OpenAI deems its voice cloning tool too risky for general release
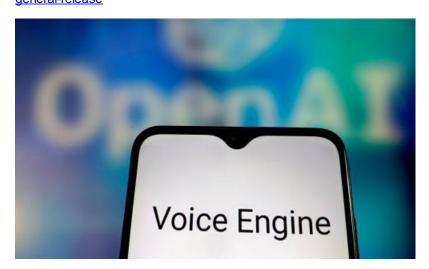
Publication Date: 2024-03-31

Author: Alex Hern

Section: Technology

Tags: OpenAI, ChatGPT, news

Article URL: https://www.theguardian.com/technology/2024/mar/31/openai-deems-its-voice-cloning-tool-too-risky-for-general-release

A new tool from OpenAI that can generate a convincing clone of anyone's voice using just 15 seconds of recorded audio has been deemed too risky for general release, as the AI lab seeks to minimise the threat of damaging misinformation in a global year of elections. Voice Engine was first developed in 2022 and an initial version was used for the text-to-speech feature built into ChatGPT, the organisation's leading AI tool. But its power has never been revealed publicly, in part because of the "cautious and informed" approach that OpenAI is taking to release it more widely. "We hope to start a dialogue on the responsible deployment of synthetic voices, and how society can adapt to these new capabilities," OpenAI said in an unsigned blogpost. "Based on these conversations and the results of these small-scale tests, we will make a more informed decision about whether and how to deploy this technology at scale." In its post the company shared examples of real-world uses of the technology from various partners who were given access to it to build into their own apps and products. Education technology firm Age of Learning uses it to generate scripted voiceovers, while "AI visual storytelling" app HeyGen offers users the ability to generate translations of recorded content in a way that is fluent but preserves the accent and voice of the original speaker. For example, generating English with an audio sample from a French speaker produces speech with a French accent. Notably, researchers at the Norman Prince Neurosciences Institute in Rhode Island used a poor-quality 15-second clip of a young woman giving a presentation at a school project to "restore the voice" that she had lost due to a vascular brain tumour. "We are choosing to preview but not widely release this technology at this time," OpenAI said, in order "to bolster societal resilience against the challenges brought by ever more convincing generative models". In the immediate future, it said: "We encourage steps like phasing out voice-based authentication as a security measure for accessing bank accounts and other sensitive information." OpenAI also called for the exploration of "policies to protect the use of individuals' voices in AI" and "educating the public in understanding the capabilities and limitations of AI technologies, including the possibility of deceptive AI content". Voice Engine generations are watermarked, OpenAI said, which allows the organisation to trace the origin of any generated audio. Currently, it added, "our terms with these partners require explicit and informed consent from the original speaker and we don't allow developers to build ways for individual users to create their own voices". But while OpenAI's tool stands out for the technical simplicity and the tiny amount of original audio required to generate a convincing clone, competitors are already available to the public. With just a "few minutes of audio", companies such as ElevenLabs can generate a complete voice clone. To try to mitigate harms, the company has introduced a "no-go voices" safeguard, designed to detect and prevent the creation of voice clones "that mimic political

candidates actively involved in presidential or prime ministerial elections, starting with those in the US and the UK".