11-15-2019

# The Legal Implications of Synthetic and Manipulated Media

Thomas E. Kadri
*University of Georgia School of Law*, tek@uga.edu

# The Legal Implications of Synthetic and Manipulated Media

**Thomas E. Kadri**

**Carnegie Endowment for International Peace – November 15, 2019**

---

Ahead of the U.S. 2020 presidential election, the Carnegie Endowment for International Peace convened more than 100 experts from three dozen organizations inside and outside Silicon Valley in private meetings to help address the challenges that synthetic and manipulated media pose for industry, government, and society more broadly. Among other things, the meetings developed a common understanding of the potential for synthetic and manipulated media circulated on technology platforms to disrupt the upcoming presidential election, generated definitions of "inappropriate" election-related synthetic and manipulated media that have informed platform content moderation policies, and equipped platforms with playbooks of effective and ethical responses to synthetic and manipulated media. Carnegie commissioned four short papers on the legal, ethical, and efficacy dimensions of election-related synthetic and manipulated media to brief meeting participants, and now the broader public, on the state-of-play.

---

Digital falsifications pose dangers for social media, governments, and society. In particular, the rise of "digitized impersonations" increasingly concern lawmakers and scholars who recognize the risks they pose to both individuals and society.[1] To address these risks, the Carnegie Endowment for International Peace convened a series of meetings aimed at reducing opportunities for digital forgeries to subvert the upcoming 2020 U.S. election. This memo informs the series by outlining the potential legal implications of synthetic or manipulated media with a view to helping platforms define what constitutes proper and improper digital falsifications in the context of the election.

Media can take various forms, and rapidly developing technology will surely lead to new types in the future. This memo focuses on just two kinds: synthetic media and manipulated media. For the purposes of this series, "synthetic media"—sometimes called deepfakes—are digital falsifications of images, video, and audio created using an editing process that is automated through AI techniques, whereas

"manipulated media" are any other digital falsification of images, video, and audio.[2]

Not all uses of synthetic or manipulated media are harmful. Indeed, they can serve many laudable purposes. Consider, for example, the enhancements they could bring in the realm of education. In teaching about the assassination of former president John F. Kennedy, these media could allow people to hear the speech he was due to give on the day of his death in his own voice, as one UK-based company has now done.[3] Similarly, imagine the powerful artistic uses of these media, such as the digital manipulation in Forrest Gump where doctored video footage portrayed three past presidents saying things they never said.[4] Synthetic and manipulated media can also enhance autonomy and equality, particularly for people with disabilities who might use the technology to virtually engage with life experiences that would be impossible in a conventional sense.[5] Moreover, these media can spur valuable political speech, as when a Belgian political party created a deepfake depicting U.S. President Donald Trump giving a fictional address where he says: "As you know I had the balls to withdraw from the Paris climate agreement. And so should you."[6] Although Trump never used those words in abandoning the agreement, the political party used this tool to "start a public debate" to "draw attention to the necessity to act on climate change."[7]
But some digital falsifications are not so salutary. Indeed, many uses could lead to grave individual and social harms—particularly in the political context. Consider this list of hypothetical scenarios catalogued by Robert Chesney and Danielle Citron:

- Fake videos could feature public officials taking bribes, displaying racism, or engaging in adultery.
- Politicians and other government officials could appear in locations where they were not, saying or doing horrific things that they did not.
- Fake videos could place them in meetings with spies or criminals, launching public outrage, criminal investigations, or both.
- Soldiers could be shown murdering innocent civilians in a war zone, precipitating waves of violence and even strategic harms to a war effort.
- A deep fake might falsely depict a white police officer shooting an unarmed black man while shouting racial epithets.
- A fake audio clip might "reveal" criminal behavior by a candidate on the eve of an election.
- Falsified video appearing to show a Muslim man at a local mosque celebrating the Islamic State could stoke distrust of, or even violence against, that community.

- A fake video might portray an Israeli official doing or saying something so inflammatory as to cause riots in neighboring countries, potentially disrupting diplomatic ties or sparking a wave of violence.
- False audio might convincingly depict U.S. officials privately "admitting" a plan to commit an outrage overseas, exquisitely timed to disrupt an important diplomatic initiative.
- A fake video might depict emergency officials "announcing" an impending missile strike on Los Angeles or an emergent pandemic in New York City, provoking panic and worse.[8]

Falsifications like these could spread with devastating effect during election season. They could erode the public's sense of trust in the news—or even in the very idea of truth—upon which an informed electorate depends. Worse still, a well-timed release of a convincing digital falsification could sway an election if enough voters believed it and the candidate had no time to debunk it effectively.

What are the potential legal responses to digital falsifications? An outright legal ban on synthetic and manipulated media would violate the First Amendment because "falsity alone" does not remove expression from First Amendment protection, and many digital falsifications would be constitutionally protected speech.[9] As a result, the mere specter of the First Amendment curtails many legislative efforts to regulate these media. Nevertheless, the following legal regimes have the potential to address certain problems posed by digital falsifications in ways consistent with the First Amendment.[10]

Intellectual Property: One potential source of legal liability could be copyright law. Because some digital falsifications rely on copyrighted content, unauthorized use of that content could lead to monetary damages and a notice-and-takedown procedure to remove it. The person who created the content usually owns the copyright, and thus a person may have a legal claim if she is depicted in synthetic or manipulated media that uses material of her own creation. Significant legal hurdles will arise, however, because defendants will argue that the falsification is "fair use" of the copyrighted material, intended for educational, artistic, or other

expressive purposes—a defense to liability under copyright law that in part turns on the question whether the falsification is sufficiently "transformed" from the original such that it receives protection.

Right of Publicity: Another option might be the right of publicity—a state-law tort that prohibits unauthorized use of a person's name, likeness, or other indicia of identity.[11] Again, however, many digital falsifications will be immune from liability under this tort because of First Amendment concerns, as well as related statutory and common-law exceptions for material that is "newsworthy" or in the "public interest."[12] Some states also restrict the tort's scope to "commercial" uses of a person's identity, such as in advertisements, meaning that many digital falsifications in the election context will not be covered. Despite these constitutional barriers, claims brought against creators of digital falsifications that inflict grave dignitary harms might survive First Amendment scrutiny, though this theory has not been tested in the courts.[13]

Defamation & False Light: A more fruitful avenue might be civil tort claims for defamation and false light, which target certain types of falsehoods. Public officials and public figures could sue if a convincing digital falsification amounted to a defamatory statement of fact made with actual malice—that is, made with knowledge that the statement was false or with reckless disregard as to its falsity.[14] Private individuals, meanwhile, need show only that the creator was negligent as to the falsity of any defamatory statement. Similarly, liability could arise if a digital falsification places a person in a "false light" by creating a harmful and false implication in the public's eye, though the victim would have to show actual malice if the falsification was related to a "matter of public concern."[15]
Intentional Infliction of Emotional Distress (IIED): A final tort that might come into play is IIED, but only if the creator of a piece of synthetic or manipulated media "intentionally or recklessly engaged in extreme and outrageous conduct that caused the plaintiff to suffer severe emotional distress."[16] This is a high bar to meet, made higher by First Amendment concerns: as in defamation, a public

figure's IIED claim resting on allegations of falsity must satisfy the strictures of actual malice, and there is also robust constitutional protection for satire and speech on matters of public concern.[17]

Criminal Law: Some digital falsifications might implicate various criminal laws. If a digital falsification targeted particular individuals by using any "interactive computer service or electronic communication system" to "intimidate" them in ways "reasonably expected to cause substantial emotional distress," the creator might have violated federal cyberstalking law.[18] Some states also make it a crime to knowingly and credibly impersonate another person online with intent to "harm, intimidate, threaten, or defraud" that person,[19] and it is a federal crime to impersonate a federal official in order to defraud others of something of value.[20] A few states also have criminal defamation laws, though prosecutions under these laws must at a minimum satisfy the same constitutional standards as the civil defamation claims discussed above.[21] Finally, some states have criminalized the intentional use of lies to impact elections, but most of these laws have been struck down as unconstitutional.[22]

Administrative Law: There may be narrow circumstances in which digital falsifications could be addressed through administrative law. The Federal Trade Commission could regulate synthetic or manipulated media that amount to deceptive or unfair commercial acts and practices, but this remit would likely cover only those media that take the form of advertising related to "food, drugs, devices, services, or cosmetics."[23] Although the Federal Communications Commission might seem like a better fit, that agency currently appears to lack both jurisdiction and interest to regulate content circulated on social media.[24] Lastly, the Federal Election Commission is empowered to regulate campaign speech, but it does not regulate the truth of campaign-related speech, nor is it likely to assert or receive this power due to the constitutional, practical, and political concerns that would accompany such efforts.[25] There are election-related rules concerning financing—for example, regulations demanding transparency of funding for political

advertisements—but social media are not currently subject to jurisdiction in this context.[26] This might change if Congress adopts the Honest Ads Act, but efforts appear to have stalled on that front for now.[27]

Five final points are essential to understanding the legal landscape around digital falsifications. First, difficulties of attribution will often impede attempts to hold creators of harmful falsifications liable; tracking down the people who create them is usually difficult and costly. Second, and relatedly, perpetrators often live outside of the United States and thus may be beyond the reach of the U.S. legal process. Third, it can be expensive and risky to bring civil claims, and victims of harmful falsifications may fear that litigation will trigger even more unwanted attention—sometimes known as the Streisand effect.[28]

Fourth, legal liability may depend on whether a digital falsification is believable, but each case will present different issues on this front. For example, if a deepfake portrayed a presidential candidate saying something racist, she would likely have to show that people reasonably believed she made the racist statements in order to successfully bring a defamation claim. If the deepfake were unbelievable, courts would more likely view it as satire or parody and thus deem it protected under the First Amendment.[29] Although the relevance of believability will depend on the type of legal claim and the facts of each case, it is safe to say that believable falsifications are both more likely to be legally problematic and less likely to receive First Amendment protection.

Last but certainly not least, Section 230 of the Communications Decency Act will largely immunize social media from most of the potential legal liability discussed in this memo. If a third party posts a digital falsification on an online platform, the platform cannot be held liable for hosting it even if the third party could be, unless hosting the content violates federal criminal or intellectual property law. At the very least, this means that platforms are not legally responsible for user-generated

falsifications that would otherwise run afoul of laws concerning the right of publicity, defamation, false light, or IIED.

Section 230 is especially important here in two respects. First, platforms cannot be sued for displaying most content republished from other sources or generated by users because the law expressly prohibits courts from treating platforms as "publishers" of that content. Second, platforms can filter and block whatever content they like without fear that they will be liable for leaving up some types of content while taking down others. This combination gives platforms wide discretion to allow or prohibit digital falsifications as they see fit. Ultimately, due to a combination of Section 230 and the First Amendment, platforms will be largely free to regulate digital falsifications however they wish as a matter of private governance of online speech.

Thomas E. Kadri is a resident fellow at Yale Law School and an adjunct professor at New York Law School.

***NOTES***

1 Robert Chesney and Danielle Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," California Law Review 107 (forthcoming 2019), draft available at https://ssrn.com/abstract=3213954. Chesney and Citron's path-breaking article on the issue of deepfakes is worth reading in full and supports much of this memo.

2 See ibid.; and James Vincent, "Why We Need a Better Definition of 'Deepfake,'" Verge, May 22, 2018, https://www.theverge.com/2018/5/22/17380306/deepfake-definition-ai-manipulation-fake-news.

3 "John F Kennedy's Lost 'Last' Speech Recreated," BBC, March 16, 2018, https://www.bbc.com/news/av/uk-scotland-43436361/john-f-kennedy-s-lost-last-speech-recreated.

4 "Forrest Gump JFK 'I Gotta Pee' Scene," YouTube, posted by Rnastershake, February 28, 2010, https://www.youtube.com/watch?v=JSEdBNslGOk; "Forrest Gump - Shot In The Buttocks," YouTube, posted by TheGameCube64Guy,

January 14, 2013, https://www.youtube.com/watch?v=mIWd3T1xjec; and "Jan. 9: Richard Nixon (with Forrest Gump)," YouTube, posted by Birthday Hall of Fame, January 9, 2015, https://www.youtube.com/watch?v=qJpDMPnmrBU.

5 Allie Volpe, "Deepfake Porn Has Terrifying Implications. But What If It Could Be Used for Good?," Men's Health, April 13, 2018, https://www.menshealth.com/sex-women/a19755663/deepfakes-porn-reddit-pornhub (giving the example of people suffering from physical disabilities interposing their faces along with those of their consenting partners into pornographic videos); see also Chesney and Citron, "Deep Fakes."

6 Hans von der Burchard, "Belgian Socialist Party Circulates 'Deep Fake' Donald Trump Video," Politico, May 21, 2018, https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video/.

7 Ibid.

8 Chesney and Citron, "Deep Fakes."

9 See United States v. Alvarez, 567 U.S. 709, 719 (2012) (plurality opinion).

10 See generally Alan Chen & Justin Marceau, "High Value Lies, Ugly Truths, and the First Amendment," Vanderbilt Law Review 68, no. 6 (2015): 1480.

11 See Thomas E. Kadri, "Fumbling the First Amendment: The Right of Publicity Goes 2–0 Against Freedom of Expression," Michigan Law Review 112 (2014): 1519, 1521.

12 See generally Thomas E. Kadri, "Drawing Trump Naked: Curbing the Right of Publicity to Protect Public Discourse," Maryland Law Review 78 (2019): 899, 928–31.

13 See ibid. at 948–58.

14 See New York Times Co. v. Sullivan, 376 U.S. 254 (1964).

15 See Time, Inc. v. Hill, 385 U.S. 374, 390 (1967).

16 See Snyder v. Phelps, 562 U.S. 443, 451 (2011)).

17 See Hustler Magazine v. Falwell, 485 U.S. 46 (1988); Snyder, 562 U.S. at 451–54.

18 18 U.S.C. § 2261A(2).

[19] See, for example, California Penal Code § 528.5 (2009).
[20] 18 U.S.C. § 912 (2009) ("Whoever falsely assumes or pretends to be an officer or employee acting under the authority of the United States or any department, agency or officer thereof, and acts as such, . . . shall be fined under this title or imprisoned not more than three years, or both.").
[21] See Eugene Volokh, "Anti-Libel Injunctions and the Criminal Libel Connection," University of Pennsylvania Law Review 167 (forthcoming 2019) (draft available at https://ssrn.com/abstract=3372064).
[22] See, for example, Susan B. Anthony List v. Driehaus, 814 F.3d 466 (6th Cir. 2016) (striking down an Ohio election-lies law as a content-based restriction of "core political speech" that lacked sufficient tailoring).
[23] 5 U.S.C. § 52.
[24] Chesney and Citron, "Deep Fakes."
[25] Ibid.
[26] Ibid.
[27] See Alexander Bolton, "McConnell Works to Freeze Support for Dem Campaign Finance Effort," Hill, March 8, 2019, https://thehill.com/homenews/senate/433154-mcconnell-works-to-freeze-support-for-dem-campaign-finance-effort.
[28] See Julie E. Cohen, "Law for the Platform Economy," UC Davis Law Review 51 (2017): 133, 149–50 ("Efforts to remove hurtful material typically backfire by drawing additional attention to it, intensifying and prolonging the unwanted exposure").
[29] See Hustler Magazine, 485 U.S. at. 57.