

Pricing

Latest models

New: Save on synchronous requests with [flex processing](#). Available on select [models](#)—see model detail pages for more.

Text tokens

Price per 1M tokens · [Batch](#) / [Flex 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