Insurance POST

Intelligence: Deepfakes – the next cyber threat?

Pamela Kokoszka
@PostPamelaK

22 Nov 2021

Indicative reading time: 10 minutes

Recent technological advances and increased dependence on video-based communication have accelerated the ability for fraudsters to create realistic audio and video fakes, also known as deepfakes, using artificial intelligence learning. Post investigates what threats deepfake technology poses for insurers and policyholders.

Deepfake is a digital manipulation of images or video that make it appear like a person did something they did not do in a “hyper-realistic way” to the point that an unaided observer cannot detect that it is fake.

With Covid-19 increasing the use of video conferencing services, many security researchers are now predicting that deepfakes could become a major security threat in the 2021-2022 period.

In March, the US Federal Bureau of Investigation also warned that “malicious actors” almost certainly will leverage the deepfake technology in the next 12 to 18 months. It expects that the technology will be used by foreign and criminal cyber actors for spearphishing and social engineering in an evolution of cyber operational tradecraft.

In fact, cyber criminals have already taken advantage of this technology. In March 2019, cyber criminals used AI-based software to impersonate the CEO of a UK-based energy firm's voice to **demand the fraudulent transfer** of $243,000 (£181,000).

Meanwhile, at the end of March 2021, a group of cyber criminals hacked into China's identity verification system to fake tax invoices using facial images purchased on the black market to create synthetic identities. The group was able to set up a shell company and issue fake tax invoices worth Y500m (£3.3m).

In the US, generally, synthetic content is considered protected speech under the First Amendment. The FBI, however, may investigate malicious synthetic content which is attributed to foreign actors or is otherwise associated with criminal activities.

In the UK, data protection laws including the *Data Protection Act* 2018 or the ***General Data Protection Regulation*** will also apply to deepfakes, even if these are generated using photographs made available by the victim, according to Kelsey Farish, associate at DAC Beachcroft. The best course of action to get unwanted deepfakes removed may be through a GDPR data subject deletion request or via the platform's terms of service or other policies, which will likely have a ban on deepfakes.

Risks

Although the **technology** has been around since late 1980s and has been steadily growing in sophistication over the years its use is still very "theoretical and hypothetical" according to Hans Allnutt, partner and head of cyber risk team at DAC Beachcroft.

Allnutt says: "What is clear is that the technology is there, as well as the potential for it to be used in **cyber** attacks. But the technology is only scratching the surface on the potential types of frauds and risks that it might create.

"Cyber criminals are very resourceful and imaginative so it is only a matter of time

FBI recommended mitigation for ransomware attack

Use multi-factor authentication.

until we see this technology used."

As the technology improves, there are concerns about how far that would go in impersonating individuals. Currently, cyber criminals can use deepfakes to portray someone using video or audio, but are unable to do so physically.

However, Brian Warszona, UK cyber deputy practice leader at Marsh, is concerned that in the future deepfakes could be physical, allowing cyber criminals to bypass biometric security including fingerprint scanning, retinal scanning or breath scanning.

He says: "Deepfakes are not impersonating all three of those [at the moment], at least that I am aware of but when we start getting into deepfakes that can fake retinal scans we remove another security layer, which is a concern."

He points out quantum computing already can be used to break through and "crack passwords and multi-factor authentication within seconds".

- Require multi-factor authentication to remotely access networks from external sources.

Implement network segmentation and filter traffic.

- Implement and ensure robust network segmentation between networks and functions to reduce the spread of the ransomware. Define a demilitarized zone that eliminates unregulated communication between networks.
- Filter network traffic to prohibit ingress and egress communications with known malicious IP addresses.
- Enable strong spam filters to prevent phishing emails from reaching end users. Implement a user training program to discourage users from visiting malicious websites or opening malicious attachments. Filter emails containing executable files to prevent them from reaching end users.
- Implement a URL blocklist and/or allowlist to prevent users from accessing malicious websites.

Scan for vulnerabilities and keep software updated.

- Set antivirus/antimalware programs to conduct regular scans of network assets using up-to-date signatures.
- Upgrade software and operating systems, applications, and firmware on network assets in a timely manner. Consider using a centralized patch management system.

Remove unnecessary applications and apply controls.

He continues: “As soon as [the technology] becomes mainstream, passwords and multi-factor authentication is useless. That’s the unknown of deepfakes.”

With the extent of the risks the deepfakes could pose still unknown, Darren Thomson, head of cyber security strategy at Cyber Cube, urges insurers to track progress in this area “closely” and ensure that “management frameworks, security strategies, analytics tools and catastrophe models take this emerging threat into consideration”.

He adds: ”The insurance market will need to consider advances in social engineering when developing attack scenarios. For example, deep fake technology could destabilize political systems, perhaps on a global basis, as communications constructed from the technology become indistinguishable from the real thing. This same technology could impact the financial markets and the reputation of large corporations.

- Remove any application not deemed necessary for day-to-day operations. Conti threat actors leverage legitimate applications—such as remote monitoring and management software and remote desktop software applications—to aid in the malicious exploitation of an organization’s enterprise.
- Investigate any unauthorized software, particularly remote desktop or remote monitoring and management software.
- Implement application listing, which only allows systems to execute programs known and permitted by the organization’s security policy. Implement software restriction policies (SRPs) or other controls to prevent programs from executing from common ransomware locations, such as temporary folders supporting popular internet browsers or compression/decompression programs.
- Implement execution prevention by disabling macro scripts from Microsoft Office files transmitted via email. Consider using Office Viewer software to open Microsoft Office files transmitted via email instead of full Microsoft Office suite applications.
- See the joint Alert, Publicly Available Tools Seen in Cyber Incidents Worldwide—developed by CISA and the cybersecurity authorities of Australia, Canada, New Zealand, and the United Kingdom—for guidance on detection and protection against malicious use of publicly available tools.

"Technology can play its part in mitigating cyber risks but to understand the nature of the threat, it is important to understand the actors behind it. Multi-disciplinary experts across data science, cyber security, software engineering, actuarial modelling, the military and commercial insurance will increasingly play their part in helping to understand the psychology and motivations behind social engineering approaches."

Until then, however, the actual impact of the deepfake technology on policyholders is not that different from a "normal" cyber attack and is comparable to a phishing attack, according to Jake Moore, cyber security specialist at antivirus company Eset.

Implement endpoint and detection response tools.

- Endpoint and detection response tools allow a high degree of visibility into the security status of endpoints and can help effectively protect against malicious cyber actors.

Limit access to resources over the network, especially by restricting RDP.

- After assessing risks, if RDP is deemed operationally necessary, restrict the originating sources and require multi-factor authentication.

Secure user accounts.

- Regularly audit administrative user accounts and configure access controls under the principles of least privilege and separation of duties.
- Regularly audit logs to ensure new accounts are legitimate users.

Source: FBI.gov

He says: "Fundamentally, there are only a few reasons why a company is targeted.

"[Deepfake] is just another tool in the toolkit that can be used to manipulate someone to handing over control or handing over codes or anything that we may have seen before on phishing emails or social engineering.

"This is a far better form of social engineering, just using technology rather than you ringing up yourself."

Moore adds that fraudsters being able to use software to make it sound and look like someone else is a "huge worry" for businesses and organisations in the future as this could lead to people giving away trade secrets or private information to the wrong person.

Security

One thing that organisations can do to protect themselves from the deepfake technology is to invest in software that can detect if a video or recording has been manipulated, but that also comes with challenges.

Moore says: “Most videos are manipulated. Every video or picture you’ve ever seen, especially in advertising, will have been manipulated in some format. It’s very difficult to say ‘was it actually a deepfake?’.”

This is especially worrying as an increasing number of organisations started to move towards video verification to “identify or verify” someone to fight the increase in phishing attacks.

According to Allnutt, if you give “absolute trust” to video verification, it will be exploited by deepfake technology so it is important to not put “all your eggs in one basket” when it comes to verification and always have a second method.

He says: “This is slightly theoretical, but sometimes you get a cyber-attack via email convincing someone to transfer an unexpected payment to a new bank account. If you were to receive such an email for many the first port of call would be to try and ring up the person and you would enhance that by also carrying out a video call. But fraudsters could get the Linked In picture of the person they are pretending to be, together with some audio voice from a phone call or a You Tube video and create a deepfake to undermine the verification process.”

One thing that policyholders could do is ensure that there is a healthy relationship within their teams, where staff are not afraid to verify the authenticity of the request from their superior.

Raf Sanchez, head of cyber services at Beazley, explains: “You can get a personal invitation to speak to your boss, you think ‘okay maybe I’m getting a pay rise’ and before you know it you’re being asked ‘can you send me the login or one time password so I can log in and check something on the HR platform’.

"They will say 'it's very urgent and they will do it at 8.30am or 6pm and they will use triggers to get you to act quickly and urgently.

"But if my boss calls me at 6pm with an urgent call I will say 'I'm really sorry but you've never done this before, I'm going to email your work email or going to call your work number and speak to you on that number?'

"If I am so scared of my boss that I will do anything they say, I am more likely to be scammed than someone who has a healthy relationship with their manager and can say 'I don't know what's going on here but I will call you on a number I know is yours'."

Sanchez urges that organisations engage with their cyber insurers and take advantage of the training and services they provide.

He says: "We do not want to wait for our clients to have a problem to call us up. We want to get involved with our client's journey to a resilient posture, but we are an insurance company and we are not sitting alongside the clients' employees as they take calls or they interact with their clients or their colleagues.

"We are somewhat removed but we are absolutely trying to engage with clients and help them solve these problems."

Sanchez adds that Beazley offers its cyber insurance policyholders discounts on anti-phishing training, which he believes is one of the most useful tools because it is a risk people have to be "constantly reminded about".

He continues: "We are not sitting back and hoping that [policyholders] don't make a claim, because, frankly, that's not a good long term business model.

"We have to be there with them, helping them and that requires engagement on both sides and we hope that they can see us as delivering value to them whether or not they have an incident."

Policies

Whether deepfakes are covered under a policy will depend on the wording and breadth of the cover according to Andrea Garcia Beltran, UK and international head of cyber underwriting at RSA.

She says: “My recommendation is that [policyholders] need to be clear about their exposure, they need to work with their insurance broker and do a gap analysis to be able to negotiate a tailor-made policy, but they need also to be mindful that not all types of deepfakes attack will be addressed by cyber insurance policy. They might actually see that there are different attacks that might be covered by different policies.”

Policies that could be triggered by deepfake attack include directors’ and officers’ insurance, political risk insurance and reputational risk policies.

But when it comes to tailored policies for deepfake attacks, according to Erin Keneally, director of cyber risk analytics for Guidewire Software, insurers are still in early stages of addressing this technology so policy language “most likely won’t be tailored specifically to deepfakes, but existing policy language that addresses fraud coverage where cyber is the cause of loss may be sufficient”.

She continues: “As to the extent that insurers encounter unexpected cyber exposures from deepfake fraud attacks, they will be tightening language and looking to measure and model this risk in order to more appropriately price coverages for this risk.”

This will, however, vary from insurer to insurer, for example, Sanchez says Beazley’s cyber insurance covers any type of cyber attacks.

He says: “Our policies cover a data breach or security incidents. And a security incident is any failing in security.”

He continues: “The triggers are wide. You don’t have to specify phishing or smishing or deepfakes. It’s covered. It’s a scam that is intended to obtain [policyholders] details. It’s covered, no wording required, no changes required, and that’s the benefit of cyber specific wording.”

Some insurers will also have specific requirements for their policyholders. Garcia Beltran says that RSA encourages their policyholders to adhere to guidelines and stay alert to cyber attacks, this includes password management and training for phishing attacks.

She adds: “Most importantly we look into the culture, the savvy culture of the company or the organisation that must start at the top, so from the board of directors, going all the way down.

"[We encourage] that they are taking FBI considerations into account [see box], and that they have a proper business continuity plan that includes different attacks, that they have done their business impact assessment so that they understand if something happens, how rapidly they recover and that they understand the impact and that incident response plan that has been also tested."

Garcia Beltran adds that companies should embed insurance notification into their incident response plan as "we believe the first 48 hours are critical".

Attacks

Farish says it is important to remember that deepfake is not "necessarily bad or evil".

She explains: "For example, the film industry is using deepfakes in really exciting and interesting ways. We actually have a client that uses deepfake technology and is working with Hollywood Studios. So if one were to, for example, say 'well actually deepfake and manipulated media is covered under insurance policy', you would have to have an asterisk next to it saying 'but only the bad kind' and then you would need to define what that actually means."

She adds that DAC Beachcroft has not "seen a lot of activity in insurance space just yet".

Farish points out that since 2017 the rate of detected deepfakes has doubled every six months, with 90% of deepfakes detected being of "sexual nature" and targeting women, with less than 4% targeting c-suite executives or politicians.

She says: "It's important to see that the insurance industry is paying attention to this area but at the moment, for the time being, it's still predominantly used as a way to harm women.

"I do genuinely believe that in the months and years to follow, we're going to see a much broader form of targeted attacks in which business leaders or politicians will be targeted much more regularly. But at the moment over 90% of deepfakes are sexual in nature and target women."

What is important to remember, according to experts, is that deepfake attacks are not new, and are an evolution of phishing attacks with the aim to acquire an identity that can be used in future attacks.

Wawrzona says: “We don’t give bad actors a lot of credit, they are thinking ahead with [technology], perfecting it before we’re even thinking about it. When that perfection comes up and it’s discovered, that’s when we start to catch up. We are always playing a step or two behind them.”

MARK AS READ