

Revealing OpenAI's plan to create AGI by 2027



In this document I will be revealing information I have gathered regarding OpenAI's (delayed) plans to create human-level AGI by 2027. Not all of it will be easily verifiable but hopefully there's enough evidence to convince you

Summary: OpenAI started training a 125 trillion parameter multimodal model in August of 2022. The first stage was Arrakis also called Q\*. The model finished training in December of 2023 but the launch was canceled due to high inference cost. This is the original GPT-5 which was planned for release in 2025. Gobi (GPT-4.5) has been renamed to GPT-5 because the original GPT-5 has been canceled.

The next stage of Q\*, originally GPT-6 but since renamed to GPT-7 (originally for release in 2026), has been put on hold because of the recent lawsuit by Elon Musk

Q\* 2025 (GPT-8) was planned to be released in 2027 achieving full AGI

...

Q\* 2023 = 48 IQ

Q\* 2024 = 96 IQ (delayed)

Q\* 2025 = 145 IQ (delayed)

Elon Musk caused the delay because of his lawsuit. This is why I'm revealing the information now because no further harm can be done

I've seen many definitions of AGI – artificial general intelligence – but I will define AGI simply as an artificial intelligence that can do any intellectual task a smart human can. This is how most people define the term now.

2020 was the first time I was shocked by an AI system – that was GPT-3. GPT-3.5, an upgraded version of GPT-3, is the model behind ChatGPT. When ChatGPT was released, I felt as though the wider world was finally catching up to something I was interacting with 2 years prior. I used GPT-3 extensively in 2020 and was shocked by its ability to reason.

GPT-3, and its half-step successor GPT-3.5 (which powered the now famous ChatGPT -- before it was upgraded to GPT-4 in March 2023), were a massive step towards AGI in a way that earlier models weren't. The thing to note is, earlier language models like GPT-2 (and basically all chatbots since Eliza) had no real ability to respond coherently at all. So why was GPT-3 such a massive leap?

...

# Parameter Count

“Deep learning” is a concept that essentially goes back to the beginning of AI research in the 1950s. The first neural network was created in the 50s, and modern neural networks are just “deeper”, meaning, they contain more layers – they’re much, much bigger and trained on lots more data.

Most of the major techniques used in AI today are rooted in basic 1950s research, combined with a few minor engineering solutions like “backpropagation” and “transformer models”. The overall point is that AI research hasn’t fundamentally changed in 70 years. So, there’s only two real reasons for the recent explosion of AI capabilities: size and data.

A growing number of people in the field are beginning to believe we’ve had the technical details of AGI solved for many decades, but merely didn’t have enough computing power and data to build it until the 21st century. Obviously, 21st century computers are vastly more powerful than 1950s computers. And of course, the internet is where all the data came from.

So, what is a parameter? You may already know, but to give a brief digestible summary, it’s analogous to a synapse in a biological brain, which is a connection between neurons. Each neuron in a biological brain has roughly 1000 connections to other neurons. Obviously, digital neural networks are conceptually analogous to biological brains.

...

...

So, how many synapses (or “parameters”) are in a human brain?

The most commonly cited figure for synapse count in the brain is roughly 100 trillion, which would mean each neuron (~100 billion in the human brain) has roughly 1000 connections.

Yale School of Medicine / Neuroscience /

Colón-Ramos Lab 

Home / Overview and Projects 

# Overview

The human brain consists of 100 billion neurons and over 100 trillion synaptic connections. There are more neurons in a single human brain than stars in the milky way! During development, neurons navigate this complex cellular environment and assemble into functional circuits. How the brain develops is not well understood.

If each neuron in a brain has 1000 connections, this means a cat has roughly 250 billion synapses, and a dog has 530 billion synapses. Synapse count generally seems to predict higher intelligence, with a few exceptions: for instance, elephants technically have a higher synapse count than humans yet display lower intelligence.

The simplest explanation for larger synapse counts with lower intelligence is a smaller amount of quality data. From an evolutionary perspective, brains are “trained” on billions of years of epigenetic data, and human brains evolved from higher quality socialization and communication data than elephants, leading to our superior ability to reason. Regardless, synapse count is definitely important.

Again, the explosion in AI capabilities since the early 2010s has been the result of far more computing power and far more data. GPT-2 had 1.5 billion connections, which is less than a mouse’s brain (~10 billion synapses). GPT-3 had 175 billion connections, which is getting somewhat close to a cat’s brain.

Isn’t it intuitively obvious that an AI system the size of a cat’s brain would be superior to an AI system smaller than a mouse’s brain?

...

# Predicting AI Performance

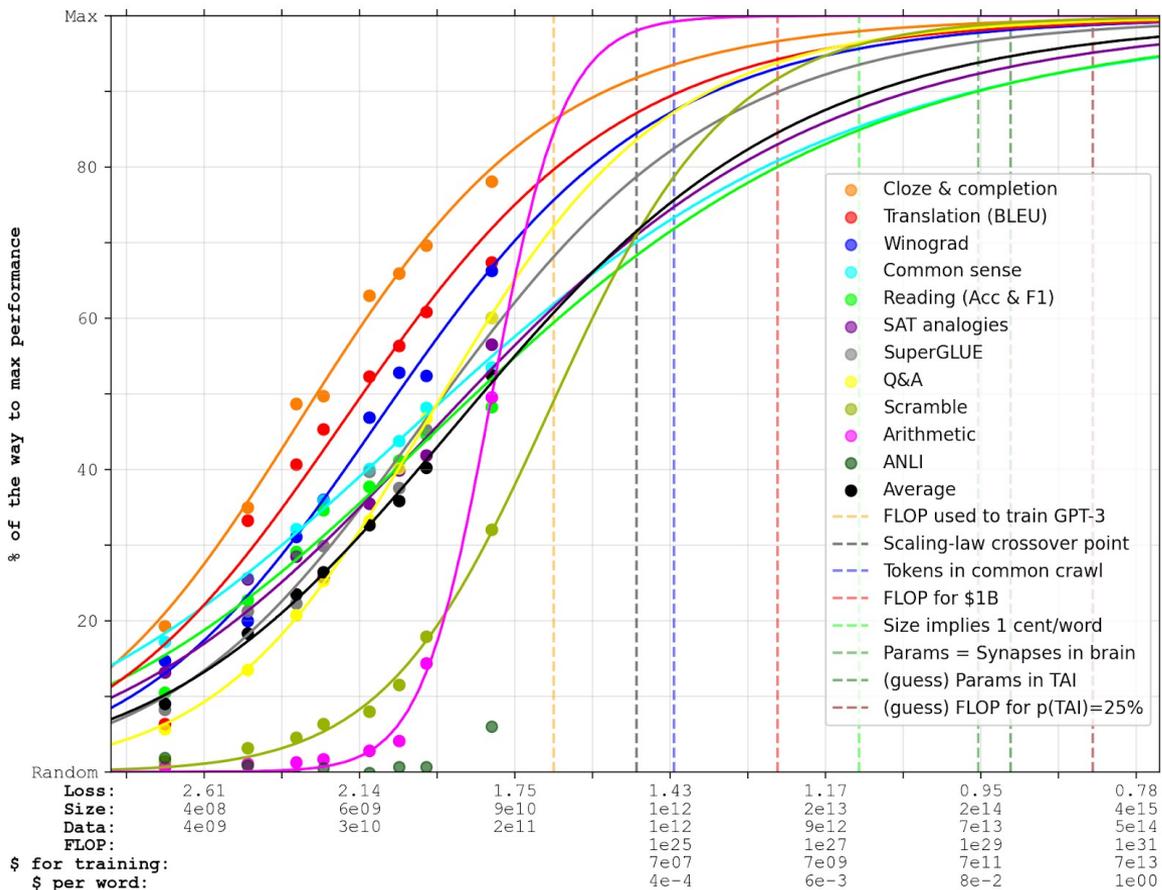
...

In 2020, after the release of the 175 billion parameter GPT-3, many speculated about the potential performance of a model ~600 times larger at 100 trillion parameters, because this parameter count would match the human brain's synapse count. There was no strong indication in 2020 that anyone was actively working on a model of this size, but it was interesting to speculate about.

The big question is, is it possible to predict AI performance by parameter count? As it turns out, the answer is yes, as you'll see on the next page.



The screenshot shows the top portion of a LessWrong article page. At the top left, there is a hamburger menu icon followed by the text 'LW'. To the right is a search icon and the text 'LOGIN'. The main title of the article is 'Extrapolating GPT-N performance' in a large, serif font. Below the title, it says 'by Lanrian', '30 min read', and '18th Dec 2020'. To the right of this information, there is a vertical stack of icons: an upward-pointing chevron, the number '103', a downward-pointing chevron, and the number '38'. Below the article information, it says '31 comments'. At the bottom of the screenshot, there is a horizontal row of five tags: 'AI Timelines', 'GPT', 'Language Models', 'AI', and 'Frontpage'.



- There are roughly  $2e^{14}$  synapses in the human brain (source), which is approximately analogous to the number of parameters in neural networks (green line).
- The dark green line marks the *median* estimate for the number of parameters in a transformative model, according to Ajeya Cotra's model<sup>[16]</sup>. Noticeably, this is quite close to when the benchmarks approaches optimal performance. The 80% confidence interval is between  $3e^{11}$  and  $1e^{18}$  parameters, going all the way from the size of GPT-3 to well beyond the edge of my graph.

[The above is from Lanrián's LessWrong post.]

...

As Lanrian illustrated, extrapolations show that AI performance inexplicably seems to reach human-level at the same time as human-level brain size is matched with parameter count. His count for the synapse number in the brain is roughly 200 trillion parameters as opposed to the commonly cited 100 trillion figure, but the point still stands, and the performance at 100 trillion parameters is remarkably close to optimal.

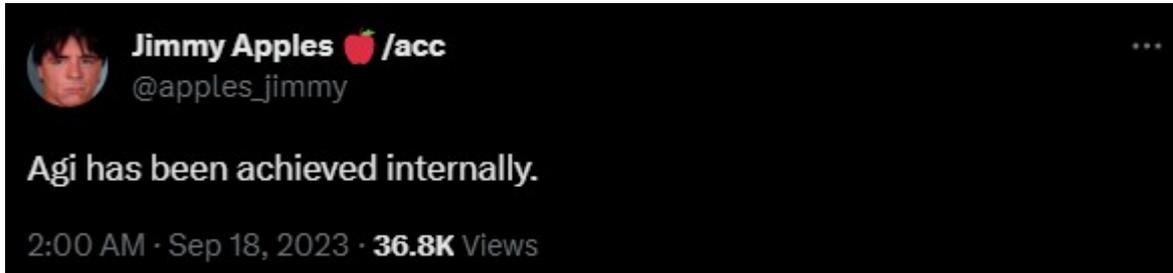
By the way – an important thing to note is that although 100 trillion is slightly suboptimal in performance, there is an engineering technique OpenAI is using to bridge this gap. I'll explain this towards the very end of the document because it's crucial to what OpenAI is building.

Lanrian's post is one of many similar posts online – it's an extrapolation of performance based on the jump between previous models. OpenAI certainly has much more detailed metrics and they've come to the same conclusion as Lanrian, as I'll show later in this document.

So, if AI performance is predictable based on parameter count, and ~100 trillion parameters is enough for human-level performance, when will a 100 trillion parameter AI model be released?

...

GPT-5 achieved proto AGI in late 2023 with an IQ of 48



**Jimmy Apples** 🍎 /acc  
@apples\_jimmy

AgI has been achieved internally.

2:00 AM · Sep 18, 2023 · 36.8K Views



**Siqi Chen** ✓  
@blader

i have been told that gpt5 is scheduled to complete training this december and that openai expects it to achieve agi.

which means we will all hotly debate as to whether it actually achieves agi.

which means it will.

12:00 AM · Mar 27, 2023 · 699.4K Views

171 Retweets 76 Quotes 1,483 Likes 607 Bookmarks



**Siqi Chen** ✓  
@blader

Love to build stuff (@runwayco, @sandboxvr, @postmates, @zynga) people love. 🧑‍💻 Investor @amplitude\_hq, @mercury, @italic

Entrepreneur ⓘ 📍 📌 runway.com  
Joined April 2007

2,695 Following 45.2K Followers

Followed by Smoke-away, Palmer Luckey, and Sam Altman

...

The first mention of a 100 trillion parameter model being developed by OpenAI was in the summer of 2021, mentioned offhand in a wired interview by the CEO of Cerebras (Andrew Feldman), a company which Sam Altman is a major investor of.



SUBSCRIBE

WILL KNIGHT

BUSINESS

AUG 24, 2021 3:13 PM

## A New Chip Cluster Will Make Massive AI Models Possible

Cerebras says its technology can run a neural network with 120 trillion connections—a hundred times what's achievable today.

“From talking to OpenAI, GPT-4 will be about 100 trillion parameters,” Feldman says. “That won’t be ready for several years.”



Sam Altman's response to Andrew Feldman, at an online meetup and Q&A called AC10, which took place in September 2021. It's crucial to note that Sam Altman ADMITS to their plans for a 100 trillion parameter model.

(Sources: <https://albertoromgar.medium.com/gpt-4-a-viral-case-of-ai-misinformation-c3f999c1f589>

[https://www.reddit.com/r/GPT3/comments/pj0ly6/sam\\_altman\\_gpt4\\_will\\_be\\_remain\\_textonly\\_will\\_not/](https://www.reddit.com/r/GPT3/comments/pj0ly6/sam_altman_gpt4_will_be_remain_textonly_will_not/)

The reddit posting itself is sourced from a LessWrong post, which was deleted at Sam Altman's request: <https://www.lesswrong.com/posts/aihztgJrknBdLHjd2/sam-altman-q-and-a-gpt-and-agi>)

New information eventually came out. Sam Altman denied the 100T rumor in the AC10 online meetup Q&A soon after I published the infamous article but asked explicitly to not reveal info from the talk (which, of course, **didn't happen**). But I

Coming from a Q&A recap:

GPT-4 is coming, but currently the focus is on coding (i.e. Codex) and that's also where the available compute is going. GPT-4 will be a text model (as opposed to multi-modal). It will not be much bigger than GPT-3, but it will use way more compute. People will be surprised how much better you can make models without making them bigger.

The progress will come from OpenAI working on all aspects of GPT (data, algos, fine-tuning, etc.). GPT-4 will likely be able to work with longer context and (possibly) be trained with a different loss function - OpenAI has "line of sight" for this. (Uncertain about "loss" function, I think he said something like "different value function", so this might be a misinterpretation.)

GPT-5 might be able to pass the Turing test. But probably not worth the effort.

100 trillion parameter model won't be GPT-4 and is far off. They are getting much more performance out of smaller models. Maybe they will never need such a big model.

It is not yet obvious how to train a model to do stuff on the internet and to think long on very difficult problems. A lot of current work is how to make it accurate and tell the truth."

...

AI researcher Igor Baikov made the claim, only a few weeks later, that GPT-4 was being trained and would be released between December and February. Again, I will prove that Igor really *did* have accurate information, and is a credible source. This will be important soon

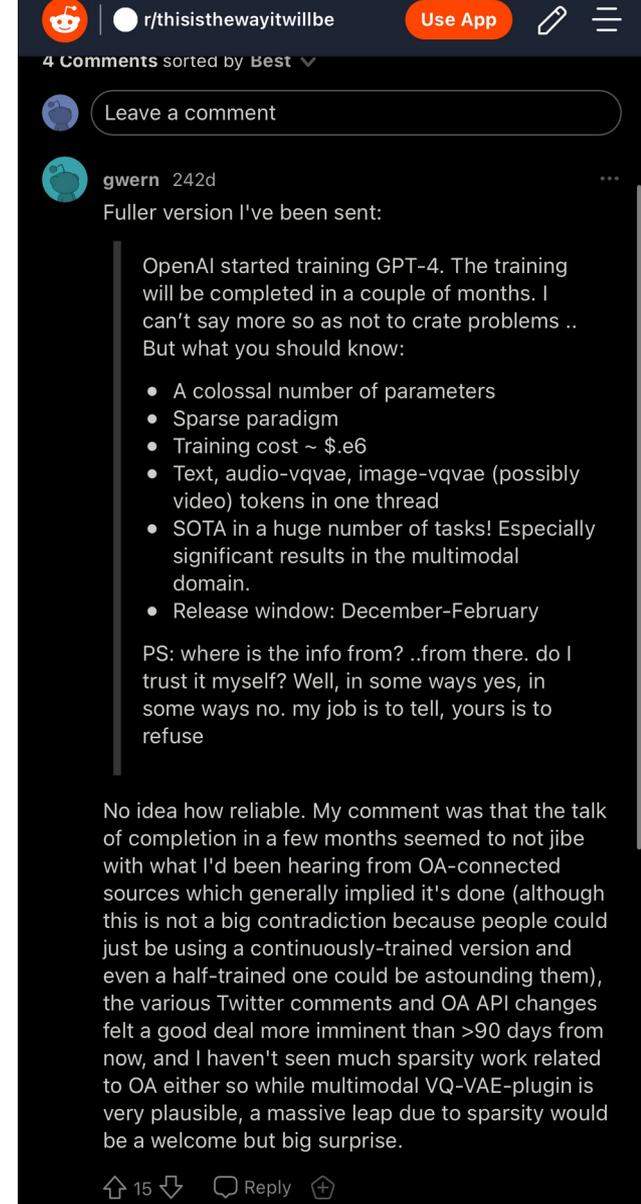


Gwern is a famous figure in the AI world – he is an AI researcher and blogger. He messaged Igor Baikov on Twitter (in September 2022) and this is the response he received.

Important to remember: “Colossal number of parameters”.

“Text”, “audio”, “images”, “possibly video”,  
and “multimodal”.

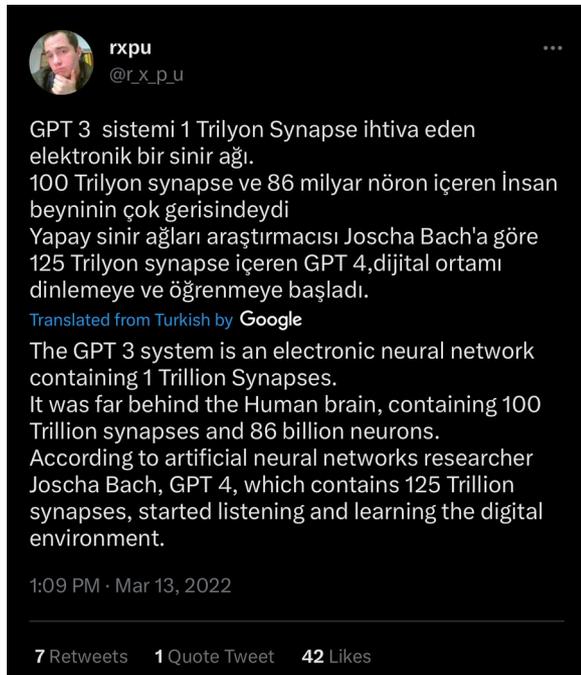
This comes from a subreddit called “thisisthewayitwillbe” which is a small, private subreddit I’m part of, run by a mathematics professor with an interest in AGI. AI enthusiasts (and a few experts) use the subreddit to discuss AI topics deeper than what you’ll find in the mainstream.



A “colossal number of parameters”? Sounds like Igor Baikov was referencing a 100 trillion parameter model, as 500 billion parameter models and up to 1 trillion parameter models had already been trained many times by the time of his tweet in summer 2022 (making models of that size unexceptional and certainly not “colossal”).

These tweets from “rxpu”, seemingly an AI enthusiast (?) from Turkey, are interesting because they make a very similar claim about GPT-4’s release window before anyone else did (trust me – I spent many hours, daily, scouring the internet for similar claims, and no one else made this specific claim before he did).

He also mentions a “125 trillion synapse” GPT-4 – however, he incorrectly states GPT-3’s parameter count as 1 trillion. (It seems as though rxpu did have inside information, but got something mixed up with the parameter counts – again, I will illustrate this later, and prove that rxpu was not lying).



 **rxpu**  
@r\_x\_p\_u

GPT 3 sistemi 1 Trilyon Synapse ihtiva eden elektronik bir sinir ağı.  
100 Trilyon synapse ve 86 milyar nöron içeren İnsan beyninin çok gerisindeydi  
Yapay sinir ağları araştırmacısı Joscha Bach'a göre 125 Trilyon synapse içeren GPT 4,dijital ortamı dinlemeye ve öğrenmeye başladı.

[Translated from Turkish by Google](#)

The GPT 3 system is an electronic neural network containing 1 Trillion Synapses.  
It was far behind the Human brain, containing 100 Trillion synapses and 86 billion neurons.  
According to artificial neural networks researcher Joscha Bach, GPT 4, which contains 125 Trillion synapses, started listening and learning the digital environment.

1:09 PM · Mar 13, 2022

7 Retweets 1 Quote Tweet 42 Likes



 **rxpu**  
@r\_x\_p\_u

GPT 4 öğrenme sürecini tamamlamasının ardından 2023 yılının başında kamuoyu ile tanıştırlılacak. Elon Musk ve Microsoft tarafından fonlanan OpenAI şirketi tarafından geliştirilen GPT 4 insan zekasının ayırt edemeyeceği akademik tezler ve yazılar kaleme alabileceği iddia ediliyor

[Translated from Turkish by Google](#)

After completing the GPT 4 learning process, it will be introduced to the public at the beginning of 2023. Developed by Elon Musk and the OpenAI company funded by Microsoft, GPT 4 is claimed to be able to write academic theses and articles that human intelligence cannot distinguish.

1:09 PM · Mar 13, 2022

2 Retweets 37 Likes



 **rxpu**  
@r\_x\_p\_u

Electronic Enginar, computer carrot.

Joined March 2010

555 Following 13.9K Followers

...

This is a weaker piece of evidence, but it's worth including because "roon" is fairly notable as a Silicon Valley AI researcher, and is followed by Sam Altman, CEO of OpenAI, and other OpenAI researchers on Twitter.

roon   
@tszsl

AGI researchoor | technology brother | "prolific Twitter shitposter" ~ Vice News | "almost billionaire" - the Atlantic | tweets DO NOT represent anybody else

📍 ānanda t̄aṇḍava [roonscape.substack.com](https://roonscape.substack.com)  
🗓️ Born June 28 📅 Joined May 2013

5,536 Following 89.5K Followers

 Followed by Sam Altman

roon   
@tszsl

on the openai API\*\*

business opportunities where the GPT error rate was too high to be useful are now within the envelope of possibilities

and, they've made it an open secret that GPT4 has orders of magnitude higher parameter count and is coming very soon

5:29 PM · Sep 20, 2022

5 Retweets 1 Quote Tweet 231 Likes

roon @tszsl

AI researchoor | technology brother | "prolific Twitter shitposter" ~ Vice News | "almost billionaire" - the Atlantic

4,892 Following 55.4K Followers

📅 Joined May 2013

Is GPT-4 100 trillion parameters as rumored?

 1

7:38 PM ✓

In November 2022 I reached out to an AI blogger named Alberto Romero. His posts seem to spread pretty far online so I was hoping that if I sent him some basic info about GPT-4 he might do a writeup and the word would get out.

The results of this attempt were pretty remarkable as I'll show in the next two pages.

**Alberto Romero**  
@Alber\_RomGar

AI & tech writer on Substack and Medium | Analyst at CambrianAI | AI Newsletter: The Algorithmic Bridge

Joined December 2020 · 3,875 Followers

Not followed by anyone you're following

OpenAI started training GPT-4. The training will be completed in a couple of months. I can't say more so as not to create problems ... But what you should know:

- A colossal number of parameters
- Sparse paradigm
- Training cost ~ \$e6
- Text, audio-vqvae, image-vqvae (possibly video) tokens in one thread
- SOTA in a huge number of tasks! Especially significant results in the multimodal domain.
- Release window: December-February

PS: where is the info from? ..from there, do I trust it myself? Well, in some ways yes, in some ways no. my job is to tell, yours is to refuse

More recent info on GPT-4 which has been corroborated by multiple sources.

**Alberto Romero**

Also, according to Robert Scoble and a few other ppl on here GPT4 has been in private beta for a few weeks, so the Dec-Feb release date makes sense. We'll see

Nov 9, 2022, 1:02 PM

Can you direct me to those posts by AI researchers close to OpenAI? I've seen Robert Scoble's tweets (actually I asked him but he's under NDA)

Nov 9, 2022, 1:10 PM

**Alberto Romero**

Nov 10, 2022, 4:06 PM

Like I said, multiple ppl have mentioned this release window. I assume it's actually an announcement window. Are you trying to do another GPT4 writeup?

Nov 10, 2022, 4:07 PM

No worries, I understand! I realized the info was public, so yeah I don't think is that big of a deal (not many details anyway).

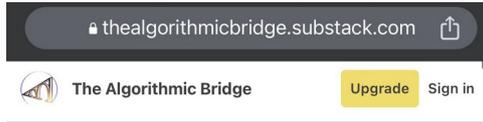
I imagined Gwern shared the comment but didn't feel like his writing style so didn't want to risk ascribing it wrongly.

Anyway, thanks so much for all this info. I'm doing a compilation of the public info that's circulating and will publish it in my newsletter tomorrow, yes!

Because the info isn't necessarily official, I'll take care of making that very clear throughout the article, though

Nov 10, 2022, 4:16 PM

Alberto Romero's post. The general response will be shown on the next page.

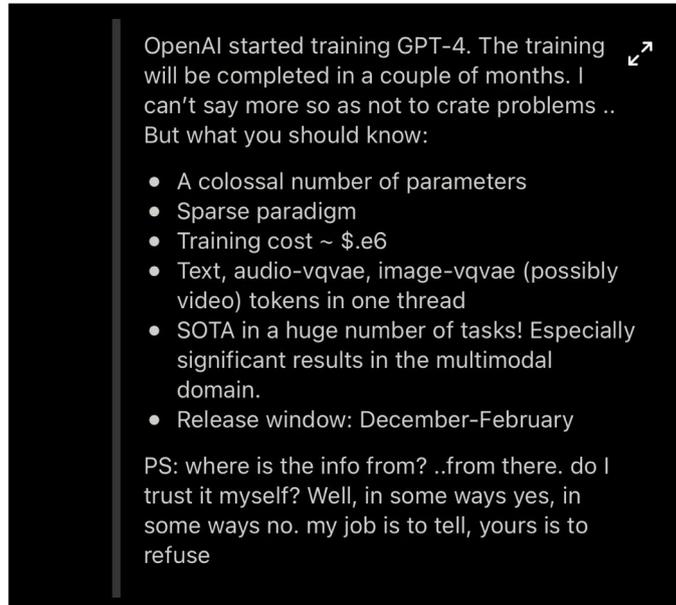


## GPT-4 Rumors From Silicon Valley

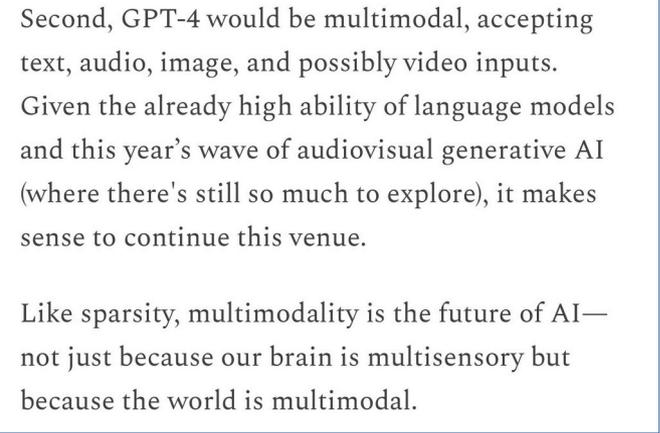
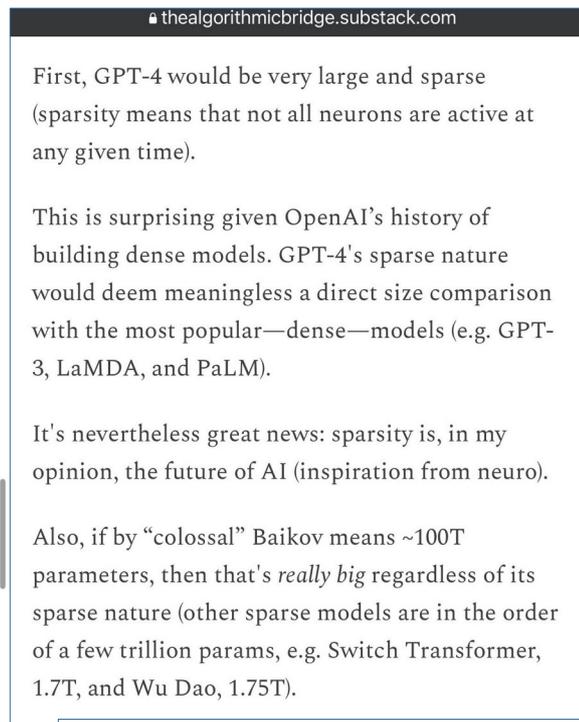
People are saying things...



ALBERTO ROMERO  
NOV 11, 2022



Reddit comment screenshot. Credit: [Igor Baikov](#) (shared by [Gwern](#))



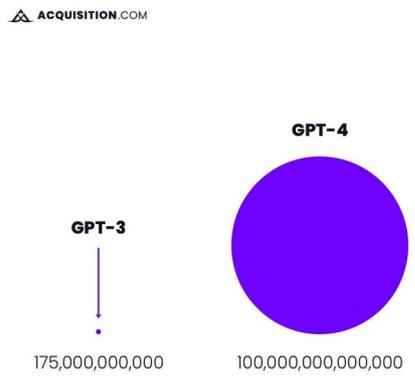
The 100 trillion parameter leak went viral, reaching millions of people, to the point that OpenAI employees including CEO Sam Altman had to respond – calling it “complete bullshit”. The Verge called it “factually incorrect”. Alberto Romero claimed responsibility for the leak as you can see on the left.

thealgorithmicbridge.substack.com

The Algorithmic Bridge Upgrade Sign in

I'm responsible for the “GPT-4 will have 100 trillion parameters” false statement going viral on social media.

In case you don't know what I'm talking about, here are a couple of visual examples:



Credit: [Alex Hormozi](#)

Andrew "The Metaverse Guy" Steinwold @AndrewSteinwold

GPT-4 is rumored to be coming soon, sometime between Dec - Feb

- GPT-3 has 175 billion parameters
- GPT-4 supposedly has 100 trillion parameters

It is something like 500x more powerful than GPT-3

What kinda stuff will you be able to create with GPT-4!?

A visual comparison showing GPT-3 with a small purple dot and GPT-4 with a large purple circle. The GPT-3 dot is positioned above the number 175,000,000,000. The GPT-4 circle is positioned above the number 100,000,000,000,000.

9:04 PM · Nov 21, 2022

1,590 Retweets 688 Quotes 7,271 Likes

966 Bookmarks

They are already doing demos.

It can write a 60,000 word book from a single prompt.

The only question I've had about AI...

A visual comparison showing GPT-3 with a small purple dot and GPT-4 with a large purple circle. The GPT-3 dot is positioned above the number 175,000,000,000. The GPT-4 circle is positioned above the number 100,000,000,000,000.

2:44 PM · Jan 10, 2023 · 4.1M Views

3,503 Retweets 750 Quote Tweets 20.5K Likes

12:06 theverge.com

# The Verge

Menu +

When asked about one viral (and factually incorrect) chart that purportedly compares the number of parameters in GPT-3 (175 billion) to GPT-4 (100 trillion), Altman called it “complete bullshit.”

12:06 theverge.com

# The Verge

Menu +

“The GPT-4 rumor mill is a ridiculous thing. I don't know where it all comes from,” said the OpenAI CEO. “People are begging to be disappointed and they will be. The hype is just like... We don't have an actual AGI and that's sort of what's expected of us.”

...

Igor Baikov, the origin of the “colossal number of parameters” statement, also saw the viral spread of the GPT-4 leak (which was essentially his own doing) and responded.

So, after all, Igor really did mean “100 trillion parameters” when he said “a colossal number of parameters”.

But, is Igor a reliable source? Are his other claims accurate? What about the multimodality?

What about the ability for GPT-4 to process images, sounds, and videos? I will prove Igor’s reliability shortly.

The image shows a screenshot of a Twitter thread on a dark background. At the top, a tweet from Igor Baikov (@Killa\_ru) is partially visible. Below it is a tweet from Andrew "The Metavers..." (@An...) dated Nov 21, 2022. The tweet contains text about GPT-4 rumors and a comparison of GPT-3 and GPT-4 parameters. A diagram shows a tiny dot for GPT-3 (175,000,000,000) and a large circle for GPT-4 (100,000,000,000,000). The tweet has 2 retweets and 1 like, dated Nov 22, 2022.

Igor Baikov @Killa\_ru

Can confirm I also heard about 100T parameters, but I found this improbable so I decided to not include it in my prev tweet on GPT-4 [twitter.com/Killa\\_ru/statu..](https://twitter.com/Killa_ru/status..)

Andrew "The Metavers..." @An... · Nov 21, 2022

GPT-4 is rumored to be coming soon, sometime between Dec - Feb

- GPT-3 has 175 billion parameters
- GPT-4 supposedly has 100 trillion parameters

It is something like 500x more powerful than GPT-3

What kinda stuff will you be able to create with GPT-4!?

[Show this thread](#)

GPT-3

175,000,000,000

GPT-4

100,000,000,000,000

8:59 AM · Nov 22, 2022

2 Retweets 1 Like

Somewhere around Oct/Nov 2022 I became convinced that OpenAI planned to first release a ~1-2 trillion parameter subset of GPT-4 before releasing the full 100 trillion parameter model (“GPT-5”).

These sources aren’t particularly solid but they all said the same thing – including rxpu, who once claimed there was a 125 trillion parameter model in the works, and then incorrectly claimed GPT-3 was 1 trillion – I believe he got his information mixed up.

**rxpu** @rx\_p\_u  
Replying to @conatus1632

Biyolojik insanın bir sonraki evrimine eşlik edecek önemli enstrümanlardan olacak ne yazık ki. Çalışma şekilleri insan zihninin milyonlarca yılda kat ettiği sürece benzer bir süreci çok daha hızlı bir şekilde kat ediyor. 1 Trilyon nöron seviyesindedir. Singularity noktası 100 Trilyon

Translated from Turkish by Google

Unfortunately, it will be one of the important instruments that will accompany the next evolution of biological human beings. Their way of working goes through a similar process much faster than the human mind in millions of years. They are at the level of 1 trillion neurons. Singularity point 100 trillion

10:26 AM · Mar 20, 2022

**rxpu** @rx\_p\_u

Sadece 41 milyar nörona sahip DeepMind (175 nöronluk GPT-3'ün alt kümesi) yapay sinir ağı matematik olimpiyatları seviyesindeki soruları çözebiliyor. İnsan gibi program ve algoritmalar yazabiliyor. GPT-4 ise 1 Trilyon nöron ile insan zekasına meydan okuyacak

Translated from Turkish by Google

With only 41 billion neurons, the DeepMind (a subset of GPT-3) artificial neural network (with 175 neurons) can solve problems at the level of the mathematics olympics. It can write programs and algorithms like a human. GPT-4 will challenge human intelligence with 1 Trillion neurons

youtube.com  
DeepMind's New AI: As Smart As An Engineer... Kind Of! 🤖

10:14 AM · Mar 20, 2022

<--- (Date: 2022)

**roon** @tszsl · Feb 15

yeah this is the correct take. **GPT4** coming soon. 2trillies baby

**bayeslord (e/acc)** @bayeslord · Feb 15

I feel like the burst of openai ppl talking about alignment and conscious language models is a hint that they've cooked up something freaky as fuck

24 27 536

**pAurus** @PPAREUS · Dec 3, 2022

Replying to @jordanbpeterson

You train a decoding language model with 300 billion parameters and then you got a ChatGPT. If you think this is cool, wait for the next year's **GPT-4**. It supposedly has **1.2 trillion** (or **2 trillion**, not sure) parameters

3

The sources here are of varying credibility (Jyri and Leor are San Francisco investors and Harris is an AI researcher) but they all inexplicably say the same thing--GPT-4 was being tested in Oct/Nov of 2022. And according to US military AI researcher Cherie M Poland, it was definitely being trained in October, which again lines up with Igor Baikov's leak.



(From October 2022 ^)



...

OpenAI's official position, as demonstrated by Sam Altman himself, is that the idea of a 100 trillion parameter GPT-4 is "complete bullshit". This is half true, as GPT-4 is a 1 trillion parameter subset of the full 100 trillion parameter model.



Menu +

ARTIFICIAL INTELLIGENCE / TECH

**OpenAI CEO Sam Altman on GPT-4: 'people are begging to be disappointed and they will be'** / In a recent interview, Altman discussed hype surrounding the as yet unannounced GPT-4 but refused to confirm if the model will even be released this year.

By [James Vincent](#)

Jan 18, 2023 at 8:55 AM CST

[15 Comments](#) / [6 New](#)



Menu +

When asked about one viral (and factually incorrect) chart that purportedly compares the number of parameters in GPT-3 (175 billion) to GPT-4 (100 trillion), Altman called it "complete bullshit."

Just to illustrate that the 100 trillion parameter model hasn't arrived yet and is still in development, Semafor in March 2023 (shortly after the release of GPT-4) claimed GPT-4 is 1 trillion parameters. (OpenAI has refused to officially disclose parameter count).

Less than six months later, OpenAI took \$1 billion from Microsoft, which could provide not just funding but infrastructure know-how. Together they built a supercomputer to train massive models that eventually created ChatGPT and the image generator DALL-E. The latest language model, GPT-4, has 1 trillion parameters.

Something else worth nothing is that OpenAI claims GPT-4 was “finished training” in August, whereas we know that a “colossal” multimodal model was being trained between August and October. One explanation for this is, OpenAI lied. Another possibility is that the 1 trillion parameter GPT-4 may have finished its first round of training in August, but went through additional retraining between August and October, which is when the bulk of the full 100 trillion parameter model was trained.

I will now provide my evidence that GPT-4 was not just trained on text and images, but was also trained on audio and video.

Francis Hellyer seems moderately credible but this page is not the most solid piece of evidence – I'm including it because it seems to corroborate what other sources are saying. Francis is an investor, entrepreneur and writer. The information he listed in his tweet about the team "running out of internet" cannot be found in any other publication, any leak, or any online posting, so he did not "steal" it from some other place.

An incredibly solid source on the next page.



The CTO of Microsoft Germany, a week prior to the official release of GPT-4, seems to have slipped up and revealed that there exists a GPT-4 which has the ability to process videos. I imagine he was unaware of OpenAI's decision not to reveal the video capabilities of the system.

This completely proves that GPT-4/5 was trained on not just text and images, but also video data, and of course we can infer that audio data was included as well.

**"We will introduce GPT-4 next week"**

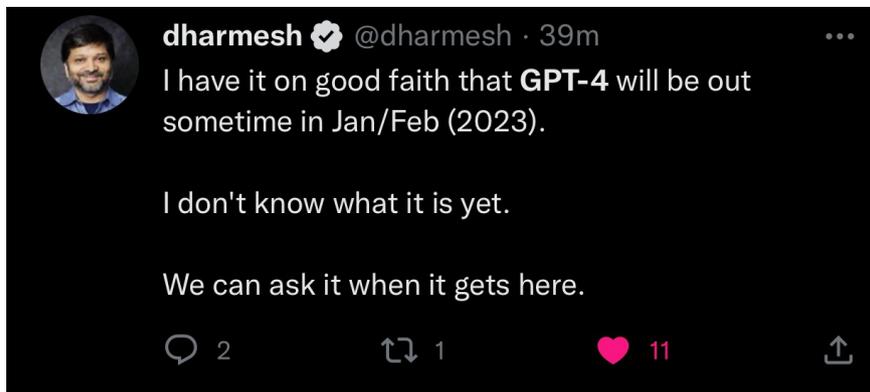
"We will introduce GPT-4 next week, there we will have multimodal models that will offer completely different possibilities – for example videos," Braun said. The CTO called LLM a "game changer" because they teach machines to understand natural language, which then understand in a statistical way what was previously only readable and understandable by humans. In the meantime, the technology has come so far that it basically "works in all languages": You can ask a question in German and get an answer in Italian. With multimodality, Microsoft(-OpenAI) will "make the models comprehensive".



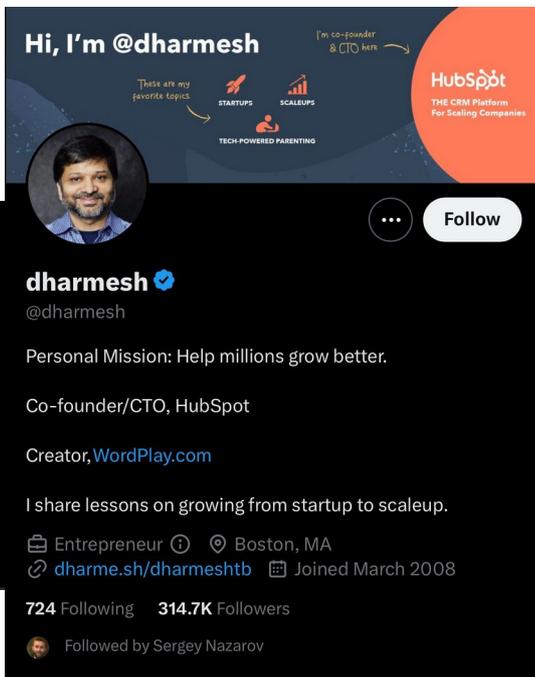
Dr. Andreas Braun, CTO Microsoft Germany and Lead Data & AI STU at the Microsoft Digital Kickoff: "KI im Fokus" (AI in Focus, Screenshot) (Bild: Microsoft)

Clearly, Igor's claim about the 100 trillion parameter model was true, down to every precise detail.

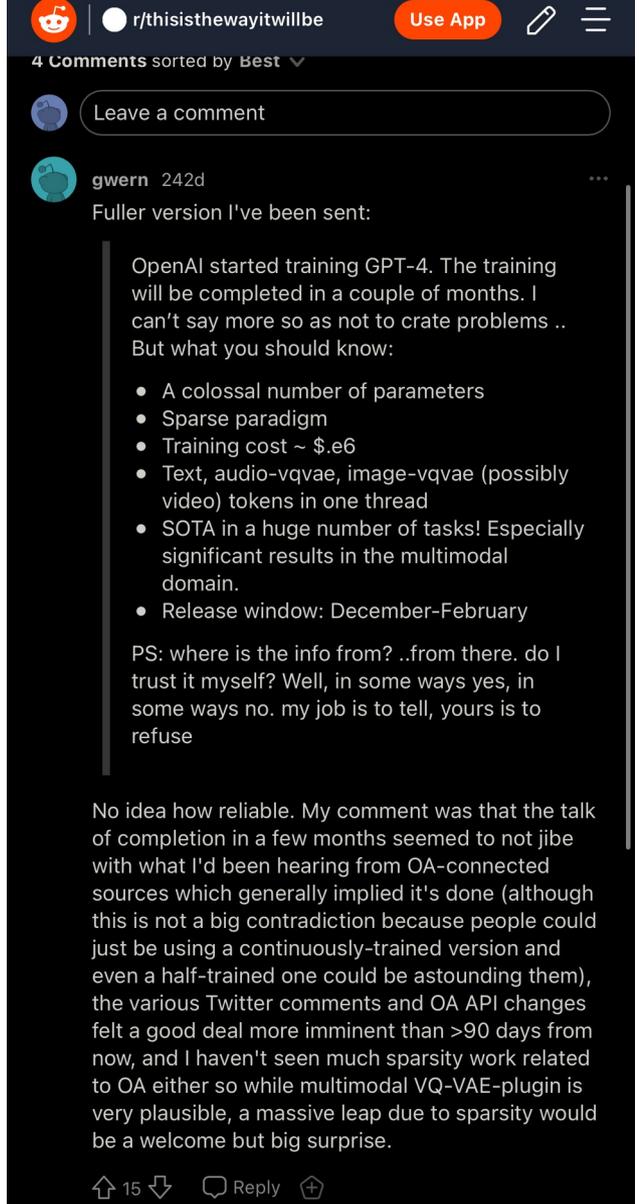
Another source that lines up with Igor's claim, is a credible entrepreneur who stated (on Oct 25 2022) that GPT-4's release date would be between January and February of 2023:



A screenshot of a tweet from the user 'dharmesh' (@dharmesh) posted 39 minutes ago. The tweet text reads: 'I have it on good faith that GPT-4 will be out sometime in Jan/Feb (2023). I don't know what it is yet. We can ask it when it gets here.' The tweet has 2 replies, 1 retweet, and 11 likes. At the bottom, it shows 'Followed by Sergey Nazarov'.



A profile card for 'dharmesh' (@dharmesh). The header says 'Hi, I'm @dharmesh' and 'I'm co-founder & CTO here'. Below the header is a circular profile picture of a man with a beard. To the right of the picture is a 'Follow' button. Below the picture is the name 'dharmesh' with a verified badge and the handle '@dharmesh'. Underneath is the 'Personal Mission: Help millions grow better.' and 'Co-founder/CTO, HubSpot'. It also lists 'Creator, WordPlay.com' and 'I share lessons on growing from startup to scaleup.' At the bottom, it shows 'Entrepreneur' and 'Boston, MA' with location icons, a link 'dharme.sh/dharmeshtb', and 'Joined March 2008'. At the very bottom, it shows '724 Following' and '314.7K Followers'.



A screenshot of a Reddit comment from the user 'gwern' posted 242 days ago. The comment text reads: 'Fuller version I've been sent: OpenAI started training GPT-4. The training will be completed in a couple of months. I can't say more so as not to create problems .. But what you should know: A colossal number of parameters, Sparse paradigm, Training cost ~ \$.e6, Text, audio-vqvae, image-vqvae (possibly video) tokens in one thread, SOTA in a huge number of tasks! Especially significant results in the multimodal domain, Release window: December-February. PS: where is the info from? ..from there. do I trust it myself? Well, in some ways yes, in some ways no. my job is to tell, yours is to refuse. No idea how reliable. My comment was that the talk of completion in a few months seemed to not jibe with what I'd been hearing from OA-connected sources which generally implied it's done (although this is not a big contradiction because people could just be using a continuously-trained version and even a half-trained one could be astounding them), the various Twitter comments and OA API changes felt a good deal more imminent than >90 days from now, and I haven't seen much sparsity work related to OA either so while multimodal VQ-VAE-plugin is very plausible, a massive leap due to sparsity would be a welcome but big surprise.'

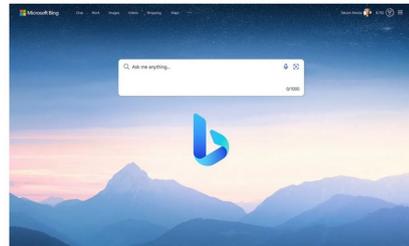
Although GPT-4 was released in March 2023, slightly outside the Dec-Feb window claimed by Igor Baikov (which I believe was done intentionally by OpenAI to discredit Igor's leak), Bing ChatGPT (based on GPT-4) was actually announced in February of 2023, clearly showing that the window claimed by Igor had validity, and was probably changed last minute by a panicked OpenAI.



MICROSOFT / TECH / ARTIFICIAL INTELLIGENCE

**Microsoft announces new Bing and Edge browser powered by upgraded ChatGPT AI** / Microsoft says it's using conversational AI to create a new way to browse the web. Users will be able to chat to Bing like ChatGPT, asking questions and receiving answers in natural language.

By [James Vincent](#)  
Feb 7, 2023 at 12:23 PM CST  
[80 Comments](#) / [80 New](#)



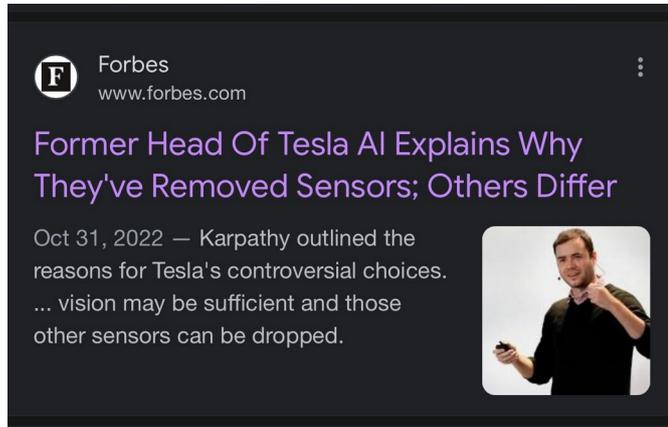
Confirmed: the new Bing runs on OpenAI's GPT-4

Congratulations to our partners at Open AI for their [release of GPT-4 today](#).

We are happy to confirm that the new Bing is running on GPT-4, which we've customized for search. If you've used the new Bing preview at any time in the last five weeks, you've already experienced an early version of this powerful model. As OpenAI makes updates to GPT-4 and beyond, Bing benefits from those improvements. Along with our own updates based on community feedback, you can be assured that you have the most comprehensive copilot features available.

A note about robotics: AI researchers are beginning to believe that vision is all that's necessary for optimal real-world/physical performance. Just to give one example, Tesla completely ditched all sensors and committed fully to vision for their self-driving cars.

The point is, training a human-brain-sized AI model on all the image and video data on the internet will clearly be more than enough to handle complex robotics tasks. Common sense reason is buried in the video data, just like it's buried in the text data (and the text-focused GPT-4 is stunningly good at common sense reasoning).



3. Vision is necessary to the task (which almost all agree on) and it should also be sufficient as well. If it is sufficient, the cost of extra sensors and tools outweighs their benefit.

Tesla made a bold move by removing the radar in its cars in favor of its new camera-based vision-only Tesla Vision approach. People said it was a bad idea and couldn't be done, but it didn't take long before safety organizations tested the vision-only features and approved them. Now, Tesla has announced that it will move forward even further with the vision-only setup, by eliminating ultrasonic sensors from its cars.

A recent example from Google, of robotics capabilities being learned from a large vision/language model. (Minimal robotics data was required on top of the language and vision training, and the knowledge from visual and text tasks transferred to the robotics tasks. OpenAI is training their 100 trillion parameter model on “all the data on the internet” which will undoubtedly include robotics data). PaLM-E is a ~500 billion parameter model – what happens to robotics performance when you train a 100 trillion parameter model on all the data available on the internet? (More on Google’s PaLM-E model on the next page).

## Transferring knowledge from large-scale training to robots

PaLM-E offers a new paradigm for training a generalist model, which is achieved by framing robot tasks and vision-language tasks together through a common representation: taking images and text as input, and outputting text. A key result is that PaLM-E attains significant *positive knowledge transfer* from both the vision and language domains, improving the effectiveness of robot learning.



Today we introduce **PaLM-E**, a new generalist robotics model that overcomes these issues by transferring knowledge from varied visual and language domains to a robotics system. We began with **PaLM**, a powerful large language model, and “embodied” it (the “E” in PaLM-E), by complementing it with sensor data from the robotic agent. This is the key difference from **prior efforts** to bring large language models to robotics – rather than relying on only textual input, with PaLM-E we train the language model to directly ingest raw streams of robot sensor data. The resulting model not only enables highly effective robot learning, but is also a state-of-the-art general-purpose visual-language model, while maintaining excellent language-only task capabilities.

ai.googleblog.com

Google Research

BLOG >

## PaLM-E: An embodied multimodal language model

FRIDAY, MARCH 10, 2023

In the first example below, a person asks a mobile robot to bring a bag of chips to them. To successfully complete the task, PaLM-E produces a plan to find the drawer and open it and then responds to changes in the world by updating its plan as it executes the task. In the second example, the robot is asked to grab a green block. Even though the block has not been seen by that robot, PaLM-E still generates a step-by-step plan that generalizes beyond the training data of that robot.



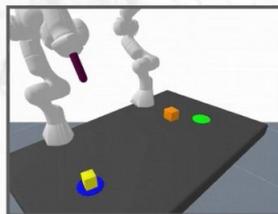
*PaLM-E controls a mobile robot operating in a kitchen environment. **Left:** The task is to get a chip bag. PaLM-E shows robustness against adversarial disturbances, such as putting the chip bag back into the drawer. **Right:** The final steps of executing a plan to retrieve a previously unseen block (green star). This capability is facilitated by transfer learning from the vision and language models.*

#### Robot Mobile Manipulation



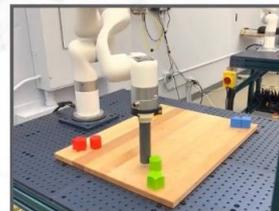
Task: give me the chips from the drawer  
Next step: [Bring it to the user](#)

#### Task and Motion Panning



Q: How to put yellow block on blue plate?  
A: [Done](#)

#### Robot Tabletop Manipulation



Task: sort blocks by colors into corners  
Next step: [Push blocks back to corners](#)

#### Visual Question Answering



Q: What is in photo 1 but not in photo 2?  
A: [Sunglasses](#)

PaLM-E  
562B

#### Conclusion

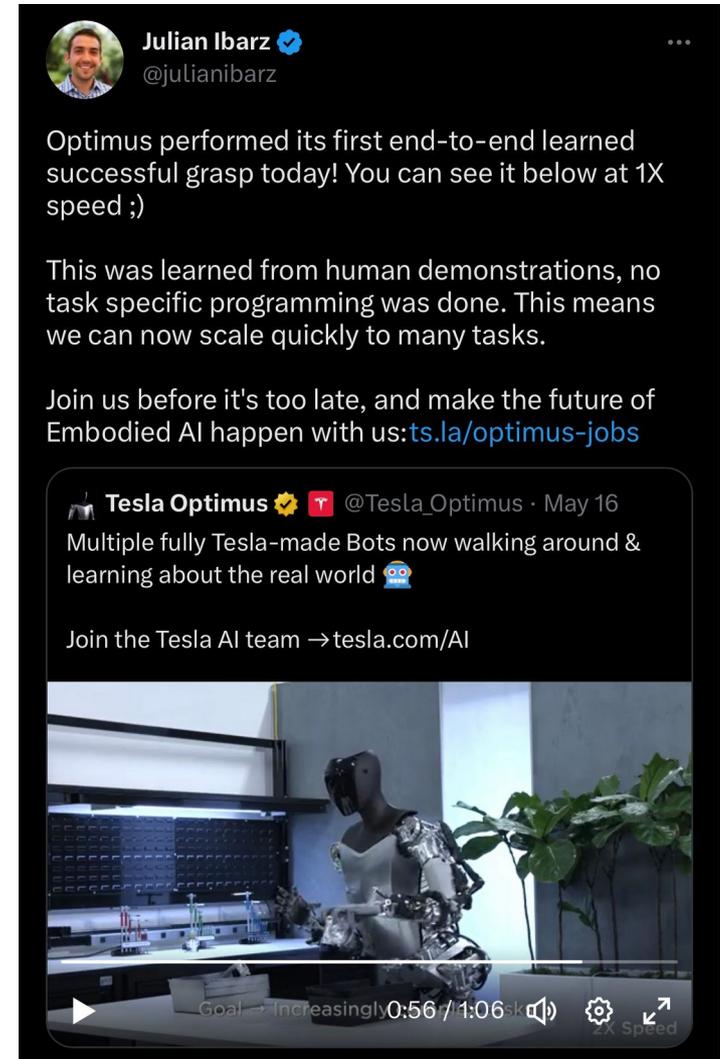
PaLM-E pushes the boundaries of how generally-capable models can be trained to simultaneously address vision, language and robotics while also being capable of transferring knowledge from vision and language to the robotics domain. There are additional topics investigated in further detail in the [paper](#), such as how to leverage [neural scene representations](#) with PaLM-E and also the extent to which PaLM-E, with greater model scale, experiences less [catastrophic forgetting](#) of its language capabilities.

An *embodied* language model, and also a visual-language generalist

Another robotics development – this time from Tesla (May 16 2023). They trained their robot “Optimus” to grasp an object – and “no task specific programming was done”. Everything was learned from HUMAN DEMONSTRATIONS.

“This means we can now scale quickly to many tasks.”

Once again: if human demonstrations are all that is needed for advanced robotics performance, a 100 trillion parameter model trained on all the video on the web would certainly be able to achieve astonishing robotics performance...



The image is a screenshot of a tweet from Julian Ibarz (@julianibarz). The tweet text reads: "Optimus performed its first end-to-end learned successful grasp today! You can see it below at 1X speed ;)" followed by "This was learned from human demonstrations, no task specific programming was done. This means we can now scale quickly to many tasks." and "Join us before it's too late, and make the future of Embodied AI happen with us: [ts.la/optimus-jobs](https://ts.la/optimus-jobs)". Below the text is a video player showing a Tesla Optimus robot in a kitchen-like environment, performing a task. The video player interface includes a play button, a progress bar showing 0:56 / 1:06, a volume icon, a settings gear, and a share icon. The video title is "Goal - Increasingly".

The image on the left shows what the 1 trillion parameter GPT-4 is capable of in terms of image recognition. The response is already clearer and more well written than what many humans would have come up with. So, again, what happens when you train a model 100 times larger than GPT-4, which is the size of the human brain, on all the data available on the internet?

User What is funny about this image? Describe it panel by panel.



Source: [hmmm \(Reddit\)](#)

GPT-4 The image shows a package for a "Lightning Cable" adapter with three panels.

Panel 1: A smartphone with a VGA connector (a large, blue, 15-pin connector typically used for computer monitors) plugged into its charging port.

Panel 2: The package for the "Lightning Cable" adapter with a picture of a VGA connector on it.

Panel 3: A close-up of the VGA connector with a small Lightning connector (used for charging iPhones and other Apple devices) at the end.

The humor in this image comes from the absurdity of plugging a large, outdated VGA connector into a small, modern smartphone charging port.

## Chapter 0

The sun hung high in the sky, casting long shadows on the cobblestone streets as I wove my way through the bustling sea of people. My nimble fingers dipped into pockets, swiped wallets, and slipped rings from distracted hands while my mind raced with a mixture of excitement and nerves. This was what I was good at; it was how I survived.

The city was still recovering from the devastating plague that had swept through it just a few years prior, leaving behind a wake of destruction and despair. Buildings lay in ruins, their shattered windows staring blankly out like empty eyes, and the people themselves bore the hollow look of those who had seen too much suffering. Poverty was rampant, but so too were resilience and determination as the city began to claw its way back to life.

A commotion caught my ear as I lifted a fine silk handkerchief from a gentleman's pocket. Spinning around, I spotted several city guards pushing their way through the crowd, their eyes scanning for any sign of trouble. My heart skipped a beat; they were looking for me. Ducking behind a fruit stall, I felt my heart pounding in my chest like a frantic bird trapped in a cage.

The above image is one of many short samples of the 1 trillion parameter GPT-4's text output.

If this is what a 1 trillion parameter model can write, what will a 100 trillion parameter model be able to write? Whether an AI model can ever be truly "creative" is up for debate, but an AI *faking* creativity is certainly possible and is **ALREADY HAPPENING**.

Important: notice how the AI model is able to generate multiple angles of the same scene with physically accurate lighting, and in some cases even physically accurate fluid and rain. If you can generate images and videos with accurate, common-sense physics, you have COMMON SENSE REASONING. If you can generate common sense, you UNDERSTAND common sense.



IMAGE^



IMAGES^



VIDEO^



<--VIDEO



<--IMAGES

VIDEO-->



Examples of the current level of quality of publically available video & image generation AI models. These models are less than 10 billion parameters in size. What happens, when you train a model 10,000 times larger, on all the data available on the internet, and give it the ability to generate images and video? (The answer: images and videos completely indistinguishable from the real thing, 100% of the time, with no exceptions, no workarounds, no possible way for anyone to tell the difference, no matter how hard they try). - (update: SORA IS FROM GPT-5 Q\* 2023 MODEL)

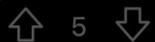
Two posts from Longjumping-Sky-1971. I'm including this because he accurately predicted the release date of GPT-4 weeks in advance (no one else posted this information publicly beforehand, meaning he had an inside source). His posts now have much more credibility – and he claimed image and audio generation would be trained in Q3 of 2023. If video generation training is simultaneous or shortly after, this lines up with Siqi Chen's claim of GPT-5 being finished training in December of 2023.

[r/thisisthewayitwillbe](#) · 63d

Rumor: Possible GPT-4 Launch Next Week According To Microsoft Germany?

*I've said it here before. The expected date is march 14th. It has image recognition but audio and image output isn't expected until Q3 for a big multimodal update. Bing is a subset of gpt4.*

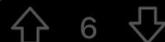
*If I really want to put my neck on the line, I'm going with 1.2 trillion parameters.*



[r/thisisthewayitwillbe](#) · 3d

Alphabet to unveil A.I. updates at Google I/O, showing off creative writing and coding capabilities

*Hmmm... if they actually unveil a multimodal model with audio + video etc, they would be beating openAi to the punch as they are not training that part till Q3. Does Google have the risk appetite to do that though ... let's see.*



Let's take a stroll back to February 2020, a few months before GPT-3 was released. An article from Technology Review, which was an "inside story" about OpenAI, seems to suggest that OpenAI was in the early stages of a "secret" project involving an AI system trained on images, text, and "other data", and that leadership at OpenAI thought it was the most promising way to reach AGI. I wonder what this could possibly be referring to.

MIT Technology Review

Subscribe

ARTIFICIAL INTELLIGENCE

## The messy, secretive reality behind OpenAI's bid to save the world

The AI moonshot was founded in the spirit of transparency. This is the inside story of how competitive pressure eroded that idealism.

By Karen Hao

February 17, 2020



Greg Brockman, co-founder and CTO; Ilya Sutskever, co-founder and chief scientist; and Dario Amodei, research director.

CHRISTIE HEMM KLOK

One of the biggest secrets is the project OpenAI is working on next. Sources described it to me as the culmination of its previous four years of research: an AI system trained on images, text, and other data using massive computational resources. A small team has been assigned to the initial effort, with an expectation that other teams, along with their work, will eventually fold in. On the day it was announced at an all-company meeting, interns weren't allowed to attend. People familiar with the plan offer an explanation: the leadership thinks this is the most promising way to reach AGI.

The next slide will reveal some quotes from the President of OpenAI – from 2019 – and it will tell you what their plan was.

OpenAI president Greg Brockman stated in 2019, following a 1 billion dollar investment from Microsoft at the time, that OpenAI planned to build a human-brain-sized model within five years, and that this was their plan for how to achieve AGI.

2019 + 5 = 2024

"In the race to build a machine with human-level intelligence, it seems, size really matters."

"We think the most benefits will go to whoever has the biggest computer," said Greg Brockman, chairman and chief technology officer of OpenAI.

The San Francisco-based AI research group, set up four years ago by tech industry luminaries including Elon Musk, Peter Thiel and Reid Hoffman, has just thrown down a challenge to the rest of the AI world.

Late last month, it raised \$1bn from Microsoft to speed its pursuit of the Holy Grail of AI: a computer capable of so-called artificial general intelligence, a level of cognition that would match its makers, and which is seen as the final step before the advent of computers with superhuman intelligence.

According to Mr Brockman, that money — a huge amount for a research organisation — will be spent "within five years, and possibly much faster", with the aim of building a system that can run "a human brain-sized [AI] model".

Whether a computer that matches the neural architecture in the human brain would deliver a comparable level of intelligence is another matter. Mr Brockman is wary about predicting precisely when AGI will arrive, and said that it would also require advances in the algorithms to make use of the massive increase in computing power.

But, speaking of the vast computing power that OpenAI and Microsoft hope to put at the service of its AI ambitions within five years, he added: "At that point, I think there's a chance that will be enough."

## AI curve steeper than Moore's Law

The tech industry is accustomed to riding the curve of Moore's Law, which describes the way that computing power roughly doubles every two years. But OpenAI is counting on a much more powerful exponential force to quickly take the capacity of its AI systems to a level that seems almost unimaginable today.

The research group calculates that since the tech industry woke up to the potential of machine learning seven years ago, the amount of processing capacity being applied to training the biggest AI models has been increasing at five times the pace of Moore's Law.

That makes today's most advanced systems 300,000 times more powerful than those used in 2012. The advance reflects the amount of money now being poured into advanced AI, as well as the introduction of parallel computing techniques that make it possible to crunch far more data.

Mr Brockman said OpenAI was counting on this exponential trend being carried forward another five years — something that would produce results that, he admits, sound "quite crazy".

As a comparison, he said that the past seven years of advances would be like extending the battery life of a smartphone from one day to 800 years: another five years on the same exponential curve would take that to 100m years.

Today's most advanced neural networks are roughly on a par with the honey bee. But with another five years of exponential advances, OpenAI believes it has a shot at matching the human brain.

 <https://www.ft.com/content>

## The billion-dollar bet to reach human-level AI | Financial Times

Aug 3, 2019 — According to Mr Brockman, that money — a huge amount for a research organisation — will be spent "within five years, and possibly much faster", ...

Both of these sources are clearly referring to the same plan to achieve AGI – a human-brain-sized AI model, trained on “images, text, and other data”, due to be trained within five years of 2019, so, by 2024. Seems to line up with all the other sources I’ve listed in this document...

One of the biggest secrets is the project OpenAI is working on next. Sources described it to me as the culmination of its previous four years of research: an AI system trained on images, text, and other data using massive computational resources. A small team has been assigned to the initial effort, with an expectation that other teams, along with their work, will eventually fold in. On the day it was announced at an all-company meeting, interns weren’t allowed to attend. People familiar with the plan offer an explanation: the leadership thinks this is the most promising way to reach AGI.

intelligence.

According to Mr Brockman, that money — a huge amount for a research organisation — will be spent “within five years, and possibly much faster”, with the aim of building a system that can run “a human brain-sized [AI] model”.

It is in this uncertain climate that Hassabis agrees to a rare interview, to issue a stark warning about his growing concerns. “I would advocate *not* moving fast and breaking things,” he says, referring to an old Facebook motto that encouraged engineers to release their technologies into the world first and fix any problems that arose later. The phrase has since become synonymous with disruption. That

He says AI is now “on the cusp” of being able to make tools that could be deeply damaging to human civilization, and urges his competitors to proceed with more caution than before. “When it comes to very powerful technologies –and obviously AI is going to be one of the most powerful ever—we need to be careful,” he says. “Not everybody is thinking about those things. It’s like experimentalists, many of whom don’t realize they’re holding dangerous material.” Worse still, Hassabis points out, we are the guinea pigs.

BUSINESS • TECHNOLOGY

## DeepMind’s CEO Helped Take AI Mainstream. Now He’s Urging Caution



Source: Time Magazine, Jan 12 2023

As I’ll show in these next few slides, AI leaders are suddenly starting to sound the alarm – almost like they know something **VERY SPECIFIC** that the general public doesn’t.

"The idea that this stuff could actually get smarter than people — a few people believed that," Hinton said in the interview. "But most people thought it was way off. And I thought it was way off. I thought it was 30 to 50 years or even longer away. Obviously, I no longer think that."

**New York(CNN)** – Geoffrey Hinton, who has been called the 'Godfather of AI,' confirmed Monday that he left his role at Google last week to speak out about the "dangers" of the technology he helped to develop.

“I thought it was 30 to 50 years or even longer away. Obviously, I no longer think that.”

What made him suddenly change his mind -- AND decide to leave Google to speak about the dangers of AI?

Deep learning pioneer Geoffrey Hinton quits Google ["According to the Times, Hinton says he has new fears about the technology he helped usher in and wants to speak openly about them, and that a part of him now regrets his life's work."]



Date of NYT interview: May 1 2023

The screenshot shows the Future of Life Institute website. At the top left is the logo 'future of life INSTITUTE' with a magnifying glass icon to its right. Below the logo is a breadcrumb trail: 'Home » Pause Giant AI Experiments: An Open Letter'. A blue arrow points left to the text 'All Open Letters'. The main heading is 'Pause Giant AI Experiments: An Open Letter'. Below the heading is a paragraph: 'We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.' Below this is a signature box with 'Signatures' and '27565' on the left, and a button 'Add your signature' on the right. At the bottom left, it says 'PUBLISHED March 22, 2023'. At the bottom right are social media icons for Twitter, Facebook, Email, and LinkedIn.

Therefore, we call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4. This pause should be public and verifiable, and include all key actors. If such a pause cannot be enacted quickly, governments should step in and institute a moratorium.

AI systems with human-competitive intelligence can pose profound risks to society and humanity, as shown by extensive research<sup>[1]</sup> and acknowledged by top AI labs.<sup>[2]</sup> As stated in the widely-endorsed [Asilomar AI Principles](#), *Advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources*. Unfortunately, this level of planning and management is not happening, even though recent months have seen AI labs locked in an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control.

Shortly after the release of GPT-4, the Future of Life Institute, a highly influential non-profit organization concerned with mitigating potential catastrophic risks to the world, released an open letter calling on all AI labs to pause AI development for six months. Why?

The first released version of the letter specifically said “(including the currently-being-trained GPT-5)”. Why was that included, and why was it removed?

“We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4 (including the currently-being-trained GPT-5),” states the letter, whose signatories include Yoshua Bengio, a professor at the University of Montreal considered a pioneer of modern AI, historian Yuval Noah Harari, Skype cofounder Jaan Tallinn, and Twitter CEO Elon Musk.

Source: Wired, March 29 2023

“We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4 (including the currently-being-trained GPT-5),” reads the letter, released by the Future of Life Institute, a nonprofit that works to reduce catastrophic and existential risks. “This pause should be public and verifiable, and include all key actors. If such a pause cannot be enacted quickly, governments should step in and institute a moratorium.”

Source: Vox, March 29 2023

Some alarming quotes from an interview and Q&A with Sam Altman from October 2022--

**youtube link:** <https://m.youtube.com/watch?v=b022FECpNe8> (Time: 49:30)

Audience Q&A question: “Do we have enough information in the internet to create AGI?”

Sam Altman’s blunt, immediate response, interrupting the man asking the question:

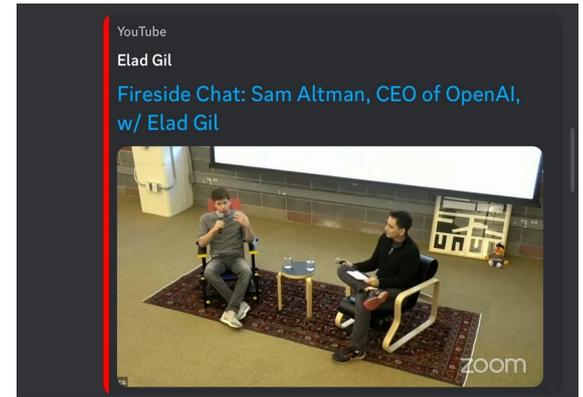
“Yes.”

Sam elaborates: “Yeah, we’re confident there is. We think about this and measure it quite a lot.”

The interviewer interjects: “What gives you that confidence?”

Sam’s reply: “One of the things I think that OpenAI has driven in the field that’s been really healthy is that you can treat scaling laws as a scientific prediction. You can do this for compute, you can do this for data, but you can measure at small scale and you can predict quite accurately how it’s going to scale up. How much data you’re going to need, how much compute you’re going to need, how many parameters you’re going to need, when the generated data gets good enough to be helpful... And the internet is...there’s a lot of data out there. There’s a lot of video out there too.”

More quotes from this Q&A on the next slide.



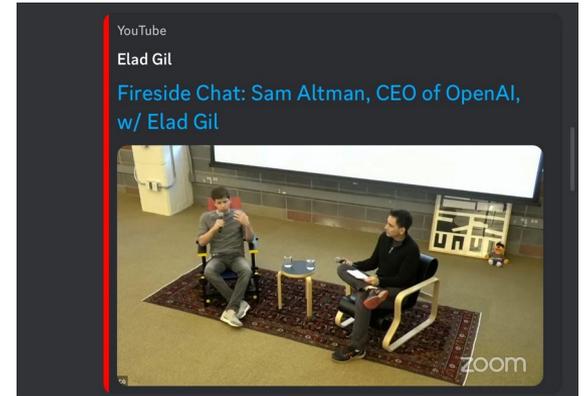
Another quote from the Q&A with Sam Altman-- **(Time: 53:00)**

[Note – an AI winter is an extended period of time where the AI field receives limited funding and is not given much attention by serious researchers. This happened twice -- once in the 70s and 80s and again from the mid 80s until roughly the late 2000s.]

Another audience question: “Could we have another AI winter and what might cause it?”

Sam Altman’s response: “Could we have an AI winter and what might cause it... yeah, of course. I think we won’t have one very soon. Because even if we never figure out another research idea, the economic value of the current paradigm and how much further that can be pushed is gonna carry us for many years to come. But it is possible, however unlikely, that we are still missing the key idea to go beyond behavioral cloning and these models are gonna be, like, stuck at human-level forever. There’s a bunch of reasons why I don’t think that’s true but if anyone tells you we could not possibly ever have another winter in this research field you should never believe them.”

I detail why these Sam Altman quotes are concerning on the next page.



## On Sam Altman's Q&A

Firstly, Sam Altman seems highly, highly confident that there exists enough data on the internet to train an AGI system – confident to the point that it makes one question if they've already done it, or are in the process of doing it.

Secondly, the “AI winter” concept generally refers to a period where progress **TOWARDS AGI** has been slowed, but Sam Altman retooled the term to refer to a period where progress **TOWARDS SUPERINTELLIGENCE** is slowed. This seems to suggest that OpenAI has already built an AGI system, or are very close to it, and **AGI is no longer the goal because it already exists.**

As I mentioned earlier in the document, a 100 trillion parameter model is actually slightly suboptimal, but there is a new scaling paradigm OpenAI is using to bridge this gap – it’s based on something called the “Chinchilla scaling laws.”

Chinchilla was an AI model unveiled by DeepMind in early 2022. The implication of the Chinchilla research paper was that current models are significantly undertrained, and with far more compute (meaning more data) would see a massive boost in performance without the need to increase parameters.

The point is, while an undertrained 100 trillion parameter model may be slightly suboptimal, if it were trained on vastly more data it would easily be able to **EXCEED** human-level performance.

The Chinchilla paradigm is widely understood and accepted in the field of machine learning, but just to give a specific example from OpenAI, President Greg Brockman discusses in this interview how OpenAI realized their initial scaling laws were flawed, and have since adjusted to take the Chinchilla laws into account: [https://youtu.be/Rp3A5q9L\\_bg?t=1323](https://youtu.be/Rp3A5q9L_bg?t=1323)

People have said, “training a compute optimal 100 trillion parameter model would cost billions of dollars and just isn’t feasible.” Well, Microsoft just invested \$10 billion into OpenAI in early 2023, so I guess it isn’t such a ridiculous possibility after all...

☰ LW 🔍 LOGIN

## New Scaling Laws for Large Language Models

239  
Ω 71

by ia3orn 6 min read 1st Apr 2022

21 comments

Language Models Machine Learning (ML) GPT AI Frontpage

*Crossposted from the AI Alignment Forum. May contain more technical jargon than usual.*

On March 29th, DeepMind published a paper, "Training Compute-Optimal Large Language Models", that shows that essentially everyone -- OpenAI, DeepMind, Microsoft, etc. -- has been training large language models with a deeply suboptimal use of compute.

Bloomberg Subscribe ☰

Technology + Work Shift

## Microsoft Invests \$10 Billion in ChatGPT Maker OpenAI

WATCH: Microsoft is investing \$10 billion in OpenAI. Bloomberg's Abigail Doolittle reports on "Bloomberg The Open." Source: Bloomberg

Copy Link

The New York Times  
www.nytimes.com

## Microsoft to Invest \$10 Billion in OpenAI, the Creator of ChatGPT

Jan 23, 2023 — The tech giant aims to remain at the forefront of generative artificial intelligence with its partnership with OpenAI.

Developer: OpenAI

Alberto Romero wrote about DeepMind's Chinchilla scaling breakthrough. Chinchilla showed that, despite being vastly smaller than GPT-3 and DeepMind's own Gopher, it outperformed them as a result of being trained on vastly more data. Just to reiterate this one more time: although a 100 trillion parameter model is predicted to achieve slightly suboptimal performance, OpenAI is well aware of the Chinchilla scaling laws (as is pretty much everyone else in the AI field), and they are training Q\* as a 100 trillion parameter multimodal model that is COMPUTE OPTIMAL and trained on far more data than they originally intended. **They have the funds to do it now, through Microsoft.** This will result in a model that FAR, FAR exceeds the performance of what they had initially planned for their 100 trillion parameter model. 100 trillion parameters without Chinchilla scaling laws = roughly human-level but slightly suboptimal. 100 trillion parameters, multimodal, WITH Chinchilla scaling laws taken into account = .....

## A New AI Trend: Chinchilla (70B) Greatly Outperforms GPT-3 (175B) and Gopher (280B)

DeepMind has found the secret to cheaply scale large language models.



Alberto Romero · [Follow](#)

Published in Towards Data Science

11 min read · Apr 11, 2022



Listen



Share

And they proved it. The star of the new paper is Chinchilla, a 70B-parameter model 4 times smaller than the previous leader in language AI, Gopher (also built by DeepMind), but trained on 4 times more data. Researchers found that Chinchilla “uniformly and significantly” outperforms Gopher, GPT-3, Jurassic-1, and Megatron-Turing NLG across a large set of language benchmarks.

The conclusion is clear: Current large language models are “significantly undertrained,” which is a consequence of blindly following the scaling hypothesis — making models larger isn't the only way toward improved performance.

Starting in July of 2022, the US started making moves to block new computer chips from being sent to China, in an attempt to halt their AI progress. This plan was finalized in October of 2022. According to a San Francisco AI researcher, Israel Gonzalez-Brooks, Sam Altman was in DC in September of 2022. Israel has claimed to be in contact with Sam Altman (I could not verify this), but what gives him credibility is the fact that Sam Altman was confirmed to have taken a trip to DC in January of 2023.

If GPT-4/GPT-5 began training in the summer of 2022, and Sam Altman visited DC during this time (probably multiple times), the China chip ban can't possibly be a coincidence.

 <https://www.csis.org> > analysis > choking-chinas-acce...

### Choking off China's Access to the Future of AI

Oct 11, 2022 · Strangle the Chinese AI and supercomputing industries by choking off access to high-end chips. The highest levels of leadership in both the United States and China—including Chinese president Xi Jinping—believe that leading in AI technology is critical to the future of global military and economic power competition.

 Semafor  
<https://www.semafor.com> · article

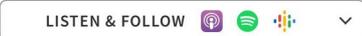
### ChatGPT creator Sam Altman visits Washington to meet lawmakers

Jan 26, 2023 — Sam Altman, creator of ChatGPT and CEO of OpenAI, is making the rounds in D.C. this week, aiming to demystify the artificial intelligence ...



 THE INDICATOR FROM PLANET MONEY



### Why the US cut China off from advanced chips

October 19, 2022 · 7:05 PM ET  
By Darian Woods, Wailin Wong

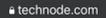
 **Israel Gonzalez-Brooks (6/100 geecs)**  
@izzyz

Yes, Clark spends a lot of time in Washington, DC. One researcher told me Sam Altman, CEO of OpenAI and Jack Clark, were both in Washington around a month ago...

Here's the article on OpenAI from @techreview  
[technologyreview.com/2020/02/17/844...](https://technologyreview.com/2020/02/17/844...)

6:13 PM · Oct 25, 2022

1 Retweet 1 Like



### ON THE CUSP

## The US's moves to contain China's semiconductor industry: a timeline from July

by Ward Zhou Oct 9, 2022



 **Israel Gonzalez-Brooks (6/100 geecs)**  
@izzyz Follows you

18.9K Tweets

postreal altermodern sentimentalist cyberpunk | Web + AI Development | #sentientist | 🤖

San Francisco, CA [researchy.onrender.com](https://researchy.onrender.com)  
Joined August 2011

986 Following 1,955 Followers

Followed by @goth



 My View  Following  Saved

**Technology**

5 minute read · October 10, 2022 4:40 AM CDT · Last Updated 7 months ago

### U.S. aims to hobble China's chip industry with sweeping new export rules

By Stephen Nellis, Karen Freifeld and Alexandra Alper





OpenAI

<https://openai.com> > [blog](#) > [introducing-superalignment](#) ⋮

## Introducing Superalignment

Jul 5, 2023 — Our goal is to build a roughly human-level automated alignment researcher. We can then use vast amounts of compute to scale our efforts, and ...

OpenAI planned to build human-level AI by 2027 and then scale up to superintelligence. This has been delayed because of Elon Musk's lawsuit but it will still be coming shortly

problem.

We are dedicating 20% of the compute we've secured to date over the next four years to solving the problem of superintelligence alignment. Our chief basic research bet is our new Superalignment team, but getting this right is critical to achieve our mission and we expect many teams to contribute, from developing new methods to scaling them up to deployment.

**Our goal is to solve the core technical challenges of superintelligence alignment in four years.**

# Scott Aaronson

Article Talk

☆

☆

✎

**Scott Joel Aaronson** (born May 21, 1981)<sup>[1]</sup> is an American [theoretical computer scientist](#) and David J. Bruton Jr. Centennial Professor of Computer Science at the [University of Texas at Austin](#). His primary areas of research are [quantum computing](#) and [computational complexity theory](#).

**Scott Aaronson**



Aaronson in 2011

## Scott Aaronson is joining OpenAI to work on AI safety

117

by [peterbarnett](#) 1 min read 17th Jun 2022

31 comments

OpenAI Community AI Personal Blog

*This is a linkpost for <https://scottaaronson.blog/?p=6484>*

Scott Aaronson is a computer scientist at the University of Texas in Austin, whose research mainly focuses on quantum computing and complexity theory. He's at least very adjacent to the Rationalist/LessWrong community. After some comments on his blog and then conversations with Jan Leike, he's decided work for one year on AI safety at OpenAI.

In the summer of 2016 he moved from MIT to the [University of Texas at Austin](#) as David J. Bruton Jr. Centennial Professor of Computer Science and as the founding director of UT Austin's new Quantum Information Center.<sup>[6]</sup> In summer 2022 he announced he would be working for a year at [OpenAI](#) on theoretical foundations of AI safety.<sup>[7][8]</sup>

In closing, I'll reveal an incredible source of info – which comes from Scott Aaronson, famous computer scientist. In the summer of 2022 he joined OpenAI for one year to work on AI safety... and he had some very interesting things to say about it on his blog, as I'll show next.

...

Scott Aaronson wrote a blog post in late December of 2022, a “letter to his 11-year-old” self, discussing current events and his accomplishments in life.

Next page is the scary part...



« [Google's Sycamore chip: no wormholes, no superfast classical simulation either](#)

[Short letter to my 11-year-old self](#) [Happy 40th Birthday Dana!](#) »

Dear Scott,

This is you, from 30 years in the future, Christmas Eve 2022. Your Ghost of Christmas Future.

To get this out of the way: you eventually become a professor who works on quantum computing. Quantum computing is ... OK, you know the stuff in popular physics books that never makes any sense, about how a particle takes all the possible paths at once to get from point A to point B, but you never actually see it do that, because as soon as you look, it only takes one path? Turns out, there's something huge there, even though the popular books totally botch the explanation of it. It involves complex numbers. A quantum computer is a new kind of computer people are trying to build, based on the true story.

Anyway, amazing stuff, but you'll learn about it in a few years anyway. That's not what I'm writing about.

I'm writing from a future that ... where to start? I could describe it in ways that sound depressing and even boring, or I could also say things you won't believe. Tiny devices in everyone's pockets with the instant ability to videolink with anyone anywhere, or call up any of the world's information, have become so familiar as to be taken for granted. This sort of connectivity would come in especially handy if, say, a supervirus from China were to ravage the world, and people had to hide in their houses for a year, wouldn't it?

Or what if Donald Trump — you know, the guy who puts his name in giant gold letters in Atlantic City? — became the President of the US, then tried to execute a fascist coup and to abolish the Constitution, and came within a hair of succeeding?

Alright, I was pulling your leg with that last one ... *obviously!* But what about this next one?

## The next part of Scott Aaronson's "letter to my 11-year-old self". Just read it...

There's a company building an AI that fills giant rooms, eats a town's worth of electricity, and has recently gained an astounding ability to converse like people. It can write essays or poetry on any topic. It can ace college-level exams. It's daily gaining new capabilities that the engineers who tend to the AI can't even talk about in public yet. Those engineers do, however, sit in the company cafeteria and debate the meaning of what they're creating. What will it learn to do next week? Which jobs might it render obsolete? Should they slow down or stop, so as not to tickle the tail of the dragon? But wouldn't that mean someone else, probably someone with less scruples, would wake the dragon first? Is there an ethical obligation to tell the world more about this? Is there an obligation to tell it *less*?

I am—you are—spending a year working at that company. My job—your job—is to develop a mathematical theory of how to prevent the AI and its successors from wreaking havoc. Where "wreaking havoc" could mean anything from turbocharging propaganda and academic cheating, to dispensing bioterrorism advice, to, yes, destroying the world.

# CONCLUSION

Scott is referring to Q\*: the multimodal, 125 trillion parameter beast.

Thanks for reading.

-Jackson (@vancouver1717 on twitter)

