GPT-4o

GPT-4o ("o" for "omni") is a multilingual, multimodal generative pre-trained transformer developed by OpenAI and released in May 2024.[1] GPT-4o is free, but ChatGPT Plus subscribers have higher usage limits.[2] It can process and generate text, images and audio.[3] Its application programming interface (API) is faster and cheaper than its predecessor, GPT-4 Turbo.[1]

Generative Pre-trained Transformer 4 Omni (GPT-4o)	
Developer(s)	OpenAI
Initial release	May 13, 2024
Preview release	ChatGPT-4o-latest (2025-03-26) / March 26, 2025
Predecessor	GPT-4 Turbo
Successor	OpenAI o1 GPT-4.5
Type	Multimodal Large language model Generative pre-trained transformer Foundation model
License	Proprietary
Website	openai.com/index/hello-gpt-4o (https://openai.com/index/hello-gpt-4o)

Background

Multiple versions of GPT-4o were originally secretly launched under different names on Large Model Systems Organization's (LMSYS) Chatbot Arena as three different models. These three models were called gpt2-chatbot, im-a-good-gpt2-chatbot, and im-also-a-good-gpt2-chatbot.[4] On 7 May 2024, OpenAI CEO Sam Altman tweeted "im-a-good-gpt2-chatbot", which was commonly interpreted as a confirmation that these were new OpenAI models being A/B tested.[5][6]

Capabilities

Examples of images generated by GPT-4o. These include a young girl at a farmer's market (left), and Karl Marx in a modern-day context (right).

When released in May 2024, GPT-4o achieved state-of-the-art results in voice, multilingual, and vision benchmarks, setting new records in audio speech recognition and translation.[7][8] GPT-4o scored 88.7 on the Massive Multitask Language Understanding (MMLU) benchmark compared to 86.5 for GPT-4.[9] Unlike GPT-3.5 and GPT-4, which rely on other models to process sound, GPT-4o natively supports voice-to-voice.[9] The Advanced Voice Mode was delayed and finally released to ChatGPT Plus and Team subscribers in September 2024.[10] On 1 October 2024, the Realtime API was introduced.[11]

When released, the model supported over 50 languages,[1] which OpenAI claims cover over 97%

of speakers.[12] Mira Murati demonstrated the model's multilingual capability by speaking Italian to the model and having it translate between English and Italian during the live-streamed OpenAI demonstration event on 13 May 2024. In addition, the new tokenizer[13] uses fewer tokens for certain languages, especially languages that are not based on the Latin alphabet, making it cheaper for those languages.[9]

GPT-4o has knowledge up to October 2023,[14][15] but can access the Internet if up-to-date information is needed. It has a context length of 128k tokens.[14]

On March 25, 2025, OpenAI released an image-generation feature that is native to GPT-4o, as an alternative to DALL-E 3. It was made available to paid users, with the rollout to free users being delayed.[16] The use of the feature was subsequently limited, with Sam Altman noting in a Tweet that "[their] GPUs were melting" from its unprecedented popularity.[17]

Corporate customization

In August 2024, OpenAI introduced a new feature allowing corporate customers to customize GPT-4o using proprietary company data. This customization, known as fine-tuning, enables businesses to adapt GPT-4o to specific tasks or industries, enhancing its utility in areas like customer service and specialized knowledge domains. Previously, fine-tuning was available only on the less powerful model GPT-4o mini.[18][19]

The fine-tuning process requires customers to upload their data to OpenAI's servers, with the training typically taking one to two hours. OpenAI's focus with this rollout is to reduce the complexity and effort required for businesses to tailor AI solutions to their needs, potentially increasing the adoption and effectiveness of AI in corporate environments.[20][18]

GPT-4o mini

On July 18, 2024, OpenAI released a smaller and cheaper version, **GPT-4o mini**.[21]

According to OpenAI, its low cost is expected to be particularly useful for companies, startups, and developers that seek to integrate it into their services, which often make a high number of API calls. Its API costs $0.15 per million input tokens and $0.6 per million output tokens, compared to $2.50 and $10,[22] respectively, for GPT-4o. It is also significantly more capable and 60% cheaper than GPT-3.5 Turbo, which it replaced on the ChatGPT interface.[21] The price after fine-tuning doubles: $0.3 per million input tokens and $1.2 per million output tokens.[22] It is estimated that its parameter count is 8B.[23]

GPT-4o mini is the default model for guests and those who have hit the limit for GPT-4o.

Controversies

Scarlett Johansson controversy

As released, GPT-4o offered five voices: Breeze, Cove, Ember, Juniper, and Sky. A similarity between the voice of American actress Scarlett Johansson and Sky was quickly noticed. On May

14, *Entertainment Weekly* asked themselves whether this likeness was on purpose.[24] On May 18, Johansson's husband, Colin Jost, joked about the similarity in a segment on *Saturday Night Live*.[25] On May 20, 2024, OpenAI disabled the Sky voice, issuing a statement saying "We've heard questions about how we chose the voices in ChatGPT, especially Sky. We are working to pause the use of Sky while we address them."[26]

Scarlett Johansson starred in the 2013 sci-fi movie *Her*, playing Samantha, an artificially intelligent virtual assistant personified by a female voice. As part of the promotion leading up to the release of GPT-4o, Sam Altman on May 13 tweeted a single word: "her".[27][28]

OpenAI stated that each voice was based on the voice work of a hired actor. According to OpenAI, "Sky's voice is not an imitation of Scarlett Johansson but belongs to a different professional actress using her own natural speaking voice."[26] CTO Mira Murati stated "I don't know about the voice. I actually had to go and listen to Scarlett Johansson's voice." OpenAI further stated the voice talent was recruited before reaching out to Johansson.[28][29]

On May 21, Johansson issued a statement explaining that OpenAI had repeatedly offered to make her a deal to gain permission to use her voice as early as nine months prior to release, a deal she rejected. She said she was "shocked, angered, and in disbelief that Mr. Altman would pursue a voice that sounded so eerily similar to mine that my closest friends and news outlets could not tell the difference." In the statement, Johansson also used the incident to draw attention to the lack of legal safeguards around the use of creative work to power leading AI tools, as her legal counsel demanded OpenAI detail the specifics of how the Sky voice was created.[28][30]

Observers noted similarities to how Johansson had previously sued and settled with The Walt Disney Company for breach of contract over the direct-to-streaming rollout of her Marvel film *Black Widow*,[31] a settlement widely speculated to have netted her around $40M.[32]

Also on May 21, Shira Ovide at *The Washington Post* shared her list of "most bone-headed self-owns" by technology companies, with the decision to go ahead with a Johansson sound-alike voice despite her opposition and then denying the similarities ranking 6th.[33] On May 24, Derek Robertson at *Politico* wrote about the "massive backlash", concluding that "appropriating the voice of one of the world's most famous movie stars — in reference [...] to a film that serves as a cautionary tale about over-reliance on AI — is unlikely to help shift the public back into [Sam Altman's] corner anytime soon."[34]

Studio Ghibli filter

Upon the launch of GPT-4o's image generation on March 2025, photographs recreated in the style of Studio Ghibli films went viral.[36] Sam Altman acknowledged the trend by changing his profile pic into a Studio Ghibli-inspired one.[37][38] The use of the Ghibli style was challenged, with the Associated Press and *The New York Times* noting that Hayao Miyazaki was critical of AI art in the 2016 documentary *Never-Ending Man: Hayao Miyazaki*.[35][39] The use of the Ghibli-filter faced further controversy when the White House's official Twitter account posted a Ghiblified version of a previously convicted Dominican immigrant getting arrested by an immigration officer.[35][40][41] North American distributor GKids responded to the trend in a press release, comparing the use of the filter to its coinciding IMAX re-release of the 1997 Studio Ghibli film, *Princess Mononoke*.[42]

An image of a migrant woman getting arrested posted by the White House's official Twitter account. The depiction of the arrest in the style of Studio Ghibli was criticized.[35]