OUCH! March 2025

The Monthly Security Awareness Newsletter for You

# Beware of Deepfakes: A New Age of Deception

## Caught off Guard: Steve's Story

Steve was at his desk when he received a frantic video call from his manager, Bela. She looked stressed in the video call, her voice hurried. "I need you to send the confidential client report to this new email right away!" she insisted. Seeing her familiar face and hearing her distinct voice, he didn't hesitate, he sent the confidential report to the new email address.

Hours later, Bela walked into his office and asked about the report. Confused, Steve mentioned the video call. Bela's expression turned to shock — she hadn't called him. The person he saw on the video wasn't Bela. It was a *deepfake*, created by a cyber-criminal to trick him.

Steve couldn't believe how real the fake call seemed. The face, the voice, everything matched his boss perfectly. He had fallen victim to a growing cyber threat where criminals use Artificial Intelligence (AI) to create highly convincing fakes.

## What is a Deepfake?

AI can create images, audio, or videos that look real. These capabilities have many legitimate uses. For instance, marketing companies use this technology to create images for ad campaigns, movie companies use it to de-age certain actors, and teachers use it to create dynamic video lessons for their students.

A deepfake is when AI is used to create fake images, audio, or videos for the purpose of deceiving others. The name "deepfake" comes from a combination of "deep learning" (a type of AI) and "fake."

Often the most damaging deepfakes are when cyber criminals create fake images, audio, or video of people that you may know, making them do things they actually never did. For example, cyber attackers may create fake pictures of famous celebrities or politicians committing a crime and spread them as fake news. Or they may clone someone's voice and use it in a call to deceive a victim's family or colleagues. What makes deepfakes especially dangerous is how easily cybercriminals can replicate anyone, making them do anything, and make it appear real.

## Three Types of Deepfakes

### 1. Image Deepfakes

These are either photos of fake people created by AI or photos of real people but showing them doing something they never did. These fake images can spread quickly and are often used to damage reputations or manipulate emotions. Deepfake images are becoming increasingly common in social media, and people or even governments are attempting to push fake stories or narratives (called fake news) to affect a certain end goal.

**2. Audio Deepfakes (Voice Cloning)**

These are fake recordings or phone calls using someone's cloned voice. Attackers can get recordings of people's voices from podcasts or YouTube, then use those recordings to replicate their voice. Once replicated, cyber attackers can then call anyone they want pretending to be that individual. For example, someone could pretend to be a manager and call an employee to ask for sensitive data, or someone could re-create a loved one's voice in an emergency call asking for money.

**3. Video Deepfakes**

These are fake videos where people's voice and actions are manipulated or recreated. Deepfake videos can be pre-recorded video, or live video such as in an online conference call. For example, cyber attackers could create a deepfake video of a CEO giving a fake announcement about their company or a politician appearing to say something they never did.

## How to Detect Deepfakes: Focus on Context

Do not try to detect deepfakes by looking for technical mistakes. Both AI and the cyber attackers who use them have become very sophisticated. Instead, focus on context. Does the image, audio, or video make sense?

**1. Trust Your Instincts**: Does something feel "off" about the interaction? Is the request urgent or unexpected? Is the person behaving strangely, even if they look and sound normal? Is someone asking for confidential information or personal data that they should not have access to? If something doesn't feel right, trust your gut and double-check before complying with their request.

**2. Watch Out for Emotional Manipulation**: Cyber attackers often create urgency or fear to make you act quickly. If a message or call makes you panic, take a breath and verify. The stronger the emotional pull, such as creating a strong sense of urgency or fear, the more likely it's a potential attack.

**3. Verify Through Another Method:** If you are concerned the person contacting you may be a deepfake, reach out to the individual using a different method. For example, for video calls or messages that you are concerned may be fake, contact the person via phone or email. If you get a voice call asking for urgent action, hang up and call back using a trusted number.

**4. Establish a Code Word or Phrase:** Agree upon a shared code word or phrase known only within a group or perhaps your family that can be used to authenticate an urgent communication.

## Guest Editor

Dhruti Mehta is an Information Security Analyst at Physicians Health Plan of Northern Indiana and President of WiCyS Northern Indiana. She is passionate about building a diverse cybersecurity workforce and bridging educational and skill gaps in the field. https://www.linkedin.com/in/dhrutimehtacyber/

## Resources

**Emotional Triggers: How Scammers Trick You:** https://www.sans.org/newsletters/ouch/emotional-triggers-how-cyber-attackers-trick-you/
**Voice Cloning:** https://www.sans.org/newsletters/ouch/phantom-voices-defend-against-voice-cloning-attacks/