

OpenAI Platform

Pricing

Latest models

New: Save on synchronous requests with [flex processing](#).

Text tokens

Price per 1M tokens · [Batch API price](#) ☐

Model	Input	Cached input	Output
gpt-4.1 ↳ gpt-4.1-2025-04-14	\$2.00	\$0.50	\$8.00
gpt-4.1-mini ↳ gpt-4.1-mini-2025-04-14	\$0.40	\$0.10	\$1.60
gpt-4.1-nano ↳ gpt-4.1-nano-2025-04-14	\$0.10	\$0.025	\$0.40
gpt-4.5-preview ↳ gpt-4.5-preview-2025-02-27	\$75.00	\$37.50	\$150.00
gpt-4o ↳ gpt-4o-2024-08-06	\$2.50	\$1.25	\$10.00
gpt-4o-audio-preview ↳ gpt-4o-audio-preview-2024-12-17	\$2.50	-	\$10.00
gpt-4o-realtime-preview ↳ gpt-4o-realtime-preview-2024-12-17	\$5.00	\$2.50	\$20.00
gpt-4o-mini ↳ gpt-4o-mini-2024-07-18	\$0.15	\$0.075	\$0.60
gpt-4o-mini-audio-preview ↳ gpt-4o-mini-audio-preview-2024-12-17	\$0.15	-	\$0.60
gpt-4o-mini-realtime-preview ↳ gpt-4o-mini-realtime-preview-2024-12-17	\$0.60	\$0.30	\$2.40
o1 ↳ o1-2024-12-17	\$15.00	\$7.50	\$60.00
o1-pro ↳ o1-pro-2025-03-19	\$150.00	-	\$600.00
o3-pro ↳ o3-pro-2025-06-10	\$20.00	-	\$80.00

Model	Input	Cached input	Output
o3 ↳ o3-2025-04-16	\$2.00	\$0.50	\$8.00
o4-mini ↳ o4-mini-2025-04-16	\$1.10	\$0.275	\$4.40
o3-mini ↳ o3-mini-2025-01-31	\$1.10	\$0.55	\$4.40
o1-mini ↳ o1-mini-2024-09-12	\$1.10	\$0.55	\$4.40
codex-mini-latest ↳ codex-mini-latest	\$1.50	\$0.375	\$6.00
gpt-4o-mini-search-preview ↳ gpt-4o-mini-search-preview-2025-03-11	\$0.15	-	\$0.60
gpt-4o-search-preview ↳ gpt-4o-search-preview-2025-03-11	\$2.50	-	\$10.00
computer-use-preview ↳ computer-use-preview-2025-03-11	\$3.00	-	\$12.00
gpt-image-1 ↳ gpt-image-1	\$5.00	\$1.25	-

[▼ All snapshots](#)

Text tokens (Flex Processing)

Price per 1M tokens

Model	Input	Cached input	Output
o3 ↳ o3-2025-04-16	\$1.00	\$0.25	\$4.00
o4-mini ↳ o4-mini-2025-04-16	\$0.55	\$0.138	\$2.20

[▼ All snapshots](#)

Audio tokens

Price per 1M tokens

Model	Input	Cached input	Output
gpt-4o-audio-preview ↳ gpt-4o-audio-preview-2024-12-17	\$40.00	-	\$80.00

Model	Input	Cached input	Output
gpt-4o-mini-audio-preview ↳ gpt-4o-mini-audio-preview-2024-12-17	\$10.00	-	\$20.00
gpt-4o-realtime-preview ↳ gpt-4o-realtime-preview-2024-12-17	\$40.00	\$2.50	\$80.00
gpt-4o-mini-realtime-preview ↳ gpt-4o-mini-realtime-preview-2024-12-17	\$10.00	\$0.30	\$20.00

▼ All snapshots

Image tokens

Price per 1M tokens

Model	Input	Cached input	Output
gpt-image-1 ↳ gpt-image-1	\$10.00	\$2.50	\$40.00

Fine-tuning

Price per 1M tokens · Batch API price ☐

Model	Training	Input	Cached input	Output
o4-mini-2025-04-16	\$100.00 / hour	\$4.00	\$1.00	\$16.00
o4-mini-2025-04-16 with data sharing	\$100.00 / hour	\$2.00	\$0.50	\$8.00
gpt-4.1-2025-04-14	\$25.00	\$3.00	\$0.75	\$12.00
gpt-4.1-mini-2025-04-14	\$5.00	\$0.80	\$0.20	\$3.20
gpt-4.1-nano-2025-04-14	\$1.50	\$0.20	\$0.05	\$0.80
gpt-4o-2024-08-06	\$25.00	\$3.75	\$1.875	\$15.00
gpt-4o-mini-2024-07-18	\$3.00	\$0.30	\$0.15	\$1.20
gpt-3.5-turbo	\$8.00	\$3.00	-	\$6.00
davinci-002	\$6.00	\$12.00	-	\$12.00
babbage-002	\$0.40	\$1.60	-	\$1.60

Tokens used for model grading in reinforcement fine-tuning are billed at that model's per-token rate. Inference discounts are available if you enable data sharing when creating the fine-tune job. [Learn more](#).

Built-in tools

The tokens used for built-in tools are billed at the chosen model's per-token rates.
GB refers to binary gigabytes of storage (also known as gibibyte), where 1GB is 2^30 bytes.

Tool	
Code Interpreter	\$0.00
File Search Storage	\$0.10 GB/day
File Search Tool Call (Responses API only*)	\$2.50 1k calls (*Does not apply on Assistants)
Web Search	Web search tool pricing is inclusive of tokens used to synthesize from the web. Pricing depends on model and search context size

Web search

Web search is a built-in tool with pricing that depends on both the model used and the search context size. The billing dashboard will report these line items as 'web search tool calls | gpt-4o' and 'web search tool calls | gpt-4o-mini'.

Model	Search context size	Cost
gpt-4.1, gpt-4o, or gpt-4o-search-preview	low	\$30.00 1k calls
	medium (default)	\$35.00 1k calls
	high	\$50.00 1k calls
gpt-4.1-mini, gpt-4o-mini, or gpt-4o-mini-search-preview	low	\$25.00 1k calls
	medium (default)	\$27.50 1k calls
	high	\$30.00 1k calls

Transcription and speech generation

Text tokensPrice per 1M tokens

Model	Input	Output	Estimated cost
gpt-4o-mini-tts	\$0.60	-	\$0.015 minute
gpt-4o-transcribe	\$2.50	\$10.00	\$0.006 minute
gpt-4o-mini-transcribe	\$1.25	\$5.00	\$0.003 minute

Audio tokensPrice per 1M tokens

Model	Input	Output	Estimated cost
gpt-4o-mini-tts	-	\$12.00	\$0.015 minute
gpt-4o-transcribe	\$6.00	-	\$0.006 minute

Model	Input	Output	Estimated cost
gpt-4o-mini-transcribe	\$3.00	-	\$0.003 minute

Other models

Model	Use case	Cost
Whisper	Transcription	\$0.006 minute
TTS	Speech generation	\$15.00 1M characters
TTS HD	Speech generation	\$30.00 1M characters

Image generation

Price per image

Please note that this pricing for GPT Image 1 does not include text and image tokens used in the image generation process, and only reflects the output image tokens cost. For input text and image tokens, refer to the corresponding sections above. There are no additional costs for DALL·E 2 or DALL·E 3.

Model	Quality	1024x1024	1024x1536	1536x1024
GPT Image 1	Low	\$0.011	\$0.016	\$0.016
	Medium	\$0.042	\$0.063	\$0.063
	High	\$0.167	\$0.25	\$0.25
Model	Quality	1024x1024	1024x1792	1792x1024
DALL·E 3	Standard	\$0.04	\$0.08	\$0.08
	HD	\$0.08	\$0.12	\$0.12
Model	Quality	256x256	512x512	1024x1024
DALL·E 2	Standard	\$0.016	\$0.018	\$0.02

Embeddings

Price per 1M tokens · Batch API price ☐

Model	Cost
text-embedding-3-small	\$0.02
text-embedding-3-large	\$0.13
text-embedding-ada-002	\$0.10

Moderation

Model	Cost
omni-moderation-latest ↳ omni-moderation-2024-09-26	Free
text-moderation-latest ↳ text-moderation-007	Free

▼ All snapshots

Other models

Price per 1M tokens · Batch API price ☐

Model	Input	Output
chatgpt-4o-latest	\$5.00	\$15.00
gpt-4-turbo ↳ gpt-4-turbo-2024-04-09	\$10.00	\$30.00
gpt-4 ↳ gpt-4-0613	\$30.00	\$60.00
gpt-4-32k	\$60.00	\$120.00
gpt-3.5-turbo ↳ gpt-3.5-turbo-0125	\$0.50	\$1.50
gpt-3.5-turbo-instruct	\$1.50	\$2.00
gpt-3.5-turbo-16k-0613	\$3.00	\$4.00
davinci-002	\$2.00	\$2.00
babbage-002	\$0.40	\$0.40

▼ All snapshots