BANK OF AMERICA INSTITUTE

BANK OF AMERICA

Transformation

When seeing is no longer believing: The dangers of deepfakes

10 July 2023

Key takeaways

- Today's cyber landscape provides ample opportunities for criminals to target consumers and businesses alike. Deepfakes - videos, audio, photos and text that are created using artificial intelligence (AI) - are one such threat as they are extremely hard to differentiate from authentic media.
- While there are legitimate uses for deepfakes, they can also put businesses at risk for exploitation and fraud, and pose a significant threat to authentication technologies. In fact, two out of three cybersecurity professionals saw malicious deepfakes used as part of a strike against businesses in 2022, a 13% increase from the previous year.
- We identify proactive steps that organizations can take to protect their businesses from deepfakes, including tightening identity verification protocols and incorporating deepfakes into incident response training and cyber awareness education.

Today's threat landscape

As discussed in Cybersecurity: Landscape, impact and what comes next, cyberattacks are becoming more frequent and expensive as trends such as digitalization, hybrid work, and the transition to the public cloud increase the potential for a cyberattack across organizations' networks. So, as the cyber landscape rapidly evolves, vigilance, innovation and knowledge sharing are more vital than ever. Businesses want their clients to trust in their ability to prepare, prevent, detect, mitigate, respond to and recover from information security threats and risks. As part of that, it's important for businesses to have the most current information available, understand the latest cyber threats and have a plan in place around protecting their employees and their companies.

Today's threat landscape continues to evolve, but common threats include:

1. **Imposter scams** – Whether through a phone call, text, email or video, imposter scams are becoming increasingly common and include:
    - **Deepfakes** – Fake videos, images and even audio generated by artificial intelligence (AI) are typically used maliciously, or to spread false information.
    - **Phishing and smishing** – Email and text messaging scams can be effective on mobile device screens that may cut off key message details.
    - **Vishing** – Cyber criminals use tactics such as pretending to be a trusted source and robocalls with urgent messages to scam people out of data and money.
2. **Malware** – Spyware, banking malware, ransomware and adware can be designed to target personal and company-issued phones and weaknesses in mobile apps.
3. **Compromised applications** – Criminals may access backdoors and weak app encryption to bypass a user's login credentials.
4. **Network spoofing** – Cyber criminals may set up fake Wi-Fi hubs to pry passwords and personal details from unsuspecting travelers.
5. **Data leakage** – Data can be lost or compromised by sending a message to the wrong recipient, or falling for spoofing, phishing or smishing attempts.
6. **Investment scams** – Scammers push for a financial investment with a promise of a payout, quick money or guaranteed returns. Such scams include cryptocurrency related investments and real estate investments, among others.

While all cybersecurity issues have the potential to significantly disrupt or undermine the credibility of an organization, in this piece, we delve into deepfakes, which are becoming increasingly common and convincing, and proving to be a major issue when it comes to eroding trust.

Deepfakes at a glance

Deepfakes, a type of synthetically modified media used to impersonate real humans, are one of the most effective and dangerous tools of disinformation. As they are increasingly being used to imitate executives and target organizations, it's important that organizations learn how to shield their companies from this emerging threat as discerning between legitimate and inauthentic content online has never been more difficult than it is today.

Manipulation of digital media has blurred the lines of reality, making it easier for cyber criminals to target and deceive individuals and businesses. In the past, media such as video, audio and photos could only be modified manually by humans with sophisticated editing skills and/or software programs. Today, digital video, audio, images and even text can be created and reshaped using artificial intelligence (AI). And while there are legitimate uses for this type of synthetic media, it is more often deployed in disinformation campaigns seeking to subvert the truth, which can damage an organization's reputation or include fraudulent requests for payment.

Deepfakes are digital content created or modified using a subset of AI known as deep learning, or deep neural networks. Deep learning algorithms are generated by a mesh network of computers that sync like a human brain to produce video, audio, photos or text depicting a fake event or spoofing an identity. Many deepfakes are convincing enough to fool most viewers, and the technology has already been successfully abused by cyber criminals for financial gain. In fact, while deepfake technology is still fairly new (first developed in 2017), deepfakes have been called one of the most dangerous AI crimes of the future.[1]

Legitimate use for deepfakes

Deepfakes are a dangerous tool in the disinformation arsenal. However, they're unlikely to be outlawed, as there are several legitimate applications for the technology. The biggest adopter of legitimate deepfakes so far is Hollywood, where the technology has been used to restore an actor's vocals, improve foreign-language dubbing, age down actors in flashbacks, or even complete works after an actor has died or retired. In addition, deepfakes have been used for educational purposes. In St. Petersburg, Florida, for example, the Dalí Museum has a deepfake video of the surrealist painter Salvador Dalí introducing his art and taking selfies with visitors.[2]

Deepfakes can also be used for legitimate corporate purposes, as a time- and cost-saving measure. For example, instead of sitting through multiple takes, a busy executive could record a video once and let deepfakes make rapid corrections. In addition, a recorded video can be translated and localized in multiple languages using deepfakes. The technology can even be used to create completely synthetic characters, replacing costly actors for professional development courses or commercials.

Deepfake risks for companies

While deepfakes can be used for sanctioned business content, organizations must acknowledge their inherent risks. In 2021, the FBI issued a warning to businesses about deepfake fraud, saying that malicious actors "almost certainly will leverage synthetic content for cyber-crime and foreign influence operations in the next 12-18 months."[3] In fact, two out of three cybersecurity professionals saw malicious deepfakes used as part of a strike against businesses in 2022, a 13% increase from the previous year, with email as the top delivery method.[4]

Because well-crafted deepfakes require high-end computing resources, time and technical skill, cyber criminals typically use them for operations against large enterprises and demand steep payments — but as technologies evolve to make deepfakes easier and cheaper to create, criminals will be able to target companies (or third-party vendors) of all sizes.

Business identity compromise risks

In 2020, threat actors used an audio deepfake to steal $35 million from a Hong Kong bank, the largest publicly disclosed amount lost to inauthentic content yet.[5] They pulled off the sophisticated heist using a newly defined threat vector called business identity compromise (BIC).

BIC uses deepfake technology to create synthetic corporate personas or imitate existing employees, often posing as a well-known, high-ranking professional in the organization. Put simply, BIC builds trust where there shouldn't be. Once it's established, criminals can seize trade secrets and patents, impact company culture with political commentary, undermine relationships with customers and partners, tank stock values, create turmoil in the supply chain and otherwise sow chaos.

---

[1] University College London, "Deepfakes' ranked as most serious AI crime threat," August 2020.

[2] Dami Li, The Verge, "Deepfake Salvador Dalí takes selfies with museum visitors," May 2019.

[3] FBI, "Malicious Actors Almost Certainly Will Leverage Synthetic Content for Cyber and Foreign Influence Operations," March 2021.

[4] VMWare, "Global Incident Response Threat Report," August 2022.

[5] Thomas Brewster, *Forbes*, "Fraudsters Cloned Company Director's Voice in $35 Million Bank Heist, Police Find," October 2021.

Both audio and video deepfakes have already been used to impersonate executives at Fortune 500 companies, including CEOs, CFOs, treasurers and other senior leadership. In some cases, deepfakes have been deployed to humiliate or harass the executive, making it seem like they said or did something that they did not, in order to damage the reputation of the company and executive.

Deepfake phishing risks

Deepfake phishing is another emerging threat for businesses, combining disinformation (in the form of deepfakes/BIC) and phishing to fool employees into making unauthorized payments or volunteering sensitive proprietary or customer information. Often, deepfake phishing begins with an audio deepfake of a trusted figure in the organization. The criminal, disguised as the figurehead, reaches out via web conferencing or voicemail, then follows up with other forms of social engineering, such as business email compromise (BEC) or dynamic voice manipulation, using a sense of urgency to pressure employees into releasing funds or data.

Authentication risks

Finally, deepfakes pose a significant threat to authentication technologies, including facial recognition and voice recognition. Researchers at the University of Chicago found that AI-generated deepfake voices were able to fool three popular real-world voice recognition systems.[6] Similarly, studies have shown that some deepfake-generating techniques have been able to trick common web-based facial recognition APIs[7] and even certain types of technologies used to unlock smartphones, though smartphones[8] that use three-dimensional mapping as part of facial recognition are not yet vulnerable to two-dimensional deepfakes.

As deepfake technology becomes more readily available, organizations with less sophisticated security capabilities — and fewer awareness and mitigation policies around deepfakes — will be at greater risk. As mobile and cloud-based deepfake applications further penetrate the market, criminal focus may shift from executives to high-net-worth individuals or professionals with access to critical infrastructure. If threat actors have access to enough input data, anyone from a power plant manager to a social media influencer could be targeted with deepfakes.

Quick take: Shallowfakes

Although true deepfakes require the use of deep learning or deep neural networks, shallowfakes (also called cheapfakes) are similar in concept but use simpler methods. For instance, they can consist of media presented out of context or doctored with simple editing tools, such as filters or airbrushing.

Shallowfakes are often used in disinformation campaigns, as well as in business scams. For example, criminals have hacked into video conference calls using previously recorded video with the sound turned off, so they could impersonate an executive while feigning audio problems. They then call back by phone, pretending to be the executive and request a wire transfer. Other examples of shallowfakes include the following:

- Audio/video speed — To cause reputational damage, scammers may slow down audio to make a speaker sound intoxicated or speed up video to make a person's actions look violent.
- Proof of identity or address — By using simple photo editing tools to doctor photo IDs, utility bills or bank statements, for example, documents can be manipulated to falsely prove identity or location.
- Documentation — Manipulated invoices, receipts or even photo evidence can be used to make fraudulent expense reports or insurance claims.

Deepfake detection

There are a few telltale signs of audio and video deepfakes:

- Audio: Listen for longer-than-usual pauses between words and sentences. The person's voice may also sound flat and lifeless. If it sounds off, it likely is.
- Video: Watch for long periods without blinking, patchy skin tones and poor lip syncing. Jawlines can sometimes reveal flaws, such as blurriness or flickering around the edge of the face. Ears may also appear a completely different skin color from the face.

However, as deepfakes evolve, it will become more and more difficult to distinguish them from authentic media. However, synthetic media created with rudimentary deepfake technology or by threat actors with lesser skills can still be detected by

[6] Emily Wenger *et al.*, University of Chicago, "Hello, It's Me: Deep Learning-based Speech Synthesis Attacks in the Real World," September 2021.
[7] Kyle Wiggers, VentureBeat, "Study warns deepfakes can fool facial recognition," March 2021.
[8] Jessica Hallman, Penn State, "Deepfakes expose vulnerabilities in certain facial recognition technology," August 2022.

human senses. Training in deepfake detection can improve the likelihood that employees will catch BIC and deepfake phishing attempts before they cause irreparable harm. Even if personnel are unable to pinpoint a specific flaw, they may experience the "uncanny valley" phenomenon, where the slight difference between a humanlike deepfake and an actual human causes discomfort or revulsion.[9]

Because deepfake threats against organizations are so new, there aren't many established technical solutions for detection or protection purposes. Those that do exist tend to be pre-profit startups that allow users to upload a suspected deepfake video, as well as photo and audio files to be analyzed for anomalies or traces of spoofing. As technology companies iterate on detection algorithms, they'll develop ever more robust models and systems to meet the need for more reliable solutions. Until then, people and organizations can trust — but verify.

Tips to mitigate deepfake risks

Deepfake incidents have only recently started targeting businesses. Here are proactive steps organizations can take to protect their businesses from deepfakes:

1. **Educate employees and partners** - Many professionals outside of cybersecurity may not have heard of deepfakes and will benefit from learning about different formats and likely threat scenarios. Show a variety of deepfake examples — legitimate and inauthentic, video and audio — to raise awareness of their risks. Additionally, leadership should reinforce their risk expectations and clarify when senior leaders can request payments and how employees can validate those requests. Remember, employees are an organization's first line of defense.

2. **Maintain cybersecurity best practices** - Cybersecurity best practices, especially those related to social engineering and fraud prevention, can help fortify companies against deepfake phishing and other disinformation campaigns.

    - Review foundational security policies with employees, especially how to spot relevant scams and sophisticated phishing techniques. This should include scrutinizing communications with skepticism and verifying their validity through secondary channels — especially those that ask for personal information or payments outside of usual billing structures, and that make such requests with insistence.

    - Empower teams to "pause" and raise a concern, so experts can validate or mitigate the item in question.

    - Provide positive reinforcement when employees spot fraud and prevent losses.

3. **Strengthen identity verification and validation protocols** - This should include strengthening login credentials and authentication methods, as well as adhering to the principle of "least privilege" — and to a lesser degree "trust, but verify."

    - Least privilege asserts that users should only be given access to those accounts deemed necessary to perform their jobs. Whereas trust, but verify can be applied upon suspicion of a false identity. Even if an employee knows the face on the screen or voice on the line, it's important to have a secondary confirmation channel.

    - If an employee receives a call asking for payments that exceed acceptable thresholds, they should be aware of their organization's clear exceptions process that requires verification — even if the request comes from the C-suite.

4. **Include deepfakes in incident response planning** - After developing deepfake awareness and reinforcing identity verification policies, businesses should incorporate deepfakes into their incident response planning. It's critical to publicly acknowledge and squash a deepfake (or shallowfake) as fast as possible — the longer a deepfake spreads without being addressed, the more people can be convinced, and the more believable it becomes. Tamping down disinformation quickly and succinctly in a press release and/or in messaging delivered through verified social media channels helps address media, employee and even vendor communications.

[9] Jeremy Hsu, *Scientific American*, "Why 'Uncanny Valley' Human Look-Alikes Put Us On Edge," April 3, 2012.

Disclaimer

Neither Bank of America nor its affiliates provide information security or information technology (IT) consulting services. This material is provided "as is," with no guarantee of completeness, accuracy, timeliness, or of the results obtained from the use of this material, and without warranty of any kind, express or implied, including, but not limited to, warranties of performance, quality and fitness for a particular purpose. This material should be regarded as general information on information security and IT considerations and is not intended to provide specific information security or IT advice nor is it any substitute for your own independent investigations. If you have questions regarding your particular IT system or information security concerns, please contact your IT or information security advisor.

Contributors

Vanessa Cook

Content Strategist, Bank of America Institute

Sources

Kristopher Fador

Chief Information Security Officer, Bank of America

Roland Chan

Cyber Crime Defense Executive, Bank of America

Disclosures

These materials have been prepared by Bank of America Institute and are provided to you for general information purposes only. To the extent these materials reference Bank of America data, such materials are not intended to be reflective or indicative of, and should not be relied upon as, the results of operations, financial conditions or performance of Bank of America. Bank of America Institute is a think tank dedicated to uncovering powerful insights that move business and society forward. Drawing on data and resources from across the bank and the world, the Institute delivers important, original perspectives on the economy, sustainability and global transformation. Unless otherwise specifically stated, any views or opinions expressed herein are solely those of Bank of America Institute and any individual authors listed, and are not the product of the BofA Global Research department or any other department of Bank of America Corporation or its affiliates and/or subsidiaries (collectively Bank of America). The views in these materials may differ from the views and opinions expressed by the BofA Global Research department or other departments or divisions of Bank of America. Information has been obtained from sources believed to be reliable, but Bank of America does not warrant its completeness or accuracy. Views and estimates constitute our judgment as of the date of these materials and are subject to change without notice. The views expressed herein should not be construed as individual investment advice for any particular client and are not intended as recommendations of particular securities, financial instruments, strategies or banking services for a particular client. This material does not constitute an offer or an invitation by or on behalf of Bank of America to any person to buy or sell any security or financial instrument or engage in any banking service. Nothing in these materials constitutes investment, legal, accounting or tax advice.

Copyright 2023 Bank of America Corporation. All rights reserved.