

Number of Parameters in GPT-4 (Latest Data)



by [Josh Howarth](#)

February 24, 2025

This is a comprehensive list of stats on the number of parameters in ChatGPT-4 and GPT-4o.

While [OpenAI hasn't publicly released the architecture of their recent models, including GPT-4 and GPT-4o](#), various experts have made estimates.

While these estimates vary somewhat, they all agree on one thing: GPT-4 is massive. It's far larger than previous models and many competitors.

However, [more parameters doesn't necessarily mean better](#).

In this article, we'll explore the details of the parameters within GPT-4 and GPT-4o.

Key Parameters Stats (Top Picks)

- ChatGPT-4 is estimated to have roughly **1.8 trillion** parameters
- GPT-4 is therefore **over ten times larger** than its predecessor, GPT-3
- ChatGPT-4o Mini could be as small as **8 billion** parameters

Instantly Analyze
Any Market

Get My Free Report 📄

Number Of Parameters In ChatGPT-4

According to multiple sources, ChatGPT-4 has approximately 1.8 trillion parameters.

This estimate first came from AI experts like George Hotz, who is also known for being the first person to crack the iPhone.

In June 2023, just a few months after GPT-4 was released, Hotz publicly explained that GPT-4 was comprised of roughly 1.8 trillion parameters. More specifically, the architecture consisted of eight models, with each internal model made up of 220 billion parameters.

GPT4 is 8 x 220B params = 1.7 Trillion params <https://t.co/DW4jrZFEn2>

ok I wasn't sure how widely to spread the rumors on GPT-4 but it seems Soumith is also confirming the same so here's the quick clip!

so yes, GPT4 is technically 10x the size of GPT3, and all the small...
<pic.twitter.com/m2YiaHGVs4>

— swyx  (@swyx) June 20, 2023

While this has not been confirmed by OpenAI, the 1.8 trillion parameter claim has been supported by multiple sources.

Shortly after Hotz made his estimation, a [report by Semianalysis](#) reached the same conclusion. More recently, [a graph](#) displayed at Nvidia's GTC24 seemed to support the 1.8 trillion figure.

And back in June 2023, Meta engineer [Soumith Chintala confirmed Hotz's description](#), tweeting:

“I might have heard the same ... info like this is passed around but no one wants to say it out loud. GPT-4: 8 x 220B experts trained with different data/task distributions and 16-inter inference.”

So, we can assert with reasonable confidence that GPT-4 has 1.76 trillion parameters. But what exactly does the breakdown – eight experts of 220 billion parameters, or 16 of 110 billion – mean?

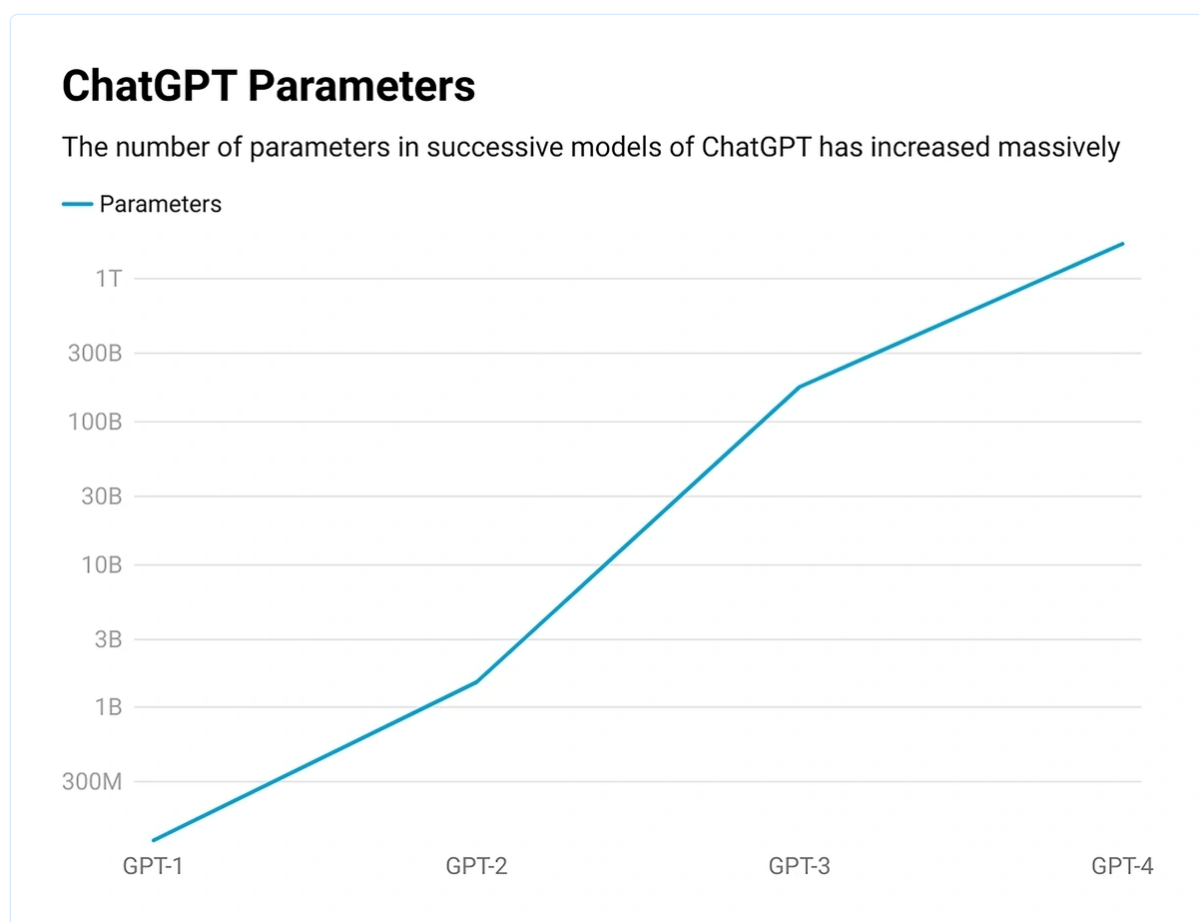
Evolution Of ChatGPT Parameters

ChatGPT-4 has 10x the parameters of GPT-3 (OpenAI)

Back when OpenAI was actually open source, it confirmed the number of parameters in GPT-3: 175 billion.

ChatGPT-4 has over 15,000x the parameters of GPT-1 (MakeUseOf)

ChatGPT-1 was made up of 117 million parameters.



What Are AI Parameters?

AI models like ChatGPT work by breaking down textual information into tokens. A token is roughly the same as three-quarters of an English word.

GPT-4o & GPT-4o mini (coming soon)

GPT-3.5 & GPT-4

GPT-3 (Legacy)

This sentence is an example to demonstrate how ChatGPT (and similar AI tools) break up English text into tokens.

Clear Show example

Tokens	Characters
24	112

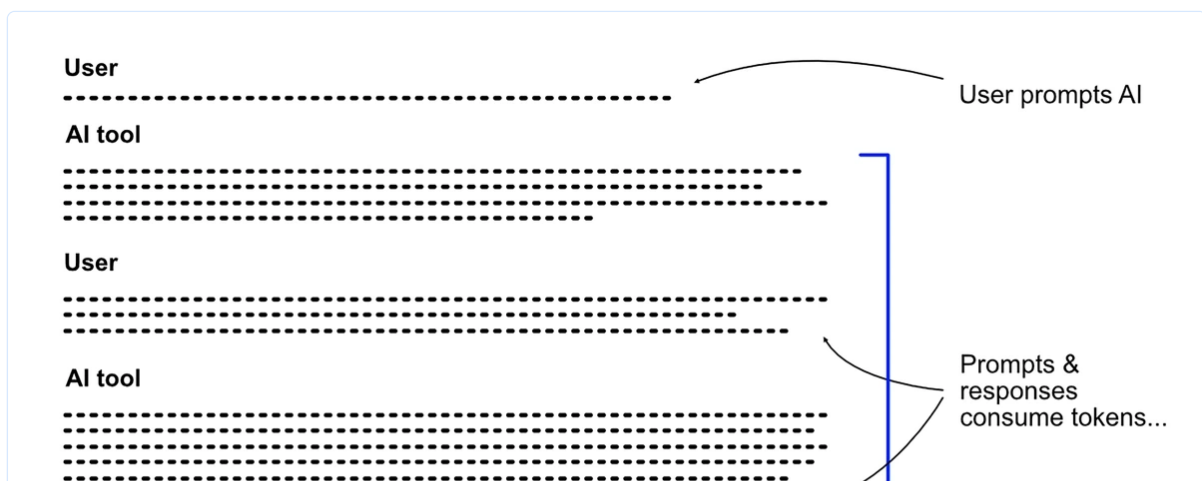
This sentence is an example to demonstrate how ChatGPT (and similar AI tools) break up English text into tokens.

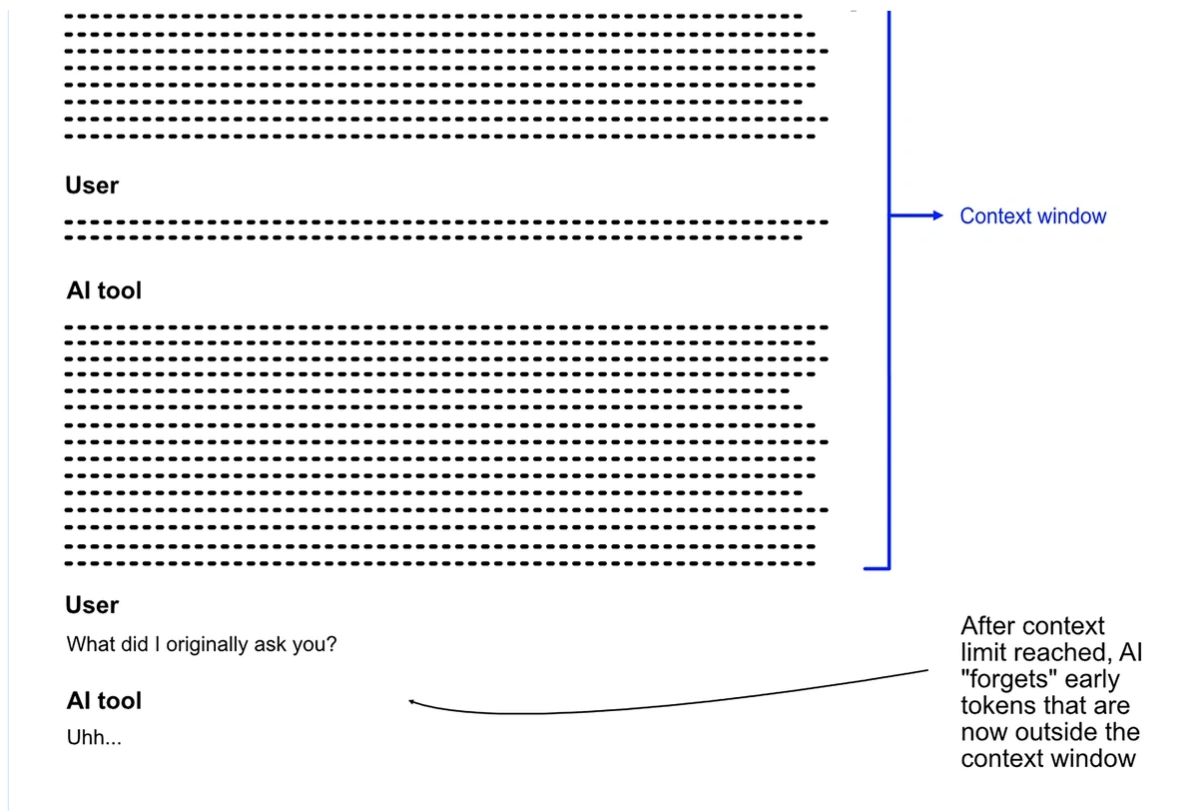
The number of tokens an AI can process is referred to as the context length or window. ChatGPT-4, for example, has a context window of 32,000 tokens. That's about 24,000 words in English.

Want to Uncover Your True Audience? 🔍

Get My Free Report 📄

Once you surpass that number, the model will start to "forget" the information sent earlier. This can lead to mistakes and hallucinations.





Parameters are what determine how an AI model can process these tokens. They are sometimes compared to neurons in the brain. **The human brain has some 86 billion neurons**. The connections and interactions between these neurons are fundamental for everything our brain – and therefore body – does. They interpret information and determine how to respond.

Research shows that adding more neurons and connections to a brain can help with learning. A jellyfish possesses just a few thousand neurons. A snake might have ten million. A pigeon, a few hundred million.

In turn, AI models with more parameters have demonstrated greater information processing ability.

This isn't always the case, however.

More parameters aren't always better

An AI with more parameters might be generally better at processing information. However, there are downsides.

The most obvious is cost. **OpenAI's Sam Altman has said that the company spent more**

than \$100 million training GPT-4. The CEO of Anthropic, meanwhile, warned that “by 2025 we may have a \$10 billion model.”

This is the main reason OpenAI recently released GPT-4o mini, a “cost-efficient small model.” Despite being smaller than GPT-4 – that is, having fewer parameters – it outperforms its predecessor on several important benchmark tests.

Why ChatGPT-4 Has Multiple Models

ChatGPT-4 is made up of eight models, each with 220 billion parameters. But what, exactly, does that mean?

Previous AI models were built using the “dense transformer” architecture. ChatGPT-3, Google PaLM, Meta LLAMA, and dozens of other early models used this formula.

ChatGPT-4, however, uses a notably different architecture.

Instead of piling all the parameters together, GPT-4 uses the “Mixture of Experts” (MoE) architecture. This isn’t a new idea, but its use by OpenAI is.

Each of the eight models within GPT-4 is composed of two “experts.” In total, GPT-4 has 16 experts, each with 110 billion parameters. As you might expect, each expert is specialized.

Therefore, when GPT-4 receives a request, it can route it through just one or two of its experts – whichever are most capable of processing and responding.

That way, GPT-4 can respond to a range of complex tasks in a more cost-efficient and timely manner. In reality, far fewer than 1.8 trillion parameters are actually being used at any one time.

ChatGPT-4o Mini has around 8 billion parameters (TechCrunch)

According to an article published by TechCrunch in July, OpenAI’s new ChatGPT-4o Mini is comparable to Llama 3 8b, Claude Haiku, and Gemini 1.5 Flash. Llama 3 8b is one of Meta’s open-source offerings, and has just 7 billion parameters. That would make GPT-4o Mini remarkably small, considering its impressive performance on various

benchmark tests.

Number Of Parameters In ChatGPT-4o

As stated above, ChatGPT-4 may have around 1.8 trillion parameters.

However, the exact number of parameters in GPT-4o is somewhat less certain than in GPT-4.

OpenAI has not confirmed how many there are. However, OpenAI's CTO [has said](#) that GPT-4o “brings GPT-4-level intelligence to everything.” If that's true, then GPT-4o might also have 1.8 trillion parameters – [an implication made by CNET](#).

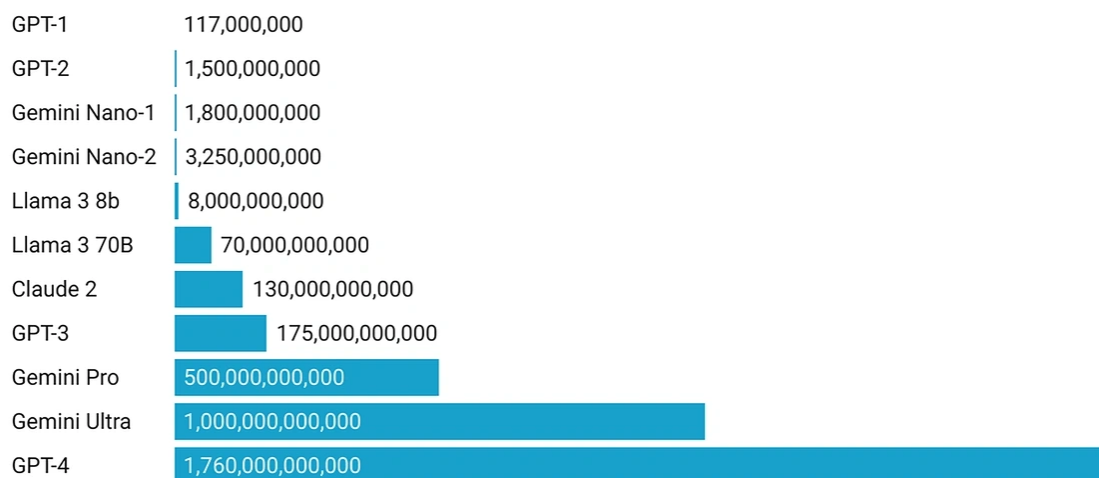
Nevertheless, that connection hasn't stopped other sources from providing their own guesses as to GPT-4o's size.

[The Times of India](#), for example, estimated that ChatGPT-4o has over 200 billion parameters.

Number Of Parameters In Other AI Models

Parameters in Selected AI Models

Some of these figures are estimates. Newer models are many times larger than their predecessors.



GPT-1	117,000,000
GPT-2	1,500,000,000
Gemini Nano-1	1,800,000,000
Gemini Nano-2	3,250,000,000
Llama 3 8b	8,000,000,000
Llama 3 70B	70,000,000,000
Claude 2	130,000,000,000
GPT-3	175,000,000,000
Gemini Pro	500,000,000,000
Gemini Ultra	1,000,000,000,000
GPT-4	1,760,000,000,000

Gemini Ultra may have over 1 trillion parameters **(Manish Tyagi)**

Google, perhaps following OpenAI's lead, has not publicly confirmed the size of its latest AI models. However, one estimate puts Gemini Ultra at over 1 trillion parameters.

Gemini Nano-2 has 3.25 billion parameters (New Scientist)

Google's Gemini Nano comes in two versions. The first, Nano-1, is made of 1.8 billion parameters. The second, Nano-2, is notably larger, at 3.25 billion parameters. Both models are condensed versions of larger predecessor models. They are intended to be used by smartphones.

Llama 2 has 70 billion parameters (Meta)

Meta's open-source model was trained on two trillion tokens of data, 40% more than Llama 1.

[Free] Steal Your Competitor's Strategy

Reveal Their Methods 📌

Claude 2 has over 130 billion parameters (Text Cortex)

Claude 2 was released by Anthropic in July 2023.

Claude 3 Opus could have 137 billion parameters (Cogni Down Under)

Unfortunately, this estimate did not provide a source.

While the 1.76 trillion figure is the most widely accepted estimate, it's far from the only guess.

ChatGPT-4 has 1 trillion parameters ([Semafor](#))

The 1 trillion figure has been [thrown around a lot](#), including by authoritative sources like reporting outlet [Semafor](#).

ChatGPT-4 has 100 trillion parameters ([AX Semantics](#))

This is according to the CEO of Cerebras, Andrew Feldman. He claims to have learned this in a conversation with OpenAI.

ChatGPT-4 was trained on 13 trillion tokens ([The Decoder](#))

According to The Decoder, which was one of the first outlets to report on the 1.76 trillion figure, ChatGPT-4 was trained on roughly 13 trillion tokens of information. This training set reportedly included text and code. It was likely drawn from web crawlers like CommonCrawl, and may have also included information from social media sites like Reddit. There's a chance OpenAI included information from textbooks and other proprietary sources.

Conclusion

Unfortunately, many AI developers – OpenAI included – have become reluctant to publicly release the number of parameters in their newer models.

Nevertheless, experts have made estimates as to the sizes of many of these models.

While models like ChatGPT-4 continued the trend of models becoming larger in size, more recent offerings like GPT-4o Mini perhaps imply a shift in focus to more cost-efficient tools.