NSI BACKGROUNDER

DECEMBER 2020

DECODING **DEEPFAKES**

By MATTHEW F. FERRARO

THIS NSI **BACKGROUNDER**

1 **DEFINES** "deepfakes".

2 **EXPLAINS** how they are made and highlights recent examples.

3 **DESCRIBES** the dangers of deepfakes to national security, politics, business, economic security, and personal privacy.

4 **OUTLINES** potential technological and legislative fixes.

WHAT ARE DEEPFAKES

- **Deepfakes Defined.** A deepfake is synthetic media (text, images, audio, or video) that is either manipulated or wholly generated by artificial intelligence (AI).[1] Put simply, it is a very convincing media forgery created by computers.
  - The term, "deepfake," is a blend of the words "deep learning" (a branch of AI) and "fake." Deepfakes are also sometimes called "machine-manipulated media," "synthetic media," "digital content forgeries," and so forth.
  - In a recent example, MIT researchers, to show how easy it is to make a deepfake, created in 2020 a video in which President Richard Nixon announced that the Apollo 11 moon landing had failed. In truth, the researchers used deepfake technology to edit a video from a different Nixon speech to make his voice and facial movements appear he was giving a fabricated address.[2]
- **Deepfake Technology.** Deepfake technology uses computer systems called "generative adversarial networks" (GANs), wherein one algorithm (the generator) creates content modeled on source data, such as existing video or audio of an individual, while a second algorithm (the discriminator) tries to spot the artificial content.[3]
  - The competition between the two networks produces a better-and-better fake until the discriminator can no longer identify the forgery.
- **Growth of Deepfakes.** The technology used to make deepfakes is getting better and more accessible, and, as a result, the number of deepfakes available online is steadily increasing. As of July 2020, there were about 50,000 deepfakes on the Internet. The number doubles every six months.[4]

DANGERS OF DEEPFAKES

- **Political Disinformation.** We are accustomed to believing that when a public official appears on camera to deliver an urgent message, the message and messenger are authentic. But a deepfake could create confusion or undermine official government conduct in the context of a political campaign, a public health emergency, a prosecution, or similar circumstances.
  - Imagine the political impact of a realistic looking fake video that circulates on the eve of an election of a candidate using a racial epithet. Likewise, imagine a believable forged video of a state official announcing that polling places had been closed. Voters could be discouraged from casting ballots, throwing doubt on the entire process.

- **National Security.** There are major national security implications of deepfakes. They could be used in a variety of ways to undermine governmental leadership, disrupt effective military communication, or even spark conflict.
  - Consider the potentially deadly repercussions of a fake video of the American president announcing a missile strike on North Korea that goes viral before it can be explained.
- **"Liar's Dividend."** Deepfakes can bestow a "Liar's Dividend"—one can successfully deny the authenticity of genuine content by claiming the content is a deepfake.[5]
  - This happened in 2018 in Gabon when opponents of the president claimed that a video of him was fake and he was incapacitated or dead. The conspiracy spread and, within a week, the military launched a coup.[6]
- **Personal Privacy.** Deepfakes pose a particular threat of intimate harassment. According to a 2019 study, over 95% of all deepfake videos are of nonconsensual pornography[7] —a nonconsenting person's face placed on the body of a pornographic performer. Most of those videos are today of famous female actors, but the technology menaces everyone's privacy, including governmental leaders and their loved ones.[8]
- **Economic Security.** Disinformation and deepfakes endanger economic security and the private sector,[9] including through stock-price manipulation, credential theft, and fraud.[10] Well-timed deepfakes could also cause systemic harm to the economy by, for example, undermining confidence in the central bank, sparking a market sell-off, or crippling a U.S. business to favor a foreign national champion.
  - For example, in March 2019, an insurer reported that one of its clients lost roughly $250,000 after a company official wired money to a telephone caller impersonating the parent company's CEO with AI-based software.[11]
  - More than 70 countries already have state-sponsored disinformation units.[12] These outfits could relatively easily use manipulated media to target the economic lifeblood of adversaries or protect national champions. For example, imagine the Chinese state promoting a convincing synthetic video of the violent crash of an American autonomous vehicle to undermine confidence in a U.S. competitor for the benefit of a Chinese automotive company.[13]

DEALING WITH DEEPFAKES

- **Technological Fixes.** There are basically two technological ways to counter deepfakes. One is to detect the phony media after it is created. For example, Microsoft launched a new tool in September 2020 to spot deepfakes by giving a confidence score to analyzed pictures and videos about whether they were manufacted artificially.[14]
  - The second way is to verify photographs at the "point of capture" in such a manner that they cannot be altered or modified after the fact.[15] That technology is also proceeding apace: in October 2020, Qualcomm and a startup called Truepic announced that it would embed a photo and video verification tool leveraging image-provenance technology within smartphone chips that will be available in some Android devices in 2021.[16]
- **Legislative Action.** Legislators in Congress and the states have moved quickly to address the dangers of manipulated media by outlawing certain uses of deepfakes and studying their implications for national and economic security.
  - Five states (California, Maryland, New York, Texas, and Virginia) have adopted laws barring certain deepfakes that target candidates for public office or deepfake pornography. About ten other states are considering similar legislation.
  - Congress passed the first national law on deepfakes in 2019. It requires intelligence agencies to report on the foreign weaponization of deepfakes and to notify Congress of deepfake activities targeting elections.[17] Congress is considering about 10 other bills on deepfakes that mostly require reports and research.
- **Other Actions.** There are non-legislative mechanisms that can address deepfakes. For example, in January 2020 the U.S. House of Representatives Committee on Ethics cautioned Representatives and their staffs against posting deepfakes because they could be in violation of the House's Code of Official Conduct.[18] Similarly, ethical guidelines, professional rules, and social norms may be developed to address deepfakes.[19]

KEY ITEMS TO WATCH

EXPONENTIAL INCREASE

- Whether deepfakes will continue to proliferate exponentially and whether the technology to create them will continue to improve and become universally accessible.

LEGISLATIVE EFFORTS

- Whether the President signs the National Defense Authorization Act for FY 2021, which would require the Departments of Defense and Homeland Security to study the threats posed by deepfakes to the military and national and economic security; whether Congress will amend federal election law to prohibit deepfakes in elections; and whether more states will bar deepfake pornography or election-related deepfakes and whether successful claims will be brought under those laws.

TECHNOLOGICAL FIXES

- Whether efforts to improve and distribute technologies to detect deepfakes or to verify images at the point of capture will be widely adopted and able to catch and stay ahead of deepfake technology.

ENDNOTES

Matthew F. Ferraro is an NSI Visiting Fellow at George Mason University's Antonin Scalia Law School. He is an attorney and former intelligence officer who writes widely on national security and legal issues. A term member of the Council on Foreign Relations, he is counsel at the international law firm Wilmer Cutler Pickering Hale and Dorr where he represents clients on matters related to defense and national security, cybersecurity, and crisis management. Previously, Mr. Ferraro held staff, policy, and operational positions with the Director of National Intelligence, Central Intelligence Agency, and other government agencies.

1 *See generally* Nina Schick, Deepfakes: The Coming Infocalypse 8 (2020).

2 Bonnie Burton, *MIT Releases Deepfake Video of 'Nixon' Announcing NASA Apollo 11 Disaster*, CNET (Jul. 20, 2020), https://www.cnet.com/news/mit-releases-deepfake-video-of-nixon-announcing-nasa-apollo-11-disaster/; Mike Well, *Apollo 11 'Disaster' Video Project Highlights Growing Danger of Deepfake Tech*, Space.com (Jul. 19, 2020), https://www.space.com/apollo-11-disaster-deepfake-video-tech.html.

3 Sarah Basford, *What Deepfakes Actually Are*, Gizmodo (Jul. 31, 2020), https://www.gizmodo.com.au/2020/07/what-are-deepfakes/; Meredith Somers, *Deepfakes, Explained*, MIT (Jul. 21, 2020), https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained.

4 Henry Ajder, *Deepfake Threat Intelligence: A Statistics Snapshot from June 2020*, Sensity (Jul. 3, 2020), https://sensity.ai/deepfake-threat-intelligence-a-statistics-snapshot-from-june-2020/ (Sensity was previously known as Deeptrace Labs).

5 Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 California Law Review 1753, 1785-1786 (2019).

6 Rob Toews, *Deepfakes Are Going To Wreak Havoc On Society. We Are Not Prepared*, Forbes (May 25, 2020), https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/.

7 Giorgio Patrini, *The State of Deepfakes 2019*, Sensity (Oct. 7, 2019), https://sensity.ai/mapping-the-deepfake-landscape/.

8 Jane Lytvynenko & Scott Lucas, *Thousands of Women Have No Idea a Telegram Network is Sharing Fake Nude Images of Them*, BuzzFeed (Oct. 20, 2020) https://www.buzzfeednews.com/article/janelytvynenko/telegram-deepfake-nude-women-images-bot (deepfake technology allowed users to create photo-realistic simulated nude images of 680,000 women without their knowledge or consent, according to tech researchers).

9 *See* Matthew F. Ferraro, Jason C. Chipman & Stephen W. Preston, *Identifying the Legal and Business Risks of Disinformation and Deepfakes: What Every Business Needs to Know*, 6 Pratt's Privacy and Cybersecurity Law Report 142, 142 (2020); Claire Atkinson, *Fake News Can Cause 'Irreversible Damage' to Companies—And Sink Their Stock Price*, NBC News (Apr. 25, 2019), https://www.nbcnews.com/business/business-news/fake-news-can-cause-irreversible-damage-companies-sink-their-stock-n995436.

10 Ferraro, et al., *supra* note 9, at 143.

11 *Id.* at 145-146 (citing reports).

12 Matthew F. Ferraro & Preston B. Golson, *The Next Gray Zone Conflict: State-Based Disinformation Attacks on the Private Sector*, Lawfare (Mar. 24, 2020), https://www.lawfareblog.com/next-gray-zone-conflict-state-based-disinformation-attacks-private-sector.

13 *Id.* (describing how a Russian company recently designed a fake electric car accident to garner attention, and several media outlets treated the bogus video as if it were real); Cat Zakrzewski, *Businesses Should be Watching Out for Deepfakes Too, Experts Warn*, Washington Post (Dec. 13, 2019), https://www.washingtonpost.com/news/powerpost/paloma/the-technology-202/2019/12/13/the-technology-202-businesses-should-be-watching-out-for-deepfakes-too-experts-warn/5df279f1602ff125ce5b2fe7/.

14 Leo Kelion, *Deepfake Detection Tool Unveiled by Microsoft*, BBC (Sept. 1, 2020), https://www.bbc.com/news/technology-53984114.

15 Mounir Ibrahim, *To Beat Deepfakes, We Need to Prove What is Real. Here's How*, World Economic Forum (Mar. 23, 2020), https://www.weforum.org/agenda/2020/03/how-to-make-better-decisions-in-the-deepfake-era/.

16 Olivia Solon, *Qualcomm Announces Photo Verification Tool*, NBC News (Oct. 15, 2020), https://www.nbcnews.com/tech/security/qualcomm-announces-photo-verification-tool-n1243550.

17 Matthew F. Ferraro, Jason C. Chipman & Stephen W. Preston, *The Federal "Deepfakes" Law*, 3 The Journal of Robotics, Artificial Intelligence & Law 229, 229 (2020).

18 Memorandum for All Members, Officers, and Employees from Committee on Ethics, Subj. Intentional Use of Audio-Visual Distortions & Deep Fakes (Jan. 28, 2020), https://ethics.house.gov/sites/ethics.house.gov/files/wysiwyg_uploaded/Deep%20Fakes%20Pink%20Sheet%20Guidance-Final.pdf.

19 *See* Matthew F. Ferraro, *Through A Straw Darkly*, N.Y.U. Journal of Legislation & Public and Policy Quorum (2020), https://nyujlpp.org/quorum/ferraro-reflections-on-nyu-deepfakes-conference/.