21st Century-Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and National Security

Robert Chesney
Danielle Keats Citron

Follow this and additional works at: https://digitalcommons.law.umaryland.edu/mlr
Part of the Constitutional Law Commons, Law and Society Commons, and the Privacy Law Commons

Recommended Citation
78 Md. L. Rev. 882 (2019)
Symposium

21ST CENTURY-STYLE TRUTH DECAY: DEEP FAKES AND THE CHALLENGE FOR PRIVACY, FREE EXPRESSION, AND NATIONAL SECURITY

ROBERT CHESNEY* & DANIELLE KEATS CITRON**

On February 1, 2019, the Maryland Law Review hosted a spectacular symposium entitled Truth Decay: Deep Fakes and the Implications for Privacy, National Security, and Democracy.1 The Law Review brought together scholars and advocates to discuss the deep-fake phenomenon and the looming challenges to privacy, civil liberties, national security, and civic trust. Before we begin our foreword to the symposium papers, we wanted to thank Executive Symposium Editor Meredith Storm, Editor-in-Chief Alexandra Botsaris, and Managing Editor Caroline Covington for their leadership, hard work, and enthusiasm.

Last spring, after telling them about our work on the topic, Meredith, Alexandra, and Caroline decided to dedicate the symposium to the deep-fake phenomenon. Of all of the cyber law issues grabbing headlines, from the loss of trust in social media companies to the daily drum beat of data breaches, they chose this topic. As the news from the weeks following the symposium showed, these women had foresight. We are grateful to the staff of the Law Review, especially Meredith, Alexandra, and Caroline, for having the vision, smarts, and devotion to bring together scholars with an array of perspectives to discuss the privacy, free expression, intellectual property, and national security challenges raised by the emergence of deep-fake technologies.

© 2019 Robert Chesney & Danielle Keats Citron.

* James Baker Chair in Law, University of Texas School of Law; Co-Founder, Lawfare.

** Morton & Sophia Macht Professor of Law, University of Maryland Francis King Carey School of Law; Affiliate Fellow, Yale Information Society Project; Affiliate Scholar, Stanford Center on Internet & Society. We are so grateful to the editors of the Maryland Law Review and to Sue Glueck for providing the Law Review with a Microsoft grant that allowed us to bring symposium participants to Baltimore. Deep thanks to Dean Donald Tobin for supporting the symposium.

On the day before the symposium, U.S. Senator Angus King appeared on MSNBC to discuss President Donald J. Trump’s response to recent testimony by U.S. intelligence chiefs. Earlier that week, in public congressional testimony, Director of National Intelligence Dan Coats and Director of the Central Intelligence Agency Gina Haspel expressed their disagreement with the President’s policy toward Syria and ISIS. President Trump responded swiftly to the video. Rather than criticizing them, the President simply asserted that neither Coats nor Haspel had disagreed with him. President Trump said, in so many words, that it was all fake news—that the officials said they agreed with his policies. When the TV host asked for a reaction to the President’s response, Senator King remarked,

Well, this reminds me of the old country song, Chris, who you going to believe, me or your own lying eyes? I mean, the testimony is there. I was there. I asked Gina Haspel very directly, is Iran in compliance with the nuclear agreement and she hemmed around a little bit but then she said yes it is. And that’s what their findings are.

The Senator was telling the hosts to watch the video of the proceedings to gauge the truth for themselves. In other words, videos do not lie.

Senator King’s response captures the belief in the truth-telling power of audio and video recordings. Even the United States Supreme Court has attested to this power: If a video shows someone driving recklessly, then that person drove recklessly. Video and audio recordings elicit a visceral response. Human beings believe what their eyes and ears are telling them. Video and audio evidence become inscribed as truths about the world around us and stories embed quickly into memory. Although human beings may be able to dismiss gossip and rumors, especially if the gossip and rumors do not accord with their worldview, watching and

2. Senator Angus S. King, Jr., Senator King’s Comments with Chris Hayes—January 31, 2019, YOUTUBE (Feb. 1, 2019), https://www.youtube.com/watch?v=ETGviBUGbAQ.


5. See id. (“Just concluded a great meeting with my Intel team in the Oval Office who told me that what they said on Tuesday at the Senate Hearing was mischaracterized by the media—and we are very much in agreement on Iran, ISIS, North Korea, etc. Their testimony was distorted press . . . .”).


8. That is, assuming nothing suggests the video has been altered or tampered in some way.

hearing video and audio recordings make it much more difficult to disregard the content. Senator King’s remarks reflect human experience: We ascribe truth-telling powers to video and audio recordings. This response to digital imagery is exactly why deep-fake technologies are so powerful.

DEEP-FAKE TECHNOLOGY

Faked video and audio recordings are not new—we have long had the ability to alter pictures and sounds. In the past, doctored videos were fairly easy to debunk. Deep-fake technology, however, is poised to produce significant change in how we differentiate what is real from what is fake. The technology can insert people’s faces and voices into video and audio doing and saying things they never did and said. Deep-fake technology—and in particular “generative adversarial network” (“GAN”) techniques—allow the wholesale creation of audio and video that is incredibly realistic and increasingly difficult to distinguish from real events.

The GAN technique involves neural network algorithms that learn to replicate patterns by sifting through large data sets. The generator algorithm mines source data (let’s say a trove of photos) for patterns, and it generates images or videos. Then, the generated video is subject to an adversarial algorithm, which looks for defects or artificial content. The generator then takes a turn, refining the video and eliminating errors. This process continues in a loop, producing highly realistic, yet fake audio and video content.

From what we have learned, the most difficult thing to replace is faces, so nothing is perfect now. But every time there is a breakthrough in the detection of deep fakes, there is a counter breakthrough with deep-fake technology that evades detection. White-hat technologists and black- or grey-hat technologists are locked in a battle. As technology emerges that detects the...

11. The GAN approach was invented by Google researcher Ian Goodfellow. See Ian Goodfellow et al., Generative Adversarial Nets, 27 ADVANCES NEURAL INFO. PROCESSING SYMS. 2672 (2014) (detailing the GAN approach); Chesney & Citron, supra note 1 (manuscript at 6).
14. Karras, supra note 13, at 1–5; Chesney & Citron, supra note 1 (manuscript at 6).
15. Thank you to Andreas Schou of Google for talking to us about the technology.
latest deep-fake technology (say the technique finds that veins in the neck are not pulsating), a better version of the deep fake hits the scene. It is unclear who will win this arms race, but for now the fight is on.

So, that is an overview (simplified, of course) of the technology. Now to discuss the concerns that drove us—a national security expert and a privacy and free speech expert—to join forces to write about deep fakes. Let’s begin by exploring the different kinds of harm that we might see for individuals and society. Then, we will contemplate the modest role for law and the potential for market responses. We will end by discussing what society can do as legal responses and market solutions evolve.

INDIVIDUAL EXPLOITATION AND SABOTAGE

Individuals can be targeted with damaging deep-fake videos. Let’s begin as the deep-fake trend did—no surprise, with deep-fake sex videos. About eight months ago, Motherboard published a story about a subreddit devoted to deep-fake sex videos of celebrities. At the time, the subreddit (now closed) had countless threads. Redditors posted deep-fake sex videos of Gal Gadot and Taylor Swift. These commentators asked each other questions about deep-fake technology. A commentator said he wanted to make a deep-fake sex video of his ex-girlfriend, but he only had thirty photographs of her. He wondered if the fake app technology available on a desktop would work. Commentators also directed one another to YouTube tutorials providing instructions on the creation of deep-fake videos.

Consider the experience of Noelle Martine. Ms. Martine was a high-school student in Australia when she began using social media to post videos and photos of herself. Someone—a stranger, she thinks—inserted her face


into pornographic images and posted them online. Shortly after, posters revealed her home address and cell phone number. Not long thereafter, a deep-fake sex video of Ms. Martin appeared, showing her performing oral sex on someone. She was inundated with death and rape threats, and strangers contacted her for sex. Keep in mind that Ms. Martin was just a high-school student. Ms. Martin contacted law enforcement and government agencies but was told that nothing could be done about the postings. Instead, law enforcement suggested she contact the webmasters of the numerous websites on which the false images were displayed and request they be taken down.21

Rana Ayyub, a journalist in India, wrote about Hindu nationalist politics. Her pieces explored concerns about corruption in Hindu national political parties. A deep-fake sex video of her appeared and went viral. The video was shared via Twitter and text messages. It appeared next to her home address alongside her phone number and one phrase: “I am available.”22 Ms. Ayyub went into hiding. She was deluged with death and rape threats. For weeks, she could not write, let alone speak. She could barely eat. She was terrified.23

Deep-fake sex videos are likely to have a disproportionate impact on women and marginalized communities, as is true of other sexual-privacy invasions.24 As Ms. Martin and Ms. Ayyub underscored, it is incredibly disturbing to be turned into a sex object. Deep-fake sex videos reduce individuals to genitalia, breasts, buttocks, and anus, creating a sexual identity not of the individual’s own making.25 If a deep-fake sex video appears in a Google search of one’s name, it can be impossible to get or keep a job. It can have a profound impact on one’s social life and sense of safety.

Deep-fake videos could also be used to sabotage corporate CEOs and their companies. Imagine the night before an Initial Public Offering, a deep-fake video appears online showing the company’s CEO soliciting a child prostitute or saying something he shouldn’t say in a way that upends the public offering of the company’s stock. Yes, the video could be debunked in a few days, but the damage has already been done. If released at a crucial time,

23. Id. .
a deep-fake video could destroy the marketplace’s faith in a CEO or company. Depending on the timing of the release, a deep-fake video can hijack people’s lives and companies’ fortunes.

NATIONAL SECURITY PERILS

Deep-fake videos can have profound implications for national security. For example, in the middle of sensitive negotiations, diplomats could be shown a deep-fake video suggesting an adversary is disingenuous. The video could scuttle the diplomacy. Far more violent possibilities are implicated as well. Imagine a deep-fake video released in Gaza showing Israeli soldiers murdering a Palestinian child. Violence could explode, with devastating consequences for regional stability.

Let’s consider the City of Baltimore the day after Freddie Gray was killed in police custody. City residents were overrun with grief, especially the black community. The black community was justifiably angry and frustrated after the death of Mr. Gray. False arrests and mistreatment by police officers were routine in Baltimore City. Imagine if a deep-fake video appeared of the police officers responsible for Mr. Gray’s death in which they said they were ordered to kill Mr. Gray. As most readers know, the day after Mr. Gray’s death was characterized by protests and civil unrest. If such a deep-fake video had appeared and gone viral, we might have seen far more violence and disruption in Baltimore. If the timing was just right and the video sufficiently inflammatory, we might have seen greater destruction of property and possibly of lives.

TRUTH DECAY AND ELECTIONS

Deep-fake technology is emerging at a difficult time. The public is increasingly distrustful of the media and public officials (at least the public officials who are not aligned with their preferences). Deep-fake videos exacerbate distrust in civic and democratic institutions.

26. Panelists Benjamin Wittes, Quinta Jurecic, and Alan Rozenshtein explored national security issues at the Maryland Law Review Symposium, including whether we might see offensive counterintelligence measures involving deep-fake content. See Maryland Carey Law, supra note 24.


29. At the symposium, panelists Jessica Silbey and Woodrow Hartzog raised important concerns about privacy and free expression related to deep fakes. Deep fakes decontextualize narratives
A deep-fake video could hijack elections, eroding faith in our democracy. What if the night before the 2018 election a deep-fake video showed a candidate locked in a tight race engaging in anti-social behavior that he never engaged in. The national security community is wrestling with concerns about deep fakes in connection with the 2020 elections. This does not mean that a deep-fake video cannot be debunked in time. But a well-timed deep fake could change an election’s outcome—the damage would be irreparable. Elections usually cannot be undone. It is easy to imagine the potential for deep disruption to our democracy.

Further, our shared sense of reality is in jeopardy. A healthy democracy requires a shared set of truths and facts for citizens to consider, debate, and ponder. Deep-fake videos undermine the possibility of having conversations about a shared reality and exacerbate the disinformation wars that disrupt democratic politics.

Then there is what we have called the “Liar’s Dividend.” The problem is not just that deep-fake content can be used to stoke social, political, and ideological divisions. As people become more aware of the existence of deep fakes, wrongdoers may find it easier to cast doubt on real recordings of their mischief. Recall that President Trump has tried to leverage this possibility. After the 2016 election, he denounced the Lester Holt interview (where he admitted that he fired FBI Director James Comey because of that “Russia thing”) and the Access Hollywood tape (where he said when you are a star you can “grab ‘em by the pussy”) as fakes. If the public were more sensitized to the issue of deep-fake content, perhaps his assertions would be believed. Our society is imperiled when people can escape accountability for their words and actions by ascribing genuine audio and video content to deep-fake technology.

and so the question they raised is how we might as citizens resituate and recontextualize the fakery rather than regulate them. See Jessica Silbey & Woodrow Hartzog, The Upside of Deep Fakes, 78 Md. L. Rev. 960 (2019); Maryland Carey Law, supra note 24.


31. Chesney & Citron, supra note 1 (manuscript at 28).


Looking to the Future

Unfortunately, there are no easy answers to this difficult problem. While several legal and technological approaches could help mitigate the threat, none will eliminate the problem. As thoughtful computer scientist and privacy policy wonk Edward Felten once explained, with hard tech policy problems, lawyers tend to point to technologists for the answers, and the technologists tend to point to the lawyers.\textsuperscript{35} We need both lawyers and technologists to tackle the deep-fake problem. And we need a heavy dose of societal resilience to make our way through these concerns.

We should be watching advances in authentication technologies.\textsuperscript{36} Companies like Truepic are working on methods of authentication.\textsuperscript{37} If those methods are adopted broadly (a big “if”), it would help quickly authenticate content. The problem is that platforms are likely to continue posting content with few restrictions on provenance. Unless platforms refuse to post content without certain authentication markers (highly unlikely given the “more eyeballs” advertising business model), there is little reason to think that a given authentication strategy will be widely adopted. As explored at the outset, because forensic technology is engaged in an arms race, there is little reason to place our hopes on a slam-dunk technical solution.

What about the law?\textsuperscript{38} Individuals could sue deep-fake creators for defamation or intentional infliction of emotional distress. But for civil claims to work, harmed individuals would need to find perpetrators, those perpetrators would have to live in the United States (for the court to have any chance of having jurisdiction over them), and individuals would have to have funds to pay for counsel. Given those hurdles, civil litigation is not a practical response.\textsuperscript{39} There are criminal laws related to impersonation that would be fruitful so long as law enforcement has training in the law and technology.\textsuperscript{40} The First Amendment would likely countenance prosecutions for deliberate


\textsuperscript{36} See Chesney & Citron, supra note 1 (manuscript at 30–31) (discussing technological solutions).


\textsuperscript{38} See Chesney & Citron, supra note 1 (manuscript at 31–45) (discussing legal solutions).

\textsuperscript{39} See id. (manuscript at 33–41) (considering the possibility of civil liability).

\textsuperscript{40} See id. (manuscript at 42–45) (addressing categories of criminal liability). Panelist Mary Anne Franks discussed the possibility of amending identity theft statutes to address deep fake sex videos. Professor Franks is currently working with members of Congress on such an idea. Senator Ben Sasse has proposed a federal criminal statute that would extend to individuals and platforms that publish deep fakes knowing that they are fake and knowing that they enable crimes or torts. See Malicious Deep Fake Prohibition Act of 2018, S. 3805, 115th Cong. (2018). Members of his staff attended the symposium, and it sparked important conversations about the free speech implications as well as the potential for platform liability.
impersonations of individuals. We can punish deliberate lies because they concern the source of who is speaking. Proscribing those lies is permissible because such lies threaten significant harm to listeners who rely on them as a proxy for reliability and credibility.41

Now to platforms.42 Under current law, online service providers enjoy a broad sweeping immunity from liability for user-generated content. As one of us has argued with fellow symposium participants Benjamin Wittes43 and Quinta Jurecic,44 federal immunity should condition the immunity on reasonable moderation practices rather than the free pass that exists today. The current interpretation of that federal statute—Section 230 of the Communications Decency Act—leaves platforms with no incentive to address destructive deep-fake content.45 That should change.46

What about the possibility of people protecting themselves with authentication services?47 Companies might begin offering alibi services—the practice of recording every aspect of one’s life. For instance, imagine an amplified Google Glass that focused on oneself at all times, providing twenty-four-hour surveillance of one’s activities. Such services would help counter the deep fake by allowing a person to refute any allegations with recorded surveillance. One could respond to a deep-fake video with the following: “I was not there saying or doing that terrible thing because my alibi services show that I was miles away doing and saying something else.” As Scott Peppet suggests, however, lifelogging would be deeply invasive and risks the broader unraveling of privacy.48 Even if only a small number of people took up lifelogging, they would produce vast reservoirs of personal information.


42. See Chesney & Citron, supra note 1 (manuscript at 36–41) (considering liability for platforms).


46. Panelists Olivier Sylvain and Stacey Dogan raised crucial questions about platform liability in the face of deep fakes. See Maryland Carey Law, supra note 24.

47. See Chesney & Citron, supra note 1 (manuscript at 53–58) (anticipating market solutions).

data in which the rest of us would find ourselves inadvertently caught, creating a massive peer-to-peer surveillance network for constantly recording our activities.

We have far more questions than answers. That is why we wanted to bring together the smartest people we know to talk about the privacy, free speech, intellectual property, and national security implications of deep-fake content. And that we did at the Maryland Law Review Spring 2019 Symposium—we owe a debt of gratitude to the participants in the discussion and to Microsoft’s Sue Glueck who ensured that we could bring everyone together at the University of Maryland Francis King Carey School of Law in Baltimore.