

 BRIEF

SYNTHETIC MEDIA: NOTABLE ATTACKS

ZEROFOX INTELLIGENCE

B-2023-04-12a

Classification: TLP:CLEAR

Criticality: HIGH

Intelligence Requirements: Deepfake, Emerging Threats

April 12, 2023



Scope Note

*ZeroFox Intelligence is derived from a variety of sources, including—but not limited to—curated open-source accesses, vetted social media, proprietary data sources, and direct access to threat actors and groups through covert communication channels. Information relied upon to complete any report cannot always be independently verified. As such, ZeroFox applies rigorous analytic standards and tradecraft in accordance with best practices and includes caveat language and source citations to clearly identify the veracity of our Intelligence reporting and substantiate our assessments and recommendations. All sources used in this particular Intelligence product were **identified prior to 4:00 PM (DST) on April 5, 2023**; per cyber hygiene best practices, caution is advised when clicking on any third-party links.*

Synthetic Media: Notable Attacks

Executive Summary

As synthetic media creation tools become increasingly sophisticated, accessible, and affordable, threat actors will be more likely to utilize this technology to enhance and diversify existing threats. A growing body of evidence collected since 2020 indicates that financial firms are becoming prime targets of synthetic media-driven attacks. In 2022, the Federal Bureau of Investigation (FBI) explicitly warned about the potential deployment of deepfakes used to commit financial fraud, identity theft, and job-related scams.

In addition to the financial sector, other industries, including the cryptocurrency market, have faced similar attacks in which deepfakes have been used to impersonate influential figures and manipulate public opinion. Furthermore, nation-state actors have adopted deepfake technology for political subterfuge and propaganda distribution, while cybercriminals have exploited it for espionage and social engineering schemes.

The rising ubiquity of synthetic media has triggered significant concerns about its potential ramifications on democratic processes. Consequently, organizations such as the FBI, the Cybersecurity and Infrastructure Security Agency (CISA), and the World Economic Forum (WEF) have issued warnings and urged increased vigilance. To address this growing threat, numerous organizations are actively working on the development and improvement of tools to counteract deepfakes and safeguard against their malevolent use.



Key Findings

- Since 2021, cybercriminals have been increasingly using synthetic media to conduct more convincing social engineering campaigns used in financial fraud scams.
- Synthetic media technology has been used in various cryptocurrency scams to create deepfake videos of prominent figures in the industry, impersonate customer service representatives of cryptocurrency exchanges, and create fake news stories and social media posts that make it difficult for cryptocurrency investors to differentiate between real and fake information.
- Nation-state actors, including those from Russia and China, have utilized synthetic media technology for political manipulation and propaganda purposes, while hackers have used it for espionage and social engineering attacks; the use of synthetic media in such attacks increases their chances of success by making fake personas more convincing and difficult to detect.
- The increasing prevalence of synthetic media technology has raised concerns about its potential impact on political campaigns, democratic processes, and geopolitical landscapes; manipulated videos can be used to spread false information, influence public opinion, and even sway election results.
- The FBI, CISA, and WEF have issued warnings about the potential risks of synthetic media technology, including the use of deepfakes for financial fraud and identity theft, as well as the manipulation of public opinion and interference with democratic processes. Various organizations, including the Defense Advanced Research Projects Agency (DARPA), Microsoft, and Facebook, are developing tools and technologies to detect and combat deepfakes.

Use of Synthetic Media by Cybercriminals

The use of synthetic media in social engineering attacks and fraud scams—particularly in deepfake vishing schemes—allows attackers to impersonate legitimate decision-makers convincingly and extract sensitive information from unsuspecting victims in a less suspicious manner. Due to the novelty of deepfake technology, cybercriminals are likely to exploit the general lack of education and awareness surrounding it. Additionally, current authentication methods may be insufficient for detecting deepfakes, which further increases the risk of businesses falling prey to these highly convincing, fraudulent interactions.

Cybercriminals are increasingly utilizing synthetic media primarily to conduct more believable social engineering campaigns in financial fraud scams. In early 2020, a bank manager was socially engineered over the phone into transferring USD 35 million to an attacker-controlled account.¹ The attackers used deepfake technology to duplicate

¹

[hXXps://www.forbes\[.\]com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=388d1da77559](https://www.forbes.com/sites/thomasbrewster/2021/10/14/huge-bank-fraud-uses-deep-fake-voice-tech-to-steal-millions/?sh=388d1da77559)



the voice of the company's director, with whom the bank manager had previously spoken. The attackers claimed that the USD 35 million was needed immediately to close an acquisition. To further backstop their story, the attackers sent convincing fake emails from the director and a lawyer, whom the attackers claimed was hired to close the phony acquisition.

Cybercriminals have increasingly employed synthetic audio to mimic bank customer service representatives, deceiving victims into divulging their account information. The complexity of these attacks has advanced significantly, as evidenced by the FBI's warning in February 2022 that threat actors have utilized deepfake audio to execute business email compromise attacks through virtual meeting platforms.²

- These attacks involve the use of deepfake technology to synthesize a voice that sounds similar to a real bank employee and social engineering tactics, such as creating a sense of urgency or offering a reward, to convince the victim to provide their sensitive information.
- The attackers may also use caller ID spoofing or other methods to make the call appear to come from a legitimate source; as audio manipulation is easier than making deepfake videos, the frequency of these attacks is expected to increase.
- It is crucial for businesses and individuals to be aware of this new type of attack vector and adopt better authentication methods to prevent falling victim to these highly convincing conversations.

The rise of deepfake technology has also created new opportunities to engage in identity theft on a larger scale by synthesizing the voice of a trusted figure to gain unauthorized access to large databases of personal information. In one scenario, an e-commerce company official may receive a deepfake phone call that appears to come from an IT administrator asking for their login credentials. The deepfake is the first phishing attribute, luring the employee into providing the login information. Once the criminals have obtained access, they can engage in real identity theft on a larger scale in the second stage—potentially compromising the personal data of thousands or even millions of individuals.

Synthetic Media and Cryptoscams

Synthetic media technology has also been leveraged to manipulate cryptocurrency investors and users to steal access to capital and accounts. For example, cybercriminals created a deepfake video of Binance Chief Communications Officer Patrick Hillmann and used it in Zoom calls to scam cryptocurrency executives.³ The scammers built the deepfake based on Hillmann's interviews and TV appearances to fool crypto project

² [hXXps://www.ic3.gov/Media/Y2022/PSA220216](https://www.ic3.gov/Media/Y2022/PSA220216)

³

[hXXps://www.theverge.com/2022/8/23/23318053/binance-comms-crypto-chief-deepfake-scam-claim-patrick-hillmann](https://www.theverge.com/2022/8/23/23318053/binance-comms-crypto-chief-deepfake-scam-claim-patrick-hillmann)



representatives into thinking he would help their tokens get listed on Binance's exchange. It is unclear how many crypto projects were affected by the scam or how much the victims may have paid for the promise of a Binance listing. Although Hillmann notes that Binance adheres to stringent cybersecurity best practices, criminals are still trying to impersonate its workers.⁴

- Cybercriminals are developing AI-generated chatbots and voice bots to impersonate customer service representatives of cryptocurrency exchanges to trick victims into giving away their login credentials and private keys.

Moreover, some cryptocurrency scams have used synthetic media to create fake news stories, social media posts, and even entire websites to promote fraudulent investment opportunities. These scams can be highly convincing, using professional-looking websites and fake endorsements from celebrities and industry experts.

Overall, the use of synthetic media in cryptocurrency scams poses a significant threat to investors, as it can be difficult to differentiate between real and fake information. Investors should be cautious of investment opportunities that promise guaranteed returns, free money, or use celebrity endorsements and should only invest through reputable exchanges and platforms.

Nation-State Actors Manipulate Political Realities

Nation-state actors—particularly from countries like Russia and China—have been documented employing deepfake technology as a tool for political manipulation and disseminating propaganda, further demonstrating the potential impact of this technology to increase the risk of miscalculation in decision-making. For example, Russian intelligence agencies have been accused of creating deepfake videos to influence the 2020 U.S. presidential election.⁵ Chinese state media has also faced allegations of employing deepfake technology to manipulate public opinion and spread propaganda, likely with the objective of bolstering favorable views of the Chinese regime. For instance, a deepfake video that was circulated on Chinese social media in 2020 appeared to show Italian people applauding and thanking China for its aid during the COVID-19 pandemic. The video was later revealed to be a manipulation, as it was created using stock footage from a video-sharing platform and then edited to include pro-China messages. Another example is a deepfake video that showed then-U.S. Speaker of the House, Nancy Pelosi, appearing to slur her words and stammer during a speech.⁶ The video was shared widely on social media platforms and was seen as an attempt to discredit Pelosi and undermine the U.S. political system. The

⁴

⁴ [hXXps://www.binance\[.\]com/en/blog/community/scammers-created-an-ai-hologram-of-me-to-scam-unsuspecting-projects-6406050849026267209](https://www.binance.com/en/blog/community/scammers-created-an-ai-hologram-of-me-to-scam-unsuspecting-projects-6406050849026267209)

⁵ [hXXps://www.reuters\[.\]com/article/us-facebook-deepfake-idUKKBN1Z60JV](https://www.reuters.com/article/us-facebook-deepfake-idUKKBN1Z60JV)

⁶ [hXXps://www.washingtonpost\[.\]com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/](https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/)



video was later revealed to be a fake, created using editing software to slow down and distort Pelosi's speech.

Highlighting the increasingly complicated nature of synthetic media and geopolitics, media outlets revealed in 2021 that APT-C-23, a component of the Hamas-linked Molerats hacking group, targeted Israel Defence Forces soldiers on social media. The goal behind the attack was to extract sensitive information from the victims' devices for espionage purposes.. The attackers reportedly generated fabricated personas of Israeli women and used voice-altering software to produce convincing audio messages of female voices, which they used to persuade the Israeli soldiers to download a mobile app. The app was designed to install malware on the soldiers' devices, giving the attackers complete control over the infected systems.⁷ The use of synthetic audio in this attack made the fake personas more convincing and increased the chances of the soldiers falling for the deception.

Deepfakes and Political Manipulation

The increasing prevalence of synthetic media technology has raised concerns about its potential impact on political campaigns, democratic processes, and geopolitical landscapes. Manipulated videos can be used to spread false information, influence public opinion, and even sway election results.

Even low-quality deepfakes could be dangerous and potentially make people question the veracity of videos in the future. In March 2022, hackers circulated a deepfake video on social media and a Ukrainian news website of Ukrainian President Volodymyr Zelenskyy that appeared to show him surrendering to Russia. Although the video was quickly debunked and removed from major social media platforms, it gained traction on Russian social media.

Similarly, in June 2022, the mayors of several European capitals, including Berlin, Madrid, and Vienna, were fooled into holding video calls with a deepfake of Kyiv mayor Vitali Klitschko.⁸ The impersonator raised controversial issues, such as bringing back Ukrainian refugees for military service, which led the mayors to suspect they were not speaking to the real Klitschko. These incidents highlight the growing concern around deepfakes and the potential for them to be used as tools for information warfare, propaganda, and manipulation.

⁷

[hXXps://www.securityweek\[.\]com/apt-group-using-voice-changing-software-spear-phishing-campaign/](https://www.securityweek.com/apt-group-using-voice-changing-software-spear-phishing-campaign/)

⁸

[hXXps://www.theguardian\[.\]com/world/2022/jun/25/european-leaders-deepfake-video-calls-mayor-of-kyiv-vitali-klitschko](https://www.theguardian.com/world/2022/jun/25/european-leaders-deepfake-video-calls-mayor-of-kyiv-vitali-klitschko)



Federal Organizations Weigh-in on Synthetic Media Technology Risks

The FBI and CISA have both issued warnings about the potential risks of synthetic media technology.

- In 2020, the FBI warned that "foreign actors" could use synthetic technology to "disrupt democratic processes" during the U.S. presidential election.⁹ The FBI said it had "identified multiple campaigns which have leveraged synthetic content" since late 2019, and the number looks set to grow. The FBI also provided guidance on how to identify deepfakes and protect oneself from cybercrime, emphasizing that people should not assume an online profile corresponds to a real person.
- In 2022, the FBI also warned of the potential for deepfakes to be used for financial fraud and identity theft, particularly in job-related scams.¹⁰ The warning was issued regarding an increase in complaints of deepfakes and stolen Personally Identifiable Information (PII) being used to apply for remote work positions. The positions identified include those with access to customer PII, financial data, and proprietary information, and some reports indicate the use of voice spoofing or deepfakes during online interviews. Stolen PII is also being used to apply for these positions, with some victims reporting that they realized their identities were used and that pre-employment background checks discovered the use of PII belonging to someone else.

CISA has issued similar warnings, stating that deepfakes could be used to manipulate public opinion, interfere with elections, or even cause physical harm.¹¹ In addition, CISA has provided guidelines and best practices for organizations to protect themselves from deepfake attacks, such as implementing multi-factor authentication and conducting regular employee training on deepfake detection.

Other bodies, such as the WEF, have also raised concerns about the risks posed by synthetic media technology. WEF has highlighted the potential for deepfakes to undermine trust in public institutions, exacerbate existing political tensions, and create new security threats.

In response to these risks, various organizations are developing tools and technologies to detect and combat deepfakes. For example, DARPA has launched the Semantic Forensics program, which aims to develop advanced technologies for analyzing and detecting synthetic media.¹² Leveraging some of the research from another DARPA

⁹ [hXXps://www.ic3.gov/Media/News/2021/210310-2.pdf](https://www.ic3.gov/Media/News/2021/210310-2.pdf)

¹⁰ [hXXps://www.ic3.gov/Media/Y2022/PSA220628](https://www.ic3.gov/Media/Y2022/PSA220628)

¹¹ [hXXps://www.cisa.gov/rumor-vs-reality](https://www.cisa.gov/rumor-vs-reality)

¹² [hXXps://www.darpa.mil/news-events/2021-03-02](https://www.darpa.mil/news-events/2021-03-02)



program, the Media Forensics (MediFor) program, the semantic detection algorithms will seek to determine whether a media asset has been generated or manipulated. Attribution algorithms will aim to automate the analysis of whether media comes from where it claims to originate, and characterization algorithms seek to uncover the intent behind the content's falsification.

Companies like Microsoft and Facebook are creating their own deepfake detectors, but these are not widely available yet. Facebook, however, has developed a tool called Rosetta to attempt to detect and remove deepfakes from its platform.¹³ It extracts text from more than a billion public Facebook and Instagram images and video frames (in a wide variety of languages), daily and in real-time, and inputs it into a text recognition model that has been trained on classifiers to understand the context of the text and the image together.

Recommendations

Despite the nascency of synthetic media within the threat landscape and the low frequency of attacks, mitigating these threats should be part of any cybersecurity plan.

Prevent

- Incorporate synthetic media education into existing cybersecurity training, including examples designed to increase workforce awareness.
- Review compliance procedures for financial transactions, providing greater latitude to challenge senior leadership requests.
- Document and track executive exposure in open and closed sources and reduce digital footprints to minimize the availability of media that fuels impersonation.
- Consume ZeroFox Synthetic Media Intelligence for ongoing awareness and recommendations for defending against synthetic media-enabled threats.

Detect

- Leverage deepfake detection technologies, such as [Sensity](#), [DuckDuckGoose](#), [Reality Defender](#), [deepware](#), or tools from Microsoft and Intel.
- Monitor corporate social media and websites for signs of manipulation.
- Inspect images used by third-party profiles for distortions, indistinct and blurry backgrounds, and other visual artifacts often found in synthetic images.

Respond

- Create a crisis response plan to neutralize and contain any incidents of deepfake impersonation or misinformation, disinformation, or malinformation (MDM) targeting the organization.

¹³

<https://ai.facebook.com/blog/rosetta-understanding-text-in-images-and-videos-with-machine-learning/>



Protect

- Apply watermarks and/or use blockchain technology to verify media content.¹⁴
- Explore digital provenance solutions through the [Coalition for Content Provenance and Authenticity \(CP2A\)](#) or the [Content Authenticity Initiative \(CAI\)](#).

¹⁴ [hXXps://c2pa\[.\]org/](https://c2pa.org/)



Appendix: Traffic Light Protocol for Information Dissemination

WHEN SHOULD IT BE USED?

Red

Sources may use

TLP:RED when information cannot be effectively acted upon by additional parties and could lead to impacts on a party's privacy, reputation, or operations if misused.

Amber

Sources may use

TLP:AMBER when information requires support to be effectively acted upon but carries risks to privacy, reputation, or operations if shared outside of the organizations involved.

HOW MAY IT BE SHARED?

Recipients may NOT share

TLP:RED with any parties outside of the specific exchange, meeting, or conversation in which it is originally disclosed.

Recipients may ONLY share

TLP:AMBER information with members of their own organization and its clients, but only on a need-to-know basis to protect their organization and its clients and prevent further harm.

Note that

TLP:AMBER+STRICT restricts sharing to the organization only.

Green

WHEN SHOULD IT BE USED?

Sources may use

TLP:GREEN when information is useful for the awareness of all participating organizations, as well as with peers within the broader community or sector.

Clear

Sources may use

TLP:CLEAR when information carries minimal or no risk of misuse in accordance with applicable rules and procedures for public release.

HOW MAY IT BE SHARED?

Recipients may share

TLP:GREEN information with peers and partner organizations within their sector or community but not via publicly accessible channels.

Recipients may share

TLP:CLEAR information without restriction, subject to copyright controls.