

Artificial Intelligence: The Release of the Upcoming Open AI Model, GPT-5

GPT-5 Features and Capabilities (Expected)

By [Arjun Sha](#)

Global Research, May 24, 2023

[Beebom](#) 19 May 2023

Region: [USA](#)

Theme: [Intelligence](#)

All Global Research articles can be read in 51 languages by activating the Translate Website button below the author's name.

To receive Global Research's Daily Newsletter (selected articles), [click here](#).

Click the share button above to email/forward this article to your friends and colleagues. Follow us on [Instagram](#) and [Twitter](#) and subscribe to our [Telegram Channel](#). Feel free to repost and share widely Global Research articles.

It has been just over two months since the launch of GPT-4, but users have started anticipating the release of GPT-5. We have already seen how capable and powerful GPT-4 is in various kinds of tests and qualitative evaluations. With many new features like [ChatGPT plugins](#) and internet browsing capability, it has gotten even better. Now, users are waiting to learn more about the upcoming Open AI model, GPT-5, the possibility of AGI, and more. So to find in-depth information about GPT-5's release date and other expected features, follow our explainer below.

When GPT-4 was released in March 2023, it was expected that OpenAI would release its next-generation model by December 2023. Siqi Chen, the CEO of Runway also [tweeted](#) saying that "*gpt5 is scheduled to complete training this December.*" However, speaking at an MIT event in April, OpenAI CEO Sam Altman said "*We are not and won't for some time*" when asked if OpenAI is training GPT-5. So the rumor of GPT-5 releasing by the end of 2023 is already quashed.

That said, experts suggest that OpenAI might come out with GPT-4.5, an intermediate release between GPT-4 and GPT-5 by October 2023, just like GPT-3.5. It's being said that GPT-4.5 will finally bring the multimodal capability, aka the ability to analyze both images and texts. OpenAI has already announced and demonstrated GPT-4's [multimodal capabilities](#) during the GPT-4 Developer livestream back in March 2023.

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.

GPT-4 Multimodal Capability

Apart from that, OpenAI currently has a lot on its plate to iron out on the GPT-4 model before it starts working on GPT-5. Currently, GPT-4's inference time is very high and it's quite expensive to run. GPT-4 API access is still hard to get by. Moreover, OpenAI just recently opened up [access to ChatGPT plugins](#) and internet browsing capability, which are still in Beta. It's yet to bring Code Interpreter for all the paying users, which is again in the Alpha phase.

While GPT-4 is plenty powerful, I guess OpenAI realizes that compute efficiency is one of the key elements for running a model sustainably. And well, add new features and capabilities to the mix, and you have a bigger infrastructure to deal with while making sure that all checkpoints are up and running reliably. So to hazard a guess, GPT-5 is likely to come out in 2024, just around Google Gemini's release, if we assume government agencies don't put a regulatory roadblock.

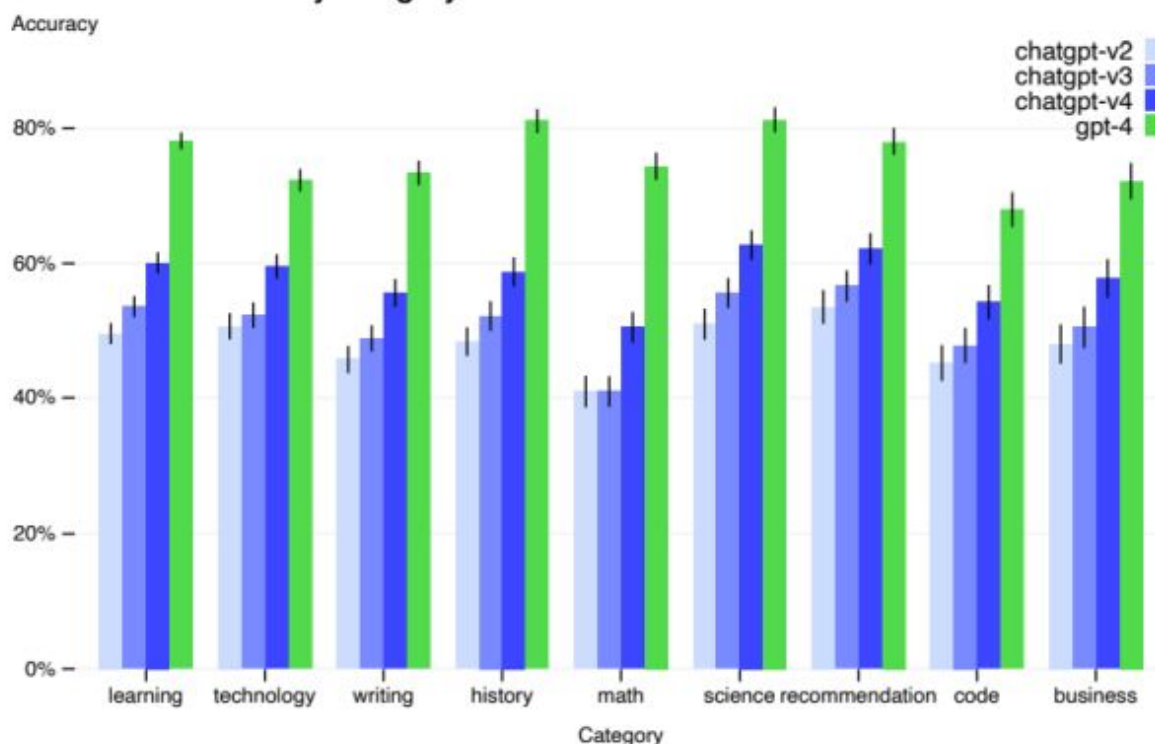
GPT-5 Features and Capabilities (Expected)

Reduced Hallucination

The hot talk in the industry is that GPT-5 will achieve AGI (Artificial General Intelligence), but we will come to that later on in detail. Besides that, GPT-5 is supposed to reduce the inference time, enhance efficiency, bring down further hallucinations, and a lot more. Let's

start with hallucination, which is one of the key reasons why most users don't readily believe in AI models.

Internal factual eval by category



GPT-4 Accuracy Test

According to OpenAI, GPT-4 scored 40% higher than GPT-3.5 in internal adversarially-designed factual evaluations under all nine categories. Now, GPT-4 is 82% less likely to respond to inaccurate and disallowed content. It's very close to touching the 80% mark in accuracy tests across categories. That's a huge leap in combating hallucination.

Now, it's expected that OpenAI would reduce hallucination to less than 10% in GPT-5, which would be huge for making LLM models trustworthy. I have been using the GPT-4 model for a lot of tasks lately, and it has so far given factual responses only. So it's highly likely that GPT-5 will hallucinate even less than GPT-4.

Compute-efficient Model

Next, we already know that GPT-4 is expensive to run (\$0.03 per 1K tokens) and the inference time is also higher. Whereas, the older GPT-3.5-turbo model is 15x cheaper (\$0.002 per 1K tokens) than GPT-4. That's because GPT-4 is trained on a massive 1 trillion parameters, which requires a costly compute infrastructure. In our recent explainer on [Google's PaLM 2 model](#), we found that PaLM 2 is quite smaller in size and that results in quick performance.

A recent [report](#) by CNBC confirmed that PaLM 2 is trained on 340 billion parameters, which is far less than GPT-4's large parameter size. Google even went on to say that bigger is not always better and research creativity is the key to making great models. So if OpenAI wants to make its upcoming models compute-optimal, it must find new creative ways to reduce the size of the model while maintaining the output quality.

Came back to a project I was working on with OpenAI GPT-4 API, noticed the API response times were pretty slow.

Tested average response on a fresh context for "Can you show me a basic scatter matplotlib example?"

GPT-3.5: 13.4 seconds

GPT-4: 44.7 seconds.

□□□

— Harrison Kinsley (@Sentdex) [May 10, 2023](#)

A huge chunk of OpenAI revenue comes from enterprises and businesses, so yeah, GPT-5 must not only be cheaper but also faster to return output. Developers are already berating the fact that GPT-4 API calls frequently stop responding and they are forced to use the GPT-3.5 model in production. It must be on OpenAI's wishlist to improve performance in the upcoming GPT-5 model, especially after the launch of Google's much-faster [PaLM 2 model, which you can try right now](#).

Multisensory AI Model

While GPT-4 has been announced as a multimodal AI model, it deals with only two types of data i.e. images and texts. Sure, the capability has not been added to GPT-4 yet, but OpenAI may possibly release the feature in a few months. However, with GPT-5, OpenAI may take a big leap in making it truly multimodal. It may also deal with text, audio, images, videos, depth data, and temperature. It would be able to interlink data streams from different modalities to create an embedding space.

Recently, Meta released [ImageBind](#), an AI model that combines data from six different modalities and open-sourced it for research purposes. In this space, OpenAI has not revealed much, but the company does have some strong foundation models for vision analysis and image generation. OpenAI has also developed CLIP (Contrastive Language-Image Pretraining) for analyzing images and DALL-E, a [popular Midjourney alternative](#) that can generate images from textual descriptions.

It's an area of ongoing research and its applications are still not clear. According to Meta, it can be used to design and create immersive content for virtual reality. We need to wait and see what OpenAI does in this space and if we will see more AI applications across various multimodalities with the release of GPT-5.

Long Term Memory

With the release of GPT-4, OpenAI brought a maximum context length of 32K tokens, which cost \$0.06 per 1K token. We have rapidly seen the transformation from the standard 4K tokens to 32K in a few months. Recently, Anthropic increased the context window from 9K to 100K tokens in its [Claude AI chatbot](#). It's expected that GPT-5 might bring long-term memory support via a much larger context length.

This can help in making AI characters and friends who remember your persona and memories that can last for years. Apart from that, you can load libraries of books and text

documents in a single context window. There can be various new AI applications due to long-term memory support and GPT-5 can make that possible.

GPT-5 Release: Fear of AGI?

In February 2023, Sam Altman wrote a [blog](#) on AGI and how it can benefit all of humanity. AGI (Artificial General Intelligence), as the name suggests, is the next generation of AI systems that is generally smarter than humans. It's being said that OpenAI's upcoming model GPT-5 will achieve AGI, and it seems like there is some truth in that.

We already have several autonomous AI agents like [Auto-GPT](#) and [BabyAGI](#), which are based on GPT-4 and can take decisions on their own and come up with reasonable conclusions. It's entirely possible that some version of AGI will be deployed with GPT-5.

In the blog, Altman says that *"We believe we have to continuously learn and adapt by deploying less powerful versions of the technology in order to minimize 'one shot to get it right' scenarios"* while also acknowledging *"massive risks"* in navigating vastly powerful systems like AGI. Before the recent [Senate hearing](#), Sam Altman also urged US lawmakers for regulations around newer AI systems.

In the hearing, Altman said, *"I think if this technology goes wrong, it can go quite wrong. And we want to be vocal about that."* Further, he added, *"We want to work with the government to prevent that from happening."* For some time, OpenAI has become quite vocal about regulations on newer AI systems that would be highly powerful and intelligent. Do note that Altman is seeking safety regulation around incredibly powerful AI systems and not open-source models or AI models developed by small startups.

regulation should take effect above a capability threshold.

AGI safety is really important, and frontier models should be regulated.

regulatory capture is bad, and we shouldn't mess with models below the threshold. open source models and small startups are obviously important.
<https://t.co/qdWHFjX4s>

— Sam Altman (@sama) [May 18, 2023](#)

It should be worth noting that Elon Musk and other prominent personalities, including Steve Wozniak, Andrew Yang, and Yuval Noah Harari, et al called for a [pause on giant AI experiments](#), back in March 2023. Since then, there has been a wide pushback against AGI and newer AI systems — more powerful than GPT-4.

If OpenAI is indeed going to bring AGI capability to GPT-5, then expect more delay in its public release. Regulation would definitely kick in and work around safety and alignment would be scrutinized thoroughly. The good thing is that OpenAI already has a powerful GPT-4 model, and it's continuously adding new features and capabilities. There is no other AI model that comes close to it, not even the [PaLM 2-based Google Bard](#).

OpenAI GPT-5: Future Stance

After the release of GPT-4, OpenAI has gotten increasingly secretive about its operations. It no longer shares research on the training dataset, architecture, hardware, training compute, and training method with the open-source community. It has been a strange flip for a company that was founded as a nonprofit (now it's capped profit) based on the principles of free collaboration.

In March 2023, speaking with [The Verge](#), Ilya Sutskever, the chief scientist of OpenAI said, *"We were wrong. Flat out, we were wrong. If you believe, as we do, that at some point, AI — AGI — is going to be extremely, unbelievably potent, then it just does not make sense to open-source. It is a bad idea... I fully expect that in a few years, it's going to be completely obvious to everyone that open-sourcing AI is just not wise."*

Now, it has become clear that neither GPT-4 nor the upcoming GPT-5 would be open-source in order to stay competitive in the AI race. However, another giant corporation, Meta has been approaching AI development differently. Meta has been releasing multiple AI models under the CC BY-NC 4.0 license (research only, non-commercial) and gaining traction among the open-source community.

Seeing the huge adoption of Meta's LLaMA and other AI models, OpenAI has also changed its stance on open source. According to recent [reports](#), OpenAI is working on a new open-source AI model that will be released to the public soon. There is no information on its capabilities and how competitive it will be against GPT-3 .5 or GPT-4, but it's indeed a welcome change.

In summation, GPT-5 is going to be a frontier model that will push the boundary of what is possible with AI. It seems likely that some form of AGI will launch with GPT-5. And if that will be the case, OpenAI must get ready for tight regulation (and possible bans) around the world. As for the GPT-5 release date, the safe bet would be sometime in 2024.

*

Note to readers: Please click the share button above. Follow us on Instagram and Twitter and subscribe to our Telegram Channel. Feel free to repost and share widely Global Research articles.

Featured image: OpenAI headquarters, Pioneer Building, San Francisco (Licensed under CC BY-SA 4.0)

The original source of this article is [Beebom](#)

Copyright © [Arjun Sha](#), [Beebom](#), 2023

[Comment on Global Research Articles on our Facebook page](#)

[Become a Member of Global Research](#)

Articles by: [Arjun Sha](#)

Disclaimer: The contents of this article are of sole responsibility of the author(s). The Centre for Research on Globalization will not be responsible for any inaccurate or incorrect statement in this article. The Centre of Research on Globalization grants permission to cross-post Global Research articles on community internet sites as long the source and copyright are acknowledged together with a hyperlink to the original Global Research article. For publication of Global Research articles in print or other forms including commercial internet sites, contact: publications@globalresearch.ca

www.globalresearch.ca contains copyrighted material the use of which has not always been specifically authorized by the copyright owner. We are making such material available to our readers under the provisions of "fair use" in an effort to advance a better understanding of political, economic and social issues. The material on this site is distributed without profit to those who have expressed a prior interest in receiving it for research and educational purposes. If you wish to use copyrighted material for purposes other than "fair use" you must request permission from the copyright owner.

For media inquiries: publications@globalresearch.ca