OpenAI Platform

< Models

GPT-5.2-Codex

Default

Our most intelligent coding model optimized for long-horizon, agentic coding tasks.

Compare

Reasoning	
Speed	
Price	$1.75 · $14
Input	
Output	

GPT-5.2-Codex is an upgraded version of GPT-5.2 optimized for agentic coding tasks in Codex or similar environments. GPT-5.2-Codex supports `low`, `medium`, `high`, and `xhigh` reasoning effort settings. If you want to learn more about prompting GPT-5.2-Codex, refer to our dedicated guide.

- 400,000 context window
- 128,000 max output tokens
- Aug 31, 2025 knowledge cutoff
- Reasoning token support

Pricing

Pricing is based on the number of tokens used, or other metrics based on the model type. For tool-specific models, like search and computer use, there's a fee per tool call. See details in the pricing page.

Text tokens

Per 1M tokens

Input
$1.75

Cached input
$0.175

Output
$14.00

Quick comparison

	Input	Cached input	Output
GPT-5.2-Codex	$1.75		
GPT-5.1 Codex	$1.25		

Modalities

Text
Input and output

Image
Input only

Audio
Not supported

Video
Not supported

Endpoints

Chat Completions

v1/chat/completions

Responses
v1/responses

Realtime
v1/realtime

Assistants
v1/assistants

Batch
v1/batch

Fine-tuning
v1/fine-tuning

Embeddings
v1/embeddings

Image generation
v1/images/generations

Videos
v1/videos

Image edit
v1/images/edits

Speech generation
v1/audio/speech

Transcription
v1/audio/transcriptions

Translation
v1/audio/translations

Moderation
v1/moderations

Completions (legacy)
v1/completions

Features

Streaming
Supported

Function calling
Supported

Structured outputs
Supported

Fine-tuning
Not supported

Distillation
Not supported

Predicted outputs
Not supported

Snapshots

Snapshots let you lock in a specific version of the model so that performance and behavior remain consistent. Below is a list of all available snapshots and aliases for GPT-5.2-Codex.

gpt-5.2-codex
↳ gpt-5.2-codex

gpt-5.2-codex

Rate limits

Rate limits ensure fair and reliable access to the API by placing specific caps on requests or tokens used within a given time period. Your usage tier determines how high these limits are set and automatically increases as you send more requests and spend more on the API.

TIER	RPM	TPM	BATCH QUEUE LIMIT
Free		Not supported	
Tier 1	500	500,000	1,500,000
Tier 2	5,000	1,000,000	3,000,000
Tier 3	5,000	2,000,000	100,000,000
Tier 4	10,000	4,000,000	200,000,000
Tier 5	15,000	40,000,000	15,000,000,000