Realtime deepfakes are a dangerous new threat. How to protect yourself

May 17 2023, by Jon Healey

Credit: Unsplash/CC0 Public Domain

You've probably seen deepfake videos on the internet that inject facsimiles of famous people into odd or funny situations—for example, a fake Tom Cruise doing "industrial cleanup," or in a truly meta effort,

an artificial Morgan Freeman hyping "the era of synthetic reality."

Now imagine receiving a phone call from someone who sounds exactly like your child, pleading for emergency help. Same technology, but no one's laughing.

Cybersecurity experts say deepfake technology has advanced to the point where it can be used in real time, enabling fraudsters to replicate someone's voice, image and movements in a call or virtual meeting. The technology is also widely available and relatively easy to use, they say. And it's getting better all the time.

"Thanks to AI tools that create 'synthetic media' or otherwise generate content, a growing percentage of what we're looking at is not authentic, and it's getting more difficult to tell the difference," the Federal Trade Commission warned.

Researchers say the technology for real-time deepfakes has been around for the better part of a decade. What's new is the range of tools available to make them.

"We know we're not prepared as a society" for this threat, said Andrew Gardner, vice president of research, innovation and AI at Gen. In particular, he said, there's nowhere to go if you're confronted with a potential deepfake scam and you need immediate help verifying it.

Real-time deepfakes have been used to scare grandparents into sending money to simulated relatives, win jobs at tech companies in a bid to gain inside information, influence voters, and siphon money from lonely men and women. Fraudsters can copy a recording of someone's voice that's been posted online, then use the captured audio to impersonate a victim's loved one; one 23-year-old man is accused of swindling grandparents in Newfoundland out of $200,000 in just three days by using this

technique.

Tools to weed out this latest generation of deepfakes are emerging too, but they're not always effective and may not be accessible to you. That's why experts advise taking a few simple steps to protect yourself and your loved ones from the new type of con.

The term deepfake is shorthand for a simulation powered by deep learning technology—artificial intelligence that ingests oceans of data to try to replicate something human, such as having a conversation (e.g., ChatGPT) or creating an illustration (e.g., Dall-E). Gardner said it's still an expensive and time-consuming proposition to develop these tools, but using them is comparatively quick and easy.

Yisroel Mirsky, an AI researcher and deepfake expert at Ben-Gurion University of the Negev, said the technology has advanced to the point where it's possible to do a deepfake video from a single photo of a person, and a "decent" clone of a voice from only three or four seconds of audio. But Gardner said the tools widely available to make deepfakes lag behind the state of the art; they require about five minutes of audio and one to two hours of video.

Regardless, thanks to sites like Facebook, Instagram and YouTube, there's plenty of images and audio for fraudsters to find.

Mirsky said it's easy to imagine an attacker looking on Facebook to identify a potential target's children, calling the son to record enough audio to clone his voice, then using a deepfake of the son to beg the target for money to get out of a jam of some kind.

The technology is becoming so efficient, he said, you can clone a face or a voice with a basic gaming computer. And the software is "really point and click," he said, easily available online and configurable with some

basic programming.

To illustrate how effective real-time deepfakes can be, the LexisNexis Risk Solutions' Government Group shared a video that David Maimon, a criminology professor at Georgia State University, grabbed from the dark web of an apparent catfishing scam in progress. It showed an online chat between an older man and a young woman who was asking for a loan so she could meet the man in Canada. But in a third window, you could see a man was actually saying the words that were coming out of the woman's mouth in a woman's voice—she was a deepfake, and he was a scammer.

This technique is known as reenactment, Mirsky and Wenke Lee of the Georgia Institute of Technology said in a paper published in 2020. It also can be used to "perform acts of defamation, cause discredibility, spread misinformation and tamper with evidence," they wrote. Another approach is replacement, where the target's face or body is placed on someone else, as in revenge porn videos.

But how, exactly, fraudsters are using the tools remains a bit of a mystery, Gardner said. That's because we only know what they've been caught doing.

Haywood Talcove, chief executive of LexisNexis Risk Solutions' Government Group, said the new technology can circumvent some of the security techniques that companies have been deploying in lieu of passwords. For example, he pointed to California's two-step online identification process, which has users upload two things: a picture of their driver's license or ID card, then a freshly snapped selfie. Fraudsters can buy a fake California ID online for a few dollars, then use deepfake software to generate a matching face for the selfie. "It's a hot knife through butter," he said.

Similarly, Talcove said that financial companies need to stop using voice-identification tools to unlock accounts. "I'd be nervous if [at] my bank, my voice were my password," he said. "Just using voice alone, it doesn't work anymore." The same goes for facial recognition, he said, adding that the technology was at the end of its useful life as a way to control access.

The Cybercrime Support Network, a nonprofit that helps individuals and businesses victimized online, often works with targets of romance scams, and it urges people to do video chats with their suitors to try to weed out scammers. Ally Armeson, the network's program director, said that just two or three years ago, they could tell clients to look for easy-to-spot glitches, like frozen images. But in recent weeks, she said, the network has been contacted by scam victims who said they'd done a video chat for 10 or 20 minutes with their supposed suitor, "and it absolutely was the person that they sent me in the photo."

She added, "The victims did say, 'The head did kind of look weird on the body, so it looked a little off.'" But it's not unusual for people to ignore red flags, she said. "They want to believe that the video is real, so they'll overlook minor discrepencies."

(Victims of romance scams in the United States reported $1.3 billion in losses last year.)

Real-time deepfakes represent a dangerous new threat to businesses too. A lot of companies are training employees to recognize phishing attacks by strangers, Mirsky said, but no one's really preparing for calls from deepfakes with the cloned voice of a colleague or a boss.

"People will confuse familiarity with authenticity," he said. "And as a result, people are going to fall for these attacks."

How to protect yourself

Talcove offered a simple and hard-to-beat way to guard against deepfakes that impersonate a family member: Have a secret code word that every family member knows, but that criminals wouldn't guess. If someone claiming to be your daughter, grandson or nephew calls, Talcove said, asking for the code word can separate real loved ones from fake ones.

"Every family now needs a code word," he said.

Pick something simple and easily memorable that doesn't need to be written down (and isn't posted on Facebook or Instagram), he said, then drill it into your family's memory. "You need to make sure they know and practice, practice, practice," Talcove said.

Gardner also advocated for code words. "I think preparation goes a long way" in defending against deepfake scams, he said.

Armeson said her network still tells people to look for certain clues on video calls, including their supposed paramour blinking too much or too little, having eyebrows that don't fit the face or hair in the wrong spot, and skin that doesn't match their age. If the person is wearing glasses, check whether the reflection they give is realistic, the network says—"deepfakes often fail to fully represent the natural physics of lighting."

She also urges people to give these simple tests: Ask the other person in the video call to turn their head around and to put a hand in front of their face. Those maneuvers can be revealing, she said, because deepfakes often haven't been trained to do them realistically.

Still, she admitted, "we're just playing defense." The fraudsters are

"always going to kind of be ahead of us," weeding out the glitches that reveal the con, she said. "It's infuriating."

Ultimately, she said, the most reliable way to smoke out deepfakes may be to insist on an in-person meeting. "We have to be really analog about it. We can't just rely on technology."

There are software tools that automatically look for AI-generated glitches and patterns in an effort to separate legitimate audio and video from fake. But Mirsky said "this potentially is a losing game" because as technology improves, the telltale signs that used to betray the fakes will go away.

Mirsky and his team at Ben-Gurion University have developed a different approach, called D-CAPTCHA, which is operates on the same principle that some websites use to stop bots from submitting forms online. A D-CAPTCHA system poses a test that's designed to flummox current real-time deepfakes—for example, asking callers to hum, laugh, sing or just clear their throat.

The system, which has yet to be commercialized, could take the form of a waiting room to authenticate guests attending sensitive virtual meetings or an app that verifies suspect callers. In fact, Mirsky said, "we can develop apps that can try to catch these suspicious calls and vet them before they're connected."

Gardner offered one other, hopeful note. The experiences people are having now with AI and apps like ChatGPT, he said, have made people quicker to question what is real and what is fake, and to look more critically at what they're seeing.

"The fact that people are having these AI conversations one-on-one on their own is, I think, helping," he said.