OpenAI

Menu

Blog

Navigating the Challenges and Opportunities of Synthetic Voices

We're sharing lessons from a small scale preview of Voice Engine, a model for creating custom voices.

March 29, 2024

Authors
OpenAI ↓

Product, Announcements

OpenAI is committed to developing safe and broadly beneficial AI. Today we are sharing preliminary insights and results from a small-scale preview of a model called Voice Engine,

which uses text input and a single 15-second audio sample to generate natural-sounding speech that closely resembles the original speaker. It is notable that a small model with a single 15-second sample can create emotive and realistic voices.

We first developed Voice Engine in late 2022, and have used it to power the preset voices available in the text-to-speech API as well as ChatGPT Voice and Read Aloud. At the same time, we are taking a cautious and informed approach to a broader release due to the potential for synthetic voice misuse. We hope to start a dialogue on the responsible deployment of synthetic voices, and how society can adapt to these new capabilities. 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.

Early applications of Voice Engine

To better understand the potential uses of this technology, late last year we started privately testing it with a small group of trusted partners. We've been impressed by the applications this group has developed. These small scale deployments are helping to inform our approach, safeguards, and thinking about how Voice Engine could be used for good across various industries. A few early examples include:

- **Providing reading assistance** to non-readers and children through natural-sounding, emotive voices representing a wider range of speakers than what's possible with preset voices. Age of Learning, an education technology company dedicated to the academic success of children, has been using this to generate pre-scripted voice-over content. They also use Voice Engine and GPT-4 to create real-time, personalized responses to interact with students. With this technology, Age of Learning has been able to create more content for a wider audience.

1. Reference audio

English   Spanish

◯

2. Generated audio

Biology   Reading   Chemistry   Math   Physics

◯

Some of the most amazing habitats on Earth are found in the rainforest. A rainforest is a place with a lot of precipitation and it has many kinds of animals trees and other plants. Tropical rainforests are usually not too far from the equator and are warm all year.

- **Translating content**, like videos and podcasts, so creators and businesses can reach more people around the world, fluently and in their own voices. One early adopter of this is HeyGen, an AI visual storytelling platform that works with their enterprise customers to create custom, human-like avatars for a variety of content, from product marketing to sales demos. They use Voice Engine for video translation, so they can translate a speaker's voice into multiple languages and reach a global audience. When used for translation, Voice Engine preserves the native accent of the original speaker: for example generating English with an audio sample from a French speaker would produce speech with a French accent.

1. Reference audio

○

2. Generated audio

Spanish   Mandarin   German   French   Japanese

○

La amistad es un tesoro universal aporta alegría apoyo y risas a nuestras vidas sin importar donde estemos en el mundo. Los verdaderos amigos están con nosotros en las buenas y en las malas compartiendo nuestras alegrías y aliviando nuestras penas. Celebremos los lazos de amistad que nos conectan a todos a través de cada idioma y cultura.

- **Reaching global communities**, by improving essential service delivery in remote settings. Dimagi is building tools for community health workers to provide a variety of essential services, such as counseling for breastfeeding mothers. To help these workers develop their skills, Dimagi uses Voice Engine and GPT-4 to give interactive feedback in each worker's primary language including Swahili or more informal languages like Sheng, a code-mixed language popular in Kenya.

1. Reference audio

Swahili   Sheng

○

2. Generated audio

Nutrition   Vitamin A   Breastfeeding

○

Lishe bora ni muhimu katika kuhakikisha kwamba watoto wanakua vizuri, kimwili na kiakili. Vyakula kama matunda, mboga, protini, kalsiamu, na vitamini mbalibali ni muhimu sana kwa ukuaji wa mifupa na maendeleo ya ubongo. Kula vizuri kunamaanisha kwamba mtoto anakuwa na mfumu wa kinga imara unaomwezesha kupambana na magonjwa. Hii ina maana kwamba, hata kama kuna mafua yanayoenea mtaani, mtoto atakuwa na uwezo mkubwa wa kukabiliana nayo. Hivyo, hakutakuwa na haja ya kumpeleka hospitalini mara kwa mara. Kwa kufanya hivyo, tunakuwa tunajenga kizazi cha watu imara. Kama unavyojua, mustakabali wa jamii yetu uko mikononi mwa vijana hawa. Ni vyema tuwape mwanzo bora maishani.

- **Supporting people who are non-verbal**, such as therapeutic applications for individuals with conditions that affect speech and educational enhancements for those with learning needs. Livox, an AI alternative communication app, powers Augmentative & Alternative Communication (AAC) devices that enable people with disabilities to communicate. By using Voice Engine, they are able to offer people who are non-verbal unique and non-robotic voices across many languages. Their users can choose speech that best represents them, and for multilingual users, maintain a consistent voice across each spoken language.

1. Reference audio

○

2. Generated audio

English   Portuguese

○

Excuse me can I get your attention? Thank you for your help. Can we watch a movie tonight? Could you please help me find my glasses? Thank you for your understanding, it means a lot to me.

- **Helping patients recover their voice**, for those suffering from sudden or degenerative speech conditions. The Norman Prince Neurosciences Institute at Lifespan, a not-for-profit health system that serves as the primary teaching affiliate of Brown University's medical school, is exploring uses of AI in clinical contexts. They've been piloting a program offering Voice Engine to individuals with oncologic or neurologic etiologies for speech impairment. Since Voice Engine requires such a short audio sample, doctors Fatima Mirza, Rohaid Ali

and Konstantina Svokos were able to restore the voice of a young patient who lost her fluent speech due to a vascular brain tumor, using audio from a video recorded for a school project.

1. Current voice

◯

2. Reference audio

◯

3. Generated audio

Talking   Ordering

◯

Hi everyone, this is what my voice sounds like using OpenAI's new text to speech model called Voice Engine. I was able to use just 15 seconds of a video that I made for a class project to be the reference audio source for the voice you hear right now. What do you think?

Building Voice Engine safely

We recognize that generating speech that resembles people's voices has serious risks, which are especially top of mind in an election year. We are engaging with U.S. and international partners from across government, media, entertainment, education, civil society and beyond to ensure we are incorporating their feedback as we build.

The partners testing Voice Engine today have agreed to our usage policies, which prohibit the impersonation of another individual or organization without consent or legal right. In addition, 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. Partners must also clearly disclose to their audience that the voices they're hearing are AI-generated. Finally, we have implemented a set of safety measures, including watermarking to trace the origin of any audio generated by Voice Engine, as well as proactive monitoring of how it's being used.

We believe that any broad deployment of synthetic voice technology should be accompanied by voice authentication experiences that verify that the original speaker is knowingly adding their voice to the service and a no-go voice list that detects and prevents the creation of voices that are too similar to prominent figures.

Looking ahead

Voice Engine is a continuation of our commitment to understand the technical frontier and openly share what is becoming possible with AI. In line with our approach to AI safety and our voluntary commitments, we are choosing to preview but not widely release this technology at this time. We hope this preview of Voice Engine both underscores its potential and also motivates the need to bolster societal resilience against the challenges brought by ever more convincing generative models. Specifically, we encourage steps like:

- Phasing out voice based authentication as a security measure for accessing bank accounts and other sensitive information

- Exploring policies to protect the use of individuals' voices in AI

- Educating the public in understanding the capabilities and limitations of AI technologies, including the possibility of deceptive AI content

- Accelerating the development and adoption of techniques for tracking the origin of audiovisual content, so it's always clear when you're interacting with a real person or with an AI

It's important that people around the world understand where this technology is headed, whether we ultimately deploy it widely ourselves or not. We look forward to continuing to engage in conversations around the challenges and opportunities of synthetic voices with policymakers, researchers, developers and creatives.

Authors

OpenAI

View all articles

Related research

View all research

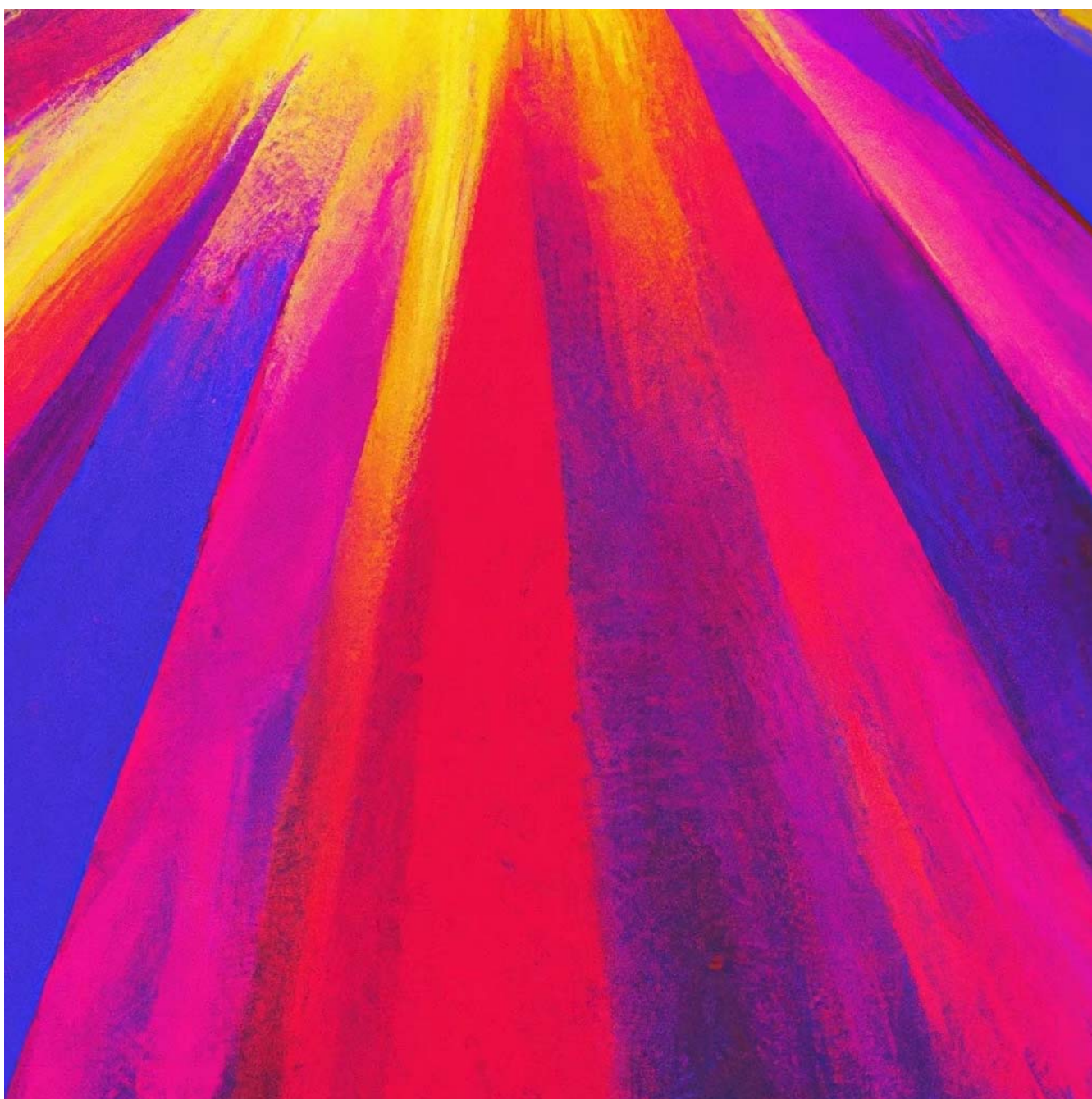## Video generation models as world simulators

Feb 15, 2024

Building an early warning system for LLM-aided biological threat creation

Jan 31, 2024

Weak-to-strong generalization

Dec 14, 2023

Practices for Governing Agentic AI Systems

Dec 14, 2023

Research   API

Overview

Overview

Index

Pricing

GPT-4

Docs

DALL·E 3

Sora

ChatGPT

Company

Overview

About

Team

Blog

Enterprise

Careers

Pricing

Charter

Try ChatGPT

Security

Customer stories

Safety

OpenAI © 2015 – 2024

Social

Terms & policies

Twitter

Privacy policy

YouTube

Brand guidelines

GitHub

SoundCloud

LinkedIn

Back to top