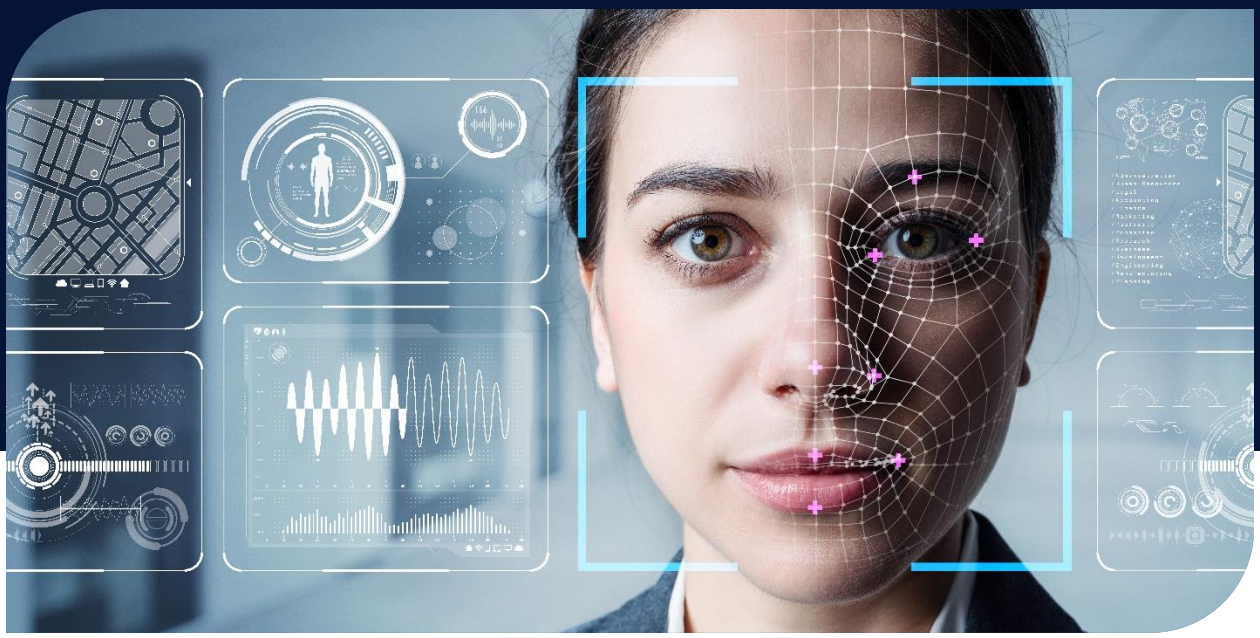
Teneo®

Deepfakes in 2024 are Suddenly Deeply Real:

An Executive Briefing on the Threat and Trends

April 2024



Deepfakes arrived on the mainstream scene around 2019, quickly bringing us bogus celebrity videos and audio in the form of Nancy Pelosi, Tom Cruise and even Keanu Reeves – mostly for laughs and entertainment at the time.

Since widespread access to and adoption of OpenAI's ChatGPT in 2023 and newly emerging generative technologies, the ability to fast-track development of hyper-realistic deepfakes has given threat actors a new tool for potential widespread fraud, manipulation and disinformation – and global organizations are taking notice.

As a result, the deepfake threat has quickly moved to the top of the risk agenda. While celebrities, politicians and high-profile executives were the original "targets" of deepfake attacks, threat actors are now also targeting everyday businesses and individuals using disinformation and fictitious media for political and criminal purposes.

The following article provides a brief history on the origins of deepfake technologies, the latest criminal developments, emerging regulation and importantly, considerations for executives and organizations.

While a deepfake attack is a cyberattack, preparedness and response are uniquely different.

How it Started

In 2014, a 27-year-old American graduate student and computer scientist named Ian Goodfellow invented a breakthrough in artificial intelligence (AI) called generative adversarial networks or "GANs." A GAN is a deep learning architecture that uses neural networks (networks which allow programs to recognize patterns) to compete against one another to create new and realistic data. That data can come in the form of text, images, video or audio, and is becoming increasingly more sophisticated in duping humans.

First, the quick technical breakdown. These networks consist of two algorithms: a "generator," which uses tens of thousands of training data points to create hyper-realistic audio, image or video content; and a "discriminator," which provides a quantified assessment of the likelihood of the media being real or fake.

Let us imagine that you are a moderately gifted artist and your child has asked you to draw a picture of the cartoon animal from their favorite kid's show. You look at 10 pictures of the cartoon character from the show's official website and try to draw it, but your child says it isn't exactly right. You look at 10 more pictures of the cartoon character and try again. Your discerning child says it is better, but not exactly what they were looking for. Frustrated, you examine 10 more pictures and watch a few minutes of the cartoon on your phone and then put the finishing touches on version three of the character. Finally, your child exclaims, "this looks exactly like the pictures! I can't tell the difference!" Deepfake accomplished.

In 2017, a purported Reddit moderator created the subreddit r/deepfakes, thus coining the term, a likely portmanteau of "deep learning" and "fake" media. The Reddit moderator used face-swapping technology, digitally manipulating photos by superimposing the faces of well-known women on body images that were not their own. In certain circles, these developments brought the concept of deepfakes "mainstream." Yet it wasn't until 2019 when more of society began paying attention as a result of the U.S. Speaker Nancy Pelosi deepfake, followed by the 2021 Tom Cruise videos, which garnered laughs but an increasing sense of nervousness about the developments in the technology and its capabilities. By 2023, the Internet reveled in celebrity deepfakes targeting Morgan Freeman, Pope Francis in a puffer jacket and Donald Trump in handcuffs. So far, in 2024, an audio deepfake of President Biden, deepfakes targeting Taylor Swift and renewed use of Trump deepfakes have made the rounds, all to disseminate disinformation and/or advance political agendas. These notable deepfakes are quickly turning the conversation toward the role of social media platforms and regulators in content moderation, disinformation detection and development of guardrails for the ethical use of AI.

In the business world, the deepfake "use case" for disinformation, however, may soon be overshadowed by the emerging and significant threat of deepfakes for financial gain at the hands of criminals. Executives and the organizations they lead are now faced with a new vector for fraud, theft and manipulation.

The Latest Targets

Elections

In a historic election year, where at least 64 countries will take to the polls, concern over the spread of disinformation to potentially manipulate or influence election outcomes is taking center stage. The threat of deepfakes is also spurring global governments and regulatory agencies into action. In the U.S. and around the globe, social media is becoming the playground for the dissemination of deepfakes, depicting politicians or would-be politicians doing or saying things to influence voters or smear reputations for political objectives. Social media platforms like Facebook and X – as well as partners in their ecosystem like OpenAI, Google and Shutterstuck – have stepped up their policies. These include adding watermarks on AI generated images, identifying and labeling watermarked and AI generated content, as well as takedowns of damaging content/disinformation. For now, the changes will help, but won't be a categorical fix to identify and label all content given the sheer breadth of content across the Internet.

While social media platforms attempt to do their part, traditional media outlets and consumers of digital media also have a role to play in mitigating the spread of disinformation. We all need to be vigilant and discerning by fact-checking multiple credible sources before liking, sharing and reposting in this highly important and unprecedented election year.

High-profile executives

An unfortunate byproduct of an elevated profile, visibility and authority is advanced risk. The aforementioned examples highlighted showcase celebrities and politicians, but threat actors are also coming after ultra-high net worth (UHNW) individuals and business executives using deepfakes as impersonation tools for financial gain. In March 2019, one of the first widely reported executive deepfakes made global headlines when fraudsters used AI to impersonate a CEO. The chief executive of a UK-based energy company believed he was on the phone with the German CEO of his parent company and took instructions to transfer $243,000 to a Hungarian bank account of what the fraudster said was a supplier's bank account. The voice deepfake was so realistic that the UK executive recognized the German accent and the specific "melody" of his boss's voice.

In August 2022, Binance Chief Communications Officer Patrick Hillman warned that threat actors used television appearances and online interviews of him to create a deepfake video which was then deployed over Zoom to trick crypto project teams. In February of this year, a finance employee based in Hong Kong working for a multinational firm joined a video conference call with a group of individuals whom he believed to be the CFO and members of the company's staff. The characters on the call were instead part of a deepfake video so convincing that the employee agreed to transfer $25 million at their direction.

While these examples seem few and far between now, the pace at which the technology is advancing would indicate that executives will continue to be targets of opportunity, with unwitting staff increasingly vulnerable as the deepfakes become more and more realistic.

Financial institutions

Deployment of AI technologies has helped banks and other financial institutions make significant strides to enable 24/7 customer service, streamline formerly complex bank processes and provide more targeted product offerings. Yet criminals are now leveraging AI and deepfakes to commit identity theft, create documents and falsify identities to open bank accounts, obtaining unauthorized access to customer accounts, funds and illegally obtained loans. Similarly, criminals are creating fake audio and video to spear-phish (a targeted malicious email attack) and social engineer customers and employees into sharing personal or financial information, or, as was the case with the Hong Kong employee, transfer money to fraudulent bank accounts.

In March, the U.S. Treasury Department issued a comprehensive report covering cybersecurity risks in the financial services sector. The report notes that "these (LLM and deepfake) technologies not only lower the barrier to entry but also complicate authenticity and verification measures." The report also recommends that firms invest in defense and detection systems, specifically calling out the challenges associated with deepfakes and misinformation.

In a recent interview with Teneo, Ben Colman, Co-Founder and CEO of Reality Defender, a leader in deepfake detection technology said "the financial services sector relies heavily on and has invested billions in voice verification platforms over telephony systems, all used for everything from authentication to transactions to sending wire transfers. Now that anyone can realistically clone a real human voice with as little as 15 seconds of existing audio, these platforms have more or less been rendered useless in the face of deepfakes."

Regulatory Developments

European Union

On Tuesday, April 9th, a majority of the European Union's political parties, including the European People's Party (EPP), the Party of European Socialists (PES) and the European Conservatives and Reformists Party (ECR), signed on to a code of conduct in preparation for the June 6-9 European Parliament elections. The objectives of the code of conduct are to support and promote fair campaigning, election integrity and resilience. While the code does not specifically use the word "deepfake," abiding parties agree to "abstain from producing, using or disseminating misleading content, in particular, any type of deceptive content using audio, images or video and generated with or without artificial intelligence to falsely or deceptively alter or fake candidates, officials or any electoral stakeholder."

In addition, through the EU Digital Services Act, the European Commission plans to issue fines to social media platforms (such as X and Facebook) for failure to adequately moderate AI generated disinformation. As part of the proposed regulation, platforms will be expected to organize dedicated teams to work with cybersecurity agencies across the EU's 27 member states in monitoring for problematic content.

United States

In February of this year, the U.S. launched a bipartisan task force to assess the potential for AI regulation. Yet with the U.S. Presidential election just seven months away, it is unclear whether appropriate federal regulation will be in place to mirror some of the advances made by the European Union. At present, states such as California, Florida, Georgia, Hawaii, Illinois, Minnesota, New York, South Dakota, Texas and Virginia have all passed legislation banning deepfakes, with penalties for dissemination of "synthetic media" without relevant disclosures including potential fines or jail time. Additional states including Alaska, Colorado, Massachusetts, Oklahoma, Nebraska, Arizona and Idaho have proposed similar bills to ban the spread of disinformation through deepfakes.

Also in February, the U.S. Federal Trade Commission (FTC) issued a notice seeking public comment on a "supplemental notice of proposed rulemaking" to the trade regulation rule on impersonation of government and business. This supplemental notice would add a "prohibition on the impersonation of individuals and extend liability for violations of the Rule to parties who provide goods and services with knowledge or reason to know that those goods or services will be used in impersonations" that violate the rules. In the same month, the Federal Communications Commission (FCC) issued a ruling restricting the use of "artificial or prerecorded voice" to encompass current AI technologies that generate human voices. This will have a direct impact on the creation and use of robocalls, such as the Biden robocall, which went viral leading up to the 2024 New Hampshire primary.

Asia Pacific

According to the Global Initiative Against Transnational Organized Crime, the APAC region experienced a 1530% increase in deepfake cases between 2022 and 2023. In the region, Vietnam saw the highest increase in deepfake fraud with 25.3%, followed by Japan at 23.4%. The Philippines felt the highest growth in deepfake cases at 4500%. In terms of regulation, China has led the charge. In January 2023, the Cyberspace Administration of China (CAC) began regulating "deep synthesis" technology to include AI generated images, text and audio, with the goal of preventing dissemination of information deemed disruptive to the economy or national security. The country also aimed to enforce the labeling of images that have been synthetically created or modified. Other APAC governments are examining potential rule making and enhanced procedures in support of fraud identification and mitigation.

These regulatory developments around the world have clearly evolved in response to the growing threat of deepfake attacks. However, oftentimes and depending on the jurisdiction, the time to market between threat identification, rulemaking and implementation lags behind technology advances and the exploitation of vulnerabilities. As such, executives will need to be proactive in determining how best to address and mitigate the deepfake threat to their business operations.

Considerations for Executives

Deepfake detection technologies

As deepfakes continue to advance in look and feel, it will become increasingly impossible for employees and customers to detect a deepfake with the naked eye or ear. Employee training and awareness is a major part of the puzzle, but organizations reliant upon voice, video or biometric verification as part of their business processes will need to implement detection and mitigation tools to get ahead of advances by threat actors. Deepfake detection platforms on the market today offer organizations the ability to upload images, text, video and/or audio files for analysis to determine whether content has been manipulated, with reports or scorecards for relative ease of review and interpretation.

Regarding his platform, Colman explained to Teneo that “we designed Reality Defender to work at the enterprise and government level, with the onus on these entities and not ordinary citizens and consumers in the detection of deepfakes.” While the onus is on the entity, the power of detection technology can give organizations a significant advantage against criminals when deployed as part of a holistic cybersecurity strategy.

Enhanced crisis management and crisis communications protocols to address deepfake attacks

While increased cyber risk in 2024 has created a renewed sense of urgency to update and test cyber crisis management and communications protocols, most have not contemplated a deepfake attack. The usual cyberattacks heretofore have included denial of service, malware, ransomware or phishing attacks. The attackers traditionally present themselves as amorphic, digital intrusions or, at best, as strange email addresses or ominous ransom notes. Deepfakes add a very real human element with faces, voices or text message structures that may be known to us – preying on familiarity, emotion, good will and trust.

Crisis management protocols will need to consider new scenarios where high-profile executives become the attack vector and employees or others fall victim to manipulation. Executive impersonation could have an immediate and highly detrimental impact to financials, operations and reputation, not to mention with stakeholders such as media, investors, employees and customers. The ability to move swiftly to verify a deepfake and communicate to the relevant stakeholder universe will be paramount.

Tabletops and crisis simulation exercises around deepfakes

Having updated crisis management and crisis communications protocols are only part of preparedness. It is imperative that organizations conduct tabletop exercises to test muscle memory around managing a crisis, considering in this instance that someone around the executive table may have been the “source” of the impersonation. Protocols, tabletops and simulations should test processes whereby the organization verifies and then clearly communicates that it has been the victim of a deepfake.

Employee training and awareness around social engineering

Most organizations have implemented business process and escalation protocols such as “four-eyes checks” or approval and override processes – all with the goals of preventing mistakes, mitigating fraud and achieving transparency and regulatory/compliance requirements. Employees need to be hyper-vigilant and aware of the potential for deepfake attacks and social engineering attempts by bad actors. There are simple and reasonable process steps we can all take to uncover potential deepfakes, including conducting additional gut check verifications. If concerned about identifying anomalous requests or social engineering incidents, organizations can develop frameworks and proprietary verification steps that align with the business, such as asking to see the room or office around the caller, hanging up and calling a colleague back on a known line or implementing a company code word.

Deepfakes are proving that, despite the advances in technology, nothing is more real than human interaction.

For more information, email: Resilience&Intelligence@teneo.com

Author

Courtney Adante
President, Security Risk Advisory
Courtney.Adante@teneo.com

Teneo
The Global CEO Advisory Firm

Teneo is the global CEO advisory firm.

We partner with our clients globally to do great things for a better future.

Drawing upon our global team and expansive network of senior advisors, we provide advisory services across our five business segments on a stand-alone or fully integrated basis to help our clients solve complex business challenges. Our clients include a significant number of the Fortune 100 and FTSE 100, as well as other corporations, financial institutions and organizations.

Our full range of advisory services includes strategic communications, investor relations, financial transactions and restructuring, management consulting, physical and cyber risk, organizational design, board and executive search, geopolitics and government affairs, corporate governance, ESG and DE&I.

The firm has more than 1,600 employees located in 40+ offices around the world.

teneo.com