TechTarget | SearchCIO

ENTERPRISES NEED TO PLAN FOR DEEPFAKE TECHNOLOGY

Politicians and Hollywood stars aren't the only ones at risk: Enterprises need to understand the dangers deepfakes pose to their brands and employees. Here's a primer.

George Lawton

The new AI-based technology known as deepfakes allows people to produce or alter video content so that it presents something that didn't, in reality, occur. Although there may be some practical and positive uses of deepfake technology -- an effort to help people with ALS is an example -- the nefarious uses of deepfakes have politicians and technology executives rattled.

Deepfakes may have been responsible, for example, for an attempted military coup in Gabon and have been used to impersonate Facebook CEO Mark Zuckerberg. Experts fear deepfake technology could be used to damage brands, create stock scares or sully the reputation of executives. Deepfake technology could also be used to compromise cybersecurity.

"As some very high-quality but obviously fake deepfakes have gone viral, I think business and government leaders have begun to realize the potential for disruption they could create," said Dave Hatter, cybersecurity consultant at Intrust IT. Politicians are calling for hearings and proposing new regulations; businesses must start grappling with deepfake technology, too.

Interviews with experts in technology, ethics and AI research make it clear that enterprises should understand the various dangers of deepfake technologies, including reputation damage and new security vulnerabilities. Enterprises may also want to develop educational campaigns to help educate staff about how to deal with deepfake attacks.

But combatting deepfakes won't be easy, they warned. Although tools for detecting deepfakes are getting better, so is the deepfake technology. In the long run, the best protection against deepfake problems is likely to be ongoing monitoring of processes.

Before getting into the enterprise risk and response to deepfakes, here is a brief introduction to the new world of deepfakes.

Photoshopping on steroids

The word deepfake, a portmanteau of deep learning and fake, takes the notion of photoshopping to another level by using an AI technique known as generative adversarial networks, or GANs. These techniques pit algorithms that generate fakes against algorithms that detect whether the result is fake or real. Over the course of several rounds of this adversarial interplay, the generative algorithms can produce results that start to fool people.

Experts point to a number of reasons for the rising concern over deepfakes. High-quality video composition tools are now more widely available than ever before. We are also a society trained to get our information in sound bites -- but not one trained to recognize deepfakes.

When we read text or look at a photo, "we likely understand that context may be missing," said Bill Bronske, senior solutions architect at Globant, a digital transformation consultancy. People understand that text can be misquoted and photos can be photoshopped. "However, we don't immediately recognize that video or audio can be as easily synthesized," he said.

Matt Price, principal research engineer at ZeroFox, a social media security firm, said that the increasingly realism of the deepfakes produced today is alarming. Before Face Swap -- the first simple deepfake creation app released in December 2017 -- deepfake creators needed to have significant computer vision knowledge, very powerful computers and a carefully crafted data set to create a convincing fake. Now, anyone with access to a powerful enough computer and moderate to low technical ability can set up one of the deepfake tools to create a bogus video, Price said.

Time will teach us to get better at recognizing deepfakes. "Just like when Photoshop came into the world and everyone [eventually] learned you cannot believe what you see, the world will adapt and be more careful when watching content and looking at its origin," said Jonathan Heimann, co-founder at Canny AI, which makes deepfake tools for video dialogue replacement and lip-syncing dubbed content in any language.

But, make no mistake, sorting out deepfakes from reality will take vigilance, he added. "Just like in [IT] security, I expect a constant struggle with those who try to fake versus those who try to detect."

Even as we battle to sort fact from fake, the damage in a sense has already been done, said Professor Robert Foehl, executive in residence of business law and ethics at Ohio University. We have, in essence, lost a window into the world.

"The biggest danger of deepfakes is that as they become absolutely indistinguishable from real video or images, society will not be able to trust the authenticity of the video and images it sees. We will not be able to ascertain the truth through use of videos or images, which is a radical departure from how we have viewed videos and images since their inception," Foehl said.

Hyper-personalization for everything from sales to physics class

There are some positive uses to the technology as well. Recently, deepfake technology was used to create a video of David Beckham telling people how they can protect themselves from Malaria in nine different languages using technology from Synthesia.io.

"Clearly, this is a huge positive coming from a well-known celebrity," said Douglas Mapuranga, CIO at Infrastructure Development Bank of Zimbabwe and president of ISACA Harare Chapter.

The technology could also be used to improve educational engagement. "Imagine sitting in a physics class and watching Einstein delivering a lecture on the theory of relativity," he added.

Moshe Kranc, CTO at Ness Digital Engineering, a custom software development company, expects to also see deepfake technology used for hyper-personalizing sales videos, customer support and employee training. At the same time, he stressed that enterprises should make it clear to the consumer that the video was artificially generated, rather than misleading the customer into thinking they paid David Beckham to specifically create an ad for that customer in Portuguese.

Deepfakes could also help enterprises edit videos for a much lower cost than today's offerings, ZeroFox's Price said. For example, an enterprise could put together a press release video and after filming, go back and edit the video for misspoken words or stuttered lines. In addition, emerging deepfake capabilities will allow enterprises to more easily dub the videos they produce rather than resorting to voiceovers or using the expensive dubbing techniques available today.

Deepfakes could be used in medical activities. In the ALS example, voice recordings of ALS patients are used to develop a vocal "clone" that allows them to continue communicating once they can no longer speak, said Dr. Martin Zizi, CEO of Aerendir, a physiological biometric technology provider. Deepfake technology could also be used to help people who have suffered strokes, or some psychological or psychiatric ailment that prevents them from communicating in the normal way. But such applications would have to be under strict ethical and legal guidelines, as "we would be dealing with literal rewiring of a human brain," he cautioned.

Dangers to the enterprise

The main danger that deepfakes pose to enterprises is misrepresentation, which can take on any number of forms, said Price. A shareholder meeting could be altered, the CEO could be shown doing or saying something that damages the brand, or an employee in uniform could be shown having a faked interaction with a customer. Deepfakes of this ilk could result in customer loss, decreases in stock prices or the company getting unwanted public attention.

The main action that enterprises can take today to protect themselves is to thoroughly monitor mentions of the company and its employees on the internet, Price said.

Currently, social networks and websites on the internet control the content that is released, so by working with those content delivery networks, enterprises can at least identify, if not remove, fake content. Robust detection of deepfakes is still difficult to do, and current research hasn't found a way to easily identify deepfake content at scale -- short of having human analysts inspect the content, Price said.

Enterprises must also protect themselves from being identified as an originator of deepfakes. Employees using deepfake technology to create fake product reviews or to smear competitors

put the company at risk. The policies that govern cybersecurity and acceptable use of an enterprise's IT assets should also apply to the production of deepfakes and/or circulation in an enterprise network, Mapuranga said.

Deepfakes in cybersecurity

Indeed, cybersecurity is another area that must adapt to the rising use of deepfake technology.

"The underlying AI technology that enables deepfakes will certainly facilitate more intensive cyber-attacks," said Steve Grobman, CTO at McAfee. "This includes recent improvements in generative adversarial networks and the lowered barriers to using advanced AI capabilities by our adversaries."

For example, deepfake technology is now being used to create high-fidelity phishing attacks where the phishing target (financial institution, healthcare provider, auction site, email provider) is indistinguishable from the real entity, said Hal Lonas, CTO at Webroot Inc., a cybersecurity and threat intelligence service.

The threat of realistic, fake digital people could also have an impact on nearly every organization that operates a digital onboarding solution, said Sanjay Gupta, vice president and general manager of corporate development at Mitek Systems, an identity verification service.

Organizations may need to implement new identity protection measures. Additionally, from an ethical perspective, these onboarding systems should be configured to not only capture the various employee biometrics, but also determine how best to safeguard the information so that it can't be used to create potential fakes.

Deepfake technology could also be used to create fake positives for facial identification software. "Access to a person's facial data, coupled with the ability to imitate their facial mannerisms using deepfakes, creates an alarming scenario where virtually anyone can be impersonated, and no one is truly safe," said Aerendir's Zizi.

Blockchain to the rescue?

Researchers are working on new tools for detecting deepfakes by looking for more subtle features related to a person's physiology. But experts warn that deepfakes will only get better over time.

"By definition, construction of deepfakes using generative adversarial networks means that over time, they will be less and less susceptible to detection," Globant's Bronske said. Combining video and audio detection techniques will be short-lived victories, he warned.

Longer-term, enterprises will need to explore other options such as use of ledgering, certification and checksums. These will necessitate complex processes and tooling like those used for electronic signing of legal documents.

Blockchain technology might play a role in the authentication of video material, said Zohar Pinhasi, CEO at MonsterCloud, a managed cyber security services provider. "Unless a piece of media can be tracked to the source, it presents a very real risk for enterprises and IT professionals who have to cope with the fallout."

Educating employees

Enterprises might want to start raising awareness among employees about the threats posed by deepfakes and the proliferation of digital impersonation, said Adam Dodge, legal and technology director at Laura's House, a domestic violence support organization.

During his trainings, he is surprised to learn that 95% of the audience has never heard of a deepfake. "We cannot expect anyone to prevent or protect themselves against deepfakes if they lack total awareness of the problem," he said.

Educating employees about deepfake technology helps them think critically about content that might be altered. Company policies should be created that:

- Help employees identify fakes
- Provide standard responses when a deepfake issue arises
- Spell out who should be notified internally of digital impersonations
- Trigger a coordinated PR response to debunk deepfakes and control the narrative

Protecting employees also is paramount, and there are several steps enterprises can take on their behalves, Dodge said. Google will de-index a deepfake from its search results and takedown requests can be issued to websites that publish fakes. Law enforcement should be notified and, if the creator of the deepfake is known, a restraining order may be appropriate. Speed is key and an action plan including these steps will help mitigate the damage to the targeted employee.

The new cyber battlefront

In just the past month, recent technical advancements now allow one to create a fairly realistic video using just one source image. Another advancement allows one to edit the audio transcript of an existing video and have those edits appear seamlessly in the produced deepfake.

"We expect that over the next few years, we will see the gradual adoption of these capabilities culminating in a tool that combines these capabilities into a single, easy-to-use package allowing for virtually anyone to create deepfakes of anything or anyone," ZeroFox's Price said.

Otavio Freire, president, CTO and co-founder at SafeGuard Cyber, a digital risk protection company, echoed the alarm: "Deepfakes will only continue becoming more sophisticated, so we'll need sophisticated tools to detect them. Deepfake technology has emerged as a new battlefront in the world of digital safety."

Matt Ferrari, CTO at ClearData, a health cloud solutions provider, remains optimistic that the good guys will keep up with the fake mongers. "I expect security tooling to expand [to] detecting deepfakes, almost like a vulnerability or malware scanner works today," he said. He added new tools will tackle identity theft, fraud and public shaming perpetrated by deepfakes. "Some of this technology is already in research."