

## Changelog

### January, 2025

Jan 31

Feature

o3-mini

o3-mini-2025-01-31

v1/chat/completions

Launched [o3-mini](#), a new small reasoning model that is optimized for science, math, and coding tasks.

### December, 2024

Dec 18

Feature

Launched [Admin API Key Rotations](#), enabling customers to programmatically rotate their admin api keys.

Updated [Admin API Invites](#), enabling customers to programmatically invite users to projects at the same time they are invited to organizations.

Dec 17

Feature

o1

gpt-4o

gpt-4o-mini

v1/fine\_tuning

v1/chat/completions

v1/realtime

Added new models for [o1](#), [gpt-4o-realtime](#), [gpt-4o-audio](#) and [more](#).

Added WebRTC connection method for the [Realtime API](#).

Added [reasoning\\_effort](#) parameter for o1 models.

Added [developer](#) message role for o1 model. Note that o1-preview and o1-mini do not support system or developer messages.

Launched Preference Fine-tuning using [Direct Preference Optimization \(DPO\)](#).

Launched beta SDKs for Go and Java. [Learn more](#).

Added [Realtime API](#) support in the [Python SDK](#).

Dec 04

Feature

Launched [Usage API](#), enabling customers to programmatically query activities and spending across OpenAI APIs.

### November, 2024

Nov 20   **Update**   v1/chat/completions

Released [gpt-4o-2024-11-20](#), our newest model in the gpt-4o series.

Nov 04   **Feature**   v1/chat/completions

Released [Predicted Outputs](#), which greatly reduces latency for model responses where much of the response is known ahead of time. This is most common when regenerating the content of documents and code files with only minor changes.

## October, 2024

Oct 30   **Feature**   gpt-4o-realtime-preview   gpt-4o-audio-preview   v1/chat/completions

Added five new voice types in the [Realtime API](#) and [Chat Completions API](#).

Oct 17   **Feature**   gpt-4o-audio-preview   v1/chat/completions

Released [new](#) [gpt-4o-audio-preview](#) [model](#) for chat completions, which supports both audio inputs and outputs. Uses the same underlying model as the [Realtime API](#).

Oct 01   **Feature**   v1/realtime   v1/chat/completions   v1/fine\_tuning

Released several new features at [OpenAI DevDay in San Francisco](#):

[Realtime API](#): Build fast speech-to-speech experiences into your applications using a WebSockets interface.

[Model distillation](#): Platform for fine-tuning cost-efficient models with your outputs from a large frontier model.

[Image fine-tuning](#): Fine-tune GPT-4o with images and text to improve vision capabilities.

[Evals](#): Create and run custom evaluations to measure model performance on specific tasks.

[Prompt caching](#): Discounts and faster processing times on recently seen input tokens.

[Generate in playground](#): Easily generate prompts, function definitions, and structured output schemas in the playground using the Generate button.

## Invalid DateTime

Invalid DateTime   **Feature**   omni-moderation-latest   v1/moderations

Released [new](#) [omni-moderation-latest](#) [moderation model](#), which supports both images and text (for some categories), supports two new text-only harm

categories, and has more accurate scores.

## Invalid DateTime

Invalid DateTime

Feature o1-preview o1-mini v1/chat/completions

Released [o1-preview](#) and [o1-mini](#), new large language models trained with reinforcement learning to perform complex reasoning tasks.

## August, 2024

Aug 29

Feature v1/assistants

Assistants API now supports [including file search results used by the file search tool, and customizing ranking behavior](#).

Aug 20

Feature gpt-4o v1/fine\_tuning

GA release for [gpt-4o-2024-08-06](#) [fine-tuning](#)—all API users can now fine-tune the latest GPT-4o model.

Aug 15

Update gpt-4o v1/chat/completions

Released [dynamic model for](#) [chatgpt-4o-latest](#)—this model will point to the latest GPT-4o model used by ChatGPT.

Aug 06

Update

Launched [Structured Outputs](#)—model outputs now reliably adhere to developer supplied JSON Schemas.

Released [gpt-4o-2024-08-06](#), our newest model in the gpt-4o series.

Aug 01

Update

Launched [Admin and Audit Log APIs](#), allowing customers to programmatically administer their organization and monitor changes using the audit logs. Audit logging must be enabled within [settings](#).

## July, 2024

Jul 24

Update

Launched [self-serve SSO configuration](#), allowing Enterprise customers on custom and unlimited billing to set up authentication against their desired IDP.

Jul 23

Update

Launched [fine-tuning for GPT-4o mini](#), enabling even higher performance for specific use cases.

Jul 18

#### Update

Released [GPT-4o mini](#), our affordable and intelligent small model for fast, lightweight tasks.

Jul 17

#### Update

Released [Uploads](#) to upload large files in multiple parts.

## June, 2024

Jun 06

#### Update

[Parallel function calling](#) can be disabled in Chat Completions and the Assistants API by passing `parallel_tool_calls=false`.

[.NET SDK](#) launched in Beta.

Jun 03

#### Update

Added support for [file search customizations](#).

## May, 2024

May 15

#### Update

Added support for [archiving projects](#). Only organization owners can access this functionality.

Added support for [setting cost limits](#) on a per-project basis for pay as you go customers.

May 13

#### Update

Released [GPT-4o](#) in the API. GPT-4o is our fastest and most affordable flagship model.

May 09

#### Update

Added support for [image inputs to the Assistants API](#).

May 07

#### Update

Added support for [fine-tuned models to the Batch API](#).

May 06

#### Update

Added `stream_options: {"include_usage": true}` parameter to the Chat Completions and Completions APIs. Setting this gives developers access to usage stats when using streaming.

May 02

Update

Added a new endpoint to delete a message from a thread in the Assistants API.

## April, 2024

Apr 29

Update

Added a new function calling option `tool_choice: "required"` to the Chat Completions and Assistants APIs.

Added a guide for the Batch API and Batch API support for embeddings models

Apr 17

Update

Introduced a series of updates to the Assistants API, including a new file search tool allowing up to 10,000 files per assistant, new token controls, and support for tool choice.

Apr 16

Update

Introduced project based hierarchy for organizing work by projects, including the ability to create API keys and manage rate and cost limits on a per-project basis (cost limits available only for Enterprise customers).

Apr 15

Update

Released Batch API

Apr 09

Update

Released GPT-4 Turbo with Vision in general availability in the API

Apr 04

Update

Added support for seed in the fine-tuning API

Added support for checkpoints in the fine-tuning API

Added support for adding Messages when creating a Run in the Assistants API

Apr 01

Update

Added support for filtering Messages by run\_id in the Assistants API

## March, 2024

Mar 29 Update

Added support for [temperature](#) and [assistant message creation](#) in the Assistants API

Mar 14 Update

Added support for [streaming](#) in the Assistants API

## February, 2024

Feb 09 Update

Added [timestamp\\_granularities](#) [parameter](#) to the Audio API

Feb 01 Update

Released [gpt-3.5-turbo-0125](#), an updated GPT-3.5 Turbo model

## January, 2024

Jan 25 Update

Released [embedding V3 models](#) and an updated GPT-4 Turbo preview

Added [dimensions](#) [parameter](#) to the Embeddings API

## December, 2023

Dec 20 Update

Added [additional\\_instructions](#) [parameter](#) to run creation in the Assistants API

Dec 15 Update

Added [logprobs](#) and [top\\_logprobs](#) [parameters](#) to the Chat Completions API

Dec 14 Update

Changed [function parameters](#) argument on a tool call to be optional

## November, 2023

Nov 30 Update

Released [OpenAI Deno SDK](#)

Nov 06 Update

Released [GPT-4 Turbo Preview](#) , updated [GPT-3.5 Turbo](#), [GPT-4 Turbo with Vision](#), [Assistants API](#), [DALL·E 3 in the API](#), and [text-to-speech API](#)

Deprecated the Chat Completions `functions` parameter in favor of `tools`

Released [OpenAI Python SDK V1.0](#)

## October, 2023

### Oct 16 Update

Added `encoding_format` parameter to the Embeddings API

Added `max_tokens` to the [Moderation models](#)

### Oct 06 Update

Added [function calling support](#) to the Fine-tuning API