

Title: GPT-4o

URL: <https://en.wikipedia.org/wiki/GPT-4o>

PageID: 76899001

Categories: Category:2024 in artificial intelligence, Category:2024 software, Category:Artificial intelligence art, Category:ChatGPT, Category:Generative pre-trained transformers, Category:Large language models, Category:Text-to-image generation

Source: Wikipedia (CC BY-SA 4.0).

GPT-4o ("o" for " omni ") is a multilingual, multimodal generative pre-trained transformer developed by OpenAI and released in May 2024. [2] It can process and generate text, images and audio. [3] [4]

Upon release, GPT-4o was free in ChatGPT , though paid subscribers had higher usage limits. [5] GPT-4o was removed from ChatGPT in August 2025 when GPT-5 was released, but OpenAI reintroduced it for paid subscribers after users complained about the sudden removal. [6]

GPT-4o's audio-generation capabilities were used in ChatGPT's Advanced Voice Mode. [7] On July 18, 2024, OpenAI released GPT-4o mini, a smaller version of GPT-4o which replaced GPT-3.5 Turbo on the ChatGPT interface. [8] GPT-4o's ability to generate images was released later, in March 2025, when it replaced DALL-E 3 in ChatGPT. [9]

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. [10] 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 . [11] [12]

Capabilities

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. [13] [14] [15] GPT-4o scored 88.7 on the Massive Multitask Language Understanding (MMLU) benchmark compared to 86.5 for GPT-4. [16] Unlike GPT-3.5 and GPT-4, which rely on other models to process sound, GPT-4o natively supports voice-to-voice. [16] The Advanced Voice Mode was delayed and finally released to ChatGPT Plus and Team subscribers in September 2024. [17] On 1 October 2024, the Realtime API was introduced. [18]

When released, the model supported over 50 languages, [2] which OpenAI claims cover over 97% of speakers. [19] 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 [20] uses fewer tokens for certain languages, especially languages that are not based on the Latin alphabet , making it cheaper for those languages. [16]

GPT-4o has knowledge up to October 2023, [21] [22] but can access the Internet if up-to-date information is needed. It has a context length of 128k tokens. [21]

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. [23] [24]

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. [25] [23]

GPT-4o mini

On July 18, 2024, OpenAI released a smaller and cheaper version, GPT-4o mini . [26]

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, [27] 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. [26] The price after fine-tuning doubles: \$0.3 per million input tokens and \$1.2 per million output tokens. [27]

GPT Image 1

On March 25, 2025, OpenAI released an image-generation model that is native to GPT-4o, as the successor to DALL-E 3 . The model was later named as GPT Image 1 (gpt-image-1) and introduced to the API on April 23. It was made available to paid users, with the rollout to free users being delayed. [28] The use of the feature was subsequently limited, with Sam Altman noting in a Tweet that "[their] GPUs were melting" from its unprecedented popularity. [29] OpenAI later revealed that over 130 million users around the world created more than 700 million images with GPT Image 1 in just the first week■. [30]

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. [31] On May 18, Johansson's husband, Colin Jost , joked about the similarity in a segment on Saturday Night Live . [32] 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." [33]

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". [34] [35]

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." [33] 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. [35] [36]

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. [35] [37]

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 , [38] a settlement widely speculated to have netted her around \$40M. [39]

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. [40] 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." [41]

Studio Ghibli filter

Upon the launch of GPT-4o's image generation (later named as GPT Image 1) in March 2025, photographs recreated in the style of Studio Ghibli films went viral. [43] Sam Altman acknowledged the trend by changing his Twitter profile picture into a Studio Ghibli-inspired one. [44] [45] The White House 's official Twitter account posted a Ghibli-style image mocking the arrest by immigration authorities of Virginia Basora-Gonzalez, a migrant previously deported after being convicted of fentanyl trafficking, which shows her crying as an immigration officer places her in handcuffs. [42] [46] [47] 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 . [48]

Sycophancy

In April 2025, OpenAI rolled back an update of GPT-4o due to excessive sycophancy , after widespread reports that it had become flattering and agreeable to the point of supporting clearly delusional or dangerous ideas . [49]

Removal with GPT-5

On August 7, 2025, OpenAI released GPT-5 . Its release was criticized as, with it, legacy GPT models were no longer available via ChatGPT, including GPT-4o, [50] except for Pro users. [51] Some users were particularly frustrated over this removal without prior warning because they used different GPT models for distinct purposes and found that GPT-5's router system left them with less control. [52] In addition, some users preferred GPT-4o's warmer and more personal tone over that of GPT-5, which they described as "flat", "uncreative" and "lobotomized", [53] and resembling an "overworked secretary". [54]

As a response, in a post on X , Sam Altman said that OpenAI would bring back the option to select GPT-4o to Plus users as well, and "[w]e [OpenAI] will watch usage as we think about how long to offer legacy models for." [52] [55] He also stated: "We for sure underestimated how much some of the things that people like in GPT-4o matter to them, even if GPT-5 performs better in most ways". [56] "Long-term, this has reinforced that we really need good ways for different users to customize things (we understand that there isn't one model that works for everyone, and we have been investing in steerability research and launched a research preview of different personalities)". [53] On August 13, 2025, Altman wrote on X that OpenAI is working on GPT-5's personality to make the model "feel warmer". [57]

See also

Apple Intelligence

List of large language models

References

v

t

e

ChatGPT in education GPT Store DALL-E ChatGPT Search Sora Whisper

in education

GPT Store

DALL-E

ChatGPT Search

Sora

Whisper

GitHub Copilot

OpenAI Codex

Generative pre-trained transformer GPT-1 GPT-2 GPT-3 GPT-4 GPT-4o o1 o3 GPT-4.5 GPT-4.1
o4-mini GPT-OSS GPT-5

GPT-1

GPT-2

GPT-3

GPT-4

GPT-4o

o1

o3

GPT-4.5

GPT-4.1

o4-mini

GPT-OSS

GPT-5

ChatGPT Deep Research

Operator

Sam Altman removal

removal

Greg Brockman

Sarah Friar

Jakub Pachocki

Scott Schools

Mira Murati

Emmett Shear

Sam Altman

Adam D'Angelo

Sue Desmond-Hellmann

Zico Kolter

Paul Nakasone

Adebayo Ogunesi

Nicole Seligman

Fidji Simo

Lawrence Summers

Bret Taylor (chair)
Greg Brockman (2017–2023)
Reid Hoffman (2019–2023)
Will Hurd (2021–2023)
Holden Karnofsky (2017–2021)
Elon Musk (2015–2018)
Ilya Sutskever (2017–2023)
Helen Toner (2021–2023)
Shivon Zilis (2019–2023)
Stargate LLC
Apple Intelligence
AI Dungeon
AutoGPT
Contrastive Language-Image Pre-training
" Deep Learning "
LangChain
Microsoft Copilot
OpenAI Five
Transformer
Category
v
t
e
Autoencoder
Deep learning
Fine-tuning
Foundation model
Generative adversarial network
Generative pre-trained transformer
Large language model
Model Context Protocol
Neural network
Prompt engineering
Reinforcement learning from human feedback
Retrieval-augmented generation
Self-supervised learning
Stochastic parrot
Synthetic data

Top-p sampling

Transformer

Variational autoencoder

Vibe coding

Vision transformer

Waluigi effect

Word embedding

Character.ai

ChatGPT

DeepSeek

Ernie

Gemini

Grok

Copilot

Claude

Gemini

Gemma

GPT 1 2 3 J 4 4o 4.5 4.1 OSS 5

1

2

3

J

4

4o

4.5

4.1

OSS

5

Llama

o1

o3

o4-mini

Qwen

Base44

Claude Code

Cursor

Devstral

GitHub Copilot

Kimi-Dev
Qwen3-Coder
Replit
Xcode
Aurora
Firefly
Flux
GPT Image 1
Ideogram
Imagen
Midjourney
Qwen-Image
Recraft
Seedream
Stable Diffusion
Dream Machine
Hailuo AI
Kling
Midjourney Video
Runway Gen
Seedance
Sora
Veo
Wan
15.ai
Eleven
MiniMax Speech 2.5
WaveNet
Eleven Music
Endel
Lyria
Riffusion
Suno AI
Udio
Agentforce
AutoGLM
AutoGPT
ChatGPT Agent

Devin AI
Manus
OpenAI Codex
Operator
Replit Agent
01.AI
Aleph Alpha
Anthropic
Baichuan
Canva
Cognition AI
Cohere
Contextual AI
DeepSeek
ElevenLabs
Google DeepMind
HeyGen
Hugging Face
Inflection AI
Krikey AI
Kuaishou
Luma Labs
Meta AI
MiniMax
Mistral AI
Moonshot AI
OpenAI
Perplexity AI
Runway
Safe Superintelligence
Salesforce
Scale AI
SoundHound
Stability AI
Synthesia
Thinking Machines Lab
Upstage
xAI

Z.ai

Category

v

t

e

History timeline

timeline

Companies

Projects

Parameter Hyperparameter

Hyperparameter

Loss functions

Regression Bias–variance tradeoff Double descent Overfitting

Bias–variance tradeoff

Double descent

Overfitting

Clustering

Gradient descent SGD Quasi-Newton method Conjugate gradient method

SGD

Quasi-Newton method

Conjugate gradient method

Backpropagation

Attention

Convolution

Normalization Batchnorm

Batchnorm

Activation Softmax Sigmoid Rectifier

Softmax

Sigmoid

Rectifier

Gating

Weight initialization

Regularization

Datasets Augmentation

Augmentation

Prompt engineering

Reinforcement learning Q-learning SARSA Imitation Policy gradient

Q-learning

SARSA
Imitation
Policy gradient
Diffusion
Latent diffusion model
Autoregression
Adversary
RAG
Uncanny valley
RLHF
Self-supervised learning
Reflection
Recursive self-improvement
Hallucination
Word embedding
Vibe coding
Machine learning In-context learning
In-context learning
Artificial neural network Deep learning
Deep learning
Language model Large language model NMT
Large language model
NMT
Reasoning language model
Model Context Protocol
Intelligent agent
Artificial human companion
Humanity's Last Exam
Artificial general intelligence (AGI)
AlexNet
WaveNet
Human image synthesis
HWR
OCR
Computer vision
Speech synthesis 15.ai ElevenLabs
15.ai
ElevenLabs

Speech recognition Whisper

Whisper

Facial recognition

AlphaFold

Text-to-image models Aurora DALL-E Firefly Flux Ideogram Imagen Midjourney Recraft Stable Diffusion

Aurora

DALL-E

Firefly

Flux

Ideogram

Imagen

Midjourney

Recraft

Stable Diffusion

Text-to-video models Dream Machine Runway Gen Hailuo AI Kling Sora Veo

Dream Machine

Runway Gen

Hailuo AI

Kling

Sora

Veo

Music generation Riffusion Suno AI Udio

Riffusion

Suno AI

Udio

Word2vec

Seq2seq

GloVe

BERT

T5

Llama

Chinchilla AI

PaLM

GPT 1 2 3 J ChatGPT 4 4o o1 o3 4.5 4.1 o4-mini 5

1

2

3

J

ChatGPT

4

4o

o1

o3

4.5

4.1

o4-mini

5

Claude

Gemini Gemini (language model) Gemma

Gemini (language model)

Gemma

Grok

LaMDA

BLOOM

DBRX

Project Debater

IBM Watson

IBM Watsonx

Granite

PanGu- Σ

DeepSeek

Qwen

AlphaGo

AlphaZero

OpenAI Five

Self-driving car

MuZero

Action selection AutoGPT

AutoGPT

Robot control

Alan Turing

Warren Sturgis McCulloch

Walter Pitts

John von Neumann

Claude Shannon

Shun'ichi Amari
Kunihiko Fukushima
Takeo Kanade
Marvin Minsky
John McCarthy
Nathaniel Rochester
Allen Newell
Cliff Shaw
Herbert A. Simon
Oliver Selfridge
Frank Rosenblatt
Bernard Widrow
Joseph Weizenbaum
Seymour Papert
Seppo Linnainmaa
Paul Werbos
Geoffrey Hinton
John Hopfield
Jürgen Schmidhuber
Yann LeCun
Yoshua Bengio
Lotfi A. Zadeh
Stephen Grossberg
Alex Graves
James Goodnight
Andrew Ng
Fei-Fei Li
Alex Krizhevsky
Ilya Sutskever
Oriol Vinyals
Quoc V. Le
Ian Goodfellow
Demis Hassabis
David Silver
Andrej Karpathy
Ashish Vaswani
Noam Shazeer
Aidan Gomez

John Schulman

Mustafa Suleyman

Jan Leike

Daniel Kokotajlo

François Chollet

Neural Turing machine

Differentiable neural computer

Transformer Vision transformer (ViT)

Vision transformer (ViT)

Recurrent neural network (RNN)

Long short-term memory (LSTM)

Gated recurrent unit (GRU)

Echo state network

Multilayer perceptron (MLP)

Convolutional neural network (CNN)

Residual neural network (RNN)

Highway network

Mamba

Autoencoder

Variational autoencoder (VAE)

Generative adversarial network (GAN)

Graph neural network (GNN)

Category