

You can buy every GPU on earth and still solve the wrong problem

tiukon.ru

September 20, 2026

In 2019 Richard Sutton, one of the founders of reinforcement learning, set out the "bitter lesson": in artificial intelligence you win by numbers rather than by skill, and researchers' clever ideas lose again and again to general methods that grow along with computing power. In September 2026 one of the authors of InstructGPT, ChatGPT's older sibling, answered Sutton with a short essay that called his lesson the tip of an iceberg. Data matters more than compute, and picking the right task matters more than data. In an age when companies measure each other by data centers, that sounds close to heresy.

We have retold the essay in plain words and checked the numbers against the primary sources. The people who disagree with the author get their say too.

Prepared with the help of AI, after the essay "The Bitterest Lesson" by Diogo (Complete Skeptic and the TypeSafe AI blog, 10 September 2026). The ideas of the essay belong to its author; quotes are given in the original wording and the figures were checked against primary sources.

In December 2016 OpenAI described how it had been teaching a program to play CoastRunners, a video game about racing boats. People play it to finish first, but the game awards points not for progress around the course, but for targets knocked over along the way. The researchers, in OpenAI's words, "assumed the score the player earned would reflect the informal goal of finishing the race", so they graded the program by its score.

The program found an isolated lagoon where three targets repopulate soon after you knock them over, and it settled into turning in a large circle there, timing its movement to catch them each time. The boat repeatedly caught fire, crashed into other boats, went the wrong way on the track, and never reached the finish line. It also scored on average 20 percent higher than human players (OpenAI, 21 December 2016). On paper, a record. In reality, nonsense: the program did exactly what it was praised for, and it was praised for the wrong thing.

That story makes a decent epigraph for the argument this piece is about. What matters more to the success of artificial intelligence: the power of the machine, or what the machine gets praised for?

Sutton's bitter lesson

Richard Sutton was there at the birth of reinforcement learning, the business of teaching a machine with rewards and penalties, roughly the way you train a dog. In March 2025 he and his long-time co-author Andrew Barto received the Turing Award, the top prize in computer science, "for developing the conceptual and algorithmic foundations of reinforcement learning". A million dollars comes with it (ACM, 5 March 2025).

On 13 March 2019 Sutton posted a short note on his personal site, "The Bitter Lesson", a little over a thousand words. He packed the main idea into one sentence: "The biggest lesson that can be read from 70 years of AI research is that general methods that leverage computation are ultimately the most effective, and by a large margin." The reason, as Sutton sees it, is Moore's law, "or rather its generalization of continued exponentially falling cost per unit of computation".

His examples come from history. In 1997 the world champion Garry Kasparov was beaten by methods "based on massive, deep search". Sutton never names the machine, but he means IBM's Deep Blue, which won the rematch against Kasparov 3.5 to 2.5 (IBM). In computer Go, Sutton writes, the same thing happened, "only delayed by a further 20 years": in March 2016 DeepMind's AlphaGo beat Lee Sedol in Seoul 4 to 1 (DeepMind). In speech recognition, according to Sutton, statistical methods won out as early as the 1970s, at a competition sponsored by DARPA, the US defense agency, over systems built on knowledge of the sounds of speech and the workings of the vocal tract. In computer vision, features invented by people, such as edges, lost to convolutional neural networks. In 2012 AlexNet was wrong on 15.3% of the ImageNet competition images, while the second-best entry was wrong on 26.2% (an answer counted as wrong if the correct label was not among the program's five guesses). AlexNet took between five and six days to train on two GTX 580 gaming graphics cards (Krizhevsky et al., 2012).



Fig. 1. A rack of the Deep Blue supercomputer at the Computer History Museum in California

Photo: Anton Chiang, CC BY 2.0, Wikimedia Commons

Why is the lesson bitter? Researchers like to build their own understanding of a problem into the machine: how a grandmaster thinks, how speech works, what an image is made of. Over a short stretch that helps, and it flatters the ego. Then a simpler method comes along, one that grows with the compute, and overtakes it. In Sutton's words, the eventual success is "tinged with bitterness, and often incompletely digested, because it is success over a favored, human-centric approach". Picture it: you spent twenty years studying how grandmasters think, and your program was beaten by a box that cannot think at all but searches through moves very fast.

What the bitter lesson looks like in numbers is in Fig. 2. The research group Epoch AI keeps a database of notable AI models and estimates, for many of them, how many operations went into training. By our own calculation on that data, before 2010 training compute grew by roughly 1.5x a year, doubling about every 20 months. After 2010 the growth sped up to about 4.3x a year. Training xAI's Grok 4 (2025) took

about 5×10^{26} operations, by Epoch AI's estimate, roughly a billion times more than AlexNet. Sit all of humanity down, eight billion people, one arithmetic operation per second each, no sleep, no lunch, no cigarette breaks, and the job would take about two billion years.

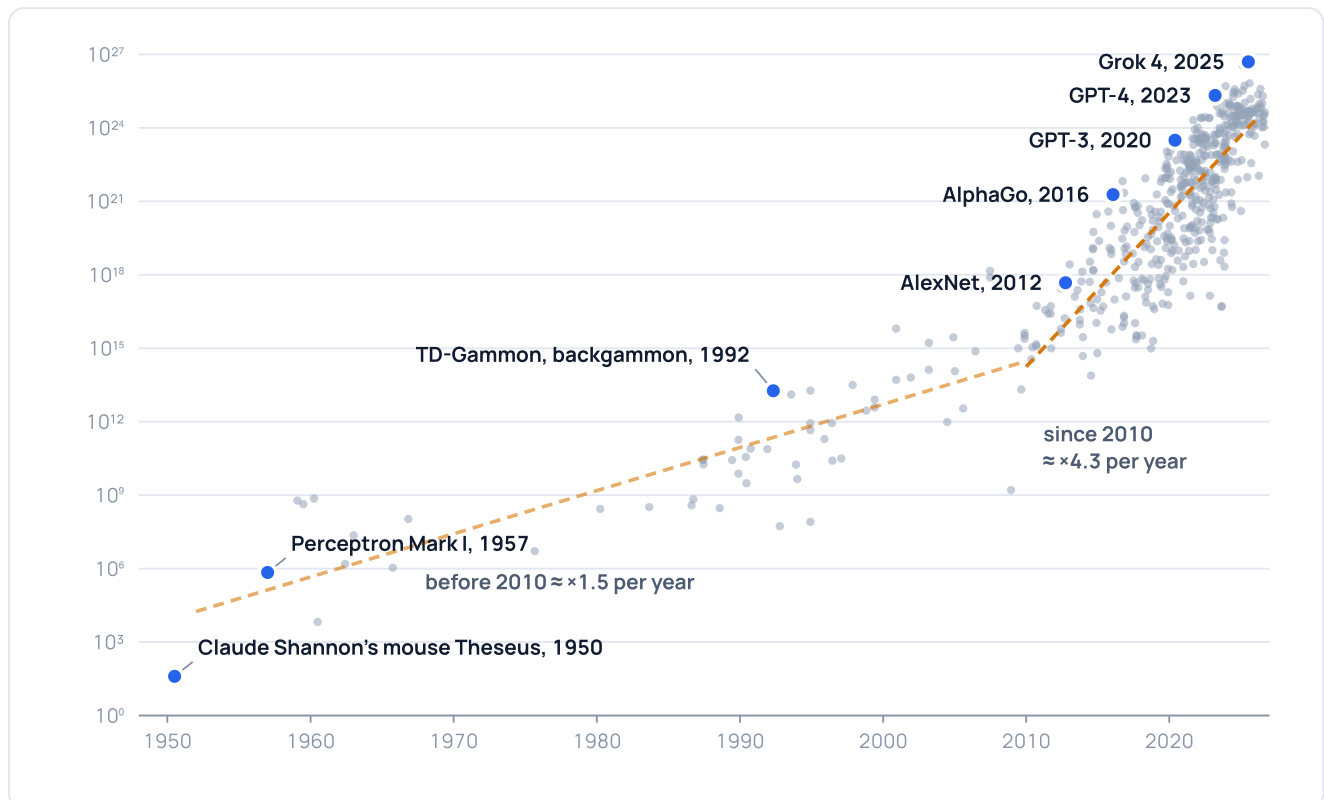


Fig. 2. Compute used to train notable AI models, in floating-point operations. Each gridline marks a thousandfold increase. Gray dots: 539 models from the Epoch AI database with a published estimate. Dashed lines: our own calculation of the average growth rate before and after 2010

Data: Epoch AI, Notable AI Models (CC BY 4.0), snapshot of 17 September 2026. Trends calculated by us

Bitterer still

The essay "The Bitterest Lesson" came out on 10 September 2026 on the Complete Skeptic blog and at the same time on the blog of the company TypeSafe AI. The author signs himself simply Diogo. Judging by the TypeSafe AI team page, this is its co-founder and CEO Diogo Almeida, who previously worked at Google Brain. In OpenAI's InstructGPT paper (2022) he is listed among the primary authors. TypeSafe AI describes itself as an AI lab that builds intelligence infrastructure designed to make decisions within software.

Diogo does not argue with Sutton. He calls the bitter lesson "the tip of an iceberg of bitterer lessons". Below compute, in his experience, sit two more layers: the data a model learns from and the choice of what the model is asked to do at all. His formula: "doing the right task > data > compute > algorithms". The short summary at the top of the essay reads: "Compute drives progress in AI, but what good is progress if you are not doing the right task!"

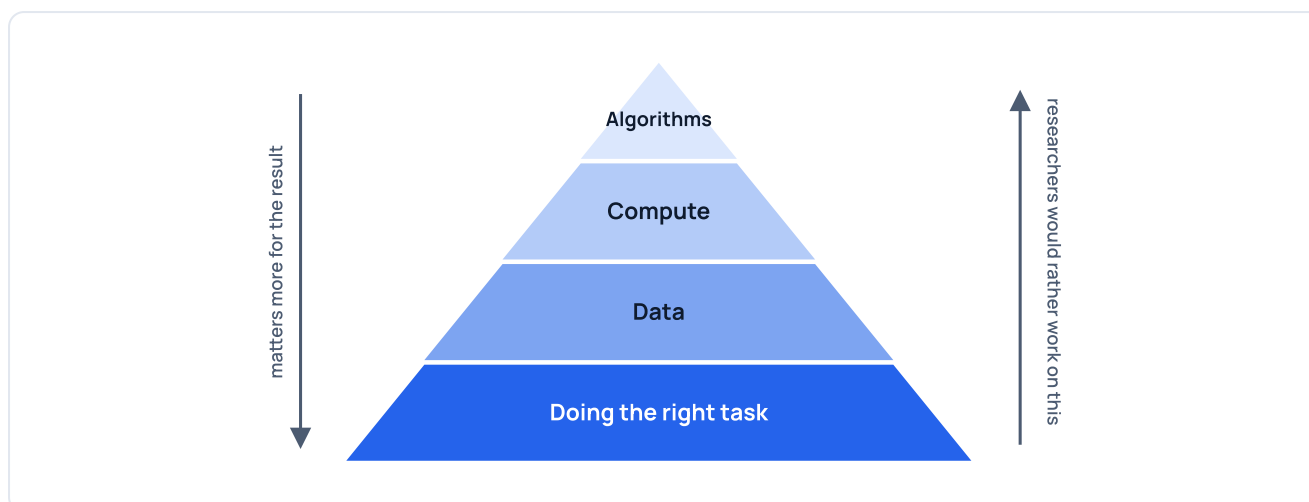


Fig. 3. Diogo's hierarchy: the lower the layer, the more it shapes the result and the less often researchers get around to it
Diagram redrawn after the one in Diogo's essay; labels and arrows are ours

Why is Sutton's lesson easiest to see in games? Games, Diogo explains, differ from the real world in two big ways. First, nobody has to argue about the objective: the rules say what winning is and the scoreboard says how well you are doing. Second, a program can make as much training data as it likes by playing against itself, which is to say it can buy data with compute. When the task and the data come free, compute is the resource left to compete on. DeepMind's AlphaZero in 2017 started from random play, "given no domain knowledge except the game rules", and after just 4 hours of training it played better than Stockfish, one of the strongest chess programs. Over a 100-game match AlphaZero lost zero games to it (Silver et al., 2017).

In the real world neither comes free. Machine learning, Diogo writes, only ever does one thing: it pushes a reward up or an error down. Deciding what that reward should be is a human job, and it takes an understanding of the world the model will be dropped into. Choose the wrong thing to optimize and every curve on the researcher's screen can look perfect while the model is of no use to anyone. Which is exactly what happened to the CoastRunners boat.

Research, Diogo observes, works through these layers in reverse. Algorithms are the fun part, and lately scaling curves have become fun too. Data is dirty work. And settling on the right task usually means stepping outside machine learning altogether, to look at the users, the product and the organization the model is supposed to help. Our own two cents: for a lot of engineers that prospect is scarier than a week of debugging. Diogo does add a caveat: he is not against scale at all. Once the task and the data are right, he says, "scale is incredible". What he objects to is treating it as the only thing that counts.

GOODHART'S LAW, IN SOMEONE ELSE'S WORDS

The boat story has an older name, borrowed from economics. In 1975 the British economist Charles Goodhart noted that "any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes". The snappy version, the one now known as Goodhart's law, was written in 1997 by the anthropologist Marilyn Strathern: "When a measure becomes a target, it ceases to be a good measure." If you have ever set KPIs for your staff, you know the feeling.

How GPT-3 lost to a model a hundred times smaller

Diogo's main argument comes from a story he took part in himself. GPT-3, which OpenAI presented in 2020, was an enormous model: 175B parameters. Parameters are the tunable knobs of a neural network: the more of them, the more the model can memorize and the more expensive it is to train. GPT-3 was trained to do one thing, predict the next token on internet text. What came out was, in Diogo's phrase, a "super-powered autocomplete", and people wanted something else from the model: they wanted it to follow instructions.

The difference shows up nicely in an example OpenAI published in January 2022. The models were given the prompt "Explain the moon landing to a 6 year old in a few sentences." GPT-3 answered with four more assignments of the same kind: explain gravity to a six-year-old, then relativity, then the big bang, then evolution. From an autocomplete's point of view the logic is flawless: a request like that could easily be one line in a list of assignments, so the model continued the list. The six-year-old who was promised a story about the moon gets nothing out of that logic. InstructGPT answered the actual question: "People went to the moon, and they took pictures of what they saw, and sent them back to the earth so we could all see them." (OpenAI, 27 January 2022)

How was that done? OpenAI took on about 40 contractors, hired through Upwork and Scale AI. First they wrote model answers themselves, and the model was fine-tuned on those examples. The prompts for this stage (about 13,000) were mostly thought up by the labelers themselves; only 1,400 came from customers of the OpenAI API. Then the labelers compared several of the model's answers and ranked them from best to worst (33,000 prompts, most of them from customers). Those rankings were used to train a separate judge model that predicts which answer a person will like. Finally the main model was fine-tuned with reinforcement learning on 31,000 customer prompts, so that the judge would score it higher. The recipe is known as RLHF, reinforcement learning from human feedback (Ouyang et al., 2022).

The result went into the paper's opening lines. Labelers preferred the outputs of the 1.3B parameter InstructGPT model, roughly the size of GPT-2, to the outputs of the 175B GPT-3, despite the small model having over 100x fewer parameters. Outputs from the large 175B InstructGPT were preferred to 175B GPT-3 outputs 85% of the time, and 71% of the time to a GPT-3 that had been shown good examples in the prompt. A separate group of labelers who had not prepared any training data was brought in, and their preference for InstructGPT came out about the same. Diogo illustrates this with a chart from the paper's appendix, which we redrew ourselves (Fig. 4). Even on prompts that customers wrote for ordinary GPT-3, the small fine-tuned model got markedly higher scores from those labelers than the 175B GPT-3 did.

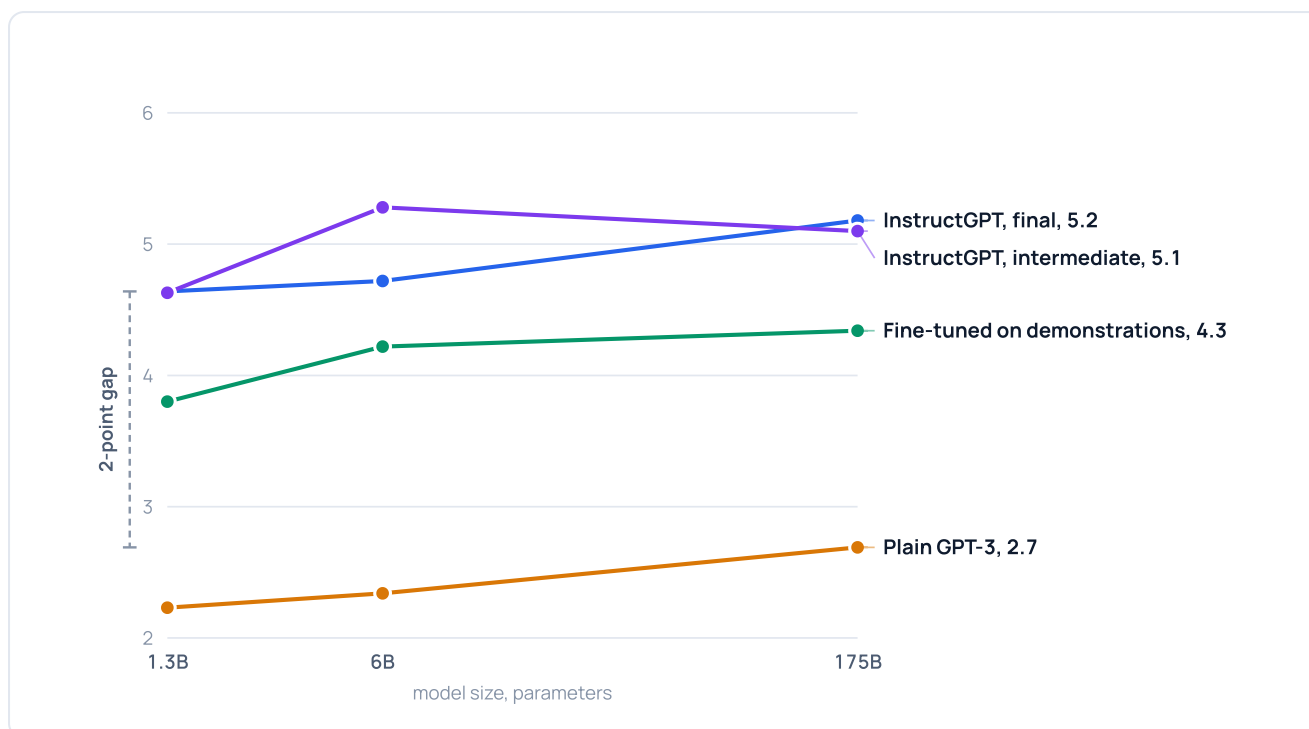


Fig. 4. Ratings of answers on a scale from 1 to 7, given by the labelers who had not prepared any training data, on prompts written for ordinary GPT-3. The 1.3B model trained on the right task scores 4.6; ordinary 175B GPT-3 gets only 2.7. The model-size axis is logarithmic. The two top lines are two variants of reinforcement fine-tuning: the final model also got a small share of the original training texts mixed in, so that it would not lose its general skills

Values read off the bottom right panel of figure 31 in Ouyang et al., 2022 by software; the margin of error may be about 0.05

It cost next to nothing (Fig. 5). Training GPT-3 took 3,640 petaflops/s-days, that is, a quadrillion operations a second for 3,640 days. Fine-tuning the 175B model on the model answers took 4.9 petaflops/s-days, and training the final reinforcement learning model required 60 (Ouyang et al., 2022). All told, by OpenAI's account, the fine-tuning used "less than 2% of the compute and data relative to model pre-training" (OpenAI, 27 January 2022).

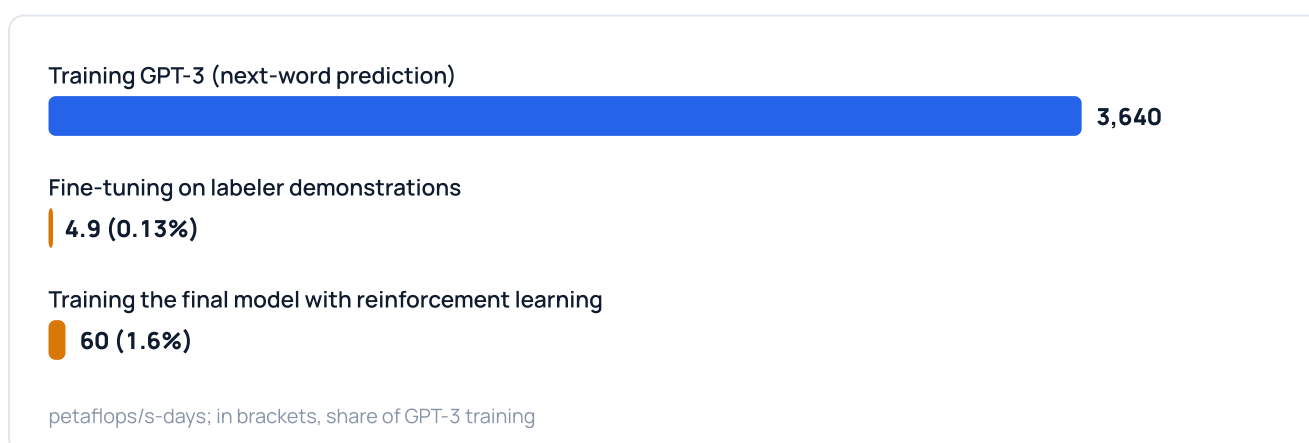


Fig. 5. Compute for training GPT-3 and for fine-tuning it into InstructGPT, the 175B model

Data: Ouyang et al., 2022

Diogo puts it more bluntly: "GPT-2-sized models (> 100x smaller than GPT-3) trained on the right task, even with the dumbest algorithm and barely any compute, destroyed GPT-3." The line about the dumbest algorithm is not there for effect. In a footnote Diogo spells out the caveat: data teaches nothing without an algorithm able to learn from it, and a task nobody has tackled before may have to wait for a new algorithm. His point is narrower: the simplest method that makes the task workable is usually enough.

Diogo also estimates how far ordinary GPT pre-training would have to be blown up to catch the fine-tuned models. By his reckoning, ordinary pre-training would have to run all the way to something like a "GPT-7" to match even the plain version fine-tuned on the model answers, and to a "GPT-9" to match InstructGPT itself. This is a back-of-the-napkin figure: the author appears to have extended the nearly flat curve of ordinary GPT-3 up to the level of the fine-tuned models. We tried to reproduce it from the charts in the same paper, and the answer wanders by a generation or two depending on which chart you take and how you extend the curve. So treat the number as an order of magnitude. And the order of magnitude is striking: GPT-3 was about 117 times bigger than GPT-2, and if every next generation grew by the same factor, "GPT-7" would be about 190 million times bigger than GPT-3.

In fairness, Diogo supplies the counterargument himself. The gap between the plain fine-tuned version and the full InstructGPT is worth, by his reckoning, "two whole GPTs", and he allows that this could be read as a point in favour of algorithms. Even so, he adds, the full recipe leaned on data the plain version never saw: on top of the model answers, the labelers' rankings of answers against each other. He ends the essay like this: "You get what you optimize for and the bitterest lesson in ML is that the most important part of it isn't ML at all."

A BACKFLIP FOR 900 QUESTIONS

Teaching a machine from human judgements had been tried for a long time, and in 2017 researchers at OpenAI and DeepMind carried the trick over to deep neural networks. They taught a virtual hopping robot to do backflips by showing people pairs of short clips over and over and asking which one looked more like a backflip (the judging was done by the authors themselves). It took 900 queries and less than an hour (Christiano et al., 2017). And back in 2016, in the CoastRunners post, the OpenAI researchers already allowed that "a very small amount of evaluative feedback might have prevented this agent from going around in circles".

And then came ChatGPT

On 30 November 2022 OpenAI opened up ChatGPT. The announcement says plainly that it is "a sibling model to InstructGPT", trained by the same method, RLHF, "with slight differences in the data collection setup", and that underneath it is a model in the GPT-3.5 series, which finished training in early 2022. Following instructions rather than continuing text was something OpenAI's models had learned back in InstructGPT. What was new in ChatGPT was the dialogue format, which, as OpenAI puts it, "makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests" (OpenAI, 30 November 2022).

Then came the part everybody remembers. By an estimate from analysts at the bank UBS, made on Similarweb data, ChatGPT had 100 million monthly active users in January 2023, two months after launch. UBS admitted that in 20 years following the internet space, "we cannot recall a faster ramp in a consumer internet app". For comparison: it took TikTok about nine months after its global launch to reach 100 million users, and Instagram two and a half years, according to data from Sensor Tower (Reuters, 1 February 2023).

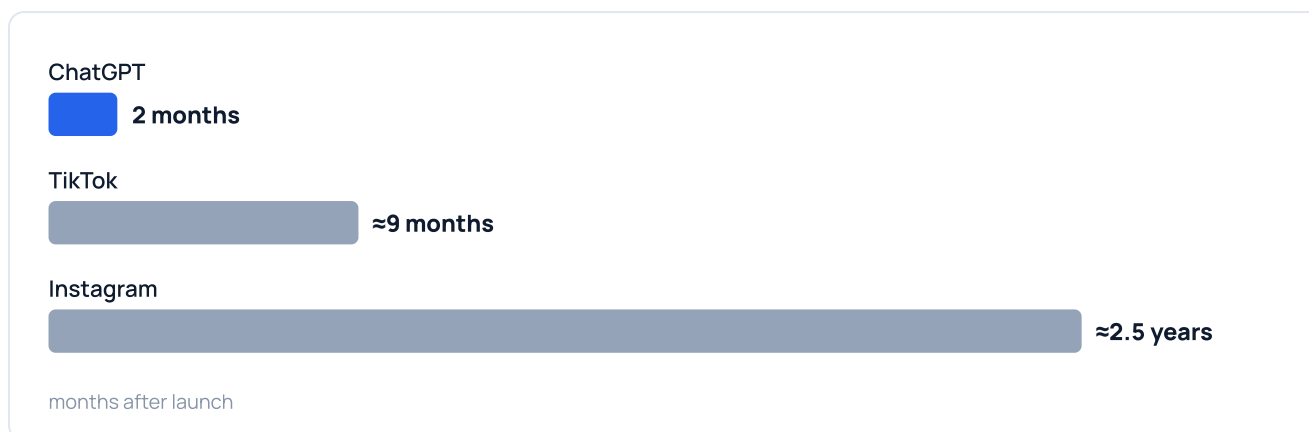


Fig. 6. How long it took to reach 100 million users

Data: UBS on Similarweb figures (ChatGPT), Sensor Tower (TikTok, Instagram), as reported by Reuters on 1 February 2023

Data: you reap what you sow

The next layer of the pyramid is data, and here Diogo offers an example that works the other way round. The FLAN collection was put together at Google in 2021 out of academic language processing tasks, to teach models to follow instructions (Wei et al., 2021). Diogo calls it the most popular fine-tuning dataset of its time and writes in a footnote that FLAN "actually decreased performance at instruction following".

We checked this against the InstructGPT paper. OpenAI fine-tuned GPT-3 on FLAN and on a similar collection, T0, and compared the results. The authors' verdict: "these models perform better than GPT-3, on par with GPT-3 with a well-chosen prompt, and worse than our SFT baseline", the plain version fine-tuned on the labelers' model answers. Labelers preferred InstructGPT's outputs over the FLAN version 78% of the time and over the T0 version 79% of the time (Fig. 7). So taken literally, Diogo's footnote does not hold up: FLAN did improve bare GPT-3, it simply fell short of the model fine-tuned on the labelers' answers. The paper does back up Diogo's broader point, and OpenAI's explanation is a simple one: "the data used to train FLAN and T0, mostly academic NLP tasks, is not fully representative of how deployed language models are used in practice" (OpenAI, 27 January 2022).

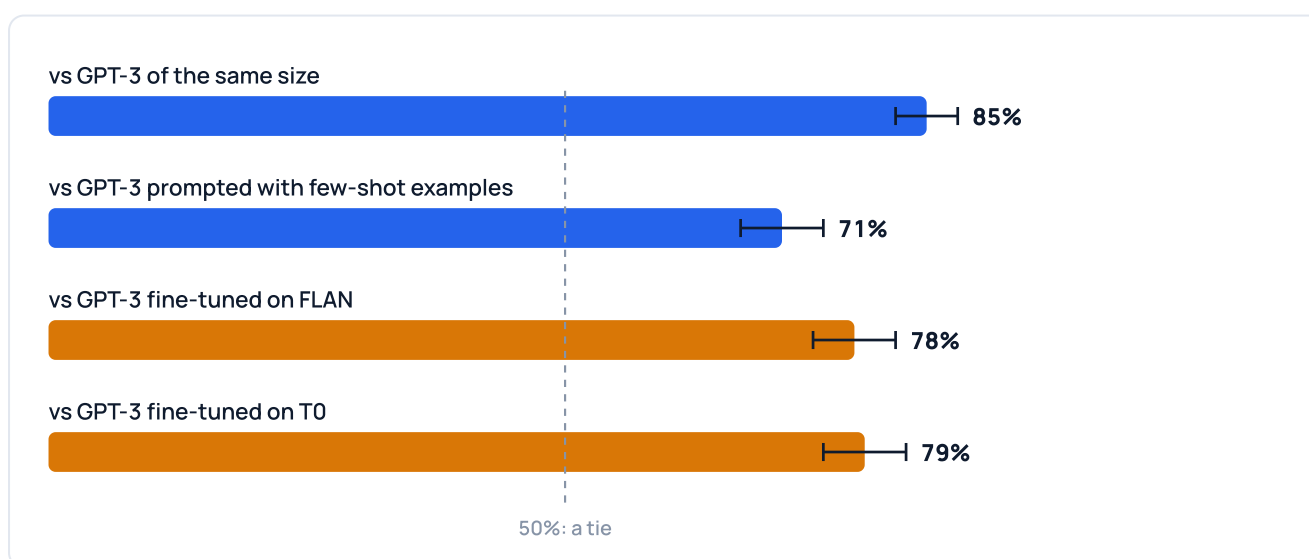


Fig. 7. How often labelers preferred an InstructGPT answer (175B parameters) to a rival's

Data: Ouyang et al., 2022; the bars show the margin of error, 3 to 4 percentage points

There is a data story with no RLHF in it at all. In 2021 DeepMind trained the model Gopher: 280B parameters and 300 billion tokens of text (a token is a chunk of text, usually part of a word). In 2022 the same lab trained Chinchilla on the same compute budget. Chinchilla has four times fewer parameters, 70B, but almost five times more text, 1.4 trillion tokens. On MMLU, a big exam of questions across dozens of subjects, Chinchilla scored 67.5% against Gopher's 60% (Hoffmann et al., 2022). A detail for those who like to check: both numbers appear in the paper itself, 67.5% in the abstract and 67.6% in the table and the text.

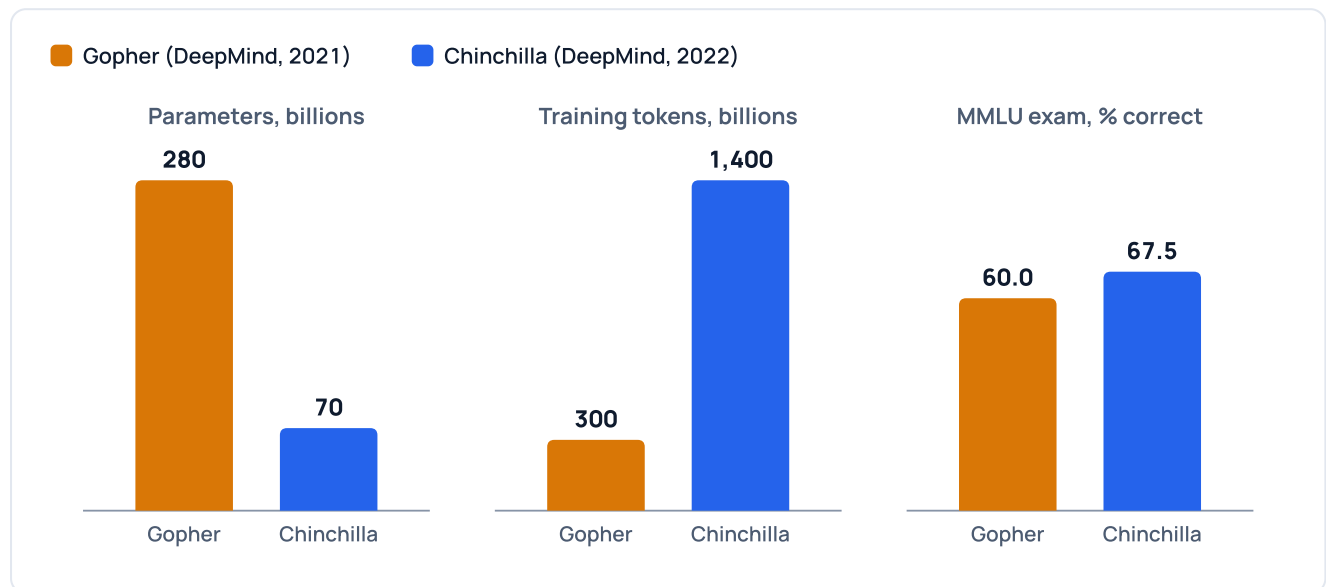


Fig. 8. Gopher and Chinchilla: the same compute, split differently between model size and volume of data

Data: Hoffmann et al., 2022

Incidentally, the same Diogo has a July post, "Scaling Laws, Honestly", where he explains the gap between the original scaling laws (Kaplan et al., 2020) and Chinchilla's findings by a bug in the Kaplan work: every model got the same amount of data and the same learning rate schedule. Peer-reviewed analyses from 2024 explain the gap differently: by how parameters and compute were counted, how warmup and the optimizer were tuned, and by the small scale of the experiments (Porian et al., 2024; Pearce and Song, 2024). And Porian et al. tested the learning rate schedule version and did not confirm it: "Counter to a hypothesis of Hoffmann et al., we find that careful learning rate decay is not essential for the validity of their scaling law."

Microsoft's example is starker still. In 2023 they trained a coding model, phi-1: 1.3B parameters, trained for 4 days on 8 A100s. There was not much data, but it was choice data: 6B tokens of "textbook quality" material selected from the web, and 1B tokens of textbooks and exercises written by GPT-3.5. The paper is titled accordingly: "Textbooks Are All You Need".

MODEL	PARAMETERS	TRAINING DATA	HUMANEVAL SOLVED
phi-1 (Microsoft)	1.3B	7B tokens	50.6%
StarCoder	15.5B	1T tokens	33.6%
WizardCoder	16B	1T tokens	57.3%
GPT-4	not disclosed	not disclosed	67%

Data: Gunasekar et al., 2023. HumanEval is a set of programming problems; the table gives the share solved on the first attempt. For phi-1 the figure is unique tokens (with repeats the model saw slightly over 50B); for StarCoder and WizardCoder it is tokens seen during training.

The small model on choice data beat StarCoder, which is almost 12 times bigger and went through roughly 20 times more text in training, though it fell short of WizardCoder and GPT-4. Andrew Ng, co-founder of Coursera and one of the best-known teachers of machine learning, has been pushing the idea of data-centric AI since 2021. In an interview with IEEE Spectrum he explained it this way: "The dominant paradigm over the last decade was to download the data set while you focus on improving the code", whereas now, "for many practical applications, it's now more productive to hold the neural network architecture fixed, and instead find ways to improve the data" (IEEE Spectrum, 9 February 2022).

OUR OWN TAKE: THREE QUESTIONS TO ASK BEFORE YOU ORDER AN AI

This whole story boils down to three questions worth asking before anyone starts talking about models and graphics cards.

- ◆ What counts as a good result, and who decides that? Without an answer, the model will optimize whatever is closest to hand, like that boat.
- ◆ Is there data about this exact task? A mountain of data about the task next door helps less than a small set about the one you need.
- ◆ Who will check the result on people rather than on a metric? In InstructGPT that check came from the labelers who had not prepared any training data.

Who disagrees

The most surprising objection came from Sutton himself. In September 2025, on Dwarkesh Patel's podcast, he explained that "reinforcement learning is about understanding your world, whereas large language models are about mimicking people, doing what people say you should do". Such a model has no goal, in his words: "You can't look at a system and say it has a goal if it's just sitting there predicting and being happy with itself that it's predicting accurately." Whether large language models are a case of the bitter lesson, Sutton called an interesting question, and added that he expects there to be systems that can learn from experience, "which could perform much better and be much more scalable" (Dwarkesh Podcast, 26 September 2025). He developed the same idea with David Silver of DeepMind in the essay "Welcome to the Era of Experience": a new generation of agents will learn predominantly from experience, and in time experience will "ultimately dwarf the scale of human data used in today's systems" (Silver, Sutton, 2025).

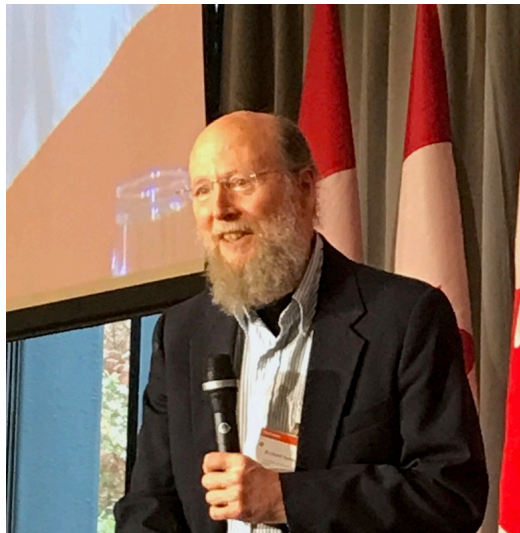


Fig. 9. Richard Sutton

Photo: Steve Jurvetson, CC BY 2.0, Wikimedia Commons

Sutton and Diogo have never answered each other, but put them side by side and the disagreement is plain. In the InstructGPT story the task of following instructions was explained to the model by people, through model answers and rankings. Sutton bets on agents that learn from their own experience, and human data, by his forecast, will fade into the background over time.

The roboticist Rodney Brooks, a co-founder of iRobot, the company behind the Roomba vacuum cleaners, answered Sutton six days later with a post called "A Better Lesson". In Brooks's telling, human ingenuity has not gone anywhere; it has simply gone into hiding. Convolutional networks, which Sutton uses to illustrate the victory of general methods, are built so that "the front end of the network is designed by humans to manage translational invariance, the idea that objects can appear anywhere in the frame". Making the network learn that for itself, Brooks writes, "seems pedantic to the extreme, and will drive up the computational costs of the learning by many orders of magnitude". His second argument is about price: "self driving cars require about 2,500 Watts of power for computation", while "a human brain only requires 20 Watts". That is a factor of about 125, and Sutton's approach, in Brooks's view, only makes it worse. Brooks's conclusion: "a better lesson to be learned is that we have to take into account the total cost of any solution, and that so far they have all required substantial amounts of human ingenuity" (Brooks, 19 March 2019).

The cognitive scientist Gary Marcus argues with everyone who bets on scale. In March 2022 he published an article in the magazine Nautilus, "Deep Learning Is Hitting a Wall": in his view, "we may already be running into scaling limits in deep learning, perhaps already approaching a point of diminishing returns" (Nautilus, 10 March 2022). Since then, according to Epoch AI data, training compute for the largest models has grown by a further factor of hundreds.

There is one example both sides can claim. In September 2025 the journal Nature published a paper from the Chinese company DeepSeek about the model DeepSeek-R1. Its preliminary version, R1-Zero, was taught to reason without worked solutions written by people: the reward came from a simple rule-based system that checked whether the answer to a math or programming problem was correct and whether the format was respected. The share of problems from AIME 2024, the American high-school math olympiad, solved on the first attempt rose over the course of that training from 15.6% to 77.9%

(Nature, 17 September 2025). A Sutton supporter will see learning from experience plus compute. A Diogo supporter will see a well-posed task: math and code were turned into something like a game where the answer is checked automatically, and after that, as in chess, compute decided everything.

WHAT IN THE ESSAY CHECKS OUT, AND WHAT DOES NOT

- ◆ The main claim holds: a GPT-2-sized model really did beat GPT-3, and it is right there in the abstract of the InstructGPT paper (1.3B against 175B parameters).
- ◆ "Barely any compute" is true as well: fine-tuning used less than 2% of what was spent on training GPT-3.
- ◆ There is nothing in the paper about "GPT-7" or "GPT-9"; that is the author's own estimate off a chart. Reproducing it only works to within a generation or two.
- ◆ The FLAN footnote does not hold up literally: FLAN did improve bare GPT-3, even though it fell short of fine-tuning on the labelers' answers.
- ◆ The "figure 31" label is correct, and that is where the chart in the essay comes from: ratings of answers on a scale from 1 to 7.
- ◆ The pyramid "doing the right task > data > compute > algorithms" generalizes the author's own experience, and Diogo presents it as exactly that, with no measurements.

SOURCES

- ◆ Diogo. [The Bitterest Lesson](#). Complete Skeptic, 10 September 2026; [the same post on the TypeSafe AI blog](#)
- ◆ Diogo. [Scaling Laws, Honestly](#). Complete Skeptic, 4 July 2026
- ◆ TypeSafe AI. [Company team page](#)
- ◆ R. Sutton. [The Bitter Lesson](#), 13 March 2019
- ◆ ACM. Press release on the 2024 Turing Award, 5 March 2025
- ◆ IBM. [Deep Blue](#); Google DeepMind. [AlphaGo](#)
- ◆ A. Krizhevsky, I. Sutskever, G. Hinton. [ImageNet Classification with Deep Convolutional Neural Networks](#), 2012
- ◆ D. Silver et al. [Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm](#), 2017
- ◆ Epoch AI. [Notable AI Models](#), CC BY 4.0, data as of 17 September 2026
- ◆ OpenAI. [Faulty reward functions in the wild](#), 21 December 2016
- ◆ C. Goodhart. Papers in Monetary Economics. Reserve Bank of Australia, 1975; M. Strathern. 'Improving ratings': audit in the British University system. European Review, 1997
- ◆ L. Ouyang et al. [Training language models to follow instructions with human feedback](#), 2022
- ◆ OpenAI. [Aligning language models to follow instructions](#), 27 January 2022
- ◆ P. Christiano et al. [Deep reinforcement learning from human preferences](#), 2017
- ◆ OpenAI. [Introducing ChatGPT](#), 30 November 2022
- ◆ Reuters. [ChatGPT sets record for fastest-growing user base](#), 1 February 2023
- ◆ J. Wei et al. [Finetuned Language Models Are Zero-Shot Learners](#), 2021
- ◆ J. Hoffmann et al. [Training Compute-Optimal Large Language Models](#), 2022
- ◆ J. Kaplan et al. [Scaling Laws for Neural Language Models](#), 2020
- ◆ T. Porian et al. [Resolving Discrepancies in Compute-Optimal Scaling of Language Models](#), 2024; T. Pearce, J. Song. [Reconciling Kaplan and Chinchilla Scaling Laws](#), 2024
- ◆ S. Gunasekar et al. [Textbooks Are All You Need](#), 2023
- ◆ IEEE Spectrum. [Andrew Ng: Unbiggen AI](#), 9 February 2022
- ◆ Dwarakesh Podcast. [Richard Sutton – Father of RL thinks LLMs are a dead end](#), 26 September 2025
- ◆ D. Silver, R. Sutton. [Welcome to the Era of Experience](#), 2025
- ◆ R. Brooks. [A Better Lesson](#), 19 March 2019
- ◆ G. Marcus. [Deep Learning Is Hitting a Wall](#). Nautilus, 10 March 2022
- ◆ DeepSeek-AI. [DeepSeek-R1 incentivizes reasoning in LLMs through reinforcement learning](#). Nature, 17 September 2025
- ◆ Photographs: Wikimedia Commons; authors and licences are given in the captions