SUBMITTED STATEMENT FOR THE RECORD OF

JEFFREY WESTLING
FELLOW, TECHNOLOGY AND INNOVATION POLICY
THE R STREET INSTITUTE

BEFORE THE

PERMANENT SELECT COMMITTEE ON INTELLIGENCE
UNITED STATES HOUSE OF REPRESENTATIVES

HEARING ON

NATIONAL SECURITY CHALLENGES OF ARTIFICIAL INTELLIGENCE, MANIPULATED MEDIA, AND DEEPFAKES

JUNE 13, 2019

NATIONAL SECURITY CHALLENGES OF ARTIFICIAL INTELLIGENCE, MANIPULATED MEDIA, AND DEEPFAKES

CHAIRMAN SCHIFF, RANKING MEMBER NUNES, AND MEMBERS OF THE COMMITTEE:

Thank you for holding this hearing on the national security challenges of artificial intelligence, manipulated media, and deepfakes. While deception is nothing new, advances in artificial intelligence technology have allowed for the creation of fake audiovisual materials that are almost indistinguishable from authentic content, especially to ordinary human senses. Because such materials can be used for a variety of nefarious purposes, this hearing is both timely and welcome.

My name is Jeffrey Westling, and I am a Fellow in Technology and Innovation Policy at the R Street Institute. R Street is a nonprofit, nonpartisan public policy research organization. Our mission is to engage in policy research and outreach to promote free markets and limited, effective government in many areas, including the technology sector. R Street has engaged in significant research on online content issues and disinformation generally,[1] and I have personally studied deepfakes for over a year, examining research on the psychology of trust and deception[2] as well as the dissemination of information online.[3]

---

[1] See, e.g., Mike Godwin, *The Splinters of Our Discontent: How to Fix Social Media and Democracy Without Breaking Them* (Zenger Press 2019); and Megan Reiss, "Disinformation in the Reagan Years and Lessons for Today," R Street Policy Study No. 164, February 2019. https://bit.ly/2Nisao1.

[2] See, e.g., Elizabeth F. Loftus, "Creating False Memories," *Scientific American* 277:3 (1997), pp. 70–75. https://bit.ly/2FNzdlC; Joel Cooper, *Cognitive Dissonance: Fifty Years of a Classic Theory* (Sage Publications 2009); C.J. Brainerd and V.F. Reyna, *The Science of False Memory* (Oxford University Press 2005), p. 52; and "'Who shared it?' How Americans Decide What News to Trust on Social Media," *American Press Institute*, Mar. 20, 2017. https://goo.gl/G5H748.

[3] See, e.g., Adrien Guille et al., "Information Diffusion in Online Social Networks: A Survey," SIGMOD Record 42:2 (June 2013). https://bit.ly/2wFYs4X; and Savvas Zannettou et al., "On the Origins of Memes by Means of Fringe Web Communities," *Proceedings of the Internet Measurement Conference* (May 31, 2018). https://bit.ly/2Ky7o4v.

This statement represents my individual views and does not necessarily reflect the views of other scholars at the R Street Institute.

Based on my research, I have found that deepfakes are not an unprecedented problem that requires an unprecedented solution. An examination of similar technological developments that have occurred in the past shows that the same fears about disinformation and the death of truth that are now prevalent in response to deepfakes arise whenever new technologies that can be used to spread disinformation emerge.[4] These developments, and the response they elicited, show that as long as the public becomes aware of the new technology and its capabilities, society can adapt to it without the need for major legislative intervention.

My research also challenges the conventional wisdom that deepfakes will likely be believed more readily, or shared more widely, than other inauthentic media. Instead, the belief in and sharing of deepfake content tends to be motivated by the same psychological factors that drive people to believe in and share other forms of media, such as adherence to pre-existing beliefs and trust in the person who shared the information.[5] Certainly, social media—and the Internet more generally—has created a different environment than that which existed in the past. But the specific technology used to spread disinformation matters less than whether the disinformation presented conforms to an existing worldview or comes from an individual the person already trusts.

The United States needs to prepare for scenarios in which bad actors use deepfakes and other technologies to spread disinformation. But instead of doing so by targeting a

[4] Jeffrey Westling, "Deep Fakes: Let's Not Go Off the Deep End," *Techdirt*, Jan. 30, 2019. https://bit.ly/2RuaqH1.

[5] Jeffrey Westling, "Deception and Trust: A Deep Look at Deep Fakes," *Techdirt*, Feb. 28, 2019. https://bit.ly/2Y2vsjr.

particular technology or the platforms by which content spreads, lawmakers should approach the problem of disinformation with an understanding that deepfakes are not an altogether new or distinct problem.

I. DEEPFAKES IN CONTEXT

Much of the concern about deepfakes stems from the fact that they appear to be an unprecedented problem, implying that they therefore require an unprecedented solution.[6] On the surface, this makes some sense, since we have not seen this type of synthetic media in the past. But countless new technological improvements over the years have generated significant skepticism or concern from the general public when they first debuted[7]; deepfakes are no different. And, as with similar developments in the past, society will adapt on its own. In light of these facts, the main goal of the Committee should be to limit the time it takes for society to adapt to this new technology.

A. Society adapted to similar technological advancements in the past

Much like with the case of deepfakes, the emergence of Adobe Photoshop and other digital photograph editing tools in the 1990s came with significant concerns about truth.[8] At the time, many believed that a photograph presented definitive proof of reality. The advent of photograph editing tools undermined this notion. And yet, despite these concerns, society was ultimately able to adapt.

In the same year that Adobe released the first commercial version of Photoshop, Newsweek published an article entitled, "When Photographs Lie," which explored the

[6] See Robert Chesney and Danielle Citron, "Deep Fakes: A Looming Crisis for National Security, Democracy and Privacy?," *Lawfare*, Feb. 21, 2018. https://bit.ly/2EP4nvf.

[7] See Adam Thierer, "Technopanics, Threat Inflation, and the Danger of Information Technology Precautionary Principle," Minnesota Journal of Law, Science, and Technology 14:1 (2013). https://bit.ly/2K3oat2.

[8] Westling, "Deep Fakes: Let's Not Go Off the Deep End." https://bit.ly/2RuaqH1.

effect that digital photograph manipulation would have on public discourse and trust in the media.[9] The article paints a bleak picture: "Take China's leaders, who last year tried to bar photographers from exposing [the leaders'] lies about the Beijing massacre. In the future, the Chinese or others with something to hide wouldn't even worry about photographers."[10] But this doomsday scenario does not track with the world as we know it today. This was due to a variety of reasons.

For one, many digital photograph manipulations were very poorly done—they often included an extra hand or an obviously misplaced object.[11] This type of obvious manipulation became the target of mockery, raising public awareness of the fact that, with the arrival of photograph editing technology, a photograph may no longer present an authentic record of events.[12]

Similarly, manipulated images also became the subject of significant controversy. The editing of models on the covers of magazines, for instance, led to body-image awareness campaigns,[13] and Time magazine's darkening of its cover photograph of O.J. Simpson raised questions about photograph manipulation and racism within the news media.[14] While the readers of these magazines may not have known about these issues at the time, the significant media backlash they caused informed the public about photograph editing practices.[15]

---

[9] Newsweek Staff, "When Photographs Lie," *Newsweek*, July 29, 1990. https://bit.ly/2Xt8dPK.

[10] Ibid.

[11] See, e.g., Katla McGlynn, "The Funniest Photoshop Fails of All Time," *The Huffington Post*, May 25, 2011. https://bit.ly/2Z8kTvV.

[12] Ibid.

[13] For example, Dove released a video in 2006 showing how much work goes into manipulating images of models, as well as showing the stark difference between the "before" and "after" images of the models. "Dove Evolution," *Dove*, last visited June 6, 2019. https://bit.ly/2OU8TbJ.

[14] See Deirdre Camody, "Time Responds to Criticism Over Simpson Cover," *The New York Times*, June 25, 1994. https://nyti.ms/2rj1biH.

[15] See Westling, "Deep Fakes: Let's Not Go Off the Deep End." https://bit.ly/2RuaqH1.

As the public became more aware of the capabilities of digital photographic manipulation, news outlets felt the need to regain the trust they once had.[16] Most notably, many of these publications established codes of ethics for publishing photos, promising their readers that their companies would not alter the images depicted in their papers or magazines.[17] Similarly, photographers themselves grew to understand that news agencies put faith in them to produce an accurate record of reality, and that if they edited their images before delivering them, they risked their credibility—not to mention their further employment.[18]

This is all to say that despite the rhetoric surrounding image manipulation and the "death of truth" that some predicted, society was able to adapt to the new technology on its own without the need for major legislation. Certainly, harms did occur because of manipulated images. [19] But overbroad legislation of manipulated images and disinformation generally could have had a dramatic, negative impact on legitimate speech by punishing borderline content or incentivizing platforms to err on the side of removing content out of fear of liability. Therefore, the role of legislators was not, and should not be, to regulate this speech, but rather to help educate consumers. When consumers understand

---

[16] For example, the Los Angeles Times quickly fired a war photographer who had combined two images of British soldiers directing Iraqi civilians into a single image that, although it was more powerful than the images on their own, was inaccurate. In response to the backlash, the newspaper republished both of the original photos and explained what had happened to readers, ultimately realizing that readers' trust was something the paper could not jeopardize. Russell Frank, "Altered Photos Break Public's Trust in Media," *Los Angeles Times,* Apr. 7, 2003. https://lat.ms/2Z6rY02.

[17] "News Values and Principles: Visuals," *Associated Press*, last visited June 13, 2019. https://bit.ly/2tCoODZ.

[18] "Code of Ethics," *National Press Photographers Association* (2018). https://goo.gl/gEZXwJ.

[19] Even today, we still see the issue of image manipulation on popular social media sites like Instagram. For example, the subreddit InstagramReality has over 400,000 subscribers with top posts achieving over 50,000 points on the platform. See "InstagramReality," *Reddit,* last visited June 6, 2019. https://bit.ly/2K1sHvW. Likewise, the popular YouTube channel H3H3Productions released a video on the subject that has received over 6 million views in a single month. H3H3Productions, "Instagram vs. Reality," *YouTube,* May 8, 2019. https://bit.ly/2wEliKj.

that what they see may not be an authentic record of events, they can better identify content that takes strides to adhere to the truth and even share content debunking fake claims with one another.[20]

B. Society will adapt to deepfakes

As with digital photograph editing technology in the 1990s, deepfakes present a new challenge to truth. Yet society will very likely adapt to these challenges without the need for broad legislative intervention.

To begin, like the manipulated photographs of the 1990s, many deepfakes and other manipulated videos are relatively poor in quality,[21] leaving glaring mistakes that are obvious to the naked eye.[22] While these types of mistakes may not rise to the level of an extra limb on the cover of a magazine, they will be easily noticeable and likely become the target of some ridicule, depending on how seriously the video is taken.

Second, and more importantly, video manipulations have already sparked significant controversy, raising awareness about the nature of truth in video. For example, a recent video purported to show House Speaker Nancy Pelosi under the influence of alcohol while she was on stage at a conference but was in fact just video of a normal speech slowed down to make her appear inebriated.[23] This not-so-deepfake was shared millions of times on Facebook; even President Donald Trump tweeted out a similar video designed to

[20] For example, despite the outrage over the photoshopped version of a Parkland shooting survivor tearing up the Constitution, the image was only "liked" and shared a total of 5,000 times before it was debunked. One tweet debunking the fake was "liked" and shared over 350,000 times. Gianluca Mezzofiore, "No, Emma Gonzalez Did Not Tear Up a Photo of the Constitution," *CNN*, Mar. 26, 2018. https://cnn.it/2G9m8Sn.

[21] See Westling, "Deep Fakes: Let's Not Go Off the Deep End." https://bit.ly/2RuaqH1.

[22] See Gregory Barber, "Deepfakes Are Getting Better, But They're Still Easy to Spot," *Wired*, May 26, 2019. https://bit.ly/2MaywZL.

[23] Drew Harwell, "Pelosi Videos, Altered to Make Her Seem Drunk, Spread Across Social Media," *The Washington Post*, May 24, 2019. https://wapo.st/2M6uFgA.

make it appear as though Speaker Pelosi was struggling through her speech.[24] Yet The Washington Post discovered the true nature of the video and published a report on it. Major cable networks also devoted significant time to covering the video's inauthenticity. Ultimately, these reports alerted millions of people to the fact that video alone can be a misrepresentation of the truth.

This type of incident prepares society for the deployment of next-generation deceptive tools and techniques.[25] While deepfakes can and will deceive some portion of the target audience, many others will look to the context and the investigative reporting done by journalists, who have a strong incentives to debunk such claims.[26] Much like with the code of ethics established by newspapers and photographers regarding image manipulation, broadcasters and cable news networks will have reason to investigate videos before disseminating them and provide transparency to their viewers in order to uphold viewer trust. While this may become less effective as fewer people rely on traditional sources of news, there is also a strong incentive for journalists and even independent fact-checkers to debunk such content.[27]

C. Promoting education about the technology

Ultimately, society will adapt to the technology without Congress needing to limit legitimate speech or inhibit technological development. However, a significant concern remains: How quickly will this process happen? Harm can and will occur while society adapts to deepfakes. To limit the length of time this process takes and, in turn, the amount

[24] Natasha Bach, "Trump-Pelosi Feud Hits New Lows With 'Doctored' Video," *Fortune,* May 24, 2019. https://bit.ly/2Wt4bKW.

[25] Jeffrey Westling, "Fool Me Once … You Can't Get Fooled Again," *Morning Consult*, June 3, 2019. https://bit.ly/2WsP9EK.

[26] Jeffrey Westling, "Can BuzzFeed Save Us From Deep Fakes?" *Morning Consult*, Oct. 23, 2018. https://bit.ly/2OyDDyO.

[27] Ibid.

of harm caused, the Committee should begin by focusing on educating the public about the dangers deepfakes can present.

In early 2018, for example, Buzzfeed released a public service announcement appearing to be from former President Barack Obama that went viral.[28] As the video develops, the viewer finds out that it does not depict the former president; in reality, the video is a deepfake, with comedian Jordan Peele providing the voice of the former president. While shocking at first, the clip left the viewer with a better understanding of the technology and what it can do.

Similarly to how Buzzfeed informed its audience of the capabilities of deepfake technology using a fake public service announcement, the Committee should focus on ways to educate the public so that the time it takes for society to adapt to the new technology is limited. As long as the adaptation period is reduced, the harms caused by the disinformation will be minimized.

To be clear, education alone will not stop the harms caused by disinformation generally, and deepfakes will certainly play a role in disinformation campaigns. However, to the extent that people desire the truth rather than information that confirms a pre-existing view, education about the capabilities of video editing technology can minimize the impact that a well-crafted deepfake will have on society.

II. IMPACT OF SOCIAL MEDIA

Since the 1990s, the internet ecosystem has evolved into a leading method for consuming content. However, unlike the traditional avenues of content distribution, many platforms allow users to post and share content that interests them. Many worry that this

[28] David Mack, "This PSA About Fake News From Barack Obama Is Not What it Appears," *BuzzFeed News*, Apr. 17, 2018. https://bit.ly/2RqVtGA.

increased connectivity, combined with a lack of editorial discretion, will be uniquely suited to spread deepfakes at alarming rates.

This fear relies on the assumption that deepfakes will be categorically more harmful than other types of disinformation online. Yet there are reasons to question this assumption. Because deepfakes operate similarly to other forms of disinformation, the Committee should not treat deepfakes as a unique problem, but rather view them as part of the greater disinformation ecosystem.

The assumption that deepfakes are categorically more harmful than other forms of disinformation relies heavily on their realistic appearance. It therefore fails to consider factors outside of the apparent authenticity of the content that drive whether an individual will trust and share that content.

The first factor is that, in an online environment, people tend to identify the person who shared the content—rather than the content's creator—as the source of that content.[29] They then assign the content a degree of credibility based not on the apparent authenticity of the content itself, but on the degree of credibility they attribute to the "source" of that content. This tendency may stem from our evolutionary history, where our ancestors tended to place trust in members of their social groups because it increased the group's relative fitness.[30] Now, however, this trust can be exploited to spread disinformation.

The second factor is that people tend to trust and share information that conforms with a pre-existing worldview. Cognitive dissonance and confirmation bias can lead individuals to trust fake content that supports their pre-existing beliefs, or to discount real

[29] "'Who shared it?' How Americans Decide What News to Trust on Social Media." https://goo.gl/G5H748.

[30] Westling, "Deception and Trust: A Deep Look at Deep Fakes." https://bit.ly/2Y2vsjr.

evidence that contradicts their pre-existing beliefs.[31] For example, many Trump supporters claimed that a photograph depicting the turnout at the National Mall on Inauguration Day 2017 showed a fuller crowd than it actually did because they knew it was a photograph of Trump's inauguration.[32] They insisted this was the case even in spite of video evidence and a side-by-side photo comparison of the crowd at the Mall indicating the contrary. Having real, photographic evidence was irrelevant to the beliefs of many of these individuals.

The third factor is that the human mind does not need much convincing to believe that something is the truth.[33] Studies have shown that even rudimentary disinformation can generate inaccurate memories in subjects.[34] Researchers in one study, for instance, found that almost 30 percent of test subjects "remembered," either partially or fully, a fake event that the researchers described to them.[35] When more broadly applied, these findings indicate that realistic fake content is unnecessary to achieve the goals that more simplistic fake content, such as a story, can achieve.

Finally, with regard to online sharing specifically, some of the most viral content is actually the most simplistic. A standalone image, for example, is far simpler than a full-fledged video clip. Yet an image's ability to quickly convey an idea helps drive the sharing of that image, often virally to millions of users.[36] This effect is compounded by the fact

[31] Cognitive dissonance occurs when an individual holds two contradictory beliefs or ideas. Cooper, p. 6. Confirmation bias is the interpretation of evidence in ways that are partial to pre-existing beliefs. Raymond S. Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology* 2:2 (1998), p. 175. https://bit.ly/1BexVUv. When combined, these theories suggest that people may not believe authentic information that contradicts their pre-existing worldviews.

[32] Mahita Gajanan, "White House's Sean Spicer Stands by False Claim That Donald Trump's Inauguration Was 'Most-Watched' Ever," *Time*, Jan. 23, 2017. http://time.com/4643927/sean-spicer-white-house-donald-trump-inauguration-press-briefing/.

[33] Brainerd and Reyna, p. 52.

[34] Loftus, pp. 70–75. https://bit.ly/2FNzdlC.

[35] Ibid.

[36] Westling, "Deception & Trust: A Deep Look at Deep Fakes." https://bit.ly/2Y2vsjr.

that humans tend to display a form of herd behavior,[37] sharing certain content when they see others do the same.

It is these factors, not the apparent authenticity of the content, that drive people's tendency to believe and share content online. These factors are not unique to deepfakes; they are common to all forms of online disinformation. Therefore, regardless of how Congress decides to address disinformation, it should be careful not to use deepfakes as a justification for overbroad regulation—especially as it relates to the role of the platforms themselves—out of the fear that this technology presents an altogether a distinct problem.[38]

III. CONCLUSION

I thank the Committee for its attention to the matters presented above. If I can be of any assistance to members of the Committee, please feel free to contact me or my colleagues at the R Street Institute.

[37] Guille et al. https://bit.ly/2wFYs4X.

[38] Jeffrey Westling and Charles Duan, "R Sheet on Deep Fakes," *The R Street Institute*, March 2019. https://bit.ly/2Ipqlnh.