Humans can detect deepfake speech only 73% of the time, study finds

Publication Date: 2023-08-02

Author: Hibaq Farah

Section: Technology

Tags: Artificial intelligence (AI), Deepfake, news

Article URL: https://www.theguardian.com/technology/2023/aug/02/humans-can-detect-deepfake-speech-only-73-of-the-time-study-finds

Humans are able to detect artificially generated speech only 73% of the time, a study has found, with the same levels of accuracy found in English and Mandarin speakers. Researchers at University College London used a text-to-speech algorithm trained on two publicly available datasets, one in English and the other in Mandarin, to generate 50 deepfake speech samples in each language. Deepfakes, a form of generative artificial intelligence, are synthetic media that is created to resemble a real person's voice or the likeness of their appearance. The sound samples were played for 529 participants to see whether they could detect the real sample from fake speech. The participants were able to identify fake speech only 73% of the time. This number improved slightly after participants received training to recognise aspects of deepfake speech. The study is the first to assess human ability to detect artificially generated speech in a language other than English. It speaks to concerns that humans are unable to consistently detect when an audio is a deepfake, despite being trained to. Kimberly Mai, first author of the study, said: "In our study, we showed that training people to detect deepfakes is not necessarily a reliable way to help them to get better at it. Unfortunately, our experiments also show that at the moment automated detectors are not reliable either. "They're really good at detecting deepfakes if they've seen similar examples during their training phase, if the speaker is the same or the clips are recorded in a similar audio environment, for example. But they're not reliable when there are changes in the test audio conditions, such as if there's a different speaker." She said it was important to improve automated deepfake speech detectors and for organisations to "think about strategies to mitigate the threat that deepfake content poses". In deepfake video, there are more clues to identify whether it has been synthetically created than in audio. This year Brad Smith, the president of Microsoft, said that his biggest concern around artificial intelligence was deepfakes. As the development of deepfakes continue, widely available sophisticated detections systems lag behind. Dr Karl Jones, the head of engineering at Liverpool John Moores University, has warned that the UK's justice system is not set up to protect against the use of deepfakes. "Deepfake speech is almost the perfect crime – because you don't know that it's been done," he said. Sam Gregory, the executive director of Witness, who has created initiatives on deepfakes, media manipulation and generative AI, has said that another threat is humans claiming that a real audio is fake and relying on the fact that there are not widespread tools available to detect whether this is the case. He said: "At Witness, we speak about a detection equity gap. The people who need the capacity to detect – journalists and factcheckers, and civil society and election officials – are the ones who don't have access to these [detection] tools. This is a huge issue that is going to get worse if we don't invest in those skills and resources. "We may not need to have detection tools available to

everyone, because that also makes them harder to be robust. But we need to think about the investment in supporting intermediaries.”