

Models

[Compare models](#)

Explore all available models and compare their capabilities.

Featured models

GPT-5.1 New

The best model for coding and agentic tasks with configurable reasoning effort.

GPT-5 mini

A faster, cost-efficient version of GPT-5 for well-defined tasks

GPT-5 nano

Fastest, most cost-efficient version of GPT-5

Frontier models OpenAI's most advanced models, recommended for most tasks.



GPT-5.1

The best model for coding and agentic tasks with configurable reasoning effort.



GPT-5 mini

A faster, cost-efficient version of GPT-5 for well-defined tasks



GPT-5 nano

Fastest, most cost-efficient version of GPT-5



GPT-5 pro

Version of GPT-5 that produces smarter and more precise responses



GPT-5

Previous intelligent reasoning model for coding and agentic tasks with configurable reasoning effort.



GPT-4.1

Smartest non-reasoning model

Open-weight models Open-weight models under a permissive Apache 2.0 license.



gpt-oss-120b

Most powerful open-weight model, fits into an H100 GPU



gpt-oss-20b

Medium-sized open-weight model for low latency

Specialized models Purpose-built for specific tasks.



Sora 2

Flagship video generation with synced audio



Sora 2 Pro

Most advanced synced-audio video generation



o3-deep-research

Our most powerful deep research model



o4-mini-deep-research

Faster, more affordable deep research model



GPT Image 1



gpt-image-1-mini



State-of-the-art image generation model



DALL-E 3

Previous generation image generation model



GPT-4o Transcribe

Speech-to-text model powered by GPT-4o



A cost-efficient version of GPT Image 1



GPT-4o mini TTS

Text-to-speech model powered by GPT-4o mini



GPT-4o mini Transcribe

Speech-to-text model powered by GPT-4o mini

Realtime and audio models

Models for audio use cases and realtime inputs and outputs.



gpt-realtime

Model capable of realtime text and audio inputs and outputs



gpt-audio

For audio inputs and outputs with Chat Completions API



gpt-realtime-mini

A cost-efficient version of GPT Realtime



gpt-audio-mini

A cost-efficient version of GPT Audio

ChatGPT models

Models used in ChatGPT, not recommended for API use.



GPT-5 Chat

GPT-5 model used in ChatGPT



ChatGPT-4o

GPT-4o model used in ChatGPT

All models

Diverse models for a variety of tasks.



GPT-5.1

The best model for coding and agentic tasks with configurable reasoning effort.



GPT-5

Previous intelligent reasoning model for coding and agentic tasks with configurable reasoning effort.



GPT-5 mini

A faster, cost-efficient version of GPT-5 for well-defined tasks



GPT-5 nano

Fastest, most cost-efficient version of GPT-5



GPT-5.1 Codex

A version of GPT-5.1 optimized for agentic coding in Codex.



GPT-5-Codex

A version of GPT-5 optimized for agentic coding in Codex



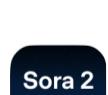
GPT-5 pro

Version of GPT-5 that produces smarter and more precise responses



Sora 2

Flagship video generation with synced audio



Sora 2 Pro

Most advanced synced-audio video generation



GPT Image 1

State-of-the-art image generation model



gpt-image-1-mini

A cost-efficient version of GPT Image 1



o3-deep-research

Our most powerful deep research model



o4-mini-deep-research

Faster, more affordable deep research model



o3-pro

Version of o3 with more compute for better responses



gpt-audio

For audio inputs and outputs with Chat Completions API



gpt-realtime

Model capable of realtime text and audio inputs and outputs



gpt-realtime-mini

A cost-efficient version of GPT Realtime



gpt-audio-mini

A cost-efficient version of GPT Audio

 o3 Reasoning model for complex tasks, succeeded by GPT-5	 o4-mini Fast, cost-efficient reasoning model, succeeded by GPT-5 mini
 GPT-4.1 Smartest non-reasoning model	 GPT-4.1 mini Smaller, faster version of GPT-4.1
 GPT-4.1 nano Fastest, most cost-efficient version of GPT-4.1	 o1-pro Version of o1 with more compute for better responses
 computer-use-preview Specialized model for computer use tool	 GPT-4o mini Search Preview Fast, affordable small model for web search
 GPT-4o Search Preview GPT model for web search in Chat Completions	 GPT-4.5 Preview (Deprecated) Deprecated large model.
 o3-mini A small model alternative to o3	 GPT-4o mini Audio Smaller model capable of audio inputs and outputs
 GPT-4o mini Realtime Smaller realtime model for text and audio inputs and outputs	 o1 Previous full o-series reasoning model
 omni-moderation Identify potentially harmful content in text and images	 o1-mini <small>Deprecated</small> A small model alternative to o1
 o1 Preview <small>Deprecated</small> Preview of our first o-series reasoning model	 GPT-4o Fast, intelligent, flexible GPT model
 GPT-4o Audio GPT-4o models capable of audio inputs and outputs	 GPT-4o mini Fast, affordable small model for focused tasks
 GPT-4o mini Audio Smaller model capable of audio inputs and outputs	 GPT-4o mini Realtime Smaller realtime model for text and audio inputs and outputs
 GPT-4o Realtime Model capable of realtime text and audio inputs and outputs	 GPT-4 Turbo An older high-intelligence GPT model
 babbage-002 <small>Deprecated</small> Replacement for the GPT-3 ada and babbage base models	 ChatGPT-4o GPT-4o model used in ChatGPT
 GPT-5.1 Codex mini Smaller, more cost-effective, less-capable version of GPT-5.1-Codex	 codex-mini-latest Fast reasoning model optimized for the Codex CLI
 DALL-E 2 Our first image generation model	 DALL-E 3 Previous generation image generation model
 davinci-002 <small>Deprecated</small> Replacement for the GPT-3 curie and davinci base models	 GPT-3.5 Turbo Legacy GPT model for cheaper chat and non-chat tasks
 GPT-4 An older high-intelligence GPT model	 GPT-4 Turbo Preview <small>Deprecated</small> An older fast GPT model
 GPT-4o mini Transcribe Speech-to-text model powered by GPT-4o mini	 GPT-4o mini TTS Text-to-speech model powered by GPT-4o mini
 GPT-4o Transcribe	 GPT-4o Transcribe Diarize



Speech-to-text model powered by GPT-4o



Transcription model that identifies who's speaking when



GPT-5.1 Chat

GPT-5.1 model used in ChatGPT.



GPT-5 Chat

GPT-5 model used in ChatGPT



GPT Image 1

State-of-the-art image generation model



gpt-oss-120b

Most powerful open-weight model, fits into an H100 GPU



gpt-oss-20b

Medium-sized open-weight model for low latency



text-embedding-3-large

Most capable embedding model



text-embedding-3-small

Small embedding model



text-embedding-ada-002

Older embedding model



text-moderation Deprecated

Previous generation text-only moderation model



text-moderation-stable Deprecated

Previous generation text-only moderation model



TTS-1

Text-to-speech model optimized for speed



TTS-1 HD

Text-to-speech model optimized for quality



Whisper

General-purpose speech recognition model

[How we use your data](#) · [Deprecated models](#)