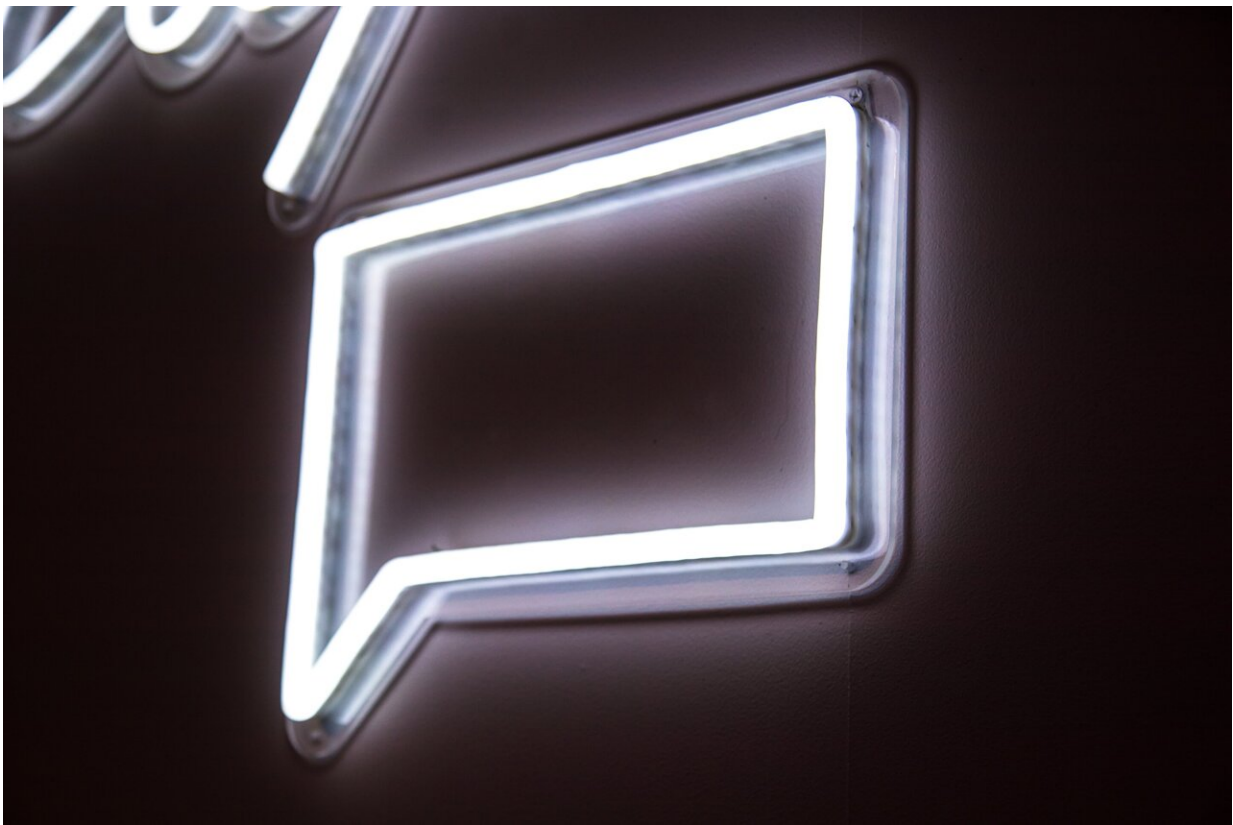


Humans unable to detect over a quarter of deepfake speech samples, research shows

August 2 2023



Credit: Unsplash/CC0 Public Domain

New research from UCL has found that humans were only able to detect artificially generated speech 73% of the time, with the same accuracy in both English and Mandarin.

The study, published in *PLOS ONE*, is the first to assess [human ability](#) to detect artificially generated [speech](#) in a language other than English.

Deepfakes are synthetic media intended to resemble a real person's [voice](#) or appearance. They fall under the category of generative artificial intelligence (AI), a type of machine learning (ML) that trains an [algorithm](#) to learn the patterns and characteristics of a dataset, such as video or audio of a real person, so that it can reproduce original sound or imagery.

While early deepfake speech algorithms may have required thousands of samples of a person's voice to be able to generate original audio, the latest pre-trained algorithms can recreate a person's voice using just a three-second clip of them speaking. Open-source algorithms are freely available and while some expertise would be beneficial, it would be feasible for an individual to train them within a few days.

Tech firm Apple recently announced software for iPhone and iPad that allows a user to create a copy of their voice using 15 minutes of recordings.

Researchers at UCL used a text-to-speech (TTS) algorithm trained on two publicly available datasets, one in English and one in Mandarin, to generate 50 deepfake speech samples in each language. These samples were different from the ones used to train the algorithm to avoid the possibility of it reproducing the original input.

These artificially generated samples and genuine samples were played for 529 participants to see whether they could detect the real thing from fake speech. Participants were only able to identify fake speech 73% of the time, which improved only slightly after they received training to recognize aspects of deepfake speech.

Kimberly Mai (UCL Computer Science), first author of the study, said, "Our findings confirm that humans are unable to reliably detect deepfake speech, whether or not they have received training to help them spot artificial content. It's also worth noting that the samples that we used in this study were created with algorithms that are relatively old, which raises the question whether humans would be less able to detect deepfake speech created using the most sophisticated technology available now and in the future."

The next step for the researchers is to develop better automated speech detectors as part of ongoing efforts to create detection capabilities to counter the threat of artificially generated audio and imagery.

Though there are benefits from generative AI audio technology, such as greater accessibility for those whose speech may be limited or who may lose their voice due to illness, there are growing fears that such technology could be used by criminals and nation states to cause significant harm to individuals and societies.

Documented cases of deepfake speech being used by criminals include one 2019 incident where the CEO of a British energy company was convinced to transfer hundreds of thousands of pounds to a false supplier by a [deepfake](#) recording of his boss's voice.

Professor Lewis Griffin (UCL Computer Science), senior author of the study, said, "With generative artificial intelligence technology getting more sophisticated and many of these tools openly available, we're on the verge of seeing numerous benefits as well as risks. It would be prudent for governments and organizations to develop strategies to deal with abuse of these tools, certainly, but we should also recognize the positive possibilities that are on the horizon."

More information: Mai KT, Warning: Humans cannot reliably detect

speech deepfakes, *PLoS ONE* (2023). DOI: [10.1371/journal.pone.0285333](https://doi.org/10.1371/journal.pone.0285333) , [journals.plos.org/plosone/arti ... journal.pone.0285333](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0285333)

Provided by University College London

Citation: Humans unable to detect over a quarter of deepfake speech samples, research shows (2023, August 2) retrieved 23 February 2024 from <https://techxplore.com/news/2023-08-humans-unable-quarter-deepfake-speech.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.