nakedsecurity.sophos.com

Scammers deepfake CEO’s voice to talk underling into $243,000 transfer

by Lisa Vaas

6-7 minutos

Any business in its right mind should be painfully aware of how much money they could bleed via skillful Business Email Compromise (BEC) scams, where fraudsters convincingly forge emails, invoices, contracts and letters to socially engineer the people who hold the purse strings.

And any human in their right mind should be at least a little freaked out by how easy it now is to churn out convincing deepfake videos – including, say, of you, cast in an adult movie, or of your CEO saying things that… well, they would simply never say.

Well, welcome to a hybrid version of those hoodwinks: deepfake audio, which was recently used in what’s considered to be the first known case of an AI-generated voice of a CEO to bilk a UK-based energy firm out of €220,000 (USD $243,000).

The Wall Street Journal reports that some time in March, the British CEO thought he had gotten a call from the CEO of his business’s parent company, which is based in Germany.

Whoever placed the call sounded legitimate. The voice had the hint of a German accent and the same “melody” that the UK CEO recognized in his boss’s voice, according to fraud expert Rüdiger Kirsch, who works with the company’s insurer, Euler Hermes Group SA. The insurer shared details of the crime with the WSJ, but it declined to identify the businesses involved.

The caller had an “urgent” request: he demanded that the British CEO transfer $243,000 to a Hungarian supplier within the hour. He complied and made the transfer.

Analysts told the WSJ that they believe that artificial intelligence- (AI)-based software was used to create a convincing imitation of the German CEO’s voice. The transfer went through, and the money was subsequently funneled into accounts in other countries.

The scammers then called back for more: Kirsch told the WSJ that the imposter called the target company three times. The transfer went through after their first call, then the attacker called a second time to lie about the money having been reimbursed to the British company. Then, they called a third time, to ask for another payment, using the same fake voice.

The British CEO had grown skeptical by that time, given that the “reimbursement” never showed up. Plus, the third call was made with an Austrian phone number. Hence, he didn’t comply with the repeated demand for money.

Joe Rogan vs. Joe Fauxgan

If you aren’t familiar with how realistic AI-generated deepfake audio has become in recent months, you can listen to this sample, produced by the AI startup Dessa and released in May. The subject of the impersonation is the popular podcaster and comedian Joe Rogan.

You can decide for yourself how accurate the deepfake audio is, or you can take this quiz that Dessa released. I guessed right on 5 out of 8 samples, but some of those correct answers were pure guesses. It’s tough to tell the difference, in short, in my and others’ experience.

As Dessa pointed out at the time, there was plenty of material to work with when it comes to training AI. As of the time its team went to work to create a fake Rogan – that would be Joe Fauxgan, as Gizmodo quipped – Rogan had released close to 1,300 episodes of his podcast, with most of those episodes being 2-3 hours long.

That’s thousands of hours worth of audio to train from. Bafflingly, though, Dessa said that its team created the Rogan replica voice with a text-to-speech deep learning system they developed called RealTalk, which generates life-like speech using *only text inputs.*

What does this all mean? Well, it means that the floodgates have opened on deepfake audio, for one thing. Dessa’s Principal Machine Learning Architect Alex

Krizhevsky:

> Human-like speech synthesis is soon going to be a reality everywhere.

It also means that we can expect more cybercrooks to pull off convincing scams like the one with the faux German CEO.

As it is, a year ago, the voice interaction identity and security infrastructure company Pindrop released a report that found that the rate of voice fraud had climbed over 350% from 2013 through 2017, with no signs of slowing down. Pindrop attributed the surge to several causes, one of which was the development of new voice technology.

In a post on Medium, Dessa said that at this point, you have to have a good deal of “technical expertise, ingenuity, computing power and data” to make models like its RealTalk perform well.

> So not just anyone can go out and do it. But in the next few years (or even sooner), we’ll see the technology advance to the point where only a few seconds of audio are needed to create a life-like replica of anyone’s voice on the planet.

From the sounds of the voice phishing (vishing) scam pulled off in March, it sounds like it isn’t years off at all. We don’t have many details, but it sounds pretty much like it’s “now.”

What to do?

If this turns out to indeed be a deepfake audio scam, we can take a page from the advice given out to avoid BEC scams to avoid these, as well. After all, both scams are after the same thing – pretending to be a business's known contacts so as to initiate fraudulent transfers. And in both types of scam, the crooks benefit from the fact that you can't see them when they're hiding behind a convincing email or a convincing-sounding phone call.

As the the FBI notes with regards to BEC, no matter how sophisticated the fraud, there's an easy way to thwart it: don't rely on email alone. In this case, we can swap in "voice" for "email".

FBI Special Agent Martin Licciardo:

> The best way to avoid being exploited is to verify the authenticity of requests to send money by walking into the CEO's office or speaking to him or her directly on the phone.

And by "directly," make sure that you're the one who places the call, as opposed to being the one who picks up the phone and potentially becomes a fraudster's mark.