ACC DOCKET
INFORMED. INDISPENSABLE. IN-HOUSE.

As Deepfakes Get Deeper, Security Risks Heighten

Technology, Privacy, and eCommerce

Cheat Sheet

- **New threat.** Deepfakes, a combination of misinformation and cyberattacks, undermine the integrity of data.
- **Faux appearance.** Since deepfakes use AI technology to overlay false information on existing content, it's difficult to identify and resolve issues companies and individuals face.
- **Business breaches.** Deepfakes have the potential to increase fraudulent transactions.
- **Mitigating deepfakes.** Companies should plan ahead and factor in consistent cybersecurity measures.

In the war of misinformation, content is a weapon. With the rampant existence of “fake news,” it has never been more difficult to discern between what is real and what is fake.

The other war we’re fighting is against cyberattacks on our information systems by nation-state actors and cyber criminals. Independent of each other, misinformation (aka disinformation) and cyberattacks are a threat in their own right. Now, a new social engineering attack has emerged that combines aspects of both: deepfakes.

There is no single legislation governing information security in the United States; rather, it is

regulated by a patchwork of industry and data-specific state and federal laws.

These information security frameworks are generally based on the “CIA” triad, referring to the three fundamental design components of information security:

1. **Confidentiality** — which protects against unauthorized access of data;
2. **Integrity** — which protects against unauthorized modification or alteration of data; and
3. **Availability** — which protects against the inaccessibility and discontinuity of data and ensures availability to authorized parties when and where needed.

Or as one cybersecurity thought leader explains, the CIA triad refers to keeping data “secure,” “clean,” and “accessible.” To date, threat actors and cyber criminals have focused on the confidentiality component of the CIA triad.

Deepfake attacks differ from your average cybersecurity attack, though, in that they attempt to violate the integrity of information — the lesser-understood component of the CIA triad. Rather than exfiltrating information or holding it ransom, deepfake attacks attempt to share it with the victim or leave it in place.

What are deepfakes?

Deepfake, the term a portmanteau of “deep learning” and “fake,” describes an increasingly broad category of digitized content created using artificial intelligence.

In a recent report, the European Parliament defined deepfakes as: “Manipulated or synthetic audio or visual media that seem authentic, and which feature people that appear to say or do something they have never said or done, produced using artificial intelligence techniques.”

All deepfakes share certain key characteristics: the use of artificial intelligence (through a type of algorithm called a generative adversarial network or GAN), automated creation and the potential to deceive.

Researchers have been developing new, innovative, and sometimes abusive technologies at a rapid pace.

In the last few years, the technology has advanced from requiring hours of processing time, hundreds of photographs or video samples, and a computer with somewhat advanced graphics capabilities to something that can be created on your cell phone with a single selfie. Recent reports describe algorithms that can create deepfakes in real-time for streaming video.

The technology has potential for positive application. In the entertainment industry, for example, it can give independent producers with lower budgets some of the same capabilities as the major motion picture studios. It can help match an actor’s mouth and facial movements to dialogue in foreign-dubbed films.

AI-generated audio can be used in digital assistant devices or to allow those who have lost their ability to speak to communicate in their own voice. Each of these use cases requires the consent of the person depicted, particularly in the case of professional performers, whose likenesses and voices are key to their livelihood.

> The technology also has significant potential for abuse when individuals' likenesses and voices are used without consent.

However, the technology also has significant potential for abuse when individuals' likenesses and voices are used without consent. This ranges from non-consensual sexual content that can be, and has been, weaponized against women, to doctored video of political figures, disinformation campaigns, or fraud.

Many deepfakes are pornographic in nature and frequently nonconsensual. Nearly all of these pornographic videos — approximately 90 percent — depict women.

Consequently, most of the legislation addressing deepfakes has dealt with non-consensual sexual content, typically in a civil context although some laws have added it to criminal invasion of privacy statutes (N.Y. Civ. Rights § 50-f, 52-c, CA civ code 1708.86, Hawaii SB309). At the federal level, legislation has been limited and largely focused on studying the threat deepfakes pose, primarily from a national security context.

There has been far less attention paid to audio deepfakes and the threats they pose. The same machine learning techniques can be used to create fake audio, replacing the time-consuming process of rearranging and combining sound fragments manually. Already, convincing voice facsimiles can be created with relatively small audio samples of an individual's speech.

Audio deepfakes have been gaining prominence since a mid-2019 report by cybersecurity firm, Symantec, of at least three incidents in which audio deepfakes were used to scam corporate finance officers into transferring large sums of money.

Since then, there has been growing attention on synthetic audio, generally, and audio deepfakes more specifically. In January 2020, the Federal Trade Commission (FTC) held its first public workshop on audio cloning technology, including discussions of its ethics, its pros and cons, and authentication, detection and mitigation.

> There has been growing attention on synthetic audio, generally, and audio deepfakes more specifically.

The workshop discussed the beneficial uses of the technology, which have driven much of their development, including their use in customer service and in creating synthetic voices for people who lose their ability to speak. It also examined how deepfakes have simplified communication-based crimes. Historically, these have been difficult to pull off; now, as the technology improves, these types of crimes will continue and proliferate.

Rapper Jay-Z made headlines last year when his company, Roc Nation, attempted to use YouTube's copyright takedown procedures to remove videos created by user Vocal Synthesis, who used audio deepfake technology to recreate Jay-Z's voice. The channel features AI-generated audio tracks of famous individuals — including past presidents, singers, and actors — reciting excerpts from a wide variety of works.

An analysis of legal issues with videos like those created by Vocal Synthesis is beyond the scope of this piece; however, there does not appear to be any viable basis for a copyright infringement claim

with audio deepfakes, which are wholly synthetized new works. While the AI algorithms do utilize copyrighted works to analyze voice patterns and create new works, there is little indication that any of the exclusive rights protected by copyright have been infringed in this context.

Deepfakes as threats to organizations

Across industries, deepfakes pose a far more insidious threat to the cybersecurity landscape than confidentiality and availability attacks.

In March 2021, the Federal Bureau of Investigations (FBI) warned against a rise of audio deepfakes. In particular, the FBI warned that foreign and criminal cyber actors use deepfake technology to create highly believable spearphishing messages — a targeted communication, typically an email, purportedly from a legitimate source aimed to persuade a specific individual to share or unwittingly allow access to personal or sensitive corporate information.

It is expected that threat actors will supplement voice spearphishing attacks with audio deepfakes and even utilize email voicemail attachments to install malware to acquire credentials or meet other objectives.

Similarly, the FBI warned of Business Identity Compromise (BIC) — a "newly defined cyberattack vector," which has evolved from Business Email Compromise (BEC). BIC use audio deepfakes to develop "synthetic corporate personas" or imitate existing employees to elicit fraudulent fund transfers.

There have already been reports of successful BIC. In 2019, a UK CEO was defrauded by a voice fake that convinced him to transfer €220,000 (approx. US$243,000) to a Hungarian supplier's bank account. He believed the request came from the parent company's chief executive and did not grow suspicious until the third call during which certain facts relating to the transfer did not line up. The threat actors are believed to have used commercial voice-generating software to carry out the attack.

> In 2019, a UK CEO was defrauded by a voice fake that convinced him to transfer €220,000 (approx. US$243,000) to a Hungarian supplier's bank account.

In 2020, a bank manager in Hong Kong was defrauded into transferring US$35 million by a voice he recognized to be the director of a company based in the UAE. The deepfake call was supplemented with emails to the bank manager from the purported director confirming the transfer.

The potential use of audio deepfakes are not limited to spearphishing attacks or BIC. Video deepfakes have already been used to bypass facial recognition technology, and it is only a matter of time until audio deepfakes bypass voice recognition technology. This biometric spoofing presents a threat to banks and other organizations that use voice recognition technology for identity verification over the phone.

Organizations can expect potentially severe and widespread reputational harm to result from deepfake attacks. Organizations can also expect significant financial harm not only in the form of fraudulent transfers or cybersecurity ransom demands, but in the transfer of certain assets or sensitive corporate information.

High-profile employees or those with decision-making authority who have video and audio that are

publicly accessible will be at highest risk. Defamatory or reputation-damaging deepfakes could also be used to extort or blackmail high-profile employees until an organization meets a threat actor's demands for payment or access to sensitive corporate information. What may have seemed far-fetched only a few years ago is now a reality.

Mitigating the harm

Though the organizational threats posed by deepfakes are evolving, if not inevitable, they are at least reasonably foreseeable. It is essential that organizations prepare for these attacks by creating and implementing administrative and technical safeguards.

Organizations should update their cybersecurity incident response plan to include a section designed to promptly respond to, and recover from, any deepfake attacks, with a special emphasis on the incident classification and escalation procedures for such attacks. The addition of deepfakes to the plan should also provide a plan for external and internal communications and information sharing since the reputational harm that can result from deepfakes can be severe.

Organizations should periodically conduct cybersecurity risk assessments to identify and mitigate potential security risks. In relation to deepfakes, organizations should assess how the confidentiality, integrity, security and availability of their information systems and personal and sensitive information would be impacted if the voice of their executives were fraudulently used to initiate authorize large payments.

Similarly, organizations should update their written information security programs (WISPs) to account for deepfakes, addressing verification of data sources and credibility. For example, organizations should have a procedure for authenticating wire transfers or other requests through traditionally secure contexts such as phone calls or video conferences.

> Organizations should have a procedure for authenticating wire transfers or other requests through traditionally secure contexts such as phone calls or video conferences.

Employee training should also include the specter of deepfakes as cyber threats, how they happen and how to spot them. Training should include how to report deepfakes or other phishing attempts, as well as a communications continuity plan in the event that email or other forms of communication are compromised.

Organizations should also update cybersecurity risk assessment protocols to account for deepfakes. Any technical safeguards should focus on preventing and detecting deepfake attacks through alternative verification methods such as water-marking audio content and the integration of media authentication tools into their electronic systems.

The risk is growing

As deepfake technology becomes more accessible and easier to use, the risk it poses to organizations grows exponentially.

Audio deepfakes have already been weaponized to scam organizations of millions of dollars and there is no doubt that the technology will continue to evolve. Organizations should revisit their

security practices in light of foreseeable risk.

Danielle S. Van Lier

Assistant General Counsel, Intellectual Property and Contracts

SAG-AFTRA

Danielle S. Van Lier is assistant general counsel, intellectual property and contracts at SAG-AFTRA. She recently earned an LL.M in Innovation, Technology and the Law from the University of Edinburgh where her dissertation focused on the threats posed by deepfakes.

Romaine Marshall

Partner

Armstrong Teasdale

Romaine Marshall is a partner and trial lawyer at Armstrong Teasdale. He has represented clients in response to hundreds of incidents involving data breaches, ransomware, malware attacks, security misconfigurations, wire fraud, software vulnerabilities, social engineering and other exploits, and in resulting litigation and regulatory investigations.

Katherine Bravo

Katherine Bravo was most recently an associate at Armstrong Teasdale with experience advising clients on state, federal and international privacy and data security laws, including in the compliance, enforcement and incident response context. Her prior experience includes working as a law clerk at the Federal Trade Commission in the Division of Privacy & Identity Protection.