

---

## OpenAI launches DALL-E 3 API, new text-to-speech models

### Description

OpenAI launched a slew of new APIs during its first-ever developer day.

DALL-E 3, OpenAI's text-to-image model, is now available via an API after first coming to [ChatGPT](#) and Bing Chat. Similar to the previous version of DALL-E (e.g. [DALL-E 2](#)), the API incorporates built-in moderation to help protect against misuse, OpenAI says.

The DALL-E 3 API offers different format and quality options and resolutions ranging from 1024×1024 to 1792×1024, with prices starting at \$0.04 per generated image. But it's somewhat limited compared to the DALL-E 2 API — at least at present.

Unlike the DALL-E 2 API, the DALL-E 3 can't be used to create edited versions of images by having the model replace some areas of a pre-existing image or create variations of an existing image. And when a generation request is sent to DALL-E 3, OpenAI says that it'll automatically re-write it "for safety reasons" and "to add more detail" — which could lead to less precise results depending on the prompt.

Elsewhere, OpenAI's now providing a text-to-speech API, Audio API, that offers six preset voices — Alloy, Echo, Fable, Onyx, Nova and Shimer — to choose from and two generative AI model variants. It's live starting today, with pricing starting at \$0.015 per input 1,000 characters.

"This is much more natural than anything else we've heard out there, which can make apps more natural to interact with and more accessible," OpenAI Sam Altman said onstage. "It also unlocks a lot of use cases like language learning and voice assistance."

Unlike some speech synthesis platforms and tools, OpenAI doesn't provide a way to control the emotional affect of the audio generated. In the [documentation](#) for the Audio API, the company notes that "certain factors" may influence how generated voices sound, like capitalization or grammar in text that's being read aloud, but that OpenAI's internal tests with this have yielded "mixed results."

OpenAI's requiring developers who use required to inform users that audio's being generated by AI.

In a related announcement, OpenAI launched the next version of its open source automatic speech recognition model, [Whisper large-v3](#), which the company claims boasts improved performance across languages. It's on GitHub, available under a permissive license.

### Date

02/08/2025

### Date Created

13/11/2023

### Author

susantwain1