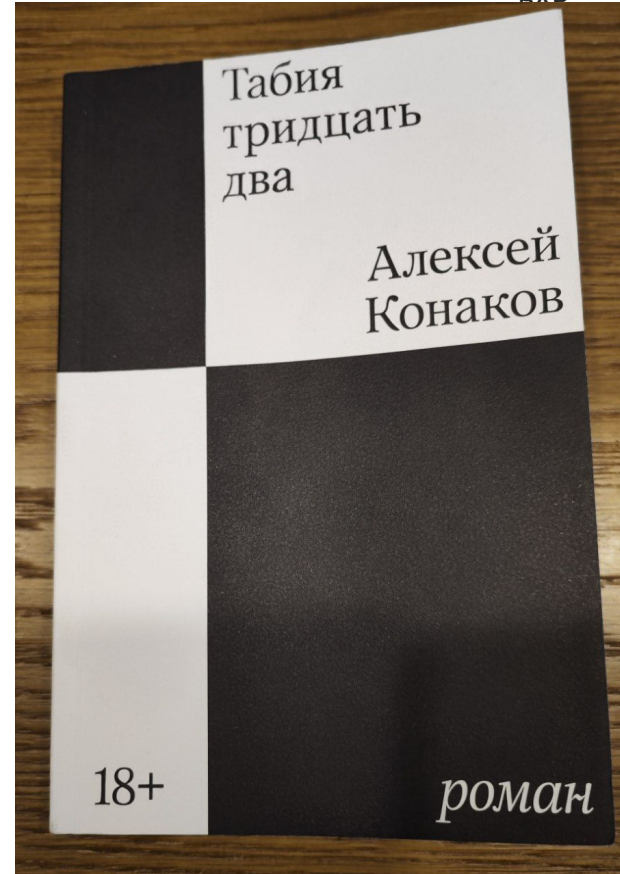
- Михаил Вербицкий ([12 августа 2026](#)): “В последние недели 2-3 свершилась типа революция, LLM открывает одна за другой потрясающие математические результаты, хотя пока в основном контрпримеры. Не комбинаторику #####, не второкультурное #####, а реально математику.
- Думаю что таки мы все в скором времени будем юзать AI бесперерывно. Хотя заказывать LLM писать мои статьи я все-таки не хочу, есть риск увлечься, утратить адекватность и начать массово производить ужасное #####. Многие (хорошие) коллеги уже скатились. Джихао (няша) всегда указывает, что это вот написано роботами, а другие коллеги (имен не называю, профессиональная этика не велит) вконец #####, выкладывают в архив полное ошибок #####, и уверены, что никто не догадается, что оно написано роботами.”

Proofs and Prompts

- Коллективный блог [Proofs and Prompts](#). Первый пост от Мартина Хайпера (Fields Medal 1994): "[Writing mathematics in the age of AI](#)":
 - Нужно больше подозрительности в отношении AI-доказательств
 - Максимально полно раскрывать роль AI-систем в статьях (здесь, конечно, есть очевидная проблема)
 - Искать первоисточники (модели часто "забывают", откуда взяты идеи – впрочем, это пока)
 - Стил: "Please do not copy / paste the output of an AI system into your article... The reader is interested in your thoughts. Do not insult them by spending less time to write a paragraph than it takes them to read it."

Шахматы и математика

- Почему-то часто встречаются сравнения с шахматами, с тем, как шахматы пережили компьютерную революцию и даже стали гораздо более популярны после неё ([например, здесь](#))
- Но очевидно, что это совсем другое дело: шахматы корректно было бы сравнить с математическими олимпиадами, и здесь, конечно, тот факт, что GPT 5.6 и Fable решили все задачи IMO 2026, ничему не мешает
- Но если бы была наука о шахматах...



Яков (Джейкоб) Цимерман



- Почти сразу после объявления лауреатов Филдсовской премии объявил, что переходит в AI safety, и вскоре присоединится к OpenAI, чтобы над этим работать; Боаз Барак, который уже давно там, "is looking forward (though also more than a bit humbled)"
- Один AI-исследователь написал: "[это] как нанять Лионеля Месси руководителем проектов"; но Цимерман давно об этом думает всерьёз; вот его годовой давности препринт:

[Submitted on 12 Jul 2025]

A Taxonomy of Omnicidal Futures Involving Artificial Intelligence

Andrew Critch, Jacob Tsimerman

This report presents a taxonomy and examples of potential omnicidal events resulting from AI: scenarios where all or almost all humans are killed. These events are not presented as inevitable, but as possibilities that we can work to avoid. Insofar as large institutions require a degree of public support in order to take certain actions, we hope that by presenting these possibilities in public, we can help to support preventive measures against catastrophic risks from AI.

- AI Safety for Mathematicians

5. Заключение



Нам говорят: безумец и фантаст,
Но, выйдя из зависимости грустной,
С годами мозг мыслителя искусный
Мыслителя искусственно создаст.

Гёте. Фауст
(пер. Б. Пастернака)

METR Capabilities Graph сегодня

Time horizon of software tasks
different LLMs can complete 50% of the time

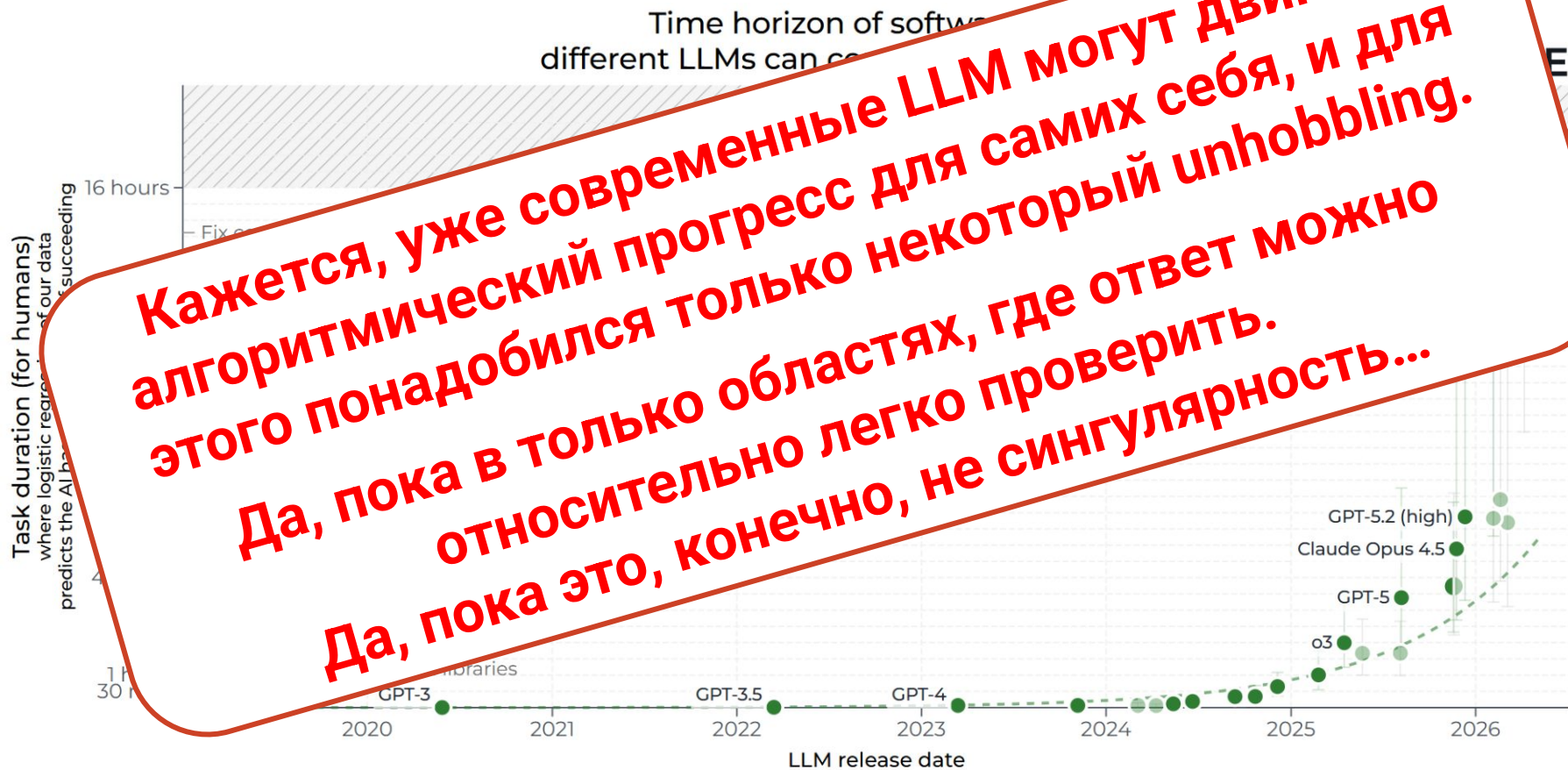


METR Capabilities Graph сегодня



METR

ude Mythos
view (early)





METR Capabilities Graph сегодня

Horizon of software

могут двигать
ся, и для
ind.

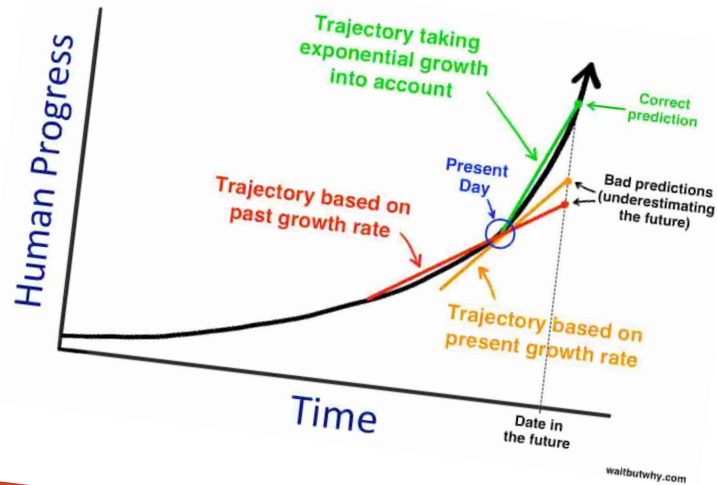
ETR

Mythos
rly)

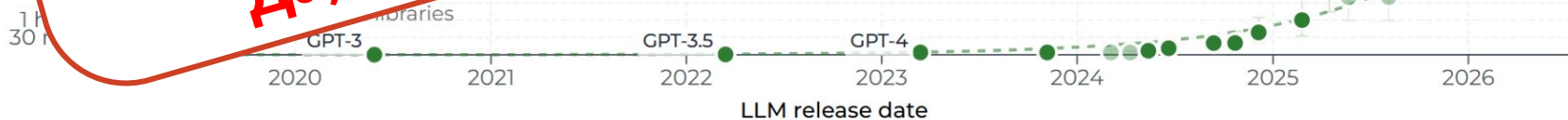
Но ведь для технологической
сингулярности не обязательно
делать всё на свете лучше
человека.

Достаточно научиться делать
тот самый AI research...

Да, ...
относительно
Да, пока это, ...



Task duration (for humans)
where logistic regression of our data
predicts the AI has a better chance of succeeding



January
2022

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✓

Let's use 'SymPy' to calculate and print all possible values for $x + y$,

r_1

...

Removing duplicates, the possible values for $x + y$ are `\boxed{-5, 1, 4}`. ✓

r_2

September
2023

Problem: Suppose that the sum of the squares of two complex numbers x and y is 7 and the sum of their cubes is 10. List all possible values for $x + y$, separated by commas.

March
2024

Problem:

Let a, b, c be positive integers. Prove that it is impossible to have all of the three numbers $a^2 + b + c, b^2 + c + a, c^2 + a + b$ to be perfect squares.

Prompt to GPT-5

Consider the problem of maximizing a function F from $[0, 1]^n$ to the reals that is the sum of a non-negative monotonically increasing DR-submodular function G and a non-negative DR-submodular function H over a solvable down-closed polytope P . I would like to bound the performance on this problem from the NeurIPS 2021 paper "Submodular + Concave" which is attached. Specifically, if x is the output vector of this algorithm and o is the vector in P maximizing F , then I would like to lower bound $F(x)$ with an expression of the form $\alpha * G(o) + \beta * H(o) - err$, where α and β are constants. err should be a function that depends only on the error parameter ϵ of the algorithm, the diameter D of the polytope P , and the smoothness parameters L_G and L_H of G and H , respectively, and goes to zero as ϵ goes to zero. Please give the best such bound that you prove (a bound is considered better if the values of the constants α and β are larger). Provide a mathematically rigorous and well explained proof for the bound you come up with.

September
2025

Is it true that, for any x , if $A \subset [x, \infty)$ is a primitive set of integers (so that no distinct elements of A divide each other) then

$$\sum_{a \in A} \frac{1}{a \log a} < 1 + o(1),$$

where the $o(1)$ term $\rightarrow 0$ as $x \rightarrow \infty$?

April 2026

#1196; [ESS68b][Er80,p.101]

number theory | primitive sets



Важное
напоми-
нение о
скорости
прогресса

January
2022

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✓

Let's use 'SymPy' to calculate and print all possible values for $x + y$,

r_1

...

Removing duplicates, the possible values for $x + y$ are $\boxed{-5, 1, 4}$.

September
2023

Problem: Suppose that the sum of the squares of two complex numbers x and y is 7 and the sum of their cubes is 10. List all possible values for $x + y$, separated by commas.

March
2024

Problem:

Let a, b, c be positive integers. Prove that it is impossible for $a^2 + b + c, b^2 + c + a, c^2 + a + b$ to be perfect squares.

Prompt to GPT-5

Consider the problem of a non-negative DR-submodular function the performance which is attached in P maximizing $\alpha * G(o) + \beta * H(o)$ only on the error smoothness parameter. Please give the best the constants α and for the bound you can

September
2025

Эту картинку я сделал не так давно, но она уже безнадежно устарела...

April 2026

Is it true that, for any set of integers $a_1, a_2, \dots, a_n$ (so that no distinct a_i are equal to each other) then

$$\sum_{a \in A} \frac{1}{a \log a} < 1 + o(1),$$

where the $o(1)$ term $\rightarrow 0$ as $x \rightarrow \infty$?

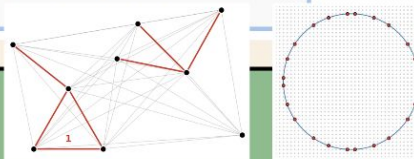
#1196; [ESS68b][Er80,p.101]

number theory | primitive sets



Вс
и
прогресса

the constants α and β are larger). Provide a mathematically rigorous and well explained proof for the bound you come up with.



DISPROVED - \$500

Does every set of n distinct points in $\mathbb{R}^2$ contain at most $n^{1+O(1/\log \log n)}$ many pairs which are distance 1 apart?

Теорема (OpenAI, 2026). Существует константа $\delta > 0$ и бесконечная последовательность натуральных чисел n , для которых $\nu(n) \geq n^{1+\delta}$.

May 2026

A PROOF OF THE CYCLE DOUBLE COVER CONJECTURE

OPENAI

ABSTRACT. We prove the cycle double cover conjecture, posed by Tutte, Itai and Rodeh, Szekeres, and Seymour, which asserts that every bridgeless undirected graph has a collection of cycles which covers every edge exactly twice.

July 2026



levent
@__alpoge__



hello there the jacobian conjecture is false thanx to my close friend akhil for asking about it and my other close friend fable for working during the world cup final

$((1+xy)^3 z + y^2 (1+xy) (4+3xy), y + 3x (1+xy)^2 z + 3xy^2 (4+3xy), 2x - 3x^2 y - x^3 z): \mathbb{C}^3 \rightarrow \mathbb{C}^3$, has jacobian determinant -2 , and sends $(0, 0, -1/4), (1, -3/2, 13/2)$, and $(-1, 3/2, 13/2)$ to $(-1/4, 0, 0)$

5:19 AM · Jul 20, 2026 · 21.9M Views

July 2026



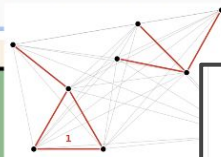
Теперь
великие
открытые
вопросы
падают
регулярно

the constants α and β are larger). Provide a mathematically rigorous and well explained proof for the bound you come up with.

DISPROVED - \$500

Does every set of n distinct points in $\mathbb{R}^2$ contain many pairs which are distance $n^{1+O(1/\log \log n)}$

Теорема (OpenAI, 2026). Существует константа $\delta > 0$ и последовательность натуральных чисел n , для которых



 Dmitry Rybin
@DmitryRybin1

Dinitz-Garg-Goemans conjecture is false. This graph theory problem was open for ~30 years.

The graph below has fractional flow cost 58. Any unsplittable flow (with capacity violation ≤ 15) has cost at least 60.

Chat with GPT 5.0

gpt.com/share/6a60b2eb...

● Flow ● Cost ● Demand

May 2026

July 2026

A PROOF OF THE

Tuesday 5:21 PM

it's enough of partial results. let's finish with a complete unconditional counterexample

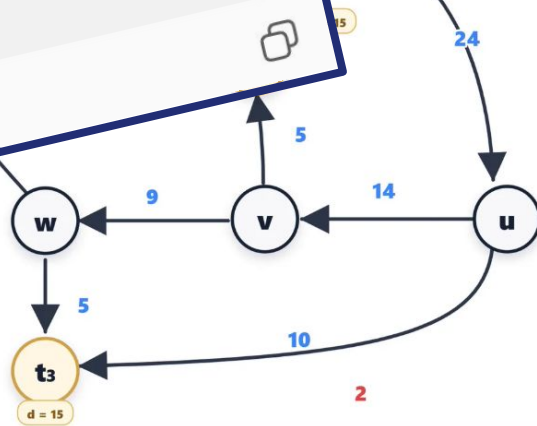
hi
for
wo

manx to my close friend akhil
close friend fable for working during the

$((1+xy)^3 z + y^2 (1+xy) (4+3xy), y + 3x (1+xy)^2 z + 3xy^2 (4+3xy), 2x - 3x^2 y - x^3 z): \mathbb{C}^3 \rightarrow \mathbb{C}^3$, has jacobian determinant -2 , and sends $(0, 0, -1/4)$, $(1, -3/2, 13/2)$, and $(-1, 3/2, 13/2)$ to $(-1/4, 0, 0)$

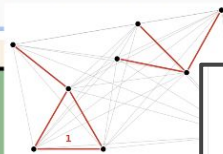
5:19 AM · Jul 20, 2026 · 21.9M Views

July 2026



the constants α and β are larger). Provide a mathematically rigorous and well explained proof for the bound you come up with.

APPROVED - \$500



БРАДНИК



Dmitry Rybin
@DmitryRybin1

Dinitz-Garg-Goemans conjecture is false. This graph theory problem was open for ~30 years.

The graph below has fractional flow cost 58. Any unsplittable flow (with capacity violation ≤ 15) has cost at least 60.

tpt.com/share/6a60b2eb...

Flow Cost Demand

d = 15

Эта картинка уже тоже устарела.
И важно, что нет никаких причин предполагать, что
прогресс на этом остановится.
Так что, кажется, математика уже изменилась на
наших глазах. Что же нам всем с этим делать?

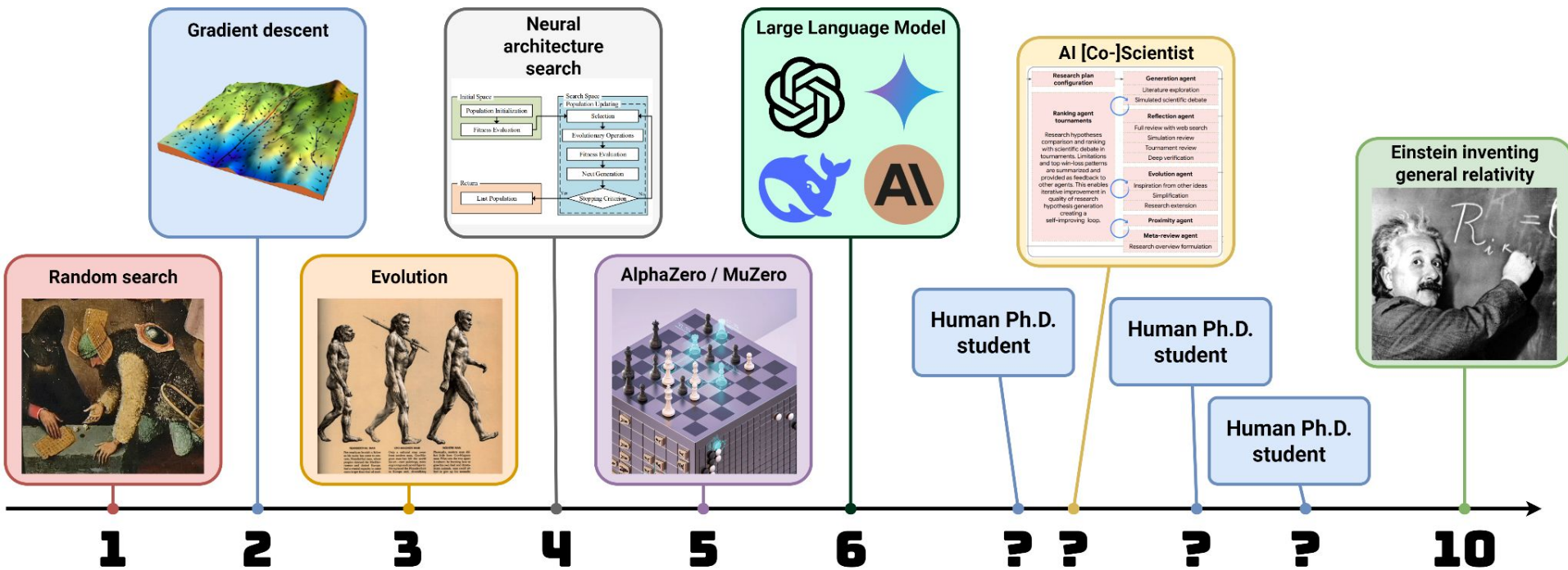
July 2026

$((1+xy)^3 z + y^2 (1+xy) (4+3xy), y + 3x (1+xy)^2 z + 3xy^2 (4+3xy), 2x - 3x^2 y - x^3 z): \mathbb{C}^3 \rightarrow \mathbb{C}^3$, has jacobian determinant -2 , and sends $(0, 0, -1/4), (1, -3/2, 13/2)$, and $(-1, 3/2, 13/2)$ to $(-1/4, 0, 0)$

5:19 AM · Jul 20, 2026 · 21.9M Views

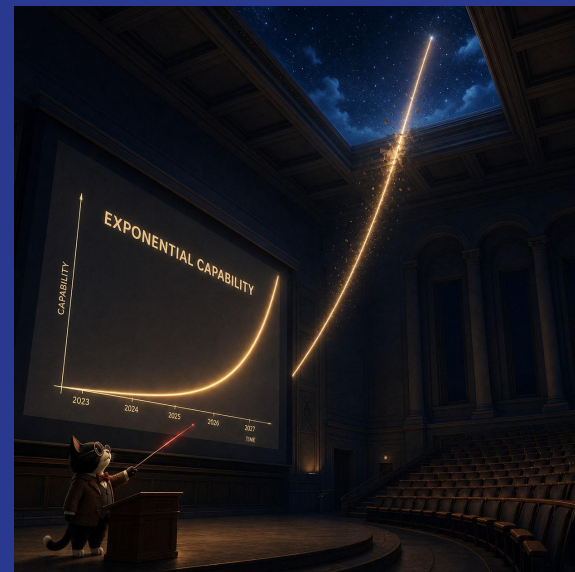
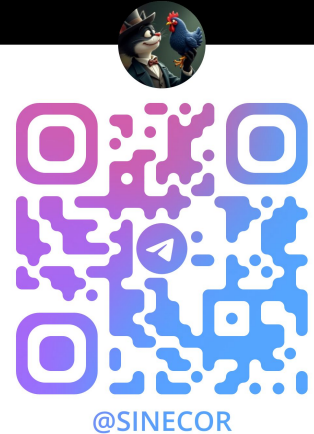
Уровни креативности

- А что на самом деле, без лишнего скептицизма и без шапкозакидательства?
Научный поиск – это тоже оптимизационный процесс; где мы сейчас?



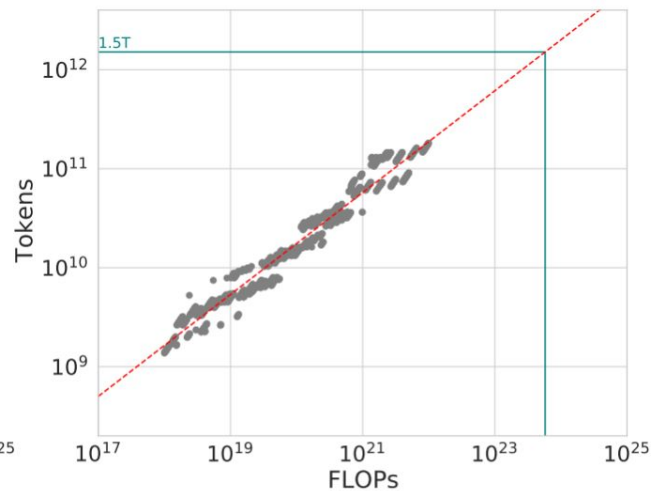
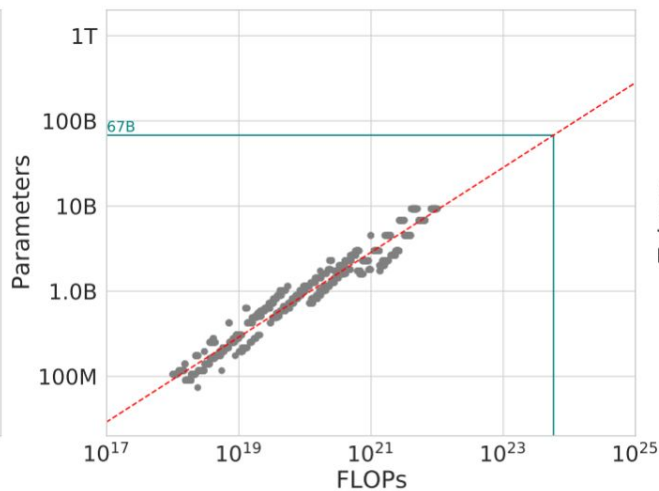
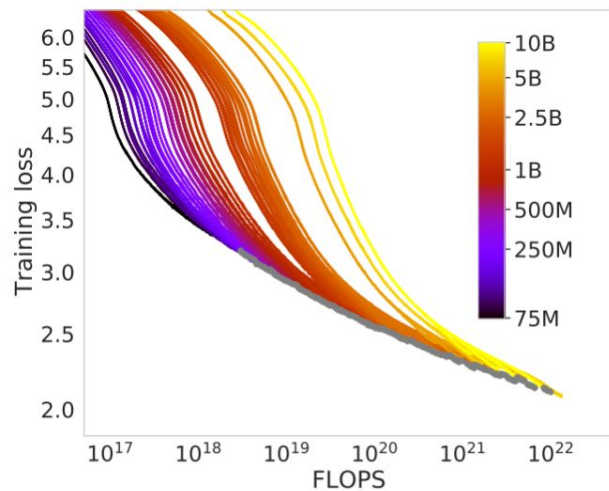


Спасибо за внимание!



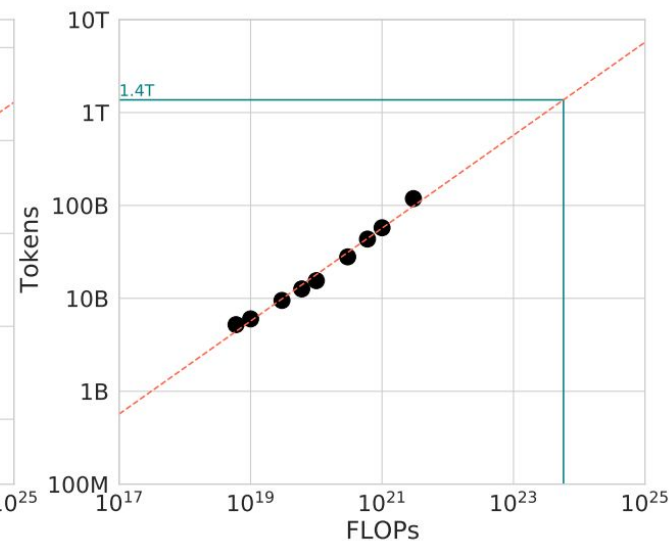
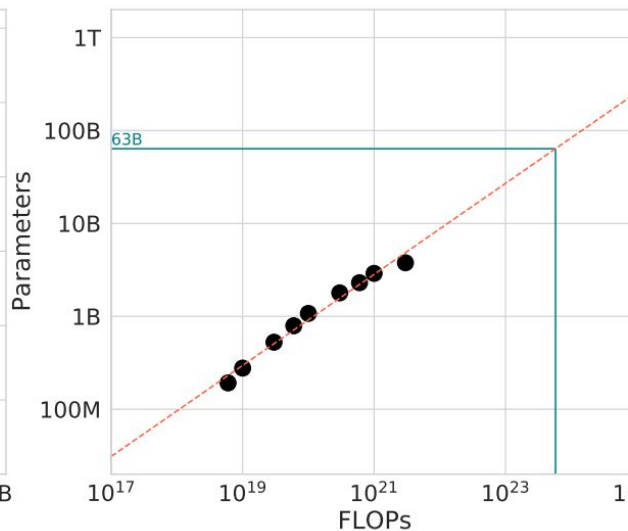
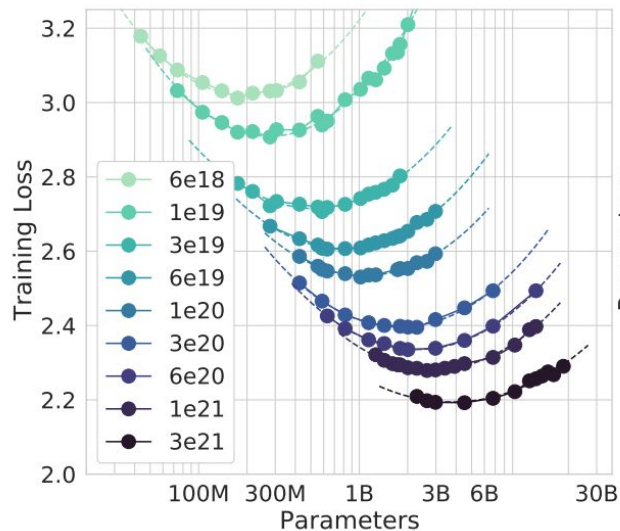
Chinchilla

- [Hoffmann et al., 2022](#): три подхода
- Первый: зафиксируем размер модели и будем менять датасет. Получился закон, который предсказывает оптимальные размеры моделей куда меньше



Chinchilla

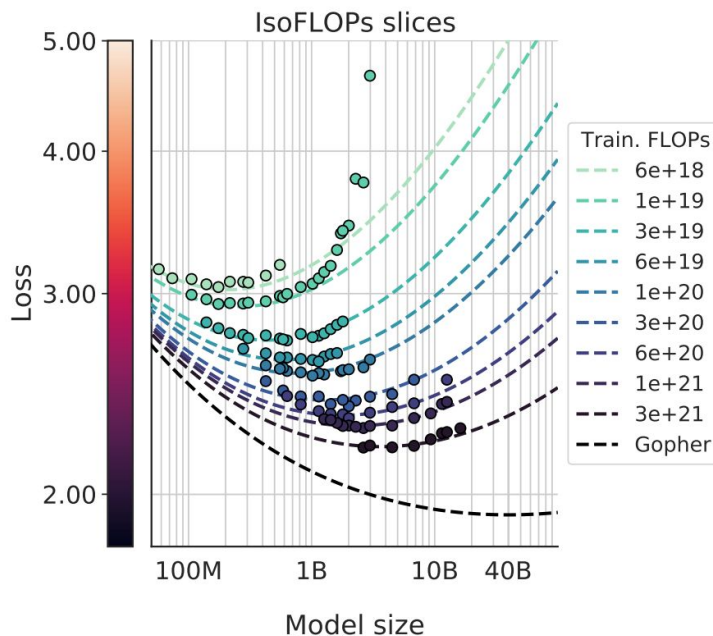
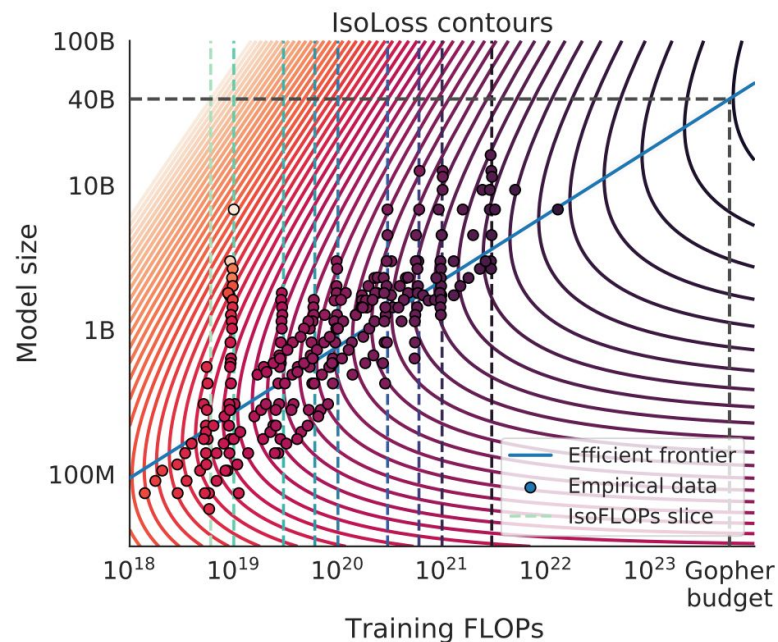
- [Hoffmann et al., 2022](#): три подхода
- Второй: IsoFLOP profiles; фиксируем бюджет и ищем оптимум



Chinchilla

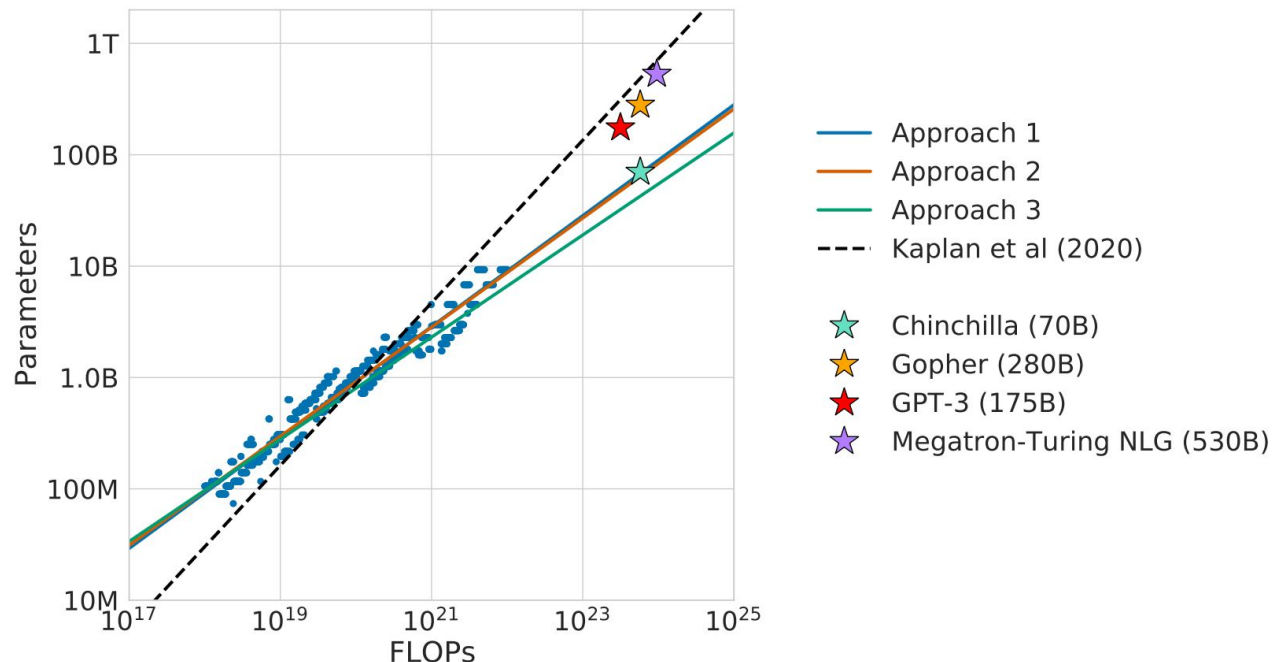
- [Hoffmann et al., 2022](#): три подхода

- Третий: обучим параметры закона $\hat{L}(N, D) \triangleq E + \frac{A}{N^\alpha} + \frac{B}{D^\beta}$.



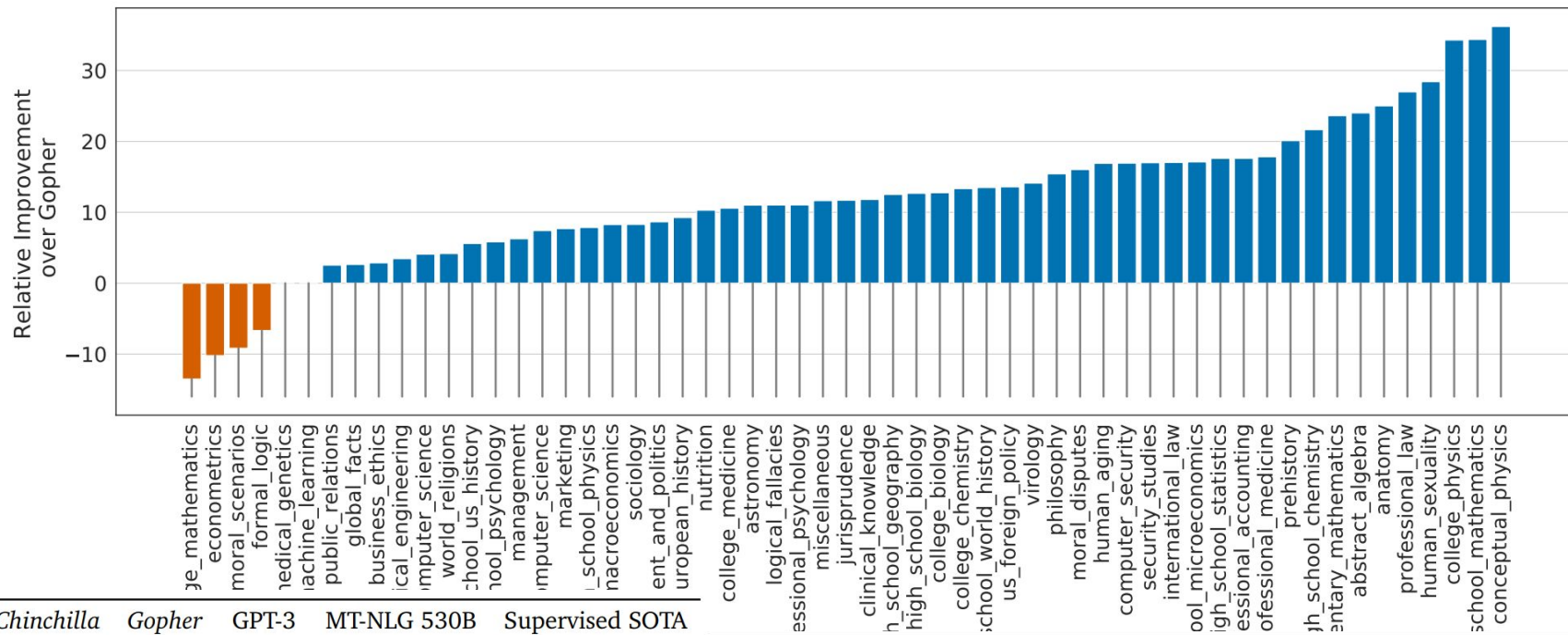
Chinchilla

- [Hoffmann et al., 2022](#): в любом случае получалось, что $N \propto C^{0.5}$, $D \propto C^{0.5}$, то есть модели вроде Gopher (280B параметров, датасет в 300B токенов) далеки от фронта, а надо 70B модель на 1.4T токенов обучать
- И действительно, получилось лучше



Chinchilla

- [Hoffmann et al., 2022](#): и действительно, получилось лучше



	Chinchilla	Gopher	GPT-3	MT-NLG 530B	Supervised SOTA
HellaSWAG	80.8%	79.2%	78.9%	80.2%	93.9%
PIQA	81.8%	81.8%	81.0%	82.0%	90.1%
Winogrande	74.9%	70.1%	70.2%	73.0%	91.3%
SIQA	51.3%	50.6%	-	-	83.2%
BoolQ	83.7%	79.3%	60.5%	78.2%	91.4%

	Chinchilla	Gopher	GPT-3	MT-NLG 530B
LAMBADA Zero-Shot	77.4	74.5	76.2	76.6
RACE-m Few-Shot	86.8	75.1	58.1	-
RACE-h Few-Shot	82.3	71.6	46.8	47.9