

Existing OpenAI models and their best uses.

Many OpenAI models have been released over the past few years, and things have become even more confusing due to the names of all the OpenAI models.

Keeping track of all the new AI models being released these days is practically a full-time job. The latest model, GPT-5.4 , was released just days after GPT-5.3 (and it's practically not even considered an official release).

Many OpenAI models have been released over the past few years, and things have become even more confusing due to the names of all the OpenAI models.

So, let's simplify things. This is a comprehensive guide to all of OpenAI's major models. We'll look at what each model offers, its cost, and how to make the most of it in your work.

Overview of OpenAI models

Model	Best for	Input	Output	Context window	Price (input/output per million tokens)	Note
GPT-5.4	Advanced reasoning and logic	Text, Images	Document	1 million tokens	\$2.50 / \$15	Leading inference model; best for programming, logic, and scientific tasks; higher token cost for tasks with long contexts.
GPT-5.4 pro	Study	Text, Images	Document	1 million tokens	\$30 / \$180	Use more computing resources for problems that require deeper reasoning.
GPT-5.4 mini	Reasoning and logical reasoning	Text, Images	Document	400,000 tokens	\$0.75 / \$4.50	A faster and cheaper alternative to GPT-5.4; supports inference effort control.

Model	Best for	Input	Output	Context window	Price (input/output per million tokens)	Note
GPT 5.4 nano	The reasoning and logic are very easy to understand.	Text, Images	Document	400,000 tokens	\$0.20 / \$1.25	GPT-5.4 is the fastest and cheapest inference model; ideal for summarizing and classifying.
GPT-5.2	Superseded by GPT-5.4	Text, Images	Document	400,000 tokens	\$1.75 / \$14	You should update any tools you built using GPT-5.2 to GPT-5.4.
GPT-5.2 pro	It has been replaced by GPT-5.4 pro.	Text, Images	Document	400,000 tokens	\$21 / \$168	You should update all the tools you built using GPT-5.2 pro to GPT-5.4 pro.
GPT-5.1	Replaced by GPT-5.2	Text, Images	Document	400,000 tokens	\$1.25 / \$10	You should update all the tools you built using GPT-5.1 to GPT-5.4.
GPT-5	Replaced by GPT-5.2	Text, Images	Document	400,000 tokens	\$1.25 / \$10	You should update all the tools you've built using GPT-5 to GPT-5.4.
GPT-5 mini	It has been replaced by GPT-5.4 mini.	Text, Images	Document	400,000 tokens	\$0.25 / \$2	You should update all the tools you've built using GPT-5 mini to GPT-5.4 mini.
GPT-5 nano	Replaced with GPT-5.4 nano	Text, Images	Document	400,000 tokens	\$0.05 / \$0.40	You should update all the tools you've built using GPT-5 nano to GPT-5.4 nano.

Model	Best for	Input	Output	Context window	Price (input/output per million tokens)	Note
GPT-5.3 Codex	Advanced Programming	Text, Images	Document	400,000 tokens	\$1.25 / \$10	Designed for code generation
GPT-4.1	Complex tasks that don't require advanced thinking.	Text, Images	Document	1,047,576 tokens	\$2 / \$8	A versatile, powerful general model.
GPT-4.1 mini	Balancing power, performance, and affordability.	Text, Images	Document	1,047,576 tokens	\$0.40 / \$1.60	A great starting point for many purposes.
GPT-4.1 nano	Optimizing speed and price	Text, Images	Document	1,047,576 tokens	\$0.10 / \$1.40	The GPT-4.1 model is the fastest and cheapest.
gpt-audio-1.5	Sound	Text, Audio	Text, Audio	128,000 tokens	Text: \$2.50 / \$10; Audio: \$32 / \$64	This API has both audio input and output.
gpt-audio-mini	Sound on a limited budget	Text, Audio	Text, Audio	128,000 tokens	Text: \$0.60 / \$2.40; Audio: \$10 / \$20	Affordable audio model
gpt-oss-120b	Open weighting model	Document	Document	131,072 tokens	Depending on the supplier	117 billion parameters; 5.1 billion active machines; Apache 2.0; downloadable, customizable, capable of single H100

Model	Best for	Input	Output	Context window	Price (input/output per million tokens)	Note
gpt-oss-20b	Average open weight model	Document	Document	131,072 tokens	Depending on the supplier	21 billion parameters; 3.6 billion active machines; Apache 2.0; open model, deployable on laptops/local computers.
GPT Image 1.5	Create images	Text, Images	Text, Images	Not applicable	Text: \$5/10; Image: \$10/\$32	Inheriting from DALL·E; belonging to the high-end imaging segment.
GPT Image 1	Create images	Text, Images	Image	Not applicable	Text: \$5; Image: \$10 / \$40	Replaced by GPT Image 1
Sora 2	Create video	Text, Images, Videos	Video	Not applicable	\$0.10 – \$0.50/second	The Pro version produces high-quality video but requires more processing time.
Sora	Create video	Text, Images, Videos	Video	Not applicable	\$20/month (Plus); \$200/month (Pro)	Creating videos is not available through the API.

GPT-5.4

The best OpenAI model for advanced inference and logic.

GPT-5.4 is OpenAI's leading inference model and one of the most powerful AI models available. Released just months after GPT-5.2 (and only days after GPT-5.3), GPT-5.4 still represents a significant performance leap.

Inference models like GPT-5.4 are best suited for logic, engineering, and scientific tasks. If you need OpenAI's most powerful model for building automated tools, generating computer code, analyzing text and image documents, or solving complex puzzles, GPT-5.4 is the right choice. It's also OpenAI's first all-purpose model

with computer capabilities and has integrated the programming capabilities of the previously standalone Codex model.

Because inference models use additional resources to solve problems, they will be more expensive for tasks that do not require advanced reasoning.

1. GPT-5.4 input: Text, images
2. GPT-5.4 output: Text
3. GPT-5.4 Context Window: 1 Million Tokens
4. GPT-5.4 entry price: \$2.50/million tokens (272k contexts); \$5/million tokens
5. Output price of GPT-5.4: \$15/million tokens

GPT-5.4 pro

The best OpenAI model for research and advanced applications.

GPT-5.4 Pro is a version of OpenAI's flagship inference model, GPT-5.4. It is designed to be the most intelligent and reliable model for solving difficult problems where financial cost and time are not significant barriers.

Response times from the GPT-5.4 Pro can be longer than with other models, and the cost of using the GPT-5.4 Pro is 12 times higher than the GPT-5.4. Unless you need its most advanced performance, this is not the right model for you.

1. Input for GPT-5.4 pro: Text, images
2. Output of GPT-5.4 pro: Text
3. GPT-5.4 pro context window: 1 million tokens
4. The entry price for GPT-5.4 pro is \$30 per million tokens.
5. Output price of GPT-5.4 pro: \$180/million tokens

GPT-5.4 mini

The best OpenAI model for affordable inference and logic.

GPT-5.4 mini is a faster and more cost-effective alternative to GPT-5.4. For well-defined responses or chatbot responses, it offers an excellent balance between power and performance.

Similar to GPT-5.4, you must set the inference effort with GPT-5.4 mini. The lower the inference effort, the faster the speed and the lower the cost.

1. Input for GPT-5.4 mini: Text, images
2. Output of GPT-5.4 mini: Text
3. GPT-5.4 mini context window: 400,000 tokens
4. The entry price for GPT-5.4 mini is \$0.75 per million tokens.
5. Output price of GPT-5.4 mini: \$4.50/million tokens

GPT-5.4 nano

The best OpenAI model for inference and logic at very low cost.

GPT-5.4 nano is OpenAI's fastest and most cost-effective inference model. It excels in tasks such as summarization and classification. Like GPT-5.4 and GPT-5.4 mini, you have to set up an inference effort with GPT-5.4 nano.

1. GPT-5.4 nano input: Text, images
2. GPT-5.4 nano output: Text
3. GPT-5.4 nano context window: 400,000 tokens
4. GPT-5.4 nano input price: \$0.20/million tokens
5. GPT-5.4 nano output price: \$1.25/million tokens

GPT-5.2

Superseded by GPT 5.4

GPT-5.2 was a major leap forward for OpenAI when it was released in December 2025, but with the arrival of GPT-5.4 just over two months later, it was superseded. Currently, it is still available in the API as a slightly more affordable option.

1. GPT-5.2 input: Text, images
2. GPT-5.2 output: Text
3. GPT-5.2 Context Window: 400,000 Tokens
4. GPT-5.2 entry price: \$1.75/million tokens
5. GPT-5.2 output price: \$14/million tokens

GPT-5.2 pro

Replaced by GPT-5.4 pro

GPT-5.2 pro is a version of OpenAI's GPT-5.2. It has been superseded by GPT-5.4 pro. Currently, it is still available in the API as a slightly more cost-effective option.

1. Input for GPT-5.2 pro: Text, images
2. Output of GPT-5.2 pro: Text
3. GPT-5.2 pro context window: 400,000 tokens
4. The entry price for GPT-5.2 pro is \$21 USD per million tokens.
5. Output price of GPT-5.2 pro: \$168/million tokens

GPT-5.1 and GPT-5

Superseded by GPT-5.4

GPT-5.1 and GPT-5 were the previous leading models. Although they are still available through the API, you should update any tools you've built using them to GPT-5.4.

1. GPT-5.1/GPT-5 input: Text, images
2. GPT-5.1/GPT-5 Output: Text
3. GPT-5.1/GPT-5 Context Window: 400,000 tokens
4. GPT-5.1/GPT-5 entry price: \$1.25/million tokens
5. GPT-5.1/GPT-5 output price: \$10 USD/million tokens

GPT-5 mini

Replaced by GPT-5.4 mini

The GPT-5 mini was a faster and more economical alternative to the GPT-5. It has since been superseded by the GPT-5.4 mini.

1. Input for GPT-5 mini: Text, images
2. Output of GPT-5 mini: Text
3. GPT-5 mini context window: 400,000 tokens
4. The input cost of GPT-5 mini: \$0.25/million tokens
5. Output price of GPT-5 mini: \$2/million tokens

GPT-5 nano

Replaced by GPT-5.4 nano

GPT-5 nano là gì?i pháp thay th? nhanh và ti?t ki?m nh?t c?a OpenAI cho GPT-5. Nó ?ã ???c thay th? b?i GPT-5.4 nano.

1. ??u vào GPT-5 nano: V?n b?n, hình ?nh
2. ??u ra GPT-5 nano: V?n b?n
3. C?a s? ng? c?nh GPT-5 nano: 400.000 token
4. Giá ??u vào GPT-5 nano: 0,05 USD/tri?u token
5. Giá ??u ra GPT-5 nano: 0,40 USD/tri?u token

GPT-5.3 Codex

?ã ???c thay th? b?i GPT-5.4

GPT-5.3 Codex là phiên b?n c?a GPT-5 ???c thi?t k? cho các tác v? l?p trình t? ??ng trong nh?ng công c? nh? Codex, Cursor và Replit. Nó phù h?p nh?t cho các tác v? l?p trình ph?c t?p và kéo dài, ch?ng h?n nh? t?o nh?ng ch?c n?ng và tính n?ng m?i, th?c hi?n các thay ??i l?n v? c? s? code, nh?ng hi?n t?i, OpenAI khuy?n ngh? s? d?ng GPT-5.4 cho t?t c? nh?ng vi?c ?ó.

??u t??ng t? c?ng ?ung v?i các phiên b?n Codex khác, bao g?m GPT-5.2 Codex, GPT-5.1 Codex, GPT-5.1 Codex max, GPT-5.1 Codex mini và GPT-5 Codex.

1. ??u vào c?a GPT-5.3 Codex: V?n b?n, hình ?nh
2. ??u ra c?a GPT-5.3 Codex: V?n b?n
3. C?a s? ng? c?nh c?a GPT-5.3 Codex: 400.000 token
4. Giá ??u vào c?a GPT-5.3 Codex: 1,75 USD/tri?u token
5. Giá ??u ra c?a GPT-5.3 Codex: 14 USD/tri?u token

GPT-4.1

Mô hình OpenAI t?t nh?t cho các tác v? ph?c t?p không yêu c?u suy lu?n nâng cao

GPT-4.1 là mô hình LLM ?a ph??ng th?c không suy lu?n thông minh nh?t c?a OpenAI

T?t nh?t nên coi GPT-4.1 nh? con dao ?a n?ng t?i ?u c?a các mô hình OpenAI. M?c dù có r?t ít vi?c nó không th? làm ???c, nh?ng có m?t s? tr??ng h?p mô hình khác có th? phù h?p h?n. Ví d?, m?t mô hình suy lu?n chuyên d?ng s? ho?t ??ng t?t h?n trên các tác v? nhi?u b??c nh? t?o code, trong khi m?t mô hình nh?, t?i ?u hóa chi phí s? mang l?i giá tr? t?t h?n cho vi?c t?o v?n b?n c? b?n.

1. ??u vào c?a GPT-4.1: V?n b?n, hình ?nh
2. ??u ra c?a GPT-4.1: V?n b?n
3. C?a s? ng? c?nh c?a GPT-4.1: 1.047.576 token
4. Giá ??u vào c?a GPT-4.1: 2 USD/tri?u token
5. Giá ??u ra c?a GPT-4.1: 8 USD/tri?u token

GPT-4.1 mini

Mô hình OpenAI t?t nh?t ?? cân b?ng s?c m?nh, hi?u su?t và giá c? ph?i ch?ng

GPT-4.1 mini là m?t mô hình AI ?a n?ng nhanh và giá c? ph?i ch?ng. Trong các bài ki?m tra hi?u n?ng, nó ho?t ??ng g?n nh? t?t b?ng GPT-4.1 nh?ng v?i giá ch? b?ng 1/5. N?u b?n không ch?c ch?n nên s? d?ng mô hình nào, GPT-4.1 mini có l? là ?i?m kh?i ??u h?p lý. Tr? khi b?n bi?t mình c?n s?c m?nh b? sung c?a GPT-4.1 ho?c quan tâm ??n vi?c t?i ?u hóa chi phí, thì ?ây là m?t l?a ch?n r?t m?nh m?.

1. ??u vào GPT-4.1 mini: V?n b?n, hình ?nh
2. ??u ra GPT-4.1 mini: V?n b?n
3. C?a s? ng? c?nh GPT-4.1 mini: 1.047.576 token
4. Giá ??u vào GPT-4.1 mini: 0,40 USD/tri?u token
5. Giá ??u ra GPT-4.1 mini: 1,60 USD/tri?u token

GPT-4.1 nano

Mô hình OpenAI t?t nh?t v? t?c ?? và giá c?

GPT-4.1 nano là mô hình GPT-4.1 nh? nh?t hi?n có. Do ?ó, nó là mô hình nhanh nh?t và r? nh?t - m?c dù nó không m?nh b?ng GPT-4.1 ho?c GPT-4.1 mini. ?ây là mô hình nên s? d?ng cho các tác v? ??n gi?n, ho?c khi b?n

c?n ??u tiên t?c ?? và hi?u qu? chi phí.

1. ??u vào c?a GPT-4.1 nano: V?n b?n, hình ?nh
2. ??u ra c?a GPT-4.1 nano: V?n b?n
3. C?a s? ng? c?nh c?a GPT-4.1 nano: 1.047.576 token
4. Giá ??u vào c?a GPT-4.1 nano: 0,10 USD/tri?u token
5. Giá ??u ra c?a GPT-4.1 nano: 1,40 USD/tri?u token

gpt-audio-1.5

Mô hình OpenAI t?t nh?t cho âm thanh

Là mô hình âm thanh ??u tiên ???c OpenAI cung c?p r?ng rãi, nó thay th? GPT-Audio và mô hình GPT-4.1 Audio c?. Nó có th? chuy?n ??i âm thanh thành v?n b?n ho?c tr? l?i b?ng gi?ng nói.

1. ??u vào c?a gpt-audio-1.5: V?n b?n, âm thanh
2. ??u ra c?a gpt-audio-1.5: V?n b?n, âm thanh
3. C?a s? ng? c?nh c?a gpt-audio-1.5: 128.000 token
4. Giá ??u vào c?a gpt-audio-1.5: 2,50 USD/tri?u token cho v?n b?n; 32 USD/tri?u token cho âm thanh
5. Giá ??u ra c?a gpt-audio-1.5: 10 USD/tri?u token cho v?n b?n; 64 USD/tri?u token cho âm thanh

gpt-audio-mini

Mô hình OpenAI t?t nh?t cho âm thanh v?i ngân sách h?n ch?

GPT-Audio mini cung c?p ??u vào và ??u ra âm thanh v?i chi phí ch? b?ng kho?ng 1/3 so v?i GPT Audio 1.5. ??i v?i các nhà phát tri?n c?n m?t mô hình âm thanh nh?ng b? h?n ch? v? ngân sách, ?ây là l?a ch?n t?t nh?t.

1. ??u vào c?a gpt-audio-mini: V?n b?n, âm thanh
2. ??u ra c?a gpt-audio-mini: V?n b?n, âm thanh
3. C?a s? ng? c?nh c?a gpt-audio-mini: 128.000 token
4. Giá ??u vào c?a gpt-audio-mini: 0,60 USD/tri?u token cho v?n b?n; 10 USD/tri?u token cho âm thanh
5. Giá ??u ra c?a gpt-audio-mini: 2,40 USD/tri?u token cho v?n b?n; 20 USD/tri?u token cho âm thanh

gpt-oss-120b

Mô hình OpenAI tr?ng s? m? t?t nh?t

gpt-oss-120b là mô hình tr?ng s? m? m?nh nh?t c?a OpenAI, và là mô hình m? m?nh nh?t có th? ch?y trên m?t GPU H100 duy nh?t. Nó ???c phát hành theo gi?y phép Apache 2.0, vì v?y b?n có th? t?i xu?ng và ch?y nó trên máy ch? c?a riêng mình ho?c ch?n t? b?t k? nhà cung c?p suy lu?n bên th? ba nào. B?n c?ng có th? tinh ch?nh mô hình ?? phù h?p v?i nhu c?u c?a mình. ?i?u này r?t quan tr?ng ??i v?i OpenAI, v?n tr??c ?ây ch? phát hành các mô hình ???c quy?n.

Mô hình gpt-oss-120b có 117 t? tham s?. Nó s? d?ng ki?n ??trúc h?n h?p các chuyên gia, do ?ó ch? có 5,1 t? tham s? ho?t ??ng cùng m?t lúc.

1. ??u vào c?a gpt-oss-120b: V?n b?n
2. ??u ra c?a gpt-oss-120b: V?n b?n
3. C?a s? ng? c?nh c?a gpt-oss-120b: 131.072 token
4. Giá ??u vào c?a gpt-oss-120b: Thay ??i tùy theo nhà cung c?p
5. Giá ??u ra c?a gpt-oss-120b: Thay ??i tùy theo nhà cung c?p

gpt-oss-20b

Mô hình OpenAI có tr?ng s? m? trung bình t?t nh?t

gpt-oss-20b là mô hình có tr?ng s? m? trung bình c?a OpenAI. Nó ???c thi?t k? ?? ch?y c?c b? và th?m chí có th? ? ch?y trên m?t laptop ?? m?nh. Gi?ng nh? gpt-oss-120b, nó ???c phát hành theo gi?y phép Apache 2.0, vì v?y b?n có th? t? t?i xu?ng ho?c ch?n t? b?t k? nhà cung c?p suy lu?n bên th? ba nào, và b?n có th? tinh ch?nh mô hình cho phù h?p v?i nhu c?u c?a mình.

gpt-oss-20b có 21 t? tham s?. Nó s? d?ng ki?n ??trúc h?n h?p các chuyên gia nên ch? có 3,6 t? tham s? ho?t ??ng cùng m?t lúc.

1. ??u vào gpt-oss-120b: V?n b?n
2. ??u ra gpt-oss-120b: V?n b?n
3. C?a s? ng? c?nh gpt-oss-120b: 131.072 token
4. Giá ??u vào gpt-oss-120b: Thay ??i tùy theo nhà cung c?p
5. Giá ??u ra gpt-oss-120b: Thay ??i tùy theo nhà cung c?p

GPT Image 1.5

Mô hình OpenAI t?t nh?t ?? t?o hình ?nh

GPT Image 1.5 là phiên b?n k? nhi?m c?a mô hình hình ?nh DALL-E c?a OpenAI, và là m?t trong nh?ng mô hình chuy?n ??i v?n b?n thành hình ?nh t?t nh?t hi?n có. Nó thay th? GPT Image 1.

Giá c? c?a OpenAI cho các mô hình hình ?nh khá ph?c t?p vì hình ?nh ???c chuy?n ??i thành token d?a trên kích th??c và ch?t l??ng c?a chúng. Tuy nhiên, nó không quá ??t. Ví d?, ch? t?n 0,10 USD ?? t?o ra nhi?u hình ?nh ch?t l??ng cao 1024 x 1024 px.

1. ??u vào c?a GPT Image 1.5: V?n b?n, hình ?nh
2. ??u ra c?a GPT Image 1.5: Hình ?nh
3. Giá ??u vào c?a GPT Image 1.5: 5\$/tri?u token v?n b?n; 1\$/tri?u token hình ?nh
4. Giá ??u ra c?a GPT Image 1.5: 10\$/tri?u token v?n b?n; 32\$/tri?u token hình ?nh

GPT Image 1

?ã ???c thay th? b?i GPT Image 1.5

GPT Image 1 là phiên bản kế nhiệm của mô hình xử lý hình ảnh DALL-E của OpenAI. Hiện tại, nó đã được thay thế bởi GPT Image 1.5, mặc dù vẫn có thể truy cập thông qua API.

1. Đầu vào của GPT Image 1: Văn bản, hình ảnh
2. Đầu ra của GPT Image 1: Hình ảnh
3. Giá đầu vào của GPT Image 1: 5\$/triệu token văn bản; 10\$/triệu token hình ảnh
4. Giá đầu ra của GPT Image 1: 40\$/triệu token hình ảnh

Sora 2

Mô hình OpenAI tạo nội dung tạo video

Sora 2 là mô hình tạo video hàng đầu của OpenAI. Nó có thể nhận kết hợp các prompt văn bản, hình ảnh tĩnh và video, tạo ra một đoạn video và âm thanh dựa trên tất cả những thứ đó. Sora 2 và Sora 2 Pro có sẵn thông qua API. Sora 2 Pro tạo ra video chất lượng và độ phân giải cao hơn những mô hình khác hiện tại.

1. Đầu vào của Sora 2: Văn bản, hình ảnh, video
2. Đầu ra của Sora 2: Video
3. Giá Sora 2: 0,10 USD/giây cho video 720p; Phiên bản Pro: 0,30 USD/giây cho video 720p và 0,50 USD/giây cho video Full HD

Sora 1

Mô hình OpenAI tạo nội dung tạo video thông qua ứng dụng

Sora 1 cung cấp sức mạnh cho ứng dụng web và ứng dụng di động Sora của OpenAI. Nó không có sẵn dưới dạng API.

Người dùng miễn phí có thể tạo 6 video 10 giây có watermark độ phân giải 720p mỗi ngày, trong khi người dùng ký ChatGPT Plus có giới hạn sử dụng cao hơn. Người dùng ký ChatGPT Pro có thể tạo video 20 giây không có watermark độ phân giải 1080p.

1. Đầu vào của Sora: Văn bản, hình ảnh, video
2. Đầu ra của Sora: Video
3. Giá Sora: Có sẵn miễn phí, như một phần của ChatGPT Plus (20 USD/tháng) và ChatGPT Pro (200 USD/tháng)

Cách chọn mô hình OpenAI phù hợp

OpenAI có rất nhiều mô hình, nhưng mỗi mô hình đều có mục đích riêng. GPT-5.4 và GPT-4.1 là những mô hình hàng đầu về suy luận và phi suy luận, với những phiên bản mini và nano khác nhau cung cấp các cân bằng khác nhau giữa giá cả, hiệu năng và tốc độ. Chắc chắn có rất nhiều sự chọn lựa khác nhau giữa các mô hình khác nhau, đặc biệt khi bạn xem xét những mô hình cũ hơn của chúng tôi thay thế. Nhìn chung, một mô hình sẽ phù hợp với bạn nhất về các cân bằng giữa sức mạnh, giá cả và khả năng phát triển liên tục.

OpenAI liên tục phát hành các mô hình mới, thay đổi giá cả và nhiều thứ khác nữa.

You finished reading the article "**Existing OpenAI models and their best uses.**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.
