Missouri Law Review

Volume 86
Issue 1 *Winter 2021*

Article 12

Winter 2021

Still Waters Run Deep(fakes): The Rising Concerns of "Deepfake" Technology and Its Influence on Democracy and the First Amendment

Lindsey Wilkerson

LAW SUMMARY

Still Waters Run Deep(fakes): The Rising Concerns of "Deepfake" Technology and Its Influence on Democracy and the First Amendment

Lindsey Wilkerson[*]

I. INTRODUCTION

"A video circulated that suggested that I was a pedophile. What do you say to that? You go on Twitter and argue you're not a pedophile? I mean, there's no excuse for that, no defense."[1] During the Obama Administration, Michael McFaul served as the U.S. ambassador to Moscow. During his term, he claimed he was the subject of Russian propaganda "[that] accused [him] of plotting to overthrow leader Vladimir Putin as well as pedophilia."[2] In February 2018, McFaul said an altered video was posted on YouTube falsely showing him walking the streets of Moscow with a convicted pedophile.[3] McFaul's claims, if true, would be the first case of a government official having their reputation attacked by "digitally constructed videos that can

[*]Bachelors of Journalism, University of Missouri, 2019; J.D. Candidate, University of Missouri School of Law, 2021; Associate Member, *Missouri Law Review*, 2019-2020. I am grateful to Professor Wells for her insight, guidance, and support during the writing of this summary, as well as the *Missouri Law Review* for its help in the editing process.

1. This quote is from former U.S. Ambassador to Russia Michael McFaul when he appeared on the CBC Radio show, "The Current." *The fight against 'deepfake' videos includes former U.S. ambassador to Russia Michael McFaul*, CBC RADIO (July 20, 2018), https://www.cbc.ca/radio/thecurrent/the-current-for-july-20-2018-1.4754632/the-fight-against-deepfake-videos-includes-former-u-s-ambassador-to-russia-michael-mcfaul-1.4754674 [https://perma.cc/87N7-8974] [hereinafter "The Current"].

2. Laura King & Sabra Ayres, *What you need to know about Michael McFaul, the ex-U.S. envoy drawn into the center of another Trump-Russia flap*, L.A. TIMES (July 19, 2018, 3:15 PM), https://www.latimes.com/world/la-fg-russia-mcfaul-20180719-story.html [https://perma.cc/QAL6-BFAN].

3. Michael McFaul, *The Smear that Killed the 'Reset': Putin Needed an American Enemy. He Picked Me.,* WASH. POST (May 11, 2018), https://www.washingtonpost.com/news/posteverything/wp/2018/05/11/feature/putin-needed-an-american-enemy-he-picked-me/ [https://perma.cc/4VAD-FQV5].

make it appear that a person is saying or doing something they never did," also known as "deepfakes."[4]

This Note explores how deepfake technology can disrupt democracy and influence elections through the protections given to political speech under the First Amendment. Part II describes deepfakes in greater detail and identifies the wide uses for deepfake technology. Part III reflects on how the federal government and states are attempting to regulate deepfakes, mainly to protect individuals from pornographic exploitation and election tampering. Finally, Part IV discusses in detail how the First Amendment creates constitutional barriers in regulating deepfakes.

II. LEGAL BACKGROUND

Deepfakes are a new and evolving form of technology that allow people to make manipulated videos that can have potentially devastating impacts if used in the wrong hands. While the technology behind deepfakes is complex, the ability to make deepfakes is only becoming more accessible as time passes. There are plenty of uses for deepfakes; they could jeopardize individual privacy, fair elections, and perhaps democracy overall. But there are some benefits from the use of deepfake technology, including the promotion of self-expression – a hallmark of the First Amendment. Because of this, deepfakes can be viewed through the lens of the First Amendment, particularly parodies and the protection of lies.

A. What are "Deepfakes"?

Deepfakes – sometimes stylized as "deep fakes" or "deep-fakes" – are videos that are digitally manipulated to make it look like a person "is realistically saying or doing something they didn't."[5] The new technology has been spreading virally on the internet for various reasons, including pornography and parody.[6] Some political and legal experts are concerned, however, about deepfake technology being used in the near future to tamper with elections across the globe.[7]

---

4. The Current, *supra* note 1.

5. Benhamin Goggin, *From porn to Game of Thrones: How deepfakes and realistic-looking fake videos hit it big*, BUSINESS INSIDER (Jun. 23, 2019, 9:45 AM), https://www.businessinsider.com/deepfakes-explained-the-rise-of-fake-realistic-videos-online-2019-6 [https://perma.cc/BE7E-T464].

6. *Id.*; *see also* Cntl Shift Face, *Better Call Trump: Money Laundering 101 [DeepFake]*, YOUTUBE (Sept. 18, 2019), https://www.youtube.com/watch?v=Ho9h0ouemWQ [https://perma.cc/HS9T-RAGP] (showing a deepfake of President Donald Trump as one of the characters in a well-known scene in the TV show "Breaking Bad").

7. Grace Shao, *Fake videos could be the next big problem in the 2020 elections*, CNBC (Jan. 17, 2020, 2:49 AM), https://www.cnbc.com/2019/10/15/deepfakes-could-be-problem-for-the-2020-election.html [https://perma.cc/KM5V-39DY] (quoting John Villasenor, a professor at UCLA focusing on artificial intelligence and

Deepfake technology has its roots in pornography; it gained popularity in 2017 when lewd videos surfaced showing celebrity faces on pornographic actors and actresses.[8] The number of deepfakes online doubled between 2018 and 2019.[9] With the increasing prominence of deepfakes, "some of the most influential people in the world, and their audiences, have become targets of deepfakers."[10] The first viral deepfake was a pornographic video that featured *Wonder Woman* lead actress Gal Gadot's face digitally transplanted on top of the face of an actual pornographic actress.[11] The video was first posted on the social media website Reddit by a user named "deepfakes," thus coining the title "deepfake" for these types of videos.[12] Deepfakes can be freakishly realistic since they are "trained on hours of footage, [and have] been specifically generated for its context, with seamless mouth and head movements and appropriate coloration."[13] As technology has progressed, deepfakes have become easier to create since anyone with "a computer, internet access, and interest in influencing an election" can make one.[14] Some apps have even been developed that allow smartphone users to create deepfakes at the touch of their fingertips.[15]

---

cybersecurity); Danielle Citron, *How deepfakes undermine truth and threaten democracy*, TED, https://www.ted.com/talks/danielle_citron_how_deepfakes_undermine_truth_and_threaten_democracy/transcript?utm_source=twitter.com&utm_medium=social&utm_campaign=tedspread [https://perma.cc/VYQ4-3BLV] (last visited Feb. 13, 2020).

8. Goggin, *supra* note 5.

9. HENRY AJDER ET. AL, DEEPTRACE, THE STATE OF DEEPFAKES: LANDSCAPE, THREATS, AND IMPACT, 1 (SEPT. 2019). Deeptrace, the organization behind this report, started as a way to understand, research, and identify deepfakes. *Id.*

10. Goggin, *supra* note 5.

11. Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All F*****d*, VICE (Dec. 11, 2017, 1:18 PM), https://www.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn [https://perma.cc/GR9C-9QNW] (The title of this article has been edited to remove profane language).

12. Goggin, *supra* note 5.

13. *Id.*; *see also* Supasorn Suwajanakorn, *Synthesizing Obama: Learning Lip Sync from Audio,* YOUTUBE (July 11, 2017), https://www.youtube.com/watch?v=9Yq67CjDqvw [https://perma.cc/GR9C-9QNW] (showing a step-by-step process used by University of Washington research students to make a deepfake of President Barack Obama).

14. Shao, *supra* note 7.

15. *See, e.g.*, Ivan Mehta, *New Deepfake App Pastes Your Face onto GIFs in Seconds*, NEXT WEB (Jan. 13, 2020, 5:04 AM), https://thenextweb.com/artificial-intelligence/2020/01/13/new-deepfake-app-pastes-your-face-onto-gifs-in-seconds/ [https://perma.cc/E4CV-DEB5]; Zak Doffman, *Chinese Deepfake App ZAO Goes Viral, Privacy Of Millions 'At Risk'*, FORBES (Sept. 2, 2019, 4:27 AM), https://www.forbes.com/sites/zakdoffman/2019/09/02/chinese-best-ever-deepfake-app-zao-sparks-huge-faceapp-like-privacy-storm/#61486e068470 [https://perma.cc/92D9-ENVZ].

B. What are the Main Uses and Threats of Deepfakes?

While deepfakes originated in the pornography industry, there are growing concerns about how they could interfere with politics and elections and threaten individual privacy. Maybe it is not surprising that with the growth and general accessibility of any video-editing software, the "Average Joe" is able to manipulate videos in order to create a false perception of another person.[16] Perhaps what is more disturbing, however, is the reaction and confusion that these altered videos can create.[17] This Part discusses the threats to democracy, elections, and individual privacy that deepfakes may cause, while also noting how they can be protected from regulation by arguments for creative freedom and self-expression.

1. Threats to Democracy and Elections

Deepfakes could potentially change the political sphere, spin elections, and increase the dissemination of "fake news." Deepfakes have been described by some as "a powerful new tool for those who might want to (use) misinformation to influence an election."[18] One possible example of this was provided by privacy law scholar Danielle Citron at a TED Talk about deepfakes.[19] She described a hypothetical where, the night before an election, a deepfake spreads online showing one of the major party candidates had fallen ill.[20] Citron claimed that the deepfake "could tip the election and shake our sense that elections are legitimate."[21] When placed in the wrong hands, "deepfakes have the potential to corrode the trust that we have in democratic institutions."[22]

Citron is not the only scholar worried about this problem; in fact, news outlets like *CNBC* and *The Guardian* have reported the same concerns.[23] They compare the spread of deepfakes to the threat of fake news in the 2016

16. Drew Harwell, *Faked Pelosi Videos, Slowed to Make Her Appear Drunk, Spread Across Social Media*, WASH. POST (May 23, 2019 3:41 PM), https://www.washingtonpost.com/technology/2019/05/23/faked-pelosi-videos-slowed-make-her-appear-drunk-spread-across-social-media/ [https://perma.cc/UX2L-RJKL] (discussing how an altered video of House Speaker Nancy Pelosi, which made it appear like she was "drunkenly slurring her words," created confusion and prompted ridicule online).

17. *Id.* (same).

18. Shao, *supra* note 7 (quoting John Villasenor, a professor at UCLA focusing on artificial intelligence and cybersecurity).

19. Citron, *supra* note 7.

20. *Id.*

21. *Id.*

22. *Id.*

23. Shao, *supra* note 7; Oscar Schwartz, *You Thought Fake News Was Bad? Deep Fakes Are Where Truth Goes to Die*, THE GUARDIAN (Nov. 12, 2018, 5:00 AM), https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth [https://perma.cc/EEK8-8B8T].

election.[24] In response to this, social media websites are making policy changes to prevent the spread of deepfakes, especially after the criticisms about the increase of fake news.[25] For example, leading up to the 2020 election and census, YouTube announced that it would not "allow election-related 'deepfake' videos and anything that aims to mislead viewers about voting procedures and how to participate in the 2020 census."[26]

Other social media websites have also established limitations to publishing deepfakes, but their limitations seem to be focused more on limiting the spread of misinformation and "fake news," rather than election tampering. Facebook first put out a statement that it had "strengthen[ed]" its policies "toward misleading manipulated videos that have been identified as deepfakes" in response to growing concerns about the dominance of its presence online.[27] Some news outlets interpreted this statement as Facebook's commitment to completely ban deepfakes from its website.[28] Similarly, Reddit has terminated some forums that were previously started by users to share deepfakes in an effort to stop the spread of deepfakes on its website.[29] On a broader scale, Twitter and Pornhub have completely banned the publication of deepfakes on their websites.[30]

While these policies are sometimes labeled as a ban, it is possible that some "loopholes" may exist around them.[31] For example, Facebook's

---

24. Shao, *supra* note 7; Schwartz, *supra* note 23.

25. *See* Dave Lee, *Matter of Fact-Checkers: Is Facebook Winning the Fake News War?*, BBC (April 2, 2019), https://www.bbc.com/news/technology-47779782 [https://perma.cc/HRU2-MSJ6] (showing that people "felt Facebook was not listening to their feedback on how to improve the tool it provides to sift through content flagged as 'fake news.'").

26. Matt O'Brien, *YouTube: No 'Deepfakes' or 'Birther' Videos in 2020 Election,* ASSOCIATED PRESS (Feb. 2, 2020), https://apnews.com/3397d5dec4972ce12cac5037eeb9f226 [https://perma.cc/8YRJ-L8VQ].

27. Monika Bickert, *Enforcing Against Manipulated Media*, FACEBOOK (Jan. 6, 2020), https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/ [https://perma.cc/8DPL-U5EH].

28. David McCabe & Davey Alba, *Facebook Says It Will Ban 'Deepfakes'*, N.Y. TIMES (Jan. 7, 2020), https://www.nytimes.com/2020/01/07/technology/facebook-says-it-will-ban-deepfakes.html [https://perma.cc/X7CA-UVSG].

29. *r/deepfakes,* REDDIT, https://www.reddit.com/r/deepfakes [https://perma.cc/7PFC-3NU8] (last visited Feb. 12, 2020) (showing that this subreddit was banned); *r/celebfakes*, REDDIT, https://www.reddit.com/r/Celebfakes/ [https://perma.cc/7XUP-K2FV] (last visited Feb. 12, 2020) (same).

30. Samantha Cole, *Twitter Is the Latest Platform to Ban AI Generated Porn*, VICE (Feb. 6, 2018, 5:12 PM), https://www.vice.com/en_us/article/ywqgab/twitter-bans-deepfakes [https://perma.cc/CJ7U-FPUW].

31. *See* Betsy Morris, *Facebook Bans Deepfakes but Permits Some Altered Content*, WALL ST. J. (Jan. 7, 2020, 5:33 PM), https://www.wsj.com/articles/facebooks-deepfake-video-ban-permits-some-altered-content-11578384519 [https://perma.cc/2WSM-NEBU]; Aaron Holmes, *Facebook Just Banned Deepfakes, but the Policy Has Loopholes — and a Widely Circulated*

deepfake policy has exceptions for "deepfakes meant as satire as well as misleading videos made with less sophisticated tools."[32] Under the First Amendment, this tracks with case law about parody speech.[33] However, in the fight against "fake news," this policy may be troubling to individuals concerned about the spread of false information.[34]

At the heart of the election tampering and "fake news" concerns rest the overall worries about generalized harm to society, including distortion of democratic discourse, eroding trust in institutions and journalism, increasing social divisions, and threats to national security.[35] Before the 2020 presidential election, there were concerns that deepfakes would be "prevalent and problematic."[36] While deepfakes were certainly circulated during the 2020 election season,[37] later reports dated closer to the election suggested that the concern over deepfake election tampering was overhyped.[38] Outside of elections, politicians on both sides of the aisle have worried about how deepfakes could threaten national security, suggesting deepfakes are a "conceivable political weapon."[39] Diplomats and ambassadors have claimed that they have been the target of deepfakes.[40] Some scholars have warned about the "Liar's Dividend," a term coined for the potential phenomenon that public figures may start claiming their missteps were actually fake news, publicized through a deepfake, rather than a truthful statement.[41]

Deepfake of Mark Zuckerberg Is Allowed to Stay Up, BUSINESS INSIDER (Jan. 7, 2020, 10:07 AM), https://www.businessinsider.com/facebook-just-banned-deepfakes-but-the-policy-has-loopholes-2020-1 [https://perma.cc/4TK4-REM8].

32. Holmes, *supra* note 31.

33. Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 50 (1988) ("At the heart of the First Amendment is the recognition of the fundamental importance of the free flow of ideas and opinions on matters of public interest and concern.").

34. Holmes, *supra* note 31 ("'Facebook wants you to think the problem is video-editing technology, but the real problem is Facebook's refusal to stop the spread of disinformation.'").

35. Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1771–86 (2019).

36. Holly Kathleen Hall, *Deepfake Videos: When Seeing Isn't Believing*, 27 CATH. U.J.L. & TECH. 51, 59 (2018).

37. David Frum, *The Very Real Threat of Trump's Deepfake*, THE ATLANTIC (April 27, 2020), https://www.theatlantic.com/ideas/archive/2020/04/trumps-first-deepfake/610750/ [https://perma.cc/J9CP-S2XH] (showing that President Donald Trump reposted deepfake of his opponent, Joe Biden, on Twitter).

38. Tim Mak & Dina Temple-Raston, *Where Are The Deepfakes In This Presidential Election?,* NPR NEWS (Oct. 1, 2020), https://www.npr.org/2020/10/01/918223033/where-are-the-deepfakes-in-this-presidential-election [https://perma.cc/53RT-TANS].

39. Hall, *supra* note 36, at 59.

40. *Id.* at 60; *see* McFaul, *supra* note 3.

41. Chesney & Citron, *supra* note 35, at 1785.

2. Threats to Individual Privacy

Some legal experts are also concerned with the use of deepfakes to harm individuals, either through exploitation or sabotage.[42] Examples of exploitation purposes could include, but are not limited to, financial blackmail, revenge porn, and possibly even "fraudulent kidnapping claims."[43] Deepfake revenge porn seems to be considered the most prominent concern in this category.[44] Revenge porn – the dissemination of pornographic videos or photos without the subject's consent or knowledge – could be transformed into a new beast by using deepfake technology.[45] Now, a person could be the subject of a porn video without having even performed the sex act being showcased in the video.[46]

3. The Defense of Deepfakes: Self-Expression

But while critics have been quick to bring up various problems with deepfakes, it is important to acknowledge that there could be some benefits to the new technology – mainly its ability to promote self-expression.[47] For example, there is some evidence that deepfake technology could be used to alter audio files and "restore the ability of persons suffering from certain forms of paralysis, such as ALS, to speak with their own voice."[48] Additionally, comedic parody deepfakes could be considered self-expression.[49] One YouTube channel dedicated to this purpose, Cntl Shift Face, has already garnered more than 300,000 subscribers since it started publishing deepfakes

---

42. *Id.* at 1772–75; *see* Citron, *supra* note 7.

43. Chesney & Citron, *supra* note 35, at 1772–73.

44. Rebecca A. Delfino, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, 88 FORDHAM L. REV. 887, 925 (2019) ("Immediately after celebrity-based pornographic deepfakes emerged in late 2017 and went viral on the internet, legal scholars and journalists raised the alarm that this conduct implicated the First Amendment protections afforded to online content."); *see* Chesney & Citron, *supra* note 35, at 1773.

45. Delfino, *supra* note 44, at 891.

46. *Id.* ("Because deepfake technology can be used to create realistic pornographic videos without the consent of the individuals depicted, and since these videos can be broadly distributed on the internet, pornographic deepfakes exist in the realm of other sexually exploitative cybercrimes such as revenge porn and nonconsensual pornography.").

47. Chesney & Citron, *supra* note 35, at 1770; *see also* Jessica Silbey & Woodrow Hartzog, *The Upside of Deep Fakes*, 78 MD. L. REV. 960, 966 (2019).

48. Chesney & Citron, *supra* note 35, at 1771

49. Cntl Shift Face, *supra* note 6. (showing a deepfake of President Donald Trump as one of the characters in a well-known scene in the TV show "Breaking Bad").

in 2019.[50] Deepfake parody videos usually consist of an actor or celebrity deepfaked into different context than normally expected, aimed at a comedic or entertaining purpose.[51] For example, in Cntl Shift Face's most popular video on YouTube, titled "Bill Hader impersonates Arnold Schwarzenegger [DeepFake]," actor Bill Hader is pictured on a late-night talk show speaking to the show's host, Conan O'Brien.[52] When Hader begins to impersonate Arnold Schwarzenegger's recognizable accent as part of a joke, Cntl Shift Face slowly morphs Hader's face into Schwarzenegger's.[53] Videos like the one described tend to go viral online, spread through various different social media platforms, like YouTube and Instagram, for example.[54]

C. Applicable First Amendment Law

This Note specifically focuses on how deepfakes could be viewed under the First Amendment. Depending on the content and the situation of which it arises, deepfakes could be viewed under several different branches of the First Amendment. First, the deepfaker could claim that the video was a political parody. Second, the deepfaker could argue that the First Amendment generally protects lies. Lastly, if the deepfake includes pornographic images, it might be covered under obscenity, child porn, or revenge porn law.

1. Politics, Parodies, and Privacy

Not all uses of deepfakes are villainous; in fact, many people claim to use the technology to comment on politics and poke fun at governmental

50. Cntl Shift Face, YOUTUBE (last visited April 11, 2020), https://www.youtube.com/channel/UCKpH0CKltc73e4wh0_pgL3g [https://perma.cc/Z3PK-ERCF].

51. *See, e.g.*, Cntl Shift Face, *supra* note 6; Cntl Shift Face, *Jim Carrey DeepFake [VFX Comparison]*, YOUTUBE (Sept. 3, 2019), https://www.youtube.com/watch?v=JbzVhzNaTdI [https://perma.cc/4DLV-WAK5] (showing a deepfake of Jim Carrey taking the place of Jack Nicholson's role in *The Shining*); Collider Videos, *Deepfake Roundtable: Cruise, Downey Jr., Lucas & More - The Streaming Wars | Above the Line*, YOUTUBE (Nov. 11, 2019), https://www.youtube.com/watch?v=l_6Tumd8EQI [https://perma.cc/2XA6-3Z92] (showing a roundtable of "celebrities" that are actually impersonators with deepfaked faces of celebrities).

52. Cntl Shift Face, *Bill Hader impersonates Arnold Schwarzenegger [DeepFake]*, YOUTUBE (May 10, 2019), https://www.youtube.com/watch?v=bPhUhypV27w [https://perma.cc/66NH-RUBU].

53. *Id.*

54. Collider Videos, *supra* note 51; @bill_posters_uk, INSTAGRAM (June 7, 2019), https://www.instagram.com/p/ByaVigGFP2U/ (showing a deepfake of Mark Zuckerberg with more than 120,000 views).

officials.[55] Because of this, it is likely that these deepfakes could be protected under the First Amendment if they are classified as parodies.[56] It is important to note that the First Amendment itself, which establishes that "Congress shall make no law . . . abridging the freedom of speech," applies only to governmental regulation of speech.[57] It does not apply to private parties.[58] So, privately owned social media websites like Facebook and Twitter have the ability to remove or block deepfakes from being posted on their platforms without violating the First Amendment rights of social media users.[59]

Parodies are "an attack on folly" and generally imitate an existing work for some comedic purpose.[60] Parodies are definitely not new; "[f]rom the early cartoon portraying George Washington as an ass down to the present day, graphic depictions and satirical cartoons have played a prominent role in public and political debate."[61] Parodies are tied closely with political speech,[62] and under the First Amendment, political speech is given a significant amount of deference by courts.[63]

There is no question that government officials, who are often the subjects of these political parodies, are considered "public figures" under the law.[64] This means they can often be subjected to "vehement, caustic, and sometimes

---

55. *See e.g.*, Cntl Shift Face, *supra* note 6 (showing a deepfake of President Donald Trump as one of the characters in a well-known scene in the TV show "Breaking Bad").

56. *See* Holmes, *supra* note 31.

57. U.S. CONST. amend. I.

58. Lata Nott, *Is your speech protected by the First Amendment?*, FREEDOM FORUM INSTITUTE, https://www.freedomforuminstitute.org/first-amendment-center/primers/basics/ [https://perma.cc/ZV43-BDJJ] (last visited Nov. 4, 2020) (outlining what the First Amendment protects).

59. *See* Sara Fischer & Ashley Gold, *All the platforms that have banned or restricted Trump so far*, AXIOS, https://www.axios.com/platforms-social-media-ban-restrict-trump-d9e44f3c-8366-4ba9-a8a1-7f3114f920f1.html [https://perma.cc/86XT-4925] (Jan. 11, 2021) (showing that many social media companies banned former President Donald Trump from their platforms because he violated the websites' policies and guidelines). *See also* Rich Barlow, *Break Up Twitter? Or Ban Trump? Where Does Social Media Go from Here?*, BU TODAY, http://www.bu.edu/articles/2021/trump-banned-on-social-media-not-first-amendment-issue/ [https://perma.cc/6ZL5-TC5K] (Jan. 11, 2021) (showing that social media companies, like Twitter, can ban users from using its website when they do not comply with the website's conditions).

60. Kyonzte Hughes, *Parody & Satire,* FREEDOM FORUM INSTITUTE (Sept. 13, 2002), https://www.freedomforuminstitute.org/first-amendment-center/topics/freedom-of-speech-2/arts-first-amendment-overview/parody-satire/ [https://perma.cc/S7AT-VRVR].

61. Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 54 (1988).

62. *Id.*

63. Stromberg v. People of State of Cal., 283 U.S. 359, 369 (1931) ("The maintenance of the opportunity for free political discussion . . . is a fundamental principle of our constitutional system.").

64. *See, e.g.* New York Times Co. v. Sullivan, 376 U.S. 254, 256 (1964).

unpleasantly sharp attacks."[65] The landmark parody case involving a public figure is *Hustler Magazine v. Falwell*, in which a well-known minister sued the magazine for emotional distress after a parody piece depicted him as an incestuous drunk.[66] The United States Supreme Court noted that the nature of parodies is that they are "not reasoned or evenhanded, but slashing and one-sided."[67]

Generally, to bring a tort claim, a public figure would need to prove the "speech could not reasonably have been interpreted as stating actual facts."[68] For libel and emotional distress cases, the public figure would also need to show that "the publication contains a false statement of fact which was made with 'actual malice,' i.e., with knowledge that the statement was false or with reckless disregard as to whether or not it was true."[69] *New York Times v. Sullivan* initially created the "actual malice" test for public officials' libel claims,[70] and *Falwell* extended it to emotional distress cases.[71] The actual malice standard has been described in more detail as "a constitutional rule that allows public figures to recover for libel or defamation only when they can prove both that the statement was *false* and that the statement was made with the requisite level of *culpability*."[72] So, in applying this analysis to deepfakes, if President Donald Trump wanted to sue the creator of a deepfake parody made about him, for example, he would have to demonstrate that the deepfake could be reasonably believed as truthful, in addition to proving that the deepfaker had "actual malice" in creating and publishing the deepfake online.

Similarly to the libel and emotional distress cases involving parodies, a victim of deepfakes would also have a difficult time bringing a false light privacy claim due to the "actual malice" standard for public figures.[73] A claim for false light privacy can be brought when a person has given "publicity to a matter concerning another that places the other before the public in a false light."[74] The standard used for determining false light privacy claims is based upon whether the perception given to the person is "offensive to a reasonable

65. *Id.* at 270. It is important to note that ordinary individuals – those that are not public figures – do not have to meet the actual malice requirement. Gertz v. Robert Welch, Inc., 418 U.S. 323, 345 (1974).

66. Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 48 (1988).

67. *Id.* at 54.

68. *Id.* at 50.

69. *Id.* at 56.

70. New York Times Co. v. Sullivan, 376 U.S. 254, 280 (1964).

71. *Falwell*, 485 U.S. at 56.

72. *Id.* at 52 (emphasis added).

73. Time, Inc. v. Hill, 385 U.S. 374, 388 (1967) (holding that a magazine publisher could not be held liable "in the absence of proof that the defendant published the report with knowledge of its falsity or in reckless disregard of the truth."); Douglas Harris, *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, 17 DUKE L. & TECH. REV. 99, 115 (2019); Restatement (Second) of Torts § 652E (AM. LAW INST. 1977) (showing that a victim must prove that the offender knew or recklessly regarded the fact that the information is false).

74. *See* Restatement (Second) of Torts § 652E (AM. LAW INST. 1977).

person."[75] It would not be an easy task to prove a false light privacy case against a deepfaker, especially since the victim would need to prove that other people actually recognized that the deepfake truly put them in a "false light."[76] This means that the public would have to understand that the deepfake was false and that it negatively portrayed the victim through its falsity.

2. The First Amendment's Protection of Lies

On the surface, the entire point of deepfakes is to spread lies – the videos show people in circumstances or positions that are simply untrue, regardless of the deepfaker's intent. The Supreme Court has been consistent in saying that "there is no constitutional value in false statements of fact."[77] However, generally, the First Amendment still protects some lies.[78] This stance "comports with the common understanding that some false statements are inevitable if there is to be an open and vigorous expression of views in public and private conversation, expression the First Amendment seeks to guarantee."[79] Lies about the government and its officials, since they relate to political speech, may be protected under this rationale as well.[80]

Regardless, there remains the availability of criminal and civil liability – like libel and slander – to punish the speaker for their lies depending on the speaker's culpability, position as a public figure, and the subject matter of the speech.[81] However, if the questionable deepfake does not fall under one of these existing exceptions, the Supreme Court has suggested that the truthfulness of speech will generally be revealed through the marketplace of ideas.[82] The reliance on the traditional "marketplace of ideas" concept in regards to deepfakes is shaky, however, if the deepfake is particularly

---

75. *See id.*

76. Harris, *supra* note 73, at 117.

77. Gertz v. Robert Welch, Inc., 418 U.S. 323, 340 (1974).

78. United States v. Alvarez, 567 U.S. 709, 718 (2012) ("Absent from those few categories where the law allows content-based regulation of speech is any general exception to the First Amendment for false statements.").

79. *Id.*

80. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 269–70 (1964).

81. *See* Dun & Bradstreet, Inc. v. Greenmoss Builders, Inc., 472 U.S. 749, 763 (1985) (showing that "actual malice" is not required if the speech is not related to a subject of "public concern"); Gertz v. Robert Welch, Inc., 418 U.S. 323, 346 (1974) (showing that the "actual malice" standard does not apply to standard individuals); Brandenburg v. Ohio, 395 U.S. 444, 447 (1969) (showing that the First Amendment does not protect speech that incites "imminent lawless action"); Chaplinsky v. State of New Hampshire, 315 U.S. 568, 572 (1942) (showing that the First Amendment does not protect "fighting words").

82. *Alvarez*, 567 U.S. at 727–728 ("The remedy for speech that is false is speech that is true. . . . The theory of our Constitution is 'that the best test of truth is the power of the thought to get itself accepted in the competition of the market.'").

convincing, misleading, and deceptive.[83] A deepfake could be so realistic that the truth may never prevail unless there is some intervention, either through private website owners or by the government.[84]

3. Obscenity, Child Porn, or Revenge Porn?

Similarly, victims of sexually explicit deepfakes would also have a difficult time proving the deepfake was a form of revenge porn. The First Amendment recognizes exceptions for obscenity and child pornography – generally, pornography cannot be banned unless it is obscene.[85] Therefore, the status quo of the law makes it difficult to sue for revenge porn unless there is an applicable state statute specifically prohibiting it or the case includes unlawful child pornography.[86] The Supreme Court outlined a three-pronged test for obscenity cases not involving children in *Miller v. California* to determine if the pornographic content had value outside of its obscenity: (1) if "'the average person, applying contemporary community standards' would find that the work, taken as a whole, appeals to the prurient interest," (2) "whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law," and (3) "whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value."[87] Proving the first and third prongs in court would likely be left to case-by-case analyses based upon the "reasonable person" standard. However, if states continue to regulate deepfakes under their statutes – as described in more detail in Part III, Subpart B – it may be easier to meet the second *Miller* test requirements, as it could make it easier to find the "applicable state law."

Additionally, some courts have already started discussions on how deepfakes will impact future child pornography cases.[88] While wide bans of

83. *See* Hall, *supra* note 36, at 52–53 ("In the past we have relied on the 'marketplace of ideas' concept, which encourages more speech as a means to uncover the truth and have the best ideas rise to the fore, rather than censor particular content. Is this argument still valid when the public cannot discern what information is true, misleading, or false?").

84. At the time of this Note, there are no laws in place that would prohibit social media websites like Facebook or Twitter from removing deepfakes on their platforms, unlike how the First Amendment could prohibit the government from doing the same.

85. Ashcroft v. Free Speech Coal., 535 U.S. 234, 240 (2002).

86. New York v. Ferber, 458 U.S. 747, 764 (1982) (showing there is no First Amendment protection for child pornography and outlining a specific test for any exceptions); *see* CAL. CIV. CODE § 1708.86 (West 2019) (showing California's revenge porn statute).

87. Miller v. California, 413 U.S. 15, 24 (1973).

88. *See* In re S.K., 215 A.3d 300, 315 n.22 (2019); United States v. Streett, No. CR 14-3609 JB, 2020 WL 231688, at *47 n.28 (D.N.M. Jan. 15, 2020). The focus on child pornography is likely because children are particularly vulnerable. *See Ferber*, 458 U.S. at 764 (showing that "kiddie porn" is not protected under the First Amendment).

child pornography have not been upheld under the First Amendment,[89] states are generally given discretion and leniency in how much they choose to regulate child pornography.[90] In *New York v. Ferber*, the Supreme Court announced a stricter standard than *Miller* in deciding child pornography cases. To be unlawful, the child pornography does not require a "prurient interest of the average person."[91] The sexual conduct shown also does not need to be portrayed in "a patently offensive manner," and the material does not need to be considered in its entirety.[92]

Lastly, if the deepfake is just a form of porn – meaning it does not include sexually explicit imagery of children and is found to not be obscene – there is no remaining way to regulate it under the First Amendment.[93] If a state has passed a revenge porn statute, however, then it is possible that the deepfake and its creator could be penalized under that law. Most states have created revenge porn statutes relying on privacy concerns and worries about intentional infliction of emotional distress.[94] Many of these statutes now include language about deepfakes.[95]

III. RECENT DEVELOPMENTS

It is not a simple task to protect individual rights and bolster democracy against threats from malicious deepfakes, some action has taken place by the federal and state government. Subpart A discusses how some federal legislation has been tossed around the House and Senate proposing to regulate deepfakes, but most attempts have been futile. However, Subpart B demonstrates that some states have successfully crafted and passed legislation to control deepfakes pertaining to election tampering and obscenity.

A. The Push and Stall in Federal Intervention

At the time of this Note, there have been numerous Congressional attempts to regulate deepfakes; only one of them has succeeded in being signed into law.[96] This Part explores four leading pieces of legislation that

89. *Ashcroft*, 535 U.S. at 258.

90. *Ferber*, 458 U.S. at 756.

91. *Id.* at 764.

92. *Id.*

93. *See* Delfino, *supra* note 44, at 925 (pointing to obscenity as one of the only First Amendment claims to make regarding pornographic deepfakes).

94. *46 States + DC + One Territory NOW have Revenge Porn Laws*, CYBER CIVIL RIGHTS INITIATIVE (last visited April 12, 2020), https://www.cybercivilrights.org/revenge-porn-laws/ [https://perma.cc/Z2DP-BS9H].

95. *See* MD. CODE ANN., CRIM. LAW § 11-208(a)(2) (West 2019); VA. CODE ANN. § 18.2-386.2(A) (West 2019); CAL. CIV. CODE § 1708.86 (West 2019).

96. NDAA for FY 2020, S.1790, Pub. L. No. 116-92, § 5709, 133 Stat. 1198, 2168 (2019).

members of Congress have put forward regarding deepfakes, which are all aimed at preventing federal political interference.

1. Malicious Deep Fake Prohibition Act of 2018

The first federal bill targeting deepfakes, the "Malicious Deep Fake Prohibition Act of 2018," was introduced by Senator Ben Sasse (R-Neb.) in December 2018.[97] The bill was read twice and sent to committee, but that is the farthest it went.[98] This is probably due to the ambitious nature of the bill, which, if passed, would have created a "new criminal offense related to the creation or distribution of fake electronic media records that appear realistic."[99] The bill was relatively short, just including sections for definitions, offenses, and exceptions.[100] The intent or the actual distribution of a deepfake would have been against federal law if the deepfaker had "actual knowledge that the audiovisual record is a deep fake" and still distributed it anyway.[101] The Act would have subjected convicted deepfakers to no more than two years in prison, unless the deepfake was found to "facilitate violence" or "affect the conduct" of government or election proceedings.[102] Interestingly, the bill had a limitation that "[n]o person shall be held liable under this section for any activity protected by the First Amendment to the Constitution of the United States."[103]

2. DEEP FAKES Accountability Act

About six months after the Malicious Deep Fake Prohibition Act of 2018 was introduced, the "DEEP FAKES Accountability Act" surfaced in the House, led by Representative Yvette D. Clarke (D-NY).[104] Representative Clarke had twenty-eight co-sponsors on the bill; all of them were members of the Democratic Party.[105] The Act was noticeably longer than the Malicious

97. Malicious Deep Fake Prohibition Act of 2018, S. 3805, 115th Cong. 2d Sess. (2018).

98. *Actions Overview S.3805 – 115th Congress (2017-2018)*, CONGRESS.GOV (last visited Mar. 13, 2020), https://www.congress.gov/bill/115th-congress/senate-bill/3805/actions?KWICView=false [https://perma.cc/WGY4-Y94Q].

99. *Summary: S.3805 – 115th Congress (2017-2018)*, CONGRESS.GOV (last visited Mar. 13, 2020), https://www.congress.gov/bill/115th-congress/senate-bill/3805.

100. S. 3805.

101. S. 3805 § 1041(b)(1)–(2).

102. S. 3805, § 1041(c)(1)–(2).

103. S. 3805, § 1041(d)(2).

104. Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019, H.R. 3230, 116th Cong. 1st Sess. (2019) [hereinafter "DEEP FAKES Accountability Act"].

105. *Cosponsors: H.R.3230 — 116th Congress (2019-2020)*, CONGRESS.GOV (last visited Mar. 13, 2020), https://www.congress.gov/bill/116th-congress/house-

Deep Fake Prohibition Act of 2018, and it covered more aspects. For example, the Act discussed digital watermarking deepfake videos so they would be more identifiable, plus requiring "clearly articulated" video and audio disclosures to show that the deepfakes were altered videos.[106]

The Act also allowed for both criminal and civil penalties for failure to disclose that the videos were altered.[107] The Act covered deepfakes with the "intent of influencing a domestic public policy debate" and "interfering in a Federal, State, local, or territorial election. . ."[108] Compared to the Malicious Deep Fake Prohibition Act of 2018, the DEEP FAKES Accountability Act carried a larger criminal sentence of up to five years.[109] For private actions, damage amounts were outlined, ranging from $50,000 to $150,000 for each violative deepfake.[110] Injunctive relief was also described as a private action remedy.[111]

Another interesting part of this Act is the creation of several task forces and coordinators to help enforce the Act. First, the Act dedicated a section to victim assistance, specifically tasking the Attorney General to place "a coordinator in each United States Attorney's Office to receive reports from the public regarding potential violations . . . relating to deep fake depictions produced or distributed by any foreign nation-state . . . and coordinate prosecutions for any violation of such section."[112] This section was likely aimed at preventing the spread of deepfake revenge porn.[113] Similarly, the Act would have created a "Deep Fakes Task Force" within the Department of Homeland Security to help "combat the national security implications of deep fakes."[114]

The DEEP FAKES Accountability Act was referred to several committees in June 2019, but it was effectively tabled by the Subcommittee on Crime, Terrorism, and Homeland Security just a few weeks after its initial introduction on the House floor.[115] The Act covered a lot of information and had "enormous loopholes," which may have caused its demise.[116] The Act

---

bill/3230/cosponsors?searchResultViewType=expanded&KWICView=false [https://perma.cc/8BKN-QH4R].

106. H.R. 3230, § 1041(a)–(e).

107. *Id.* at § 1041(f).

108. *Id.* at § 1041(f)(1)(B)(iv).

109. *Id.* at § 1041(f).

110. *Id.* at § 1041(g)(2).

111. *Id.* at § 1041(g)(3).

112. *Id.* at § 1042(a).

113. *See* Devin Coldewey, *DEEPFAKES Accountability Act would impose unenforceable rules — but it's a start*, TECHCRUNCH (June 13, 2019), https://techcrunch.com/2019/06/13/deepfakes-accountability-act-would-impose-unenforceable-rules-but-its-a-start/ [https://perma.cc/9586-9NNU].

114. H.R. 3230, § 918 sec. 7 (a).

115. *All Actions H.R.3230 — 116th Congress (2019-2020)*, CONGRESS.GOV (last visited Mar. 13, 2020), https://www.congress.gov/bill/116th-congress/house-bill/3230/all-actions [https://perma.cc/9586-9NNU].

116. *See* Coldewey, *supra* note 113.

was criticized by deepfake regulation activists and critics alike; some said the Act was overly broad and threatened individuals' First Amendment protections, and others said there were too many exceptions to where it could not be effective in preventing the spread of deepfakes.[117]

3. Deepfake Report Act of 2019

Just a month after the DEEP FAKES Accountability Act, the "Deepfake Report Act of 2019" was introduced by Senator Rob Portman (R-Ohio).[118] The House sent the Act to a subcommittee in October 2019 but never reconsidered it.[119] The Act would have required "the Department of Homeland Security to report at specified intervals on the state of digital content forgery technology."[120] This Act reached farther than just deepfakes; according to the Act's summary, "digital content forgery" includes all "artificial intelligence and machine learning techniques, to fabricate or manipulate audio, visual, or text content with the intent to mislead."[121] The Deepfake Report Act is similar in some way to its predecessor, the Malicious Deep Fake Prohibition Act of 2018. It is much shorter and narrower in scope than the DEEP FAKES Accountability Act.[122] The Deepfake Report Act only required the Department of Homeland Security to "produce a report on the state of digital content forgery technology" and conduct relevant public hearings to gather more information about the topic.[123] Media coverage at the time of this Act's introduction was limited, so it is uncertain why this Act specifically failed.

4. National Defense Authorization Act for Fiscal Year 2020

The omnibus National Defense Authorization Act for Fiscal Year 2020 ("NDAA for FY 2020") is the only successfully passed legislation so far regarding deepfakes.[124] The Act is lengthy and complex, but one section is dedicated to a "report on deepfake technology, foreign weaponization of

117. *See* Mathew Ingram, *Legislation aimed at stopping deepfakes is a bad idea,* COLUM. JOURNALISM REV. (July 1, 2019), https://www.cjr.org/analysis/legislation-deepfakes.php [https://perma.cc/N2PX-46TS].

118. *S.2065 – Deepfake Report Act of 2019*, CONGRESS.GOV (last visited Mar. 13, 2020), https://www.congress.gov/bill/116th-congress/senate-bill/2065 [https://perma.cc/QRS4-YJML].

119. *Actions Overview S.2065 — 116th Congress (2019-2020),* CONGRESS.GOV (last visited Mar. 13, 2020), https://www.congress.gov/bill/116th-congress/senate-bill/2065/actions?KWICView=false [https://perma.cc/42GK-5YVG].

120. *Summary: S.2065 – 116th Congress (2019-2020)*, CONGRESS.GOV (last visited Mar. 13, 2020), https://www.congress.gov/bill/116th-congress/senate-bill/2065 [https://perma.cc/S7L5-ACPU].

121. *Id.*

122. Deepfake Report Act of 2019, S.2065, 116th Cong. 1st Sess. (2019).

123. S.2065, sec. 3. (a), (c)(2).

124. S.1790 § 5709.

deepfakes, and related notifications."[125] Within 180 days of the Act's passage, the Director of National Intelligence must consult with appropriate intelligence leaders and develop a report to submit to the congressional intelligence committees consisting of "the potential national security impact of machine-manipulated media (commonly known as 'deepfakes')" and the uses of deepfakes by "foreign governments to spread disinformation."[126] Specifically, the report is required to include information about the "technical abilities" of China and Russia in creating and disseminating deepfakes, particularly for uses relating to "discrediting political opponents or disfavored populations . . ."[127] The section also requests research about how the United States could quickly identify and defend against a deepfake attack.[128] The bill was signed into law on December 20, 2019.[129] The report from the Director of National Intelligence, then, is due by June 17, 2020 (180 days after the Act's passage). As of November 18, 2020, the report had not been completed and was still under development.

B. States Are Taking the Matters Into Their Own Hands

States have started the discussion about deepfakes in two ways: preventing the dissemination of certain sexually explicit deepfakes and preventing deepfake election tampering. First, some states have tackled deepfakes when it relates to sexually explicit content.[130] For example, Maryland amended its statute prohibiting child pornography to include deepfake technology.[131] It now prohibits deepfakes that are "[i]ndistinguishable from an actual and identifiable child" and contain "a computer-generated image that has been created, adapted, or modified to appear as an actual and identifiable child."[132] However, the statute expressly mentions that it does not cover typical parody art forms, such as "drawings, cartoons[,] sculptures[,] or paintings."[133]

Virginia and California amended their revenge porn statutes to penalize the dissemination of deepfakes if they are used for revenge porn purposes.[134] California's revenge porn statute describes photos and videos that include

125. *Id.*

126. *Id.* at § 5709(a)(1).

127. *Id.* at § 5709(a)(2)(A)–(B).

128. *Id.* at § 5709(a)(2)(C).

129. *All Actions S.1790 — 116th Congress (2019-2020)*, CONGRESS.GOV (last visited Mar. 15, 2020), https://www.congress.gov/bill/116th-congress/senate-bill/1790/actions [https://perma.cc/KZ5P-YGGP].

130. MD. CODE ANN., CRIM. LAW § 11-208 (West 2019); VA. CODE ANN. § 18.2-386.2(A) (West 2019); CAL. CIV. CODE § 1708.86 (West 2020).

131. MD. CODE ANN., CRIM. LAW § 11-208 (West 2019).

132. *Id.* at § 11-208(a)(2).

133. *Id.* at § 11-208(a)(3).

134. VA. CODE ANN. § 18.2-386.2(A) (West 2019); CAL. CIV. CODE § 1708.86 (West 2020).

"[c]omputer-generated nude body parts as the nude body parts of the depicted individual," and "[t]he depicted individual engaging in sexual conduct in which the depicted individual did not engage."[135] There are exceptions to liability under California's law, for example, if the altered videos are "[a] matter of legitimate public concern," "[a] work of political or newsworthy value or similar work," or "[c]ommentary, criticism, or disclosure that is otherwise protected by the California Constitution or the United States Constitution."[136]

Some states have started to regulate deepfakes if they are used to interfere with elections. California and Texas have both enacted these type of laws.[137] In September 2019, Texas was the first state to take steps to regulate deepfakes in the election context.[138] The amendment of the state's election code punishes a person who intends to "injure a candidate or influence the result of an election" by creating a deepfake video and publishes or distributes it within 30 days of an election.[139] There are no exceptions to the law, which has garnered some criticism from legal experts about its constitutionality under the First Amendment.[140]

Just a month later, California passed legislation that amended the state's election code to prohibit deepfakes published "within 60 days of an election" if they were distributed "with actual malice . . . [and] with the intent to injure the candidate's reputation or to deceive a voter into voting for or against the candidate, unless the media includes a disclosure stating that the media has been manipulated."[141] The law does not use the term "deepfake" to describe the "deceptive audio or visual media," although news outlets have recognized that was the legislature's intent in passing the law.[142] The law also allows candidates that are the victims of deceptive deepfakes to sue for injunctive relief or monetary damages.[143] Unlike Texas, the California law includes exceptions for parody deepfakes and paid-for broadcast advertisements.[144]

---

135. CAL. CIV. CODE § 1708.86(a)(6)(A)–(C) (West 2020).

136. *Id.* at § 1708.86(c)(1)(B)(i)–(iii).

137. Shao, *supra* note 7.

138. Kenneth Artz, *Texas Outlaws 'Deepfakes'—but the Legal System May Not Be Able to Stop Them*, LAW.COM (Oct. 11, 2019), https://www.law.com/texaslawyer/2019/10/11/texas-outlaws-deepfakes-but-the-legal-system-may-not-be-able-to-stop-them/ [https://perma.cc/YL56-UVNZ].

139. TEX. ELEC. CODE ANN. § 255.004 (West 2019).

140. Artz, *supra* note 138.

141. A.B. 730, 2019-2020 Leg., Reg. Sess. (Cal. 2019) (as described in the Legislative Counsel's Digest). The law is now codified under CAL. ELEC. CODE § 20010 (West 2019).

142. A.B. 730; Kari Paul, *California makes 'deepfake' videos illegal, but law may be hard to enforce*, THE GUARDIAN (Oct. 7, 2019), https://www.theguardian.com/us-news/2019/oct/07/california-makes-deepfake-videos-illegal-but-law-may-be-hard-to-enforce [https://perma.cc/3QKL-BG47].

143. A.B. 730.

144. *Id.*

Other states have attempted to cover privacy more generally than just deepfakes but have not succeeded. For example, New York discussed legislation that would have made it fraud to create a digital replica of someone without their consent.[145] The intent of this bill was to protect the state's long-standing "right to publicity."[146] For example, creating a deepfake of an actor post-mortem would require the deepfaker to get the approval of the late actor's estate before using his or her likeness, otherwise the deepfaker would be criminally liable.[147] Similarly, Massachusetts introduced a bill that would have made the creation of a deepfake with the intent to distribute a crime of identity fraud.[148] Missouri legislators have not passed – let alone introduced – any laws regulating deepfakes. However, Missourians have not been completely silent about the issue; the Missouri School of Journalism, for example, hosted a national collegiate innovation competition which "tasked teams with developing tools to help verify photos, videos or audio content to help the industry fight against deepfakes and fabricated content."[149]

IV. DISCUSSION

Outside of the world of pornography, the impact of deepfakes on politics and democracy is perhaps the largest threat.[150] Given the previous summary of how deepfakes have been discussed by governments and the media nationwide, it is not surprising that deepfakes carry a negative connotation for a growing American population that feels democracy is threatened by fake news.[151] When attempting to analyze deepfakes under the fabric of the First Amendment, there are two main categories to divide political deepfakes into:

---

145. N.Y. Assemb, B. A08155, 2017-2018 Leg., Reg. Sess. (N.Y. 2017). The bill was not passed through the state Senate. *A08155 Actions*, NEW YORK STATE ASSEMBLY (last visited Nov. 25, 2020), https://nyassembly.gov/leg/?default_fld=&leg_video=&bn=A08155&term=2017&Actions=Y [https://perma.cc/N3SZ-4GRC].

146. *A08155 Chamber Transcript*, NEW YORK STATE ASSEMBLY (June 18, 2018), https://www2.assembly.state.ny.us/write/upload/transcripts/2017/6-18-18.html#08155 [].

147. *Id.*

148. H.3366, 2019-2020 Leg., 191st Sess. (Mass. 2020).

149. *Student team wins journalism innovation competition with deepfake fighting tool*, UNIV. OF MO. SYS. (Feb. 13, 2020), https://www.umsystem.edu/stories/student-team-wins-journalism-innovation-competition-deepfake-fighting-tool [https://perma.cc/TCG4-DTDK].

150. *See* Chesney & Citron, *supra* note 35, at 1778; Shao, *supra* note 7.

151. *See* Michael Dimock, *An update on our research into trust, facts and democracy*, PEW RESEARCH CENTER (June 5, 2019), https://www.pewresearch.org/2019/06/05/an-update-on-our-research-into-trust-facts-and-democracy/ [https://perma.cc/67EJ-UP6E] ("Nearly seven-in-ten (68%) say made-up news and information greatly affects Americans' confidence in government institutions, and roughly half (54%) say it is having a major impact on Americans' confidence in each other.").

(1) those that are made for comedic purposes but have negative consequences, and (2) those that are designed to spread misinformation and lies.[152] Depending on which category the deepfake falls under, the action against it might differ. For example, deepfakes that are parodies might be handled through a civil case for libel,[153] while a deepfake that was created to dismantle fair elections might be prosecuted under a statute.[154] Regulating deepfakes is certainly not restricted to this binary, as other avenues have actively been explored by legislators and scholars alike.[155]

A. Expanding on Deepfakes as Political Speech and Parodies

If a deepfake teases a political candidate or government official, the easiest argument that the deepfaker could make in defense of their video is that it is a form of political speech intended to be a parody. This argument has already been anticipated – and permitted – by social networking websites, state statutes, and federal legislation.[156] But what if the deepfake is so convincing that it does not appear to be a parody to the "reasonable person"? For a parody to really achieve its goal of being humorous, it "requires the audience recognize both the subject of the parody and the parodist's mocking distortions."[157] Perhaps there is a comparison that could be made between a convincing deepfake and a good impersonation. Alec Baldwin's impersonation of President Donald Trump on *Saturday Night Live* is a great example of a parody.[158] If Baldwin, during his impersonation of Trump, said

152. For the purpose of the discussion in this Note, the author declines to discuss pornographic deepfakes and obscenity law and instead will focus on deepfakes as they pertain to political speech.

153. David Greene, *We Don't Need New Laws for Faked Videos, We Already Have Them*, ELECTRONIC FRONTIER FOUND. (Feb. 13, 2018), https://www.eff.org/deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-already-have-them [https://perma.cc/E2JZ-KRBS].

154. *See e.g.,* TEX. ELEC. CODE ANN. § 255.004 (West 2019).

155. *See* Edvinas Meskys et al., *Regulating deep fakes: legal and ethical considerations*, 15 J. OF INTELL. PROP. L & PRAC. 24, 30–31 (2020); Matthew F. Ferraro, *Deepfake Legislation: A Nationwide Survey—State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media*, WILMERHALE (Sept. 25, 2019), https://www.wilmerhale.com/en/insights/client-alerts/20190925-deepfake-legislation-a-nationwide-survey [https://perma.cc/A28X-NYR8].

156. *See e.g.* MD. CODE ANN., CRIM. LAW § 11-208(a)(3) (West 2019) (allowing a parody exception for manipulated child pornography cases); S. 3805 § 1041(d)(2) (granting a First Amendment exception for prosecution under the Act); Holmes, *supra* note 31 (showing that Facebook makes an exception for deepfakes categorizes as satire).

157. Beth Warnken Van Hecke, *But Seriously, Folks: Toward A Coherent Standard of Parody As Fair Use*, 77 MINN. L. REV. 465, 465–66 (1992).

158. Saturday Night Live, *Impeachment Cold Open – SNL*, YOUTUBE (Sept. 28, 2019), https://www.youtube.com/watch?v=xR25izGfrmQ [https://perma.cc/JH46-F96K].

something nonfactual and defamatory, could Trump bring a claim for libel or emotional distress against him? Probably not, unless the statement made by Baldwin was said with "actual malice" and could not reasonably believed to be a joke.[159]

Taking it a step further, if a deepfaker took Baldwin's voice impersonation and layered it over the top of a deepfake video that looked more convincingly like Trump, then would Trump have a claim? Maybe – depending on the context – but still probably not. If the deepfake was very realistic, maybe the reasonable person might actually believe that Trump said the falsehoods. Trump, however, would still need to prove that the deepfaker had "actual malice" behind making the video. The "actual malice" standard cannot possibly cover all the bases here; after all, it is possible for a deepfake to mislead audiences without the deepfaker having knowledge or reckless disregard that the deepfake really contained untruths.[160] Maybe the deepfaker just wanted to innocently make Baldwin's impression resemble Trump even more, or maybe the deepfaker truly wanted to confuse people and hurt Trump's reputation and sway an election; this would have to be determined on a case-by-case basis.

For a real-life example, consider the controversy surrounding the "Yes Men," a group of pranksters that use parody and satire to comment on political and social issues. In 2009, The Yes Men published a fake press release that resembled one from the U.S. Chamber of Commerce, which falsely announced that the Chamber had changed its stance on climate regulation.[161] The Yes Men also hosted a faux press conference to accompany the press release.[162] The Chamber of Commerce sued the Yes Men in response, claiming the fake press release caused public confusion after some news outlets treated it like it was legitimate.[163] Ultimately, the Chamber dropped

159. *See* Hustler Magazine, Inc. v. Falwell, 485 U.S. 46, 50, 56 (1988).

160. *See* James Vincent, *Facebook's problems moderating deepfakes will only get worse in 2020*, THE VERGE (Jan. 15 2020, 12:36 PM), https://www.theverge.com/2020/1/15/21067220/deepfake-moderation-apps-tools-2020-facebook-reddit-social-media ("Facebook said it will remove 'manipulated misleading media' . . . But the company noted that this does not cover "parody or satire" or misleading edits made using traditional means, like last year's viral video of House Speaker Nancy Pelosi supposedly slurring her words.").

161. The parody press release is still accessible online. *U.S. Chamber Supports Climate Bill*, THE YES PEOPLE (Oct. 19, 2009), https://theyesmen.org/project/chamber/fakerelease [https://perma.cc/QP37-2JFF].

162. Anne C. Mulkern, *U.S. Chamber Sues Activists Over Climate Stunt*, N.Y. TIMES (Oct. 27, 2009), https://archive.nytimes.com/www.nytimes.com/gwire/2009/10/27/27greenwire-us-chamber-sues-activists-over-climate-stunt-50982.html [https://perma.cc/J45M-UZJS].

163. *Id.*; Chris Good*, The Yes Men Get Sued*, THE ATLANTIC (Oct. 27, 2009), https://www.theatlantic.com/politics/archive/2009/10/the-yes-men-get-sued/29131/ [https://perma.cc/3D8Z-JHVY].

the lawsuit after it extended on for four years.[164] This instance illustrates that parodies can be highly protected by the First Amendment. Arguing against parodies for libel claims can be even trickier, especially for public figures because of the heightened standards of "actual malice."

B. Deepfakes as a Vessel to Spread Political Misinformation

Outside of deepfakes made by innocent parodies, there is a concern that deepfake technology could be used to purposely disseminate lies and deceit.[165] So far, there are no proven cases of deepfakes being used to interfere with elections, although there have been some claims that deepfakes have been used to tamper with international politics.[166] As discussed in Part III, there have been some federal and state attempts to regulate deepfakes to prevent intrusion into elections.[167] However, some of those attempts, while passed successfully, have come under fire for infringing on individuals' First Amendment rights.[168]

If a deepfake is used to falsely stir up excitement and panic about an upcoming election, should the First Amendment still protect that speech? While precedent seems to indicate that it is protected,[169] it could be argued that, under certain circumstances, deceitful speech can have exacerbated harmful effects to society and should be regulated.[170] However, since the value of deceitful speech is already low,[171] if there was a need to block a harmful deepfake, a court could potentially classify it into one of the existing First Amendment exceptions – granted it actually fits into the preexisting

164. *U.S. Chamber Cries Uncle, Withdraws Lawsuit*, THE YES MEN (June 13, 2013), https://theyesmen.org/chambercriesuncle. *See also Yes Men Mourn Lawsuit Withdrawal*, THE YES MEN (last visited Mar. 14, 2020), https://theyesmen.org/lawsuitwithdrawal [https://perma.cc/P349-MR79].

165. *See* BuzzfeedVideo, *You Won't Believe What Obama Says In This Video!*, YOUTUBE (April 17, 2018), https://www.youtube.com/watch?v=cQ54GDm1eL0 [https://perma.cc/JK49-23BG] (showing a very convincing deepfake of President Barack Obama, created by Buzzfeed to show how the technology can be used to depict incorrect information).

166. McFaul, *supra* note 3; Nic Ker, *Is the political aide viral sex video confession real or a Deepfake?*, MALAYMAIL (June 12, 2019), https://www.malaymail.com/news/malaysia/2019/06/12/is-the-political-aide-viral-sex-video-confession-real-or-a-deepfake/1761422 [https://perma.cc/G5VB-U4LR] (describing how a Malaysian cabinet minister was subjected to controversy after an alleged pornographic deepfake went viral).

167. *See* TEX. ELEC. CODE ANN. § 255.004 (West 2019); H.R. 3230.

168. *See* Artz, *supra* note 138; Ingram, *supra* note 117.

169. United States v. Alvarez, 567 U.S. 709, 718 (2012).

170. *See* Helen Norton, *(at Least) Thirteen Ways of Looking at Election Lies*, 71 OKLA. L. REV. 117, 125–26 (2018).

171. Chesney & Citron, *supra* note 35, at 1791 ("Lies about the source of speech—whether a public official is actually speaking—do not serve free speech values.").

boundaries of one of those exceptions.[172] Perhaps the main exception that would be applied to controversial deepfakes would be the restrictions laid out for obscenity. Since deepfakes have a background in falsified pornographic videos, it may not be surprising if a deepfaker tried to sway an election by releasing a video of a candidate engaged in a sex act.[173]

Absent some exigent circumstances, the First Amendment will probably protect the lies spewed from misleading deepfakes, especially if they relate to politics.[174] However, just because the First Amendment protects certain speech does not mean deepfakes are excluded from private regulations or bans.[175] One solution is to leave the regulation up to private platforms like YouTube, Facebook, and Twitch since these platforms already have "the most advanced technologies to detect immoral, illegal or malicious content, flag it and remove it."[176] But under this model, it is possible that deepfakes could be regulated more than they normally would be under the First Amendment; private platforms can completely ban deepfakes from their websites, but under the First Amendment, it is assumed that an outright ban on deepfakes would be unconstitutional.[177] By leaving this regulation to private companies, the companies get complete discretion on which deepfakes have value and which are unacceptable, a role that typically the government plays – especially in regards to election tampering. However, it is likely these online platforms do not really want to take on the task of policing this content.[178] Although there is no direct incentive in place encouraging these platforms for remove deepfakes, there is increasing public pressure for them to regulate it, particularly after the wave of "fake news" concerns during the 2016 and 2020 election seasons.[179]

---

172. *Id.* ("Some deep fakes will fall into those categories and thus could be subject to regulation. This includes defamation of private persons, fraud, true threats, and the imminent-and-likely incitement of violence.").

173. *See* Ker, *supra* note 166 (describing how a Malaysian cabinet minister was subjected to controversy after an alleged pornographic deepfake went viral).

174. *See* New York Times Co. v. Sullivan, 376 U.S. 254, 269–70 (1964).

175. Meskys, *supra* note 154, at 31.

176. *Id.*

177. *Id.*; Chesney & Citron, *supra* note 35, at 1788–89; *see* Cole, *supra* note 30.

178. Olivier Sylvain, KNIGHT FIRST AMEND. INST., DISCRIMINATORY DESIGNS ON USER DATA (2018), https://knightcolumbia.org/content/discriminatory-designs-user-data [https://perma.cc/7WX3-VN36].

179. McCabe & Alba, *supra* note 28 ("False information spread furiously on the platform during the 2016 campaign, leading to widespread criticism of [Facebook]."); Sam Dean, *How Facebook and Twitter plan to handle election day disinformation*, LA TIMES (Nov. 2, 2020), https://www.latimes.com/business/story/2020-11-02/election-day-2020-disinformation-facebook-instagram-twitter-youtube; Chesney & Citron, *supra* note 35, at 1795 ("Online platforms already have an incentive to screen content, thanks to the impact of moral suasion, market dynamics, and political pressures.").

Additionally, while regulation gets its footing, some experts in privacy and technology have suggested that political campaigns create an eight-step plan in case a "deepfake emergency" takes place during election season:

> 1. Issue a statement that the candidate will not knowingly disseminate fake or manipulated media of opponents and urge campaign supporters to abide by the same commitment . . . 2. Get familiar with the terms of service and community guidelines for social media platforms on this issue, as well as the processes to report inappropriate content . . . 3. Designate a team ready to manage an incident . . . 4. Obtain a briefing on key trends and threats from knowledgeable experts . . . 5. Conduct an internal red teaming exercise to prepare for the range of ways a fake could target the candidate or campaign . . . 6. Establish relationships with company officials that will be helpful during an incident . . . 7. Establish procedures to quickly access original video and/or audio footage . . . 8. Prepare contingency web content or templates that could be swiftly used to counter false claims.[180]

While these suggestions do not directly regulate deepfakes, they provide a framework that could help society achieve a better understanding of how deepfakes operate and how they can be used to skew politics and disrupt the political process. According to the scholars behind these suggestions, by campaigns getting ahead of the threat of a "deepfake emergency," it helps candidates be on the battle lines, defending truthfulness over falsehoods.[181]

C. What Happens When Deepfakes Reach Courts?

At the time of this Note, only a few courts have mentioned deepfakes, and their discussion has been limited to the footnotes of some cases regarding the technology's impact on child pornography law.[182] Most state statutes currently addressing deepfakes have mainly focused on preventing tampering in the political process, but courts will probably see cases involving deepfakes relating to pornography first.[183] When this occurs, courts should look to

180. Katherine Charlet & Danielle Citron, *Campaigns Must Prepare for Deepfakes: This Is What Their Plan Should Look Like*, CARNEGIE ENDOWMENT FOR INTERNATIONAL PEACE (Sept. 5, 2019), https://carnegieendowment.org/2019/09/05/campaigns-must-prepare-for-deepfakes-this-is-what-their-plan-should-look-like-pub-79792 [https://perma.cc/RK7A-HH3T].

181. *See id.* ("A fake video of a candidate saying she prefers Coke to Pepsi is no big deal, but one where the candidate falsely appears saying or doing something despicable could endanger the candidacy and the democratic process.").

182. *See* In re S.K., 215 A.3d 300, 315 n.22 (2019); United States v. Streett, No. CR 14-3609 JB, 2020 WL 231688, at *47 n.28 (D.N.M. Jan. 15, 2020).

183. Robert Chesney & Danielle Citron, *21st Century-Style Truth Decay: Deep Fakes and the Challenge for Privacy, Free Expression, and National Security*, 78 MD.

comparable revenge porn cases and statutes for guidance.[184] Revenge porn cases can be compared to non-consensual pornographic deepfakes because both situations usually involve a person who intentionally distributes pornographic media without the consent of the victim, and the media would harm the victim's reputation.[185] Following this, it tracks that victims would likely bring false light privacy tort claims from these circumstances.[186] Alternatively, some prosecutors could bring charges against deepfakers if the content and dissemination of the deepfake fell under applicable revenge porn statutes.[187] In California, for example, a mix of these could be possible since its revenge porn statute now includes an exception that is very comparable to the elements laid out in the leading false light privacy case, *Time, Inc. v. Hill*.[188] Courts should be prepared to hear both of these types of cases, whether it be a civil or criminal case, and be willing to interpret the applicable precedent against the challenges brought by deepfake technology discussed in previous sections, such as parodies and other First Amendment concerns.

V. CONCLUSION

While deepfakes are a new phenomenon, they are quickly gaining popularity without an end in sight.[189] While there can be some innocence in altering videos for laughs and entertainment,[190] deepfake technology brings new challenges that have not been fully realized.[191] Lawmakers are slowly

---

L. REV. 882, 885–86 (2019) (showing two examples of private individuals that were sexually exploited by deepfake videos).

184. *See* Harris, *supra* note 73, at 119–20 ("Fairly recent nonconsensual pornography statutes may be the most effective legal recourse against uploaders of personal deepfakes featuring nonconsenting individuals.").

185. Russell Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 400 (2019) ("[C]ourts may want to consider whether non-consensual pornography of any kind, revenge porn or deepfakes, should be the foundation of a new exception to the First Amendment's broad protections.").

186. Harris, *supra* note 73, at 115.

187. *See* MD. CODE ANN., CRIM. LAW § 11-208(a)(2) (West 2019); VA. CODE ANN. § 18.2-386.2 (West 2019); CAL. CIV. CODE § 1708.86 (West 2019) (showing some revised revenge porn state statutes that include deepfakes).

188. CAL. CIV. CODE § 1708.86(c)(1)(B)(ii) (West 2019) (showing that "[a] work of political or newsworthy value or similar work" can be protected); Time, Inc. v. Hill, 385 U.S. 374, 377 (1967) (holding that when speech relates to matters of public interest, it is subjected to higher standards before the creator is subject to liability).

189. Meskys, *supra* note 155, at 24.

190. *See* Collider Videos, *Tom Cruise IMPOSSIBLE BURGER Challenge! (Deepfake Parody)*, YOUTUBE (Feb. 28, 2020), https://www.youtube.com/watch?v=ntnveCh691M [https://perma.cc/579X-U8YJ] (showing a deepfaked video of Tom Cruise going grocery shopping for the "Impossible Burger").

191. *See* Chesney & Citron, *supra* note 183, at 885 ("It is unclear who will win this arms race, but for now the fight is on.").

but surely learning more about the technology and its impact on politics and individual privacy.[192] In an age of "fake news," the need to understand deepfakes is pressing if we want to ensure trust in democracy and preserve the need for truth over lies.[193] Suppression of deepfake technology may mean suppression of rights.[194] But an absence of regulation leaves the nation vulnerable to election tampering and political dismantlement.[195] In this sense, the still waters of deepfake technology really do run… well, deep.

192. *See, e.g.*, S.1790 § 5709.

193. Schwartz, *supra* note 23 (quoting scholar Danielle Citron saying, "I'm starting to see how a well-timed deep fake could very well disrupt the democratic process.").

194. Meskys, *supra* note 155, at 29 ("Creative deep fakes could be considered a constitutive part of free speech.").

195. Chesney & Citron, *supra* note 35, at 1778.