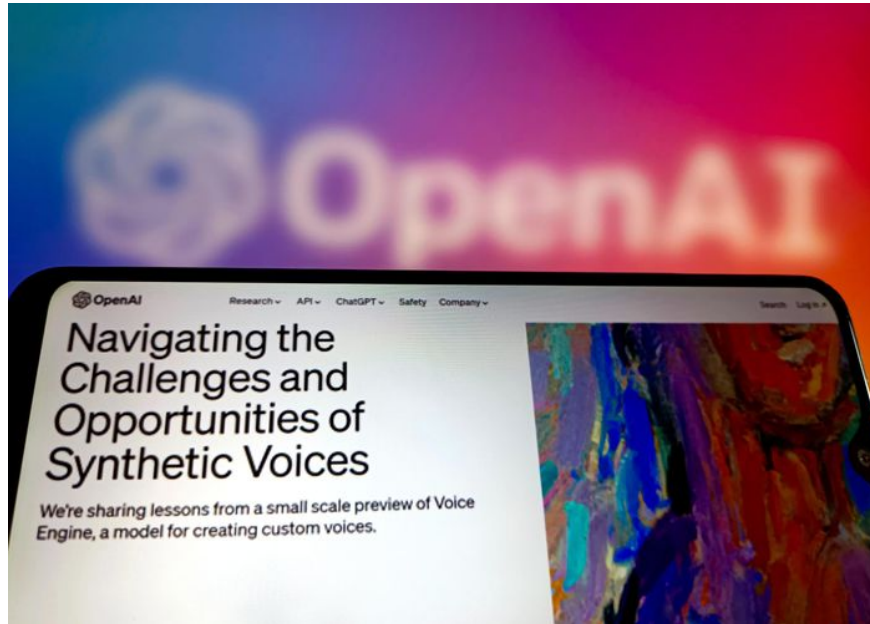


Behind OpenAI's voice imitation tool

OpenAI's Voice Engine tool can clone a voice from a 15-second sample file, which poses many risks if released widely.



As deepfake becomes more and more popular, OpenAI just introduced Voice Engine. Developed for 2 years, this tool can clone voices from a 15-second sample file provided by the user.

Similar to the Sora video creation model, Voice Engine has not been widely released. According to OpenAI, this gives the company time to analyze and prevent actions from taking advantage of the tool for malicious purposes.

"We want to make sure people are happy with how the tool is deployed. We understand the tool's potential for harm, and we take measures to mitigate that," said Jeff Harris, member of the product team. product at OpenAI, said.

How Voice Engine works

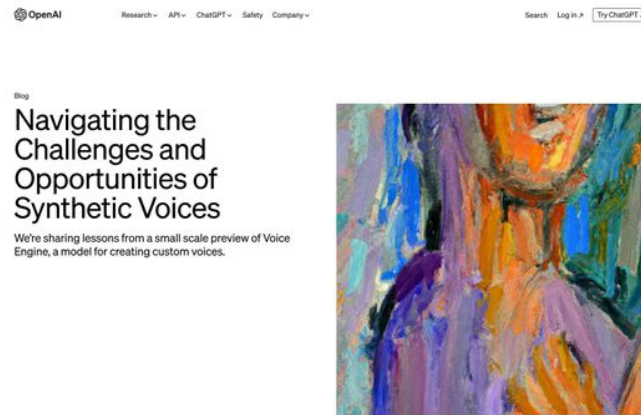
According to Harris, the generative model behind Voice Engine has been in use quietly for some time.

This is the model used for the "read aloud" feature in ChatGPT, as well as being used by Spotify since September 2023 to dub some podcasts in many languages.

OpenAI representatives said the model training data combines public and copyrighted voice recordings. According to TechCrunch, companies keep data origins confidential to avoid the possibility of being sued for

intellectual property violations. This is the situation many AI training companies are facing, including OpenAI.

Voice Engine's models are not fine-tuned, or trained based on the user. To recognize and generate speech, this tool combines a diffusion model with a transformer.



"We take a small audio sample, then create an actual voice that matches the original voice. The provided voice sample is discarded after the operation is complete," Harris explains.

According to OpenAI representatives, this model will analyze data and voice characteristics from the sample file and then combine the provided text to create a suitable voice.

This is not actually a new way of doing things. Companies that provide voice cloning services such as ElevenLabs, Replica Studios, Papercup, and even Big Tech groups such as Google and Microsoft also use this technique.

Voice Engine will not be free

According to the plan, OpenAI will charge Voice Engine fees. In one document, the tool costs \$15 per million characters, equivalent to about 162,500 words. The HD (high resolution) voice option is more than twice as expensive but the difference is unclear.

The above fee means the sound lasts about 18 hours, cheaper than competitors. For example, ElevenLabs charges \$11 for 100,000 characters per month.

Currently, Voice Engine does not have the ability to adjust voice tone, pitch or rhythm. Still, Harris said the nature of the sample's voice will be factored into the results. For example, if the original voice has an excited tone, the tool will "mimic" it in a similar manner.

The appearance of voice imitation tools has severely affected voice actors. Professional voice actors face the risk of having their voices cloned, while basic voice acting jobs face the risk of being replaced by AI.

Many AI voice cloning companies are trying to balance the benefits. Last year, Replica Studios signed an agreement with the Screen Actors Guild - American Federation of Television and Radio Artists (SAG-AFTRA) to create and license voice replicas of its members.

Meanwhile, ElevenLabs opens its voice marketplace, allowing users to create, verify, and publicly share their voices. When someone uses it, the voice owner will receive money for every 1,000 characters.

For Voice Engine, OpenAI will initially rely on "explicit permission" from the person whose voice is being cloned.

When used, the tool will "clearly reveal" AI-generated voices, and will not duplicate the voices of minors, deceased people or political figures.

Cannot be widely released yet

Not only threatening voice actors, voice cloning apps have been abused to defame or defraud.

On 4chan, many accounts use ElevenLabs to share hateful messages, imitating voices of famous people like actress Emma Watson.

Voice cloning tools are also a "hot" topic as America prepares for a presidential election. In January, a phone campaign used a fake image of President Joe Biden to discourage New Hampshire residents from voting.

For Voice Engine, Harris shares some abuse prevention policies. First, the tool is currently only available to a small group of developers, about 10 people, for testing.

OpenAI is prioritizing "low-risk" and "socially beneficial" use cases such as healthcare and serving people with disabilities.

Age of Learning, an educational technology company uses Voice Engine to create voiceovers from actors. In addition, the storytelling application HeyGen also uses tools for voice translation.

The voice created by Voice Engine will be "watermarked" using a technique developed by OpenAI, can be embedded in the recording file and cannot be heard.

"Given any audio, we can easily listen and determine whether it was created using our system or not.

To date, the source code of the tool is still closed. We are curious about making it public, but of course that comes with the risk of being exploited and sabotaged," Harris emphasized.

OpenAI plans to invite experts to the Red Teaming Network group to develop analysis strategies and reduce risks for the model.

Depending on test results and feedback from the public, OpenAI may release Voice Engine to more developers. However, at this time, the company cannot make any promises.

However, Harris also revealed the upcoming phase of Voice Engine. Specifically, OpenAI is testing a security mechanism that allows users to read random text to authenticate the "owner's voice", and clearly explain how their cloned voice is used.

"Our strategy for moving forward with real-life voice matching technology will depend on the experience from the trial, the safety issues that have not yet been discovered and the risk reduction measures we take.

We don't want people to confuse artificial voices with real human voices," Harris emphasized.

You finished reading the article "**Behind OpenAI's voice imitation tool**" edited by the [TipsMake](#) team. We hope this article has provided you with many useful tech tips and tricks. You can search for similar articles on tips and guides. Thank you for reading and for following us regularly.
