Tech Xplore

Deepfake of principal's voice is the latest case of AI being used for harm

April 29 2024, by Ben Finley

Baltimore County Police Chief Robert McCullough and other local officials speak at a news conference in Towson, Md., April 25, 2024. The most recent criminal case to involve artificial intelligence has emerged from a high school in Baltimore County, Maryland. That's where police say a principal was framed by a fake recording of his voice. Credit: Kim Hairston/The Baltimore Sun via AP, file

The most recent criminal case involving artificial intelligence emerged last week from a Maryland high school, where police say a principal was framed as racist by a fake recording of his voice.

The case is yet another reason why everyone—not just politicians and celebrities—should be concerned about this increasingly powerful deep-fake technology, experts say.

"Everybody is vulnerable to attack, and anyone can do the attacking," said Hany Farid, a professor at the University of California, Berkeley, who focuses on digital forensics and misinformation.

Here's what to know about some of the latest uses of AI to cause harm:

AI has become very accessible

Manipulating recorded sounds and images isn't new. But the ease with which someone can alter information is a recent phenomenon. So is the ability for it to spread quickly on social media.

The fake audio clip that impersonated the principal is an example of a subset of artificial intelligence known as generative AI. It can create hyper-realistic new images, videos and audio clips. It's cheaper and easier to use in recent years, lowering the barrier to anyone with an internet connection.

"Particularly over the last year, anybody—and I really mean anybody—can go to an online service," said Farid, the Berkeley professor. "And either for free or for a few bucks a month, they can upload 30 seconds of someone's voice."

Those seconds can come from a voicemail, social media post or surreptitious recording, Farid said. Machine learning algorithms capture what a person sounds like. And the cloned speech is then generated from words typed on a keyboard.

The technology will only get more powerful and easier to use, including for video manipulation, he said.

What happened in Maryland?

Authorities in Baltimore County said Dazhon Darien, the athletic director at Pikesville High, cloned Principal Eric Eiswert's voice.

The fake recording contained racist and antisemitic comments, police said. The sound file appeared in an email in some teachers' inboxes before spreading on social media.

The recording surfaced after Eiswert raised concerns about Darien's work performance and alleged misuse of school funds, police said.

The bogus audio forced Eiswert to go on leave, while police guarded his house, authorities said. Angry phone calls inundated the school, while hate-filled messages accumulated on social media.

Detectives asked outside experts to analyze the recording. One said it "contained traces of AI-generated content with human editing after the fact," court records stated.

A second opinion from Farid, the Berkeley professor, found that "multiple recordings were spliced together," according to the records.

Farid told The Associated Press that questions remain about exactly how that recording was created, and he has not confirmed that it was fully AI-generated.

But given AI's growing capabilities, Farid said the Maryland case still serves as a "canary in the coal mine," about the need to better regulate this technology.

Why is audio so concerning?

Many cases of AI-generated disinformation have been audio.

That's partly because the technology has improved so quickly. Human ears also can't always identify telltale signs of manipulation, while discrepancies in videos and images are easier to spot.

Some people have cloned the voices of purportedly kidnapped children over the phone to get ransom money from parents, experts say. Another pretended to be the chief executive of a company who urgently needed funds.

During this year's New Hampshire primary, AI-generated robocalls impersonated President Joe Biden's voice and tried to dissuade Democratic voters from voting. Experts warn of a surge in AI-generated disinformation targeting elections this year.

But disturbing trends go beyond audio, such as programs that create fake nude images of clothed people without their consent, including minors, experts warn. Singer Taylor Swift was recently targeted.

What can be done?

Most providers of AI voice-generating technology say they prohibit harmful usage of their tools. But self enforcement varies.

Some vendors require a kind of voice signature, or they ask users to recite a unique set of sentences before a voice can be cloned.

Bigger tech companies, such as Facebook parent Meta and ChatGPT-maker OpenAI, only allow a small group of trusted users to experiment with the technology because of the risks of abuse.

Farid said more needs to be done. For instance, all companies should require users to submit phone numbers and credit cards so they can trace back files to those who misuse the technology.

Another idea is requiring recordings and images to carry a digital watermark.

"You modify the audio in ways that are imperceptible to the human auditory system, but in a way that can be identified by a piece of software downstream," Farid said.

Alexandra Reeve Givens, CEO of the Center for Democracy & Technology, said the most effective intervention is law enforcement action against criminal use of AI. More consumer education also is needed.

Another focus should be urging responsible conduct among AI companies and social media platforms. But it's not as simple as banning Generative AI.

"It can be complicated to add legal liability because, in so many instances, there might be positive or affirming uses of the technology," Givens said, citing translation and book-reading programs.

Yet another challenge is finding international agreement on ethics and guidelines, said Christian Mattmann, director of the Information Retrieval & Data Science group at the University of Southern California.

"People use AI differently depending on what country they're in," Mattmann said. "And it's not just the governments, it's the people. So culture matters."

Citation: Deepfake of principal's voice is the latest case of AI being used for harm (2024, April 29) retrieved 16 August 2024 from https://techxplore.com/news/2024-04-deepfake-principal-voice-latest-case.html