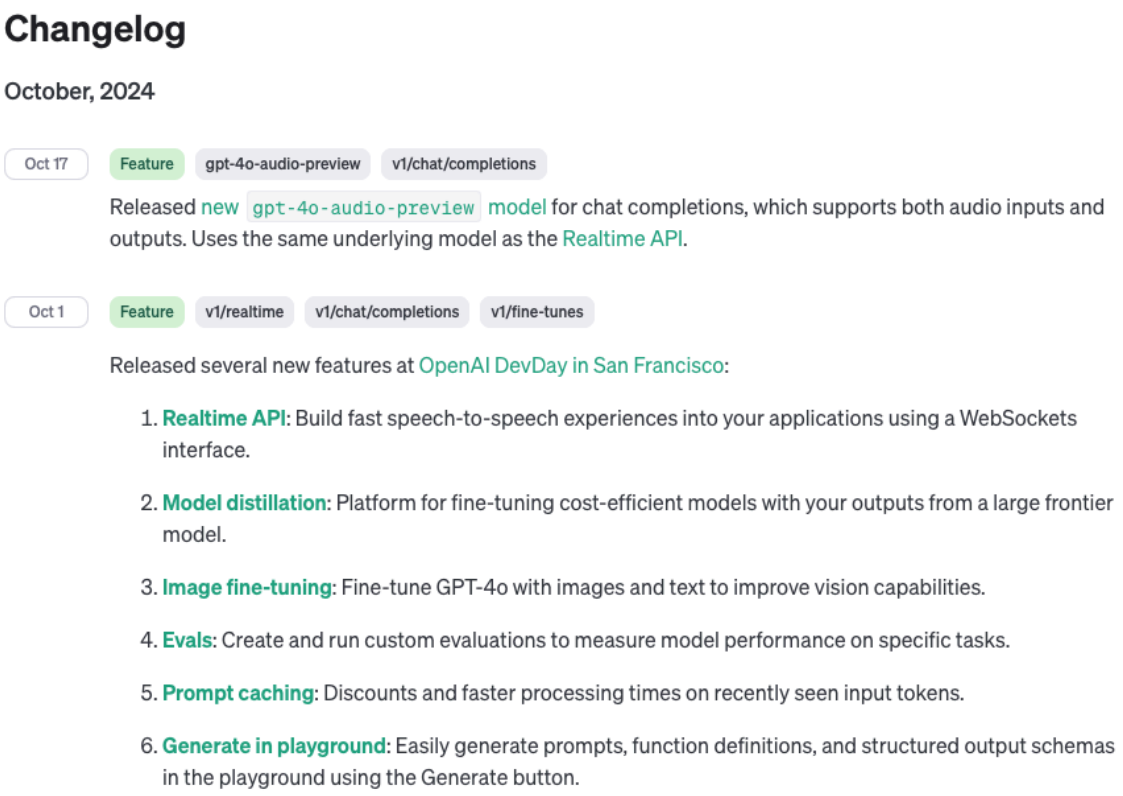
Exploring the new **gpt-4o-audio-preview** model for chat completions

---



Changelog

October, 2024

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 1 — Feature — v1/realtime — v1/chat/completions — v1/fine-tunes

Released several new features at OpenAI DevDay in San Francisco:

1. **Realtime API**: Build fast speech-to-speech experiences into your applications using a WebSockets interface.
2. **Model distillation**: Platform for fine-tuning cost-efficient models with your outputs from a large frontier model.
3. **Image fine-tuning**: Fine-tune GPT-4o with images and text to improve vision capabilities.
4. **Evals**: Create and run custom evaluations to measure model performance on specific tasks.
5. **Prompt caching**: Discounts and faster processing times on recently seen input tokens.
6. **Generate in playground**: Easily generate prompts, function definitions, and structured output schemas in the playground using the Generate button.

Link: https://platform.openai.com/docs/changelog

The audio model:

Audio generation

In addition to generating text and images, some models enable you to generate a spoken audio response to a prompt, and to use audio inputs to prompt the model. Audio inputs can contain richer data than text alone, allowing the model to detect tone, inflection, and other nuances within the input.

You can use these audio capabilities to:

- Generate a spoken audio summary of a body of text (text in, audio out)
- Perform sentiment analysis on a recording (audio in, text out)
- Async speech to speech interactions with a model (audio in, audio out)

Link: https://platform.openai.com/docs/guides/audio

What can you do with this audio completions call:

1) Generate a spoken audio summary of a body of text (text in, audio out)

- text in → text + audio out
- audio in → text + audio out
- text + audio in → text + audio out

2) Perform sentiment analysis on a recording (audio in, text out)

- audio in → text + audio out
- audio in → text out
- text + audio in → text + audio out
- text + audio in → text out

3) Async speech to speech interactions with a model (audio in, audio out)

- audio in → text + audio out
- text + audio in → text + audio out

And, you can have multi-turn conversations!

```
curl "https://api.openai.com/v1/chat/completions" \
    -H "Content-Type: application/json" \
    -H "Authorization: Bearer $OPENAI_API_KEY" \
    -d '{
        "model": "gpt-4o-audio-preview",
        "modalities": ["text", "audio"],
        "audio": { "voice": "alloy", "format": "wav" },
        "messages": [
            {
                "role": "user",
                "content": "Is a golden retriever a good family dog?"
            },
            {
                "role": "assistant",
                "audio": {
                    "id": "audio_abc123"
                }
            },
            {
                "role": "user",
                "content": "Why do you say they are loyal?"
            }
        ]
    }'
```

Links

Sample audio file “BAK.wav”
Link:
https://www.kaggle.com/datasets/crischir/sample-wav-audio-files?resource=download

Sample audio file “meeting.mp3”
Link:
https://www.kaggle.com/code/caesarlupum/speech-recognition-timealignedspectrogramsn

Work Kit – Gumroad
For access to this google sheet, all URLs, all code used, all audio samples, everything, I will link in my gumroad. It’s free or you can donate.

Fun SNL Skit Audio Generation on Reddit
Link:
https://www.reddit.com/r/OpenAI/comments/1g68p6u/gpt4oaudiopreview_generates_a_skit_with_sound/

Asks:

1. Subscribe to my channel
2. Fill out this survey: https://forms.gle/otAr1xUamgyYZE5y7