WHITE PAPER

October 30, 2024

Unmasking AI Scams: Protecting Yourself in the Digital Age

Written by:
Mikinzi Strykul

RELI Group

RELI Group

Unmasking AI Scams: Protecting Yourself in the Digital Age

In today's digital landscape, artificial intelligence (AI) is being weaponized in new and alarming ways, with scams involving deepfakes and voice cloning becoming especially prevalent. These malicious tactics—ranging from lifelike videos and photos to eerily realistic voice replicas—are being used to deceive individuals and businesses alike. Scammers leverage AI's sophistication to create scenarios that exploit trust, often by impersonating family members, colleague, or public figures to steal sensitive information and money. This white paper examines the rise of AI scams, provides real-world examples of their devastating effects, and offers strategies to help individuals and organizations protect themselves. By understanding the risks and learning practical defense tactics, we can begin to shield ourselves from this growing AI-enabled threat.

The two main malicious uses for AI are deepfakes and voice cloning. Deepfakes are videos or photos that replicate a person. Threat actors often use these to make it appear as though the victim is saying or doing things they have never done. Voice cloning, or an audio deepfake, replicates a person's voice and enables a threat actor to create speech that sounds like the original speaker. This includes replicating their tone, intonation and pronunciation. These scams are called vishing, or voice phishing, where a cyber criminal calls the target and pretends to be a family member or friend.

Today, more than ever, scammers are exploiting advancements in AI to make scams more convincing and realistic. Researchers, policymakers and tech experts rank the malicious use of deepfakes, including voice cloning, as the most alarming AI threat today (Köbis, Doležalová & Soraperra, 2021).

Researchers, policymakers and tech experts rank the malicious use of deepfakes, including voice cloning, as the most alarming AI threat today.

Threat actors may target family members or friends of a victim, who is supposedly calling for help. Scammers try to interact with the target long enough to record their voices. By doing this, they can make an AI clone to access financial accounts or other services that use voice recognition. Threat actors may also target their victims by posing as a celebrity, CEO or politician. Audio deepfakes are used along with vishing techniques to create a convincing situation that may cause a victim to give vital information to a scammer, such as bank account information or social security numbers. According to a study by McAfee, scammers can use just a three-second clip of audio to clone a person's voice (DeVon, 2024).

With the rise of the internet, phishing has emerged as a leading social engineering tactic. AI has further enhanced these methods, enabling automated and highly convincing attacks. Malicious actors now use AI to create phishing messages that closely mimic human communication. Deepfake technology has also been rapidly advancing. Previously, photo alterations were primarily done using Photoshop, making it easier to detect edited images. Now, deepfake technology has blurred the line between what is real and what is not. Deepfake scams are the most popular scamming tactic used in the United States, and these scams have real consequences. According to the Federal Trade Commission, in 2022 $8.8 billion was lost due to deepfake scams. Raising awareness about these scams is the crucial first step in protecting potential victims.

Deepfake Detection

Deepfakes rely on people believing everything they see or hear is real, and refusing to acknowledge that it could be fake. Research found that humans are not able to consistently detect audio deepfakes. Regardless of language, humans tend to rely on the naturalness of the voice to determine if it is fake or not (Mai, Bray, Davies, & Griffin, 2023). As speech synthesis algorithms improve and sound more natural, it will become more difficult for humans to detect if a voice is a deepfake. Another study tested if people could determine deepfakes within a set of pictures of human faces. Subjects tended to be confident in their ability to differentiate the real pictures from AI-generated images, but their confidence was often misplaced (Köbis, Doležalová & Soraperra, 2021). Just like voice cloning, it is becoming increasingly harder for humans to detect deepfakes.

Deepfake Examples and Their Consequences

Between 2020 and now, AI has advanced tremendously. However, even in 2020, audio deepfakes were used to pull off high-profile scams.

According to the Federal Trade Commission, in 2022 $8.8 billion was lost due to deepfake scams.

Bolster AI reported a 94 percent increase in phishing and scam pages since 2020 (Garimella, 2024).

Many recent examples exist of the public falling for deepfake scams. Several include:

1. A bank manager in Hong Kong received a phone call from a scammer, replicating the voice of his company's director, whom they had a working relationship with. The "director" requested the bank manager authorize transfers totaling $35 million (Mai, Bray, Davies, & Griffin, 2023). Because of their existing relationship, the bank manager transferred $400,000 before realizing something was wrong.

2. Deepfakes that mimic celebrities are used to lend credibility to AI-generated scams. For example, a fake AI-generated video of Taylor Swift was used to promote a Le Creuset cookware set (Yang, 2024). The ad claims that due to a packaging error, 3,000 cookware sets were being given away for free to Taylor Swift's loyal fans. The company denied participation.

3. Mr. Beast, a popular YouTuber known for giving away money, has also been targeted by multiple deepfake scams. One example claimed that his fans could receive an iPhone 15 Pro for just $2. The advertisement used an AI version of Mr. Beast, stating that a limited amount of people who clicked a link would win. Mr. Beast has responded to these deepfakes, stating, "Lots of people are getting this deepfake scam ad of me...are social media platforms ready to handle the rise of AI deepfakes? This is a serious problem," (Rosenblatt, 2023).

Deepfakes are not only utilized in social engineering scams but also exploited in malware attacks. Researchers recently found a sophisticated trojan called GoldPickaxe, designed to steal facial biometric data and create deepfakes of victims (New iOS Trojan "GoldPickaxe" Steals Facial Recognition Data, 2024). This malware targets Android and iOS devices by impersonating government officials, convincing victims to use the messaging app Line. Victims are then tricked into downloading a Trojan-laden app disguised as a "digital pension" application or other government service. Once activated, the Trojan requests the victim's ID documents, intercepts SMS messages, and proxies traffic through the infected device. It also prompts victims to record a video as a "confirmation method" within the fake app. This recorded video is then used to create a deepfake, which cyber criminals can deploy along with other collected data to bypass banking logins.

Protecting Against Deepfakes

As technology continues to evolve, traditional security awareness is no longer enough. For corporations, specific measures should be taken to train employees to identify inconsistencies in deepfakes. Incorporating enterprise AI detection technology can help automatically detect deepfake scams. To protect yourself and others from voice cloning and deepfake scams, consider the following precautions:

- Do not trust the voice on the phone or the phone number. Scammers can impersonate caller ID, so don't trust it solely based on the displayed number. Always call the person who supposedly needs help to verify the story. If they can't be reached, try to get in touch with other family members or friends.

- Create a secret password to use with family and friends in case of situations like this. Asking the person on the phone if they know the password will help you figure out if an emergency is real or a scam. Consider asking questions that only the genuine person would know the answer to and are not public information.

- Let unknown calls go to voicemail, then call back to verify identity. Scammers only need a few seconds of your voice to clone it. To further protect yourself, change your voicemail to the default system message. This way, scammers can't use your personalized voicemail audio for voice cloning.

- Some AI companies are allowing users to upload audio clips to check if they were generated by the company. Use this if you are unsure if an audio clip has been faked.

For corporations, specific measures should be taken to train employees to identify inconsistencies in deepfakes. Incorporating enterprise AI detection technology can help automatically detect deepfake scams.

Strengthening AI Defenses

AI systems that can create deceptive deepfakes should be held to robust risk assessment requirements. Policymakers can combat AI scams by prioritizing funding to research AI detection tools to make these scams less effective. Organizations can utilize automated defense tools to continuously monitor their security posture for AI threats. In addition, implementing AI scam training across the company can increase awareness of current threats.

While these methods won't prevent threat actors from using AI in scams, providing tools that can detect AI can boost awareness in potential victims.

The consequences of utilizing AI for social engineering scams are becoming increasingly concerning. While there isn't a singular solution to completely stop AI social engineering scams, it is essential to implement measures to prevent the misuse of AI. By staying informed about current scams, individuals and organizations can better protect themselves against these sophisticated threats. Collaboration between technology companies, policymakers and researchers is crucial to develop effective AI tools and regulations. As AI evolves, so must our strategies to combat malicious use. Raising awareness about the dangers of AI scams is a vital step in creating a safer digital environment for everyone.

References

DeVon, C. (2024, January 24). Scammers can use AI tools to clone the voices of you and your family—how to protect yourself. CNBC. https://www.cnbc.com/2024/01/24/how-to-protect-yourself-against-ai-voice-cloning-scams.html

Garimella, A. (2024, March 12). 2024 State of Phishing & Online Scams: Statistics, Facts, Trends & Recommendations. Bolster.ai. https://bolster.ai/blog/2024-state-of-phishing-statistics-online-scams

Ianzito, C. (2024, April 3). AI Fuels New, Frighteningly Effective Scams. AARP; AARP. https://www.aarp.org/money/scams-fraud/info-2024/ai-scams.html

Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice - People cannot detect deepfakes but think they can. IScience, 24(11), 103364. https://doi.org/10.1016/j.isci.2021.103364

Mai, K. T., Bray, S. D., Davies, T., & Griffin, L. D. (2023). Warning: Humans cannot reliably detect speech deepfakes. PLOS ONE, 18(8), e0285333–e0285333. https://doi.org/10.1371/journal.pone.0285333

New iOS Trojan "GoldPickaxe" Steals Facial Recognition Data. (2024, February 15). Hackread - Latest Cybersecurity, Tech, Crypto & Hacking News. https://hackread.com/ios-trojan-goldpickaxe-steal-facial-recognition-data/

Park, P. S., Goldstein, S., O'Gara, A., Chen, M., & Hendrycks, D. (2024). AI deception: A survey of examples, risks, and potential solutions. Patterns, 5(5), 100988. https://doi.org/10.1016/j.patter.2024.100988

Rosenblatt, K. (2023, October 3). MrBeast calls TikTok ad showing an AI version of him a "scam." NBC News. https://www.nbcnews.com/tech/mrbeast-ai-tiktok-ad-deepfake-rcna118596

Saeidi, M. (2024, May 17). Voice cloning scams are a growing threat. Here's how you can protect yourself. - CBS New York. Www.cbsnews.com. https://www.cbsnews.com/newyork/news/ai-voice-clone-scam/

Social-Engineer. (2024, February 20). Artificial Intelligence: The Evolution of Social Engineering. Security through Education. https://www.social-engineer.org/social-engineering/artificial-intelligence-the-evolution-of-social-engineering/

Think you know what the top scam of 2023 was? Take a guess. (2024, February 5). Consumer Advice. https://consumer.ftc.gov/consumer-alerts/2024/02/think-you-know-what-top-scam-2023-was-take-guess

Yang, A. (2024, January 10). That Taylor Swift AI-generated Le Creuset ad is not real. NBC News; NBC News. https://www.nbcnews.com/tech/taylor-swift-ai-generated-le-creuset-ad-not-real-rcna133285