GAO | Science, Technology Assessment, and Analytics

SCIENCE & TECH SPOTLIGHT:

DEEPFAKES

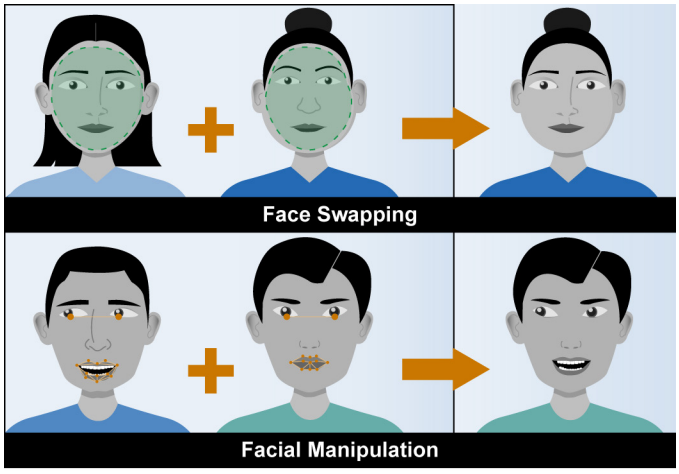
FEBRUARY 2020

WHY THIS MATTERS

Deepfakes are powerful tools that can be used for exploitation and disinformation. Deepfakes could influence elections and erode trust but so far have mainly been used for non-consensual pornography. The underlying artificial intelligence (AI) technologies are widely available at low cost, and improvements are making deepfakes harder to detect.

/// THE TECHNOLOGY

What is it? A deepfake is a video, photo, or audio recording that seems real but has been manipulated with AI. The underlying technology can replace faces, manipulate facial expressions, synthesize faces, and synthesize speech. Deepfakes can depict someone appearing to say or do something that they in fact never said or did.



Source: GAO. | GAO-20-379SP

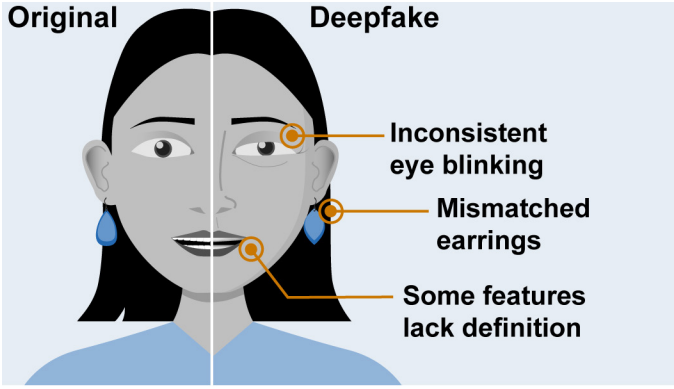
Figure 1. Deepfake videos commonly swap faces or manipulate facial expressions. In face swapping, the face on the left is placed on another person's body. In facial manipulation, the expressions of the face on the left are imitated by the face on the right.

While deepfakes have benign and legitimate applications in areas such as entertainment and commerce, they are commonly used for exploitation. According to a recent report from the company Deeptrace, much of deepfake content online is pornographic, and deepfake pornography disproportionately victimizes women. Further, there is concern about potential growth in the use of deepfakes for other purposes, particularly disinformation. Deepfakes could be used to influence elections or incite civil unrest, or as a weapon of psychological warfare. They could also lead to disregard of legitimate evidence of wrongdoing and, more generally, undermine public trust in audiovisual content.

How does it work? Deepfakes rely on artificial neural networks, which are computer systems modeled loosely on the human brain that recognize patterns in data. Developing a deepfake photo or video typically involves feeding hundreds or thousands of images into the artificial neural network, "training" it to identify and reconstruct patterns—usually faces.

Deepfakes use different underlying AI technologies—notably autoencoders and generative adversarial networks (GANs). An autoencoder is an artificial neural network trained to reconstruct input from a simpler representation. A GAN is made up of two competing artificial neural networks, one trying to produce a fake, the other trying to detect it. This competition continues over many cycles, resulting in a more plausible rendering of, for example, faces in video. GANs generally produce more convincing deepfakes but are more difficult to use.

Researchers and internet companies have experimented with several methods to detect deepfakes. These methods typically also use AI to analyze videos for digital artifacts or details that deepfakes fail to imitate realistically, such as blinking or facial tics.



Source: GAO; conceived from DARPA image at https://www.darpa.mil/news-events/2019-09-03a. | GAO-20-379SP

Figure 2. Examples of characteristics that may indicate a deepfake.

How mature is it? Anyone with basic computer skills and a home computer can create a deepfake. Computer applications are openly available on the internet with tutorials on how to create deepfake videos. However, to develop a somewhat realistic deepfake, these applications generally still require hundreds or thousands of training images of the faces to be swapped or manipulated, making celebrities and government leaders the most common subjects. More convincing deepfakes created with GANs require more advanced technical skills and resources. As artificial neural network technologies have advanced rapidly in parallel with more powerful and abundant computing, so has the ability to produce realistic deepfakes.

/// OPPORTUNITIES

There are some potentially benign or beneficial uses of this technology:

- **Entertainment.** Voices and likenesses can be used in a movie to achieve a creative effect or maintain a cohesive story when the entertainers themselves are not available.
- **E-commerce.** Retailers could let customers use their likenesses to virtually try on clothing.
- **Communication.** Speech synthesis and facial manipulation can make it appear that a person is authentically speaking another language.

/// CHALLENGES

- **Data needs for detection.** Deepfake detection tools must generally be trained with large and diverse data sets to reliably detect deepfakes. Technology companies and researchers have released data sets to help train detection tools, but the current data sets are not sufficient by themselves. Detection tools must be constantly updated with data of increasing sophistication to ensure that they continue to be effective at detecting manipulated media.
- **Detection is not yet automated.** Current tools cannot perform a complete and automated analysis that reliably detects deepfakes. Research programs are currently working on means to automatically detect deepfakes, provide information on how they were created, and assess the overall integrity of digital content.
- **Adaptation to detection.** Techniques used to identify deepfakes tend to lead to the development of more sophisticated deepfake techniques. This "cat and mouse" situation means detection tools must be regularly updated to keep pace.
- **Detection may not be enough.** Even a perfect detection technology may not prevent a fake video from being effective as disinformation, because many viewers may be unaware of deepfakes or may not take the time to check the reliability of the videos they see.
- **Inconsistent social media standards.** The major social media companies have different standards for moderating deepfakes.
- **Legal issues.** Proposed laws or regulations addressing deepfake media may raise questions regarding an individual's freedom of speech and expression and the privacy rights of individuals falsely portrayed in deepfakes. Moreover, potential federal legislation aimed at combating deepfakes could face enforcement challenges.

/// POLICY CONTEXT AND QUESTIONS

Any policy response seeking to address deepfakes would likely face constitutional and other legal challenges along with the technical challenges of detection. Key policy questions include:

- What is the maturity of deepfake detection technology? How much progress have federal programs and public-private partnerships made in developing such technology? What expertise will be required to ensure detection keeps pace with deepfake technology?
- What rights do individuals have to their privacy and likenesses? What rights do creators of deepfakes have under the First Amendment? What policy options exist regarding election interference? What policy options exist regarding exploitation and image abuse, such as non-consensual pornography?
- What can be done to educate the public about deepfakes? Should manipulated media be marked or labeled? Should media be traceable to its origin to determine authenticity?
- What should the roles of media outlets and social media companies be in detecting and moderating content that has been altered or falsified?

/// SELECTED GAO WORK

- Technology Assessment: Artificial Intelligence: Emerging Opportunities, Challenges, and Implications, GAO-18-142SP

/// SELECTED REFERENCES

Ajder, Henry et al. *The State of Deepfakes: Landscape, Threats, and Impact.* Amsterdam, Netherlands: Deeptrace, 2019.

Barrett, Paul M. *Disinformation and the 2020 Election: How the Social Media Industry Should Prepare.* New York, N.Y.: NYU Stern Center for Business and Human Rights, 2019.

Centre for Data Ethics and Innovation. *Deepfakes and Audio-visual Disinformation.* London, United Kingdom: 2019.

Collins, Aengus. *Forged Authenticity: Governing Deepfake Risks.* Lausanne, Switzerland: EPFL International Risk Governance Center, 2019.

Library of Congress. Congressional Research Service. *Deep Fakes and National Security.* IF11333. Washington, D.C.: Oct. 14, 2019.

Westerlund, Mika. "The Emergence of Deepfake Technology: A Review." *Technology Innovation Management Review*, vol. 9, no. 11 (2019): pp. 39-52.

GAO SUPPORT:

GAO meets congressional information needs in several ways, including by providing oversight, insight, and foresight on science and technology issues. GAO staff are available to brief on completed bodies of work or specific reports and answer follow-up questions. GAO also provides targeted assistance on specific science and technology topics to support congressional oversight activities and provide advice on legislative proposals.

Timothy M. Persons, PhD, Chief Scientist, personst@gao.gov

Staff Acknowledgments: Karen Howard (Director), Laura Holliday (Assistant Director), Sushil Sharma (Assistant Director), Chi Mai (Analyst-in-Charge), Adam Brooks (Analyst), Anika McMillon, and Ben Shouse.