A World of Deepfakes!

Real vs. Deepfakes?

Office of Information Technology
Division of Information Security
SEPTEMBER 2024

What are Deepfakes?

Deepfakes are videos, photos, and even audio recordings, that are digitally altered. This manipulation of images and sound makes it difficult to identify what is authentically real. Therefore, digital mass consumers should gain a closer understanding of deepfakes as we encounter them almost daily, and might not recognize them.

Where Does One Come Across a Deepfake?

Deepfakes can be found in:

- **Entertainment:** Realistic special effects in movies, TV, and gaming
- **Education and training:** Realistic AI (Artificial Intelligence) generated simulations for training purposes
- **Fake news:** Convincing but false news videos and audio recordings
- **Impersonations:** Bad actors impersonating individuals in videos or audio recordings without consent
- **Fraud:** Attacks that deceive individuals into revealing sensitive information
- **Ethical issues:** Using AI to generate people's likeness or voice without their consent. Like impersonations, this leads to violation of privacy
- **Social media and websites:** Fake photos and videos generated using AI technology

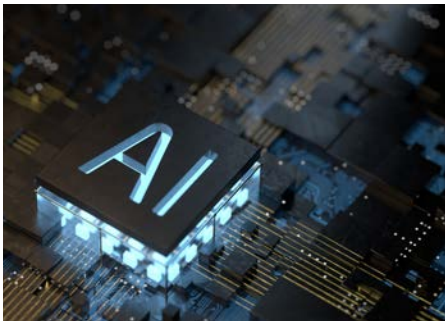
Deepfakes are All Around Us!

Digital consumers should be aware of the manipulation of reality through deepfakes since it has become difficult to tell what is "fake" or AI-generated, and what is not. One example of a deepfake technique is face swapping. An AI model can swap faces within videos and apply them to another person's face. We see this through filters in applications like Instagram. The images and videos posted can be extremely convincing, and lead viewers to believe what they are viewing is real.

Deepfakes are also created using voice cloning. Voice cloning convincingly replicates someone's voice to spread false information or trick someone into trusting a bad actor. This can lead people to give information to someone they should not. Cybercriminals can use deepfake audio tactics to generate familiar voices to make their attacks seem legitimate. This can be used in many ways, such as ransomware crimes and phishing attacks.

How to Spot Deepfakes, and Verify Information?

Deepfakes can be used to steal information and trick you into believing something that is not real. How, then, does one spot a deepfake? How does one avoid privacy violations, theft, or misinformation? It is difficult to spot deepfakes because they are generated using highly advanced technology to make one think that the person and/or information is valid. An email, text, direct



message, phone call, or other form of digital contact received from someone that seems familiar or known, but with closer attention to the number, email address, and other details might indicate a deepfake. If the tone of the message is demanding and aggressive, it might be a form of phishing. These are signs of a cybercriminal using deepfakes to try to gain the receiver's trust to commit a cybercrime. Cybercriminals will demand urgent action, and pressure the receiver to give them the information or ransom they are seeking.

Protection from Deepfakes

As we think about how to spot deepfakes, we should also be thinking about how to protect ourselves. An article written by the National Cybersecurity Alliance gives tips on how one can protect themselves from deepfake attacks, for example:

- Exercise a high level of privacy and avoid sharing too much on the internet (especially on various social media sites). The more one shares online, the easier it is for cybercriminals to use that information for deepfakes. Also, only share personal information online with people you trust. Avoid having a public account if you post personal information, high-quality images, and photos.
- Take advantage of your account's or website's privacy settings. There are online tools that will protect your account from cybercriminals.
- Watermark photos, videos, and other content you are displaying. Most of the time, this will discourage cybercriminals from using them. You can watermark by adding text to your image or video. You can use applications like Canva or Adobe Photoshop. Some social media applications like Instagram will allow you to write over photos before posting.
- Learn more about deepfakes and AI. This will help you spot red flags.
- Use multi-factor authentication as an extra layer of protection.
- Create long and unique passwords that will be hard for someone to guess.

Reporting Deepfakes or Suspicious Activities

As we continue to learn about deepfake activity, a digital consumer should keep in mind that cybercriminals can also use AI to commit malicious cybercrimes. If we are informed what could be a deepfake, we might be able to detect a malicious attack. For this reason, we ask everyone to stay vigilant and always report any incidents or suspicious activities via email to incident@ihs.gov.

NOTE: The links and products in this document are for informational purposes only and do not signify an endorsement.

Please contact cybersecurity@ihs.gov with any questions or comments about this newsletter.