NOTES

"The New Weapon of Choice":[1] Law's Current Inability to Properly Address Deepfake Pornography

Deepfake technology uses artificial intelligence to realistically manipulate videos by splicing one person's face onto another's. While this technology has innocuous usages, some perpetrators have instead used it to create deepfake pornography. These creators use images ripped from social media sites to construct—or request the generation of—a pornographic video showcasing any woman who has shared images of herself online. And while this technology sounds complex enough to be relegated to Hollywood production studios, it is rapidly becoming free and easy-to-use. The implications of deepfake pornography seep into all facets of victims' lives. Not only does deepfake pornography shatter these victims' sexual privacy, its online permanency also inhibits their ability to use the internet and find a job. Although much of the scholarship and media attention on deepfakes has been devoted to the implications of deepfakes in the political arena and the attendant erosion of our trust in the government, the implications of deepfake pornography are equally devastating. This Note analyzes the legal remedies available to victims, concludes that none are sufficient, and proposes a new statutory and regulatory framework to provide adequate redress.

1. Makena Kelly, *Congress Grapples with How to Regulate Deepfakes*, VERGE (June 13, 2019, 1:30 PM), https://www.theverge.com/2019/6/13/18677847/deep-fakes-regulation-facebook-adam-schiff-congress-artificial-intelligence [https://perma.cc/5AKJ-GKGZ] ("As Rep. Val Demings (D-FL) put it, 'the internet is the new weapon of choice.' ").

THE NEXT ITERATION OF REVENGE PORNOGRAPHY

One late night, eighteen-year-old Noelle Martin performed a reverse Google image[2] search on herself, only to discover that hundreds of images of her face had been grafted onto the bodies of pornography actresses engaged in sexual acts.[3] She had never even had a boyfriend—much less shared nude photographs of herself.[4] These falsified photos and videos were accompanied by her name and home address, all of which could be found by simply searching her name on the internet.[5]

2. Reverse Google image allows someone to Google an image to find related images online. *Find Related Images with Reverse Image Search*, GOOGLE, https://support.google.com/websearch/answer/1325808?co=GENIE.Platform%3DAndroid&hl=en (last visited June 28, 2020) [https://perma.cc/QJX6-5XGN].

3. Kristi Melville, *The Insidious Rise of Deepfake Porn Videos—And One Woman Who Won't Be Silenced*, AUSTL. BROADCASTING COMPANY (Aug. 29, 2019), https://www.abc.net.au/news/2019-08-30/deepfake-revenge-porn-noelle-martin-story-of-image-based-abuse/11437774 [https://perma.cc/NA26-H2G3].

4. *Id.*

5. Danielle Keats Citron, *Sexual Privacy*, 128 YALE L.J. 1870, 1923 (2019).

Like an ever-growing number of women, Martin had been victimized by deepfake pornography.[6] Deepfake technology uses artificial intelligence to realistically manipulate videos by splicing one person's face onto another's.[7] Consider this technology to represent the next iteration of revenge pornography: instead of leaking a nude image initially shared privately, the perpetrator can create—or request the creation of[8]—a pornographic video starring any woman who has shared images of herself, clothed or not, on social media. And while this technology sounds complex enough to be relegated to Hollywood production studios, it is rapidly becoming free and easy-to-use. Deepfake pioneer Hao Li predicted in September 2019 that deepfake technology will evolve to seamless portrayals and "will be accessible to everyday people in 'half-a-year to a year.' "[9]

Sexual privacy is at the pinnacle of privacy values and thus requires both acknowledgement and defensive tactics, like other recognized privacy violations.[10] Currently, no adequate legal solution exists that directly provides redress for the majority of victims of nonconsensual deepfake pornography.

This Note analyzes the existing legal remedies available to non-celebrity victims of deepfake pornography and concludes that none are sufficient to provide adequate redress, ultimately demonstrating the need for statutory intervention. Part I discusses deepfake pornography's technical background and its rise to prominence. Part II explains why currently available solutions fail to properly address deepfake pornography. Part III proposes civil legislation that imposes liability on both producers of deepfake pornography and the websites that knowingly harbor it.

---

6. Varying sources use "deepfake," "deep-fake," and "deep fake." For clarity, I will use "deepfake" throughout this Note.

7. *See infra* Section I.C.

8. *See infra* notes 40–45 and accompanying text.

9. Kevin Stankiewicz, *'Perfectly Real' Deepfakes Will Arrive in 6 Months to a Year, Technology Pioneer Hao Li Says*, CNBC (Jan. 17, 2020, 2:51 AM), https://www.cnbc.com/2019/09/20/hao-li-perfectly-real-deepfakes-will-arrive-in-6-months-to-a-year.html [https://perma.cc/8DJA-836Z].

10. Citron, *supra* note 5, at 1881 (naming "health privacy, financial privacy, communications privacy, children's privacy, educational privacy, and intellectual privacy" as other legally protected areas).

I. WHAT IS DEEPFAKE PORNOGRAPHY?

A. The Devastating Consequences of Deepfake Pornography

The repercussions of deepfake pornography seep into all facets of victims' lives. Not only does deepfake pornography shatter victims' sexual privacy, its online permanency also inhibits their ability to find a job or even simply use the internet without harassment.[11] Although much of the scholarship and media attention on deepfakes has been devoted to implications in the political arena and the attendant erosion of our trust in the government, deepfake pornography poses an equal risk to victims' lives.[12] The same issues of blackmail and extortion that present themselves with deepfakes of major political figures also apply in the personal sphere.[13]

Nonconsensual deepfake pornography "exclusively targets and harms women."[14] For example, one software application that produces nude images, DeepNude, will not function properly when presented with male images; it simply pastes a vulva over the male genitalia.[15] The technology itself, therefore, is not only weaponized to debase women—it also cannot be used to debase men.[16] Deepfake production and publication have rapidly proliferated, with an almost one hundred percent increase in the number of deepfake videos online between December 2018 and September 2019. Ninety-six percent of those videos are pornographic.[17] One hundred percent of the pornography on websites dedicated to deepfakes victimizes women.[18]

---

11. HENRY AJDER ET AL., THE STATE OF DEEPFAKES: LANDSCAPE, THREATS, AND IMPACT 6 (2019) [hereinafter DEEPTRACE] (citing Letter from Danielle Citron, Professor of Law, Boston Univ., to author (Aug. 16, 2019) (on file with author)) (describing the ramifications as "say[ing] to individuals that their bodies are not their own and [by] mak[ing] it difficult to stay online, get or keep a job, and feel safe"). *See infra* notes 21–32 and accompanying text.

12. Jesselyn Cook, *Here's What It's Like to See Yourself in a Deepfake Porn Video*, HUFFPOST U.S. (June 23, 2019, 12:00 PM), https://www.huffingtonpost.co.uk/entry/deepfake-porn-heres-what-its-like-to-see-yourself_n_5d0d0faee4b0a3941861fced?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cudGhldmVyZ2UuY29tLzIwMTkvNi8yNy8xODc2MDg5Ni9kZWVwZmFrZS1udWRlLWFpLWFwcC13b21lbi1kZWVwbnVkZS1ub24tY29uc2Vuc3VhbC1wb3Jub2dyYXBoeQ&guce_referrer_sig=AQAAAM9Hiy_RIBgzrHYZPWSEBfHW2gFeNU_Cu9ALNZuvvNpBkH9rJNt40cHnc36aaEwXSDiVA8voObzgwQNgtBLfGtNhldx5zkPkNLphGBeyRAEBJVK2R81A9JEIEUKKiU79ZYXDTQMUdsb2n1JLBuvccYh7bZz-19wGEi2s-rs6mtq [https://perma.cc/SM4G-MM4Y].

13. *See* Holly Kathleen Hall, *Deepfake Videos: When Seeing Isn't Believing*, 27 CATH. U. J.L. & TECH. 51, 52 (2018) (describing the potential ramifications of deepfakes).

14. DEEPTRACE, *supra* note 11, at 2.

15. James Vincent, *New AI Deepfake App Creates Nude Images of Women in Seconds*, VERGE (June 27, 2019, 6:23 AM), https://www.theverge.com/2019/6/27/18760896/deepfake-nude-ai-app-women-deepnude-non-consensual-pornography [https://perma.cc/6P6L-K4M5].

16. *See id.*

17. DEEPTRACE, *supra* note 11, at 1.

18. *Id.* at 2.

Deepfake pornography annihilates victims' sexual privacy[19] and inherently strips women of their humanity, "creating a sexual identity" they play no role in devising.[20] One woman compared appearing in a deepfake pornography video to "digital rape."[21] Deepfake pornograhy's intensely personal nature is used to intentionally attack women. For example, feminist media critic Anita Sarkeesian was featured in a "hardcore" deepfake pornography video, which garnered more than thirty thousand views on Pornhub.[22] Anonymous users celebrated the creation, gleefully commenting that "THIS is the deepfake we need and deserve, if no other reason other than principal [*sic*]" and that "[s]he attacked us first . . . . She just had to open her smarmy mouth."[23]

The loss of agency suffered by victims after being targeted by a deepfake pornography video is akin to physical assault, and the privacy invasions it threatens are painful and enduring.[24] One victim described her discomfort with "[b]eing violated in such an intimate way" as feeling like she was "being fetishized."[25] Another woman deemed the inability to control the spread of the deepfake as "grotesque" because the videos "are so horribly believable."[26] She has started to question her ability to freely use the internet: "As these videos get more prolific and realistic, is this something we're just going to be expected to accept as the cost of being online?"[27]

After Indian journalist Rana Ayyub's face was inserted into a pornographic video, her phone was inundated with violent messages from men "threaten[ing] to tear [her] clothes and drag [her] out of the country," propositioning her, sending her nude images of themselves,

---

19. Professor Citron defines sexual privacy as "the social norms (behaviors, expectations, and decisions) that govern access to, and information about, individuals' intimate lives" and "both descriptive and normative." Citron, *supra* note 5, at 1874.

20. *Id.* at 1921 (describing how deepfake pornography "reduce[s] individuals to genitalia, breasts, buttocks, and anuses").

21. Megan Farokhmanesh, *Is It Legal to Swap Someone's Face into Porn Without Consent?*, VERGE (Jan. 30, 2018, 2:39 PM), https://www.theverge.com/2018/1/30/16945494/deepfakes-porn-face-swap-legal [https://perma.cc/2LL8-BSD3].

22. Drew Harwell, *Fake-Porn Videos Are Being Weaponized to Harass and Humiliate Women: 'Everybody Is a Potential Target,'* WASH. POST (Dec. 30, 2018, 9:00 AM), https://www.washingtonpost.com/technology/2018/12/30/fake-porn-videos-are-being-weaponized-harass-humiliate-women-everybody-is-potential-target/ [https://perma.cc/NJ3X-V5MN].

23. *Id.*

24. *See id.*; *see also* Citron, *supra* note 5, at 1926 ("The emotional harm is severe and lasting, and the psychological distress can be overwhelming. Victims have difficulty concentrating, eating, and working. They experience anxiety and depression. They contemplate suicide." (footnotes omitted)).

25. Cook, *supra* note 12.

26. *Id.*

27. *Id.*

and urging others "to gang-rape [her]."[28] The anxiety induced by the video, as well as the constant phone notifications, sent her to the hospital. "This is a lot more intimidating than a physical threat," she said.[29] "This has a lasting impact on your mind. And there's nothing that could prevent it from happening to me again."[30]

Deepfake pornography videos also pose a substantial threat to a victim's job prospects if they remain on the internet, tied to her name. Eighty percent of employers perform internet searches on job candidates, and "in around seventy percent of cases, those results have a negative impact."[31] One Google search could uncover a deepfake sex tape in which the victim did not participate, permanently affecting her ability to find a job.[32]

B. Origins of Deepfake Pornography and Where Deepfake Technology Is Going

The term "deepfake" derives from a Reddit[33] user with the username "[u/]deepfakes."[34] In November 2017,[35] u/deepfakes posted on Reddit a video that purportedly featured actress Gal Gadot having sex with her stepbrother.[36] Fans of u/deepfakes's video created a subreddit dedicated exclusively to deepfake videos (r/deepfakes), which amassed

28. Rana Ayyub, *In India, Journalists Face Slut-Shaming and Rape Threats*, N.Y. TIMES (May 22, 2018), https://www.nytimes.com/2018/05/22/opinion/india-journalists-slut-shaming-rape.html [https://perma.cc/258B-9BFF].

29. Harwell, *supra* note 22.

30. *Id.*

31. Citron, *supra* note 5, at 1927.

32. *See id.*, at 1928:

> Companies may refuse to interview or hire women and minorities because their search results include nude images or deep-fake sex videos. Social norms about sexual modesty and gender stereotypes explain why women and minorities are more likely to suffer harm in the job market than heterosexual white men. Women—and especially nonwhite women—may be perceived as immoral sluts for engaging in sexual activity.

(footnote omitted).

33. Reddit is a popular online forum that hosts discussion topics, called "subreddits," in which users ("Redditors") can comment and vote. *About*, REDDIT, https://www.redditinc.com (last visited June 28, 2020) [https://perma.cc/7X59-VQZ9].

34. *See* Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All Fucked*, VICE: MOTHERBOARD (Dec. 11, 2017, 1:18 PM), https://www.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn [https://perma.cc/6M2V-7UD7] (discussing r/deepfakes, the subreddit where the first deepfake pornography videos were posted).

35. DEEPTRACE, *supra* note 11, at 3.

36. *See* Cole, *supra* note 34 ("There's a video of Gal Gadot having sex with her stepbrother on the internet. But it's not really Gadot's body, and it's barely her own face. It's an approximation, face-swapped to look like she's performing in an existing incest-themed porn video.").

more than fifteen thousand subscribers within two months[37] and eventually boasted ninety thousand subscribers.[38] U/deepfakes and other Redditors posted deepfake pornography featuring female celebrities such as Scarlett Johansson, Maisie Williams, Taylor Swift, and Aubrey Plaza.[39]

Celebrity deepfakes soon gave way to deepfake pornography starring people outside of the public eye. Discussions spread from Reddit to other forum-based websites like 4chan, 8chan, and Voat.[40] A cottage industry was born as certain deepfake producers began creating videos by request.[41] For example, most of the posts in a subreddit dubbed "doppelbangher" requested deepfakes of people the requestor knew in real life: a " 'friend's stepmom,' 'coworker of mine,' 'college friend,' 'a friend of mine and my crush,' and 'hottest girl in engineering.' "[42]

This deepfake commodification includes both online businesses dedicated to creating and selling individualized deepfake videos and individual creators.[43] Although prices vary based on the quality and duration of the requested video, a deepfake can be bought for as little as $2.99.[44] The requestor must typically provide at least 250 photos of the victim, usually ripped from photos the latter has posted to social media sites.[45]

Although Reddit eventually banned r/deepfakes, citing the subreddit as a violation of its community standards, the damage was done.[46] Pornography websites dedicated exclusively to deepfakes have

37. *See* Samantha Cole, *We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now*, VICE: MOTHERBOARD (Jan. 24, 2018, 12:13 PM), https://www.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley [https://perma.cc/CP49-8YCG] (discussing the exponential growth in deepfake pornography since the creation of r/deepfakes).

38. Samantha Cole, *Reddit Just Shut Down the Deepfakes Subreddit*, VICE: MOTHERBOARD (Feb. 7, 2018, 12:35 PM), https://www.vice.com/en_us/article/neqb98/reddit-shuts-down-deepfakes [https://perma.cc/VPJ5-8MYR].

39. Cole, *supra* note 34.

40. DEEPTRACE, *supra* note 11, at 4.

41. *See* Harwell, *supra* note 22 (describing how anonymous users on deepfakes discussion boards and private chats have requested deepfakes).

42. *See* Samantha Cole, *People Are Using AI to Create Fake Porn of Their Friends and Classmates*, VICE: MOTHERBOARD (Jan. 26, 2018, 1:00 PM), https://www.vice.com/en_us/article/ev5eba/ai-fake-porn-of-friends-deepfakes [https://perma.cc/YU3Z-8CYZ].

43. DEEPTRACE, *supra* note 11, at 5.

44. *Id.*

45. *Id.*

46. u/TheRedCow, */r/Deepfakes Has Been Banned*, REDDIT, https://www.reddit.com/r/SFWdeepfakes/comments/7vy36n/rdeepfakes_has_been_banned/ (last visited July 15, 2020) [https://perma.cc/9WYE-6SU9].

already emerged,[47] and despite new policies banning nonconsensual deepfakes, websites such as Pornhub, Gfycat, and Twitter still host deepfake pornography.[48] Deepfake technology spread like wildfire, with an almost one hundred percent increase in the number of deepfake videos online over a ten-month period.[49] And of the deepfakes online today, the vast majority—ninety-six percent—contain pornographic content.[50]

Without a specific legal response, there is no indication that deepfake technology will slow down. On the contrary, "[s]oon, it's going to get to the point where there is no way that we can actually detect [deepfakes] anymore . . . ."[51] If the progress of technology in general is any indication, deepfake software will only grow easier to locate and use and, if unchecked, will result in more sophisticated deepfakes and a rapidly expanding list of victims. Edward Delp, a media forensics expert working with the Defense Advanced Research Projects Agency ("DARPA"),[52] described this proliferation as an "arms race," warning that "[a]s the people making these videos get more and more sophisticated with their tools, we're going to have to get more and more sophisticated with ours."[53] For example, while unblinking eyes were originally a key indicator of a deepfake, some videos now contain manipulated faces that seem to blink organically.[54]

47. *See* DEEPTRACE, *supra* note 11, at 6 ("We found that the deepfake pornography ecosystem is almost entirely supported by dedicated deepfake pornography websites, which host 13,254 of the total videos we discovered.").

48. *See* Samantha Cole, *Gfycat's AI Solution for Fighting Deepfakes Isn't Working*, VICE: MOTHERBOARD (June 19, 2018, 9:00 AM), https://www.vice.com/en_us/article/ywe4qw/gfycat-spotting-deepfakes-fake-ai-porn [https://perma.cc/8FPW-YWAK] [hereinafter Cole, *Gfycat*] (discussing how this website has implemented anti-deepfake pornography terms of use); Samantha Cole, *Pornhub Is Banning AI-Generated Fake Porn Videos, Says They're Nonconsensual*, VICE: MOTHERBOARD (Feb. 6, 2018, 1:50 PM), https://www.vice.com/en_us/article/zmwvdw/pornhub-bans-deepfakes [https://perma.cc/G3LR-63DB] [hereinafter Cole, *Pornhub*] (same); Samantha Cole, *Twitter Is the Latest Platform to Ban AI-Generated Porn*, VICE: MOTHERBOARD (Feb. 6, 2018, 5:12 PM), https://www.vice.com/en_us/article/ywqgab/twitter-bans-deepfakes [https://perma.cc/36D6-8RR3] [hereinafter Cole, *Twitter*] (same).

49. DEEPTRACE, *supra* note 11, at 1.

50. *Id.*

51. Stankiewicz, *supra* note 9 (alteration in original) (quoting Hao Li).

52. DARPA is a branch of the U.S. Department of Defense. *See* Cook, *supra* note 12.

53. *Id.*

54. Jesselyn Cook, *Deepfake Videos and the Threat of Not Knowing What's Real*, HUFFPOST US (June 12, 2019, 4:03 PM), https://www.huffpost.com/entry/deepfake-videos-and-the-threat-of-not-knowing-whats-real_n_5cf97068e4b0b08cf7eb2278 [https://perma.cc/ZKH7-RBSA].

C. How Deepfakes Work

Deepfake technology originated at the University of Montreal,[55] where a team introduced the idea in 2014 "by comparing it to the duel between counterfeiters and the police, with both sides driven 'to improve their methods until the counterfeits are indistinguishable.' "[56] This technology is called Generative Adversarial Networks, or GANs, and sets two models in opposition:[57] the first—the discriminative algorithm—classifies the input data,[58] while the other—the generative model—creates "data" identical to "the dataset."[59] As applied to deepfakes,

> the generator constructs new video frames, while the discriminator tries to discern whether the frame, with its superimposed subject, is authentic (say, an actual video frame of the original actor) or fake (a doctored video frame of the actor in a compromising position). If the discriminator cannot tell the real images from the false images, a human may not be able to either.[60]

The tools needed to create deepfakes are open-source and free to the general public, allowing developers to create programs like FakeApp[61] and DeepNude.[62] These software remove the previous technological barriers to entry and allow anyone with "a computer and a robust collection of photos" to create fake pornographic videos starring any woman with an online presence.[63] All that is required is clicking a button and feeding photos of clothed women into the DeepNude software for it to pop out fake photos of those women nude.[64]

---

55. Russell Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 343 (2019).

56. Harwell, *supra* note 22.

57. *See* Spivak, *supra* note 55, at 342–45 (discussing the technology behind deepfakes).

58. This "classification" process is analogous to how discriminative algorithms predict whether an incoming email is spam or not. *See id.* at 342 (discussing the technology behind deepfakes).

59. *Id.* at 343. To continue the analogy of an email filter: "Instead of predicting a label given certain features, it attempts to predict features given a certain label." *Id.*

60. *Id.* at 345.

61. *See* Kevin Roose, *Here Come the Fake Videos, Too*, N.Y. TIMES (Mar. 4, 2018), https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html [https://perma.cc/XP7M-E5P6] (FakeApp is "a program that was built by an anonymous developer using open-source software written by Google" that "makes it free and relatively easy to create realistic face swaps and leave few traces of manipulation.").

62. *See* DEEPTRACE, *supra* note 11, at 8 ("DeepNude is a computer app that enables users to 'strip' photos of clothed women. . . . These algorithms cannot perform similar translations on images of men, having been specifically trained on images of women."). Although the creators of DeepNude took the official website down, the software continues to be available from other sites. *Id.*

63. Harwell, *supra* note 22.

64. Vincent, *supra* note 15.

To craft a seamless deepfake, the producer must find the most performer that looks to most similar to their intended victim, although this process has evolved toward automation.[65] Browser-based applications now allow a producer "to upload a photo" of the intended victim, "and the website outputs the most comparable adult performer" (called a "faceset"). Then, the producer can easily locate the best match for an unbroken video.[66] Then, all the producer has to do is download pornographic videos of that adult performer from a website like Pornhub to use as the base of the deepfake.[67] Combined with open-source tools that are able to rip a victim's photos from her social media page in one swoop, this process has become much less labor-intensive.[68]

II. THE INADEQUACY OF THE CURRENT LEGAL REGIME

As of this writing, no victim of deepfake pornography has challenged her injury in court. Nevertheless, the increasing ease-of-use, availability, and proliferation of deepfake software will lead to more creations—resulting inevitably in litigation challenging both deepfake creation and deepfake dissemination. Although many existing legal claims might provide adequate redress in highly specific circumstances, none are sufficient to address deepfake pornography at large, revealing the need for a new solution.

Section A discusses how the First Amendment and section 230 of the Communications Decency Act ("CDA") both act as powerful barriers to recovery for victims and will, in most cases, likely preclude victims from suing producers and platforms. Subsections B.1 and B.2 explain that, while tort and copyright law could provide a cause of action for some victims, their inability to address the vast majority of situations nullifies their viability. Subsection B.3 analyzes the inadequacy of state revenge pornography statutes and federal legislation. Although those statutes seem to provide the most analogous protection for deepfake victims, they are also geared toward protecting against the revelation of real images. Meanwhile, deepfakes are not exactly real, yet not exactly fake—exposing a tension that forestalls recovery under state revenge pornography statutes.[69] Finally,

65. Douglas Harris, Note, *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, 17 DUKE L. & TECH. REV. 99, 100–01 (2019).

66. *Id.* at 101.

67. *See* Cole, *supra* note 37 (discussing the process by which producers create pornographic deepfakes).

68. Harris, *supra* note 65, at 101.

69. *See* Emma Grey Ellis, *People Can Put Your Face on Porn—and the Law Can't Help You*, WIRED (Jan. 26, 2018, 7:00 AM) https://www.wired.com/story/face-swap-porn-legal-limbo [https://perma.cc/H8GY-6EC8] ("And it's the very artifice involved in these videos that provides

Subsection B.5 explores how many video-hosting websites have attempted to prevent deepfakes from appearing on their sites, although their solutions have limited practicability and sustainability.

A. Barriers to Potential Legal Remedies

Many legal remedies will fail to serve as feasible solutions to litigating deepfake pornography due to the protections of the First Amendment, which could provide producers with a powerful defense. Moreover, section 230 of the CDA presumptively bars most, if not all, potential litigation against the websites that host deepfake pornography.

1. The First Amendment and Obscenity

Any new law that addresses deepfake pornography must be narrowly tailored to avoid implicating the First Amendment, which prohibits state action impinging on freedom of speech.[70] The Supreme Court has zealously defended First Amendment protections, upholding only those restrictions that protect other fundamental rights.[71] In *Reed v. Town of Gilbert*, the Court held that content-based constraints on speech are presumptively invalid, shifting the burden to the government to prove that its laws were narrowly tailored to serve a compelling governmental interest.[72] A law is deemed content-based if it regulates material based on "the topic discussed or the idea or message expressed."[73]

The First Amendment, though, does not protect absolutely all speech. In particular, it does not protect obscene material.[74] If deepfake pornography is determined to fall within the Court's definition of obscenity, then regulations prohibiting its nonconsensual usage can likely survive the strict scrutiny imposed by *Reed*, if narrowly tailored.[75] But while deepfake pornography is certainly obscene by most layperson's standards, the Supreme Court does not judge obscenity

---

enormous legal cover for their creators. 'It falls through the cracks because it's all very betwixt and between,' says Danielle Citron, a law professor at the University of Maryland . . . .").

70. U.S. CONST. amend. I.

71. *See, e.g.*, Palko v. Connecticut, 302 U.S. 319, 327 (1937) (Justice Cardozo lauded freedom of expression as "the matrix, the indispensable condition, of nearly every other form of freedom.").

72. 135 S. Ct. 2218, 2226 (2015).

73. *Id.* at 2227.

74. *See* Miller v. California, 413 U.S. 15, 23 (1973) ("This much has been categorically settled by the Court, that obscene material is unprotected by the First Amendment.").

75. 135 S. Ct. at 2227.

solely based on general public opinion.[76] In *Miller v. California*, the Court articulated a framework to determine if a given piece is obscene:

> The basic guidelines for the trier of fact must be: (a) whether "the average person, applying contemporary community standards" would find that the work, taken as a whole, appeals to the prurient interest, (b) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law, and (c) whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value.[77]

Whether the Court would deem deepfake pornography obscene is difficult to determine; the Court has yet to be confronted with this issue as of this writing. The closest the Court came to addressing deepfake pornography was *Ashcroft v. Free Speech Coalition*, a 2002 case in which the Court examined whether the Child Pornography Prevention Act of 1996 ("CPPA") abridged freedom of speech.[78] The challenged provision prohibited virtually created child pornography.[79] Although the Court acknowledged the inherent and universal abhorrence of child sexual abuse,[80] it nevertheless struck down the provision on the grounds that it restricted freedom of speech[81] and failed to account for the *Miller* framework.[82] In his majority opinion, Justice Kennedy emphasized the distinction between the actual creation of the material and its substance.[83] He distinguished between virtual child pornography and child abuse based on the harm involved.[84] In his opinion, virtual child pornography, unlike child abuse, did not involve an underlying crime or result in actual victims.[85] Justice Kennedy also noted that, "[e]ven where there is an underlying crime . . . the Court has not allowed the suppression of speech in all cases."[86]

While deepfake pornography is similar to simulated child pornography in that it depicts an act that did not actually happen, it differs in that it causes actual harm—emotional, physical, and

---

76. The Supreme Court declined to follow Justice Stewart's articulation of obscenity in his concurrence from *Jacobellis v. Ohio*: "I know it when I see it." 378 U.S. 184, 197 (1964) (Stewart, J., concurring).

77. 413 U.S. at 24 (citation omitted).

78. 535 U.S. 234, 239 (2002).

79. *Id.*

80. *Id.* at 244 ("The sexual abuse of a child is a most serious crime and an act repugnant to the moral instincts of a decent people.").

81. *Id.* at 257.

82. *Id.* at 246 ("The CPPA, however, extends to images that appear to depict a minor engaging in sexually explicit activity without regard to the *Miller* requirements.").

83. *Id.* at 250.

84. *Id.* at 254.

85. *Id.* at 250–51 (holding that the CPPA overreached by "prohibit[ing] speech that records no crime and creates no victims by its production").

86. *Id.* at 254; *see also* Miller v. California, 413 U.S. 15, 24 (1973) (establishing the *Miller* framework for assessing obscenity).

financial.[87] But the *Free Speech Coalition* majority's distinction between production that causes real harm and production that does *not* cause real harm reframes the obscenity inquiry to one that distinguishes between reality and the appearance of reality.[88] The child pornography depicted in *Free Speech Coalition* did not involve the likenesses of actual children, while deepfake pornography does portray the likenesses of actual women.[89] Deepfake pornography, while closer to the appearance of reality than reality itself, does cause actual harm and differs from virtually created child pornography in that it depicts real victims. Legislation proscribing deepfake pornography could thus be narrowly tailored to avoid invoking the First Amendment concerns articulated in *Free Speech Coalition*.[90]

Another consideration is whether the underlying pornographic video itself is obscene. When the Court struck down the CPPA provision banning virtual child pornography, it emphasized the reality of the production over the portrayal. Whether deepfake pornography is obscene, therefore, may hinge on whether the underlying pornographic video is deemed obscene.[91] Pornography featuring consenting adults is not per se obscene.[92] Complicating this issue is that *Miller* left regulation of obscenity up to the states, citing the differences among states in what residents consider palatable.[93] Per the Court's holding in *Miller*, then, pornography must depict "patently offensive 'hard core' sexual conduct" as "specifically defined by the regulating state law" to be considered obscene.[94] Yet what constitutes non-hardcore sexual conduct and hardcore sexual conduct remains nebulous. This leaves the legal status of pornography—as in, whether it is obscene—unresolved. But internet pornography constitutes at least a billion-dollar industry in the United States; thus, its proliferation suggests that at least some commodified pornography falls into the non-hardcore sexual conduct category.[95]

---

87. *See supra* Section I.A.

88. *See* Harris, *supra* note 65, at 106 (questioning "whether obscenity lies in the reality of thing deemed obscene or in the depiction of what registers as real").

89. *See id.*

90. *Id.*

91. *See* Spivak, *supra* note 55, at 361 (explaining that because deepfakes are simply images superimposed onto existing video any obscenity is derived from the video itself).

92. *See, e.g.*, Roth v. United States, 354 U.S. 476, 487 (1957) ("[S]ex and obscenity are not synonymous.").

93. Miller v. California, 413 U.S. 15, 33 ("People in different States vary in their tastes and attitudes, and this diversity is not to be strangled by the absolutism of imposed uniformity.").

94. *Id.* at 27.

95. Ross Benes, *Porn Could Have a Bigger Economic Influence on the US than Netflix*, QUARTZ (June 20, 2018), https://qz.com/1309527/porn-could-have-a-bigger-economic-influence-on-the-us-than-netflix/ [https://perma.cc/L4VZ-EPLA].

Most states have adopted obscenity laws mirroring the *Miller* framework.[96] California, the state in which the vast majority of pornographic videos are filmed, legalized non-obscene pornography in *People v. Freeman*.[97] The film at issue depicted sexual intercourse but was still deemed non-obscene.[98] This holding could mean that if deepfake producers use legal pornography as the underlying video of their creations, the deepfakes will likely not be considered obscene, and at least some producers will therefore enjoy First Amendment protections (at least in California).[99] This protection exposes the need for legislation that is narrowly tailored and serves a compelling state interest in order to survive strict scrutiny.[100]

The First Amendment imposes a hefty burden to overcome in regulating deepfake pornography. The Court's holding in *Free Speech Coalition* facially appears to render any would-be regulation of deepfake pornography unconstitutional—after all, if the Court was unwilling to uphold a law prohibiting virtual child pornography, which is repugnant to virtually all persons, would it not also be unwilling to uphold a law regulating deepfake pornography?[101] But the distinction between actual harm and no harm that underpinned Justice Kennedy's holding may actually bolster regulation of deepfake pornography.[102] Indeed, there *is* actual harm resulting from deepfake pornography.[103] Any new law, therefore, must be narrowly tailored to and hyperfocused on providing remedies for the harm inflicted on victims. That way, the First Amendment's salience as a defense is greatly diminished.

---

96. *See, e.g.*, ALA. CODE § 13A-12-200.1(17) (1998) (defining obscene material along the lines discussed in *Miller*); ARK. CODE ANN. § 5-68-403(2) (1961) (same); CAL. PENAL CODE § 311(a) (West 1961) (same); COLO. REV. STAT. § 18-7-101(2) (1981) (same); GA. CODE ANN. § 16-12-80(b) (1878) (same); 720 ILL. COMP. STAT. 5/11-20(b) (1988) (same); KAN. STAT. ANN. § 21-6401(f)(1) (2010) (same); MO. REV. STAT. §§ 573.010(11)(a)-(c) (1977) (same); N.Y. PENAL LAW § 235.00 (McKinney 2003) (same); OHIO REV. CODE ANN. § 2907.01(F) (LexisNexis 1974) (same); TENN. CODE ANN. § 39-17-901(10) (1989) (same); TEX. PENAL CODE ANN. § 43.21(a)(2) (West 1974) (same); VA. CODE ANN. § 18.2-372 (1950) (same).

97. *See* 758 P.2d 1128, 1129 (Cal. 1988) (reversing Freeman's pandering conviction on First Amendment grounds); Melia Robinson, *How LA's 'Porn Valley' Became the Adult Entertainment Capital of the World*, BUS. INSIDER (Mar. 6, 2016, 1:33 PM), https://www.businessinsider.com/how-porn-valley-came-to-be-2016-3 [https://perma.cc/H3J5-6BKW] (describing the rise in popularity of California pornographic filming).

98. *Freeman*, 758 P.2d at 1129.

99. *See* Harris, *supra* note 65, at 105 (noting that not all pornographic deepfakes can be considered obscene under the *Miller* test and thus at least some are constitutionally protected).

100. *See, e.g.*, Reed v. Town of Gilbert, 135 S. Ct. 2218, 2226 (2015) (imposing strict scrutiny on content-based regulations of speech).

101. Ashcroft v. Free Speech Coal., 535 U.S. 234, 237–38 (2002).

102. *Id.* at 250.

103. *See supra* Section I.A (discussing the harms involved with deepfake pornography).

2. Section 230 of the Communications Decency Act and Publishers' Hyperimmunity

Websites that host deepfake pornography uploaded by third parties enjoy immunity under section 230 of the Communications Decency Act ("CDA"), leaving victims with no other option than to sue the video producers.[104] And since producers can vanish from the internet and become impossible to track down, section 230 effectively precludes victims from seeking redress of any kind for the myriad consequences of deepfake pornography.[105]

The CDA was included as part of the Telecommunications Act of 1996, with the aim of "mak[ing] the internet safer for kids" and tackling worries about pornography.[106] Section 230 of the CDA, an amendment proposed by Representatives Christopher Cox and Ron Wyden,[107] had the twin goals of encouraging the development of the internet[108] while also guaranteeing robust prosecution of cyber "obscenity, stalking, and harassment" laws.[109] Section 230, in relevant part, immunizes online platforms from civil liability for third-party posts on their sites and for efforts taken to screen the content posted on their sites.[110] Specifically, it provides:

> (c) Protection for "Good Samaritan" blocking and screening of offensive material
>
> (1) Treatment of publisher or speaker. No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider.
>
> (2) Civil liability. No provider or user of an interactive computer service shall be held liable on account of—

---

104. *See* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1792 (2019) ("The attribution problem arises . . . because the metadata relevant for ascertaining a deep fake's provenance might be insufficient to identify . . . who generated it. . . . A careful distributor . . . may take pains to be anonymous, including . . . using technologies like Tor. When . . . employed, the IP addresses connected to posts may be impossible to find . . . ." (footnote omitted)).

105. *Id.*

106. *Fostering a Healthier Internet to Protect Consumers: Hearing Before the Subcomms. on Commc'ns & Tech. and Consumer Prot. & Commerce of the H. Comm. on Energy & Commerce*, 116th Cong. 3 (2019) [hereinafter *Fostering a Healthier Internet to Protect Consumers*] (statement of Danielle Keats Citron, Professor of Law, Boston University School of Law).

107. Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 406 (2017).

108. Specifically, section 230 sought to "promote . . . development of the Internet . . . preserve [its] vibrant and competitive free market . . . encourage . . . technologies which maximize user control over what information is received . . . [and] remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access to . . . inappropriate online material." 47 U.S.C. § 230(b)(1)-(4) (2012).

109. *Id.* § 230(b)(5).

110. *Id.* § 230(c).

> (A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or
>
> (B) any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).[111]

Section 230 provides sweeping, far-reaching protections for websites.[112] No matter what users post—whether it is constitutionally protected or not—online platforms are not liable for the post or the damage it causes.[113] Even if the Supreme Court deems deepfake pornography obscene, the platforms on which the videos are hosted are protected from liability by section 230—leaving only the producer of the video potentially liable for the harm.[114]

A New York Supreme Court case, *Stratton Oakmont, Inc. v. Prodigy Services Co.*,[115] provided the impetus for the adoption of section 230.[116] In that 1995 case, Stratton Oakmont, a now-defunct securities investment banking firm,[117] sued Prodigy, a website that hosted online bulletin boards, for a post by a third party that accused the firm of committing criminal fraud.[118] At issue was whether Prodigy could be held liable for the user's allegations.[119] The court held that, by "actively" using software to screen posts for offensive language and employees to remove distasteful posts, Prodigy made decisions regarding content on its site and thus acted as a publisher.[120] Prodigy therefore incurred liability as a publisher through its partial attempts to regulate offensive content on its site.[121]

Following the *Prodigy* decision, lawmakers worried that holding websites accountable for only partially screening third-party posts would render the opposite effect: a complete lack of screening in order to evade liability for improperly screening.[122] Under *Prodigy*, by

---

111. *Id.*

112. Citron & Wittes, *supra* note 107, at 408 (describing section 230 as "an immunity from liability that is far more sweeping than anything the law's words, context, and history support").

113. *Id.*

114. *Id.*

115. No. 31063/94, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995).

116. Citron & Wittes, *supra* note 107, at 404–05 (citing *Stratton Oakmont, Inc.*, 1995 WL 323710).

117. *See* THE WOLF OF WALL STREET (Paramount 2013) (chronicling the rise and fall of Stratton Oakmont and its founder, Jordan Belfort).

118. *Stratton Oakmont, Inc.*, 1995 WL 323710, at *1.

119. *Id.*

120. *Id.* at *4.

121. Citron & Wittes, *supra* note 107, at 405.

122. *Id.*

declining to monitor any content on their websites, publishers could escape liability altogether.[123] To prevent this result, Senators J. James Exon and Slade Gordon introduced the CDA in 1995 to "incentivize the adoption of new technologies and policies that would restrict access to offensive material."[124] Section 230 was then added to immunize so-called "Good Samaritan" providers.[125]

Ironically, although this protection was intended to shield publishers who attempt to constrain access to objectionable material,[126] courts have crafted a much broader construction: an encompassing protection from liability, unmoored from what its drafters envisioned.[127] Courts have interpreted section 230 to immunize websites from liability "even though they republished content knowing it might violate the law, encouraged users to post illegal content, changed their design and policies for the purpose of enabling illegal activity, or sold dangerous products."[128] As a result, section 230 has been lionized to mythic status as courts yield to websites' near-complete immunity from potential liability incurred from users' posts.[129]

In sum, the First Amendment and section 230 erect nearly insurmountable barriers to relief for victims of deepfake pornography. Producers of deepfake pornography can simply vanish from the internet—or take precautions to ensure that they cannot be tracked down.[130] But if a victim locates the producer, the producer will likely be able to use the First Amendment as a defense against allegations.[131] In most cases, then, this ability to evade detection leaves no other actionable option for victims aside from the websites that host deepfake pornography.[132] But, under section 230, the online platforms enjoy

---

123. *Id.*

124. *Id.* at 405–06.

125. 47 U.S.C. § 230(c)(2)(A) (2012); *see also supra* notes 106–111 and accompanying text (discussing the adoption, purposes, and operation of Section 230).

126. 47 U.S.C. § 230(c)(2)(A).

127. Citron & Wittes, *supra* note 107, at 408 (describing section 230 as "an immunity from liability that is far more sweeping than anything the law's words, context, and history support").

128. *Id.* (footnotes omitted).

129. *See, e.g.*, *CDA 230: The Most Important Law Protecting Internet Speech*, ELECTRONIC FRONTIER FOUND., https://www.eff.org/issues/cda230 (last visited July 16, 2020) [https://perma.cc/8PNF-W7B6] ("In short, CDA 230 is perhaps the most influential law to protect the kind of innovation that has allowed the Internet to thrive since 1996.").

130. *See* Chesney & Citron, *supra* note 104, at 1792 ("The attribution problem arises . . . because the metadata relevant for ascertaining a deep fake's provenance might be insufficient to identify . . . who generated it. . . . A careful distributor . . . may take pains to be anonymous, including . . . using technologies like Tor. When . . . employed, the IP addresses connected to posts may be impossible to find . . . ."(footnote omitted)).

131. *See supra* Section II.A.1 (discussing how First Amendment protections might shield producers of deepfake pornography).

132. Chesney & Citron, *supra* note 104, at 1792.

near-unlimited immunity from any user posts.[133] The combination of the First Amendment and section 230, therefore, effectively precludes victims from locating a defendant and seeking redress.

B. Possible Remedies

Even assuming a victim can track down a defendant, the existing remedies will still likely only cover a minority of victims, winnowing down the availability of recovery even further. This Section assesses the potential likelihood of stating a claim or succeeding in court using tort, copyright, and state nonconsensual pornography laws and discusses the efficacy of corporate social governance. It also considers both expired and pending federal legislation drafted specifically to address deepfake pornography.

1. The Pitfalls of Tort Law

Tort law could present a viable method to address deepfake pornography on a case-by-case basis, but it suffers from fundamental issues, which disqualify it from serving as a sufficient solution. Although privacy-based torts are the most logical claim because deepfake pornography assuredly poses a distinct intrusion into a victim's sexual privacy in the eyes of society,[134] the specific invasion posed by deepfake pornography is not easily encompassed by these types of torts. Generally, an individual's right to privacy is "the right to be let alone"[135] and invokes liability for harm incurred from encroachment on that entitlement to privacy.[136] Four subtypes constitute privacy tort law: intrusion on seclusion,[137] wrongful appropriation,[138] false light,[139] and public disclosure of private fact.[140]

An intrusion on seclusion claim imposes liability on an individual who "intentionally" invades another's privacy if the invasion would be considered "highly offensive" under a reasonableness standard.[141] While one would think deepfake pornography certainly

133. 47 U.S.C. § 230(c) (2012).

134. *See* Chesney & Citron, *supra* note 104, at 1794 (discussing the application of privacy-based torts to deepfakes).

135. RESTATEMENT (SECOND) OF TORTS § 652A cmt. a (AM. LAW INST. 1977).

136. *See id.* § 652A(1) (invoking liability "for the resulting harm to the interests of the other" when their right to privacy is invaded).

137. *Id.* § 652A(2)(a).

138. *Id.* § 652A(2)(b).

139. *Id.* § 652A(2)(d).

140. *Id.* § 652A(2)(c).

141. *Id.* § 652B.

constitutes an intentional interference of a highly offensive nature,[142] an intrusion on seclusion claim applies only to places in which a defendant could reasonably expect privacy.[143] The internet is not conceptualized as a space where people can reasonably expect privacy,[144] however, and since most deepfakes are created using images the victim herself has posted online,[145] the success of an intrusion on seclusion claim is limited. The limitations on an intrusion on seclusion claim expose the legal tension between an actual portrayal and a false depiction.[146]

Wrongful appropriation and right of publicity[147] claims both require a tortfeasor to have benefitted commercially from unlawfully using a victim's likeness. [148] But not every producer of deepfake pornography makes money from the creation. Although some deepfake producers do create pornographic videos at the request of clients and thus derive a financial benefit from the creation,[149] many others create the videos for their own gratification. The financial benefit requirement in these tort claims makes them an awkward fit for many victims, severely curtailing the effectiveness of these claims in policing the creation of these videos.[150]

---

142. *Id.*

143. Chesney & Citron, *supra* note 104, at 1795.

144. *See id.*; *see also* Jonathan Zittrain, *A World Without Privacy Will Revive the Masquerade*, ATLANTIC (Feb. 7, 2020), https://www.theatlantic.com/technology/archive/2020/02/we-may-have-no-privacy-things-can-always-get-worse/606250/ [https://perma.cc/VT94-ZKU2] (discussing Clearview AI, a company that "scraped billions of photos from social-networking and other sites on the web—without permission from the sites in question, or the users who submitted them—and built a comprehensive database of labeled faces primed for search by facial recognition." Clearview defends this collection by arguing that the data is public.).

145. *See supra* note 45 and accompanying text.

146. *See supra* notes 78–95 and accompanying text; *infra* Section II.B.3.a.

147. The right-of-publicity tort "permits compensation for the misappropriation of someone's likeness for commercial gain." Chesney & Citron, *supra* note 104, at 1794.

148. *See* RESTATEMENT (SECOND) OF TORTS § 652C cmt. d (AM. LAW INST. 1977) ("It is only when the publicity is given for the purpose of appropriating to the defendant's benefit the commercial or other values associated with the name or the likeness that the right of privacy is invaded.").

149. *See supra* notes 43–45 and accompanying text.

150. *See* Chesney & Citron, *supra* note 104, at 1794 ("The commercial-gain element sharply limits the utility of this model: the harms associated with deep fakes do not typically generate direct financial gain for their creators."); Spivak, *supra* note 55, at 383:

> Wrongful appropriation cases, particularly those involving digital images of one's likeness, are almost always using the victim's likeness to endorse or advertise a particular product. A deepfake, thus, presents an atypical fact pattern because deepfakers may not be attempting to create their own commercial benefit like the typical defendant in a wrongful appropriation case.

(footnote omitted).

False light torts punish tortfeasors who publicly, knowingly, and offensively put the victim "in a false light."[151] But if the deepfake at issue is obviously inauthentic or bears a label warning of its fakeness, a false light claim may not apply simply because the portrayal cannot be taken seriously as an accurate depiction. And while this tort may seem to be the best claim—what is deepfake pornography if not the quintessential case for placing the victim in a false light?—its faults rest with the fact that deepfake technology is still in its infancy. Many deepfake videos are too low-quality to be considered truly authentic, which inhibits the efficacy of a false light claim without reducing the consequences the video could still wreak on the victims' lives.[152] While the technology will inevitably evolve to a point that renders this concern moot,[153] waiting for more realistic and seamless deepfakes to arrive before addressing the issue will leave increasing numbers of women without legal redress and is therefore not a sustainable solution.

Similarly, for a public disclosure tort claim to attach, the revealed private disclosure must be both offensive under a reasonableness standard[154] and true.[155] And while very few could argue that deepfake pornography is not highly offensive under a reasonableness standard, defendants could successfully allege that this tort is inapplicable because deepfakes are not genuine depictions.[156] Further, the privacy line is eroded by the fact that the vast majority of deepfakes are produced using photographs a victim has posted to social media or otherwise online and are publicly available.[157] The internet is not regarded as a space in which users should expect privacy.[158] Indeed, Facebook itself has argued that its members cannot reasonably expect privacy when using its network.[159]

---

151. RESTATEMENT (SECOND) OF TORTS § 652E (AM. LAW INST. 1977).

152. *See* Harris, *supra* note 65, at 117 ("In these personal deepfakes, the face may glitch by not following the head properly, be fixed into only one position, or not be properly rendered to look three-dimensional.").

153. *See supra* note 9 and accompanying text.

154. RESTATEMENT (SECOND) OF TORTS § 652D(a) (AM. LAW INST. 1977).

155. *See id.* § 652D (Special Note on Relation of § 652D to the First Amendment of the Constitution) ("This Section provides for tort liability involving a judgment for damages for publicity given to *true statements of fact*." (emphasis added)).

156. *See supra* note 69 and accompanying text.

157. *See* Chesney & Citron, *supra* note 104, at 1794 ("[U]sing a person's face in a deep-fake video does not amount to the disclosure of *private* information if the source image was publicly available.").

158. *See supra* note 144 and accompanying text.

159. *See* Hannah Albarazi, *Facebook Says Social Media Users Can't Expect Privacy*, LAW360 (May 29, 2019), https://www.law360.com/articles/1164091/facebook-says-social-media-users-can-t-expect-privacy [https://perma.cc/D865-ARPR] (" 'There is no invasion of privacy at all, because there is no privacy,' [Facebook counsel Orin Snyder] argued.").

Tort law presents other claims beyond the privacy realm that hold some facial promise but ultimately pose too many hurdles for victims to successfully pursue. For example, intentional infliction of emotional distress[160] could prove fruitful because deepfake pornography violates society's conception of sexual privacy and decency.[161] Intentional infliction of emotional distress, however, also possesses an intent requirement: a defendant must "know[] that [the] conduct is substantially certain to cause harm."[162] This intent element constrains the number of intentional infliction of emotional distress claims that ultimately succeed.[163] Given that many deepfakes are created solely for a producer's or client's own gratification[164] and that most producers do not think that the victim will ever discover the video, proving a producer intended emotional harm or could have reasonably expected his actions to cause emotional harm will be difficult.[165] Further, as with other claims, the First Amendment would again operate as an effective defense.[166] Relying on intentional infliction of emotional distress claims to provide remedies to victims will bar many victims from successfully seeking legal redress.[167]

---

160. Intentional infliction of emotional distress requires the precipitating behavior to constitute "extreme and outrageous conduct." RESTATEMENT (THIRD) OF TORTS § 46 (AM. LAW INST. 2012).

161. Chesney & Citron, *supra* note 104, at 1794; *see also* RESTATEMENT (THIRD) OF TORTS § 46 cmt. d (AM. LAW INST. 2012) ("Under the 'extreme and outrageous' requirement, an actor is liable only if the conduct goes beyond the bounds of human decency such that it would be regarded as intolerable in a civilized community.").

162. RESTATEMENT (THIRD) OF TORTS § 46 cmt. a (AM. LAW INST. 2012).

163. By the Restatement's own admission, conduct that reaches the level of "extreme and outrageous . . . describes a very small slice of human behavior." *Id.*

164. *See supra* notes 43–45 and accompanying text.

165. *See* Harris, *supra* note 65, at 112:

> The majority of the Producers who share a video online with friends or the general public will likely not know that any emotional distress is imminent because they do not expect that the Victim will watch the video or that the Victim will even learn of its existence. . . . IIED claims, thus, appear to be limited to instances where the Producer intentionally sends the deepfake to the Victim or informs her of its circulation on the internet. The threat of IIED claims will not effectively diminish publications of deepfakes.

166. *See* RESTATEMENT (THIRD) OF TORTS § 46 cmt. f (Am. Law Inst. 2012):

> Communicative conduct that is constitutionally protected may cause emotional harm. If an actor's conduct is sufficient for liability under this Section but is protected by the First Amendment, liability cannot be imposed. The Supreme Court has long held that the First Amendment imposes limits on the extent to which state tort law, regardless of the specific tort claim, may impose liability for communicative conduct;

supra Section II.A.1.

167. *See* RESTATEMENT (THIRD) OF TORTS § 46 cmt. a (AM. LAW INST. 2012) ("Courts have played an especially critical role in cabining this tort . . . These limits are essential in preventing this tort from being so broad as to intrude on important countervailing policies, while permitting its judicious use for the occasions when it is appropriate.").

Defamation, a tort imposing liability for public statements that result in reputational harm, also fails as an option due to the intent requirement.[168] Defamation is an umbrella term for two torts—slander and libel—which the Restatement (Second) of Torts acknowledges are "impossible to define and difficult to describe with precision."[169] Deepfakes may fit under libel, as that subtype includes defamatory broadcast via radio or television,[170] and the internet can be analogized to those media. But libel is typically governed by state statutes,[171] which, in some jurisdictions, require a victim to show that the producers intended emotional distress.[172] Again, intent to cause emotional distress is difficult to prove, considering many producers have no idea the victim will ever discover the video.[173] Many producers neither intend emotional distress nor reasonably know that their deepfake winds up in the victim's hands.[174]

2. The Limitations of Copyright Law

If a deepfake uses copyrighted content (for example, if the underlying pornography is copyrighted), then the owner of the copyright could argue that the deepfake producer infringed on the owner's copyright based on the alteration.[175] But this solution provides relief for only the copyright owner, not for the victim.[176] A victim might be able to assert an infringement claim if a deepfake used photographs she herself posted, but the likelihood of victory is unclear.[177] A producer of a deepfake will likely assert "fair use" as a defense, which allows "the unlicensed use of copyright-protected works" in certain contexts.[178] Courts determine fair use on an individual basis by balancing various factors.[179] When assessing the purpose and character of the use, courts

168. RESTATEMENT (SECOND) OF TORTS § 559 (AM. LAW INST. 1977); *see also supra* note 32 and accompanying text.

169. RESTATEMENT (SECOND) OF TORTS § 568 cmt. b (AM. LAW INST. 1977).

170. *Id.* § 568A. Slander, in contrast, "consists of the publication of defamatory matter by spoken words, transitory gestures or by any form of communication other than those [covered by libel]." *Id.* § 568(2). Whether a communication constitutes slander or libel depends, in part, on the "area of dissemination." *Id.* § 568(3).

171. *Id.* § 568A cmt. b.

172. Ellis, *supra* note 69.

173. *Id.*

174. *See supra* notes 43–45 and accompanying text.

175. Farokhmanesh, *supra* note 21.

176. *Id.*

177. Chesney & Citron, *supra* note 104, at 1793.

178. *More Information on Fair Use*, COPYRIGHT.GOV, https://www.copyright.gov/fair-use/more-info.html (last updated Apr. 2020) [https://perma.cc/D43H-JHV6].

179. *See id*. The factors courts look at, as mandated by section 107 of the Copyright Act, are: "[p]urpose and character of the use," "[n]ature of the copyrighted work," "[a]mount and

consider whether the the new piece constitutes a "transformative" use, which is use that injects new elements without "substitut[ing] for the original use of the work."[180] Transformative uses are more likely than not to be considered fair use of copyrighted material.[181] Even if a victim can assert a copyright over the photos used in creating a deepfake, modifying an original pornographic video to create something wholly new with someone else's face is certainly "transformative." Courts may then yield to deepfake pornography as fair use.[182]

3. State and Federal Legislation

a. The Tension with State Nonconsensual Pornography Statutes

Forty-six states, the District of Columbia, and Guam now have nonconsensual pornography statutes (commonly known as "revenge porn" statutes),[183] which criminalize the dissemination of nude images that the depicted individual did not consent to.[184] Tennessee, for example, criminalizes revenge pornography as a Class A Misdemeanor under its unlawful exposure statute.[185] But nonconsensual pornography and nonconsensual deepfakes hold legally significant differences, which make it difficult to regulate deepfake pornography under these existing state statutes.

Nonconsensual pornography statutes regulate revenge porn as privacy violations,[186] while deepfake pornography exists in a strange purgatory. Deepfakes are not fully "real" in that they depict an act that never actually happened. They are not—legally speaking—a privacy violation because they are generally produced using photographs the

---

substantiality of the portion used in relation to the copyrighted work as a whole," and "[e]ffect of the use upon the potential market for or value of the copyrighted work." *Id.*

180. *Id.*

181. *See id.* ("Additionally, 'transformative' uses are more likely to be considered fair.").

182. *See* Harris, *supra* note 65, at 109 ("[P]ublishing personal deepfakes makes fair use of another's copyrighted images because it is transformative.").

183. *See 46 States + DC + One Territory Now Have Revenge Porn Laws*, CYBER C.R. INITIATIVE, https://www.cybercivilrights.org/revenge-porn-laws/ (last visited July 16, 2020) [https://perma.cc/5A27-X46S]. The states without revenge porn laws are Massachusetts, South Carolina, Wyoming, and Mississippi.

184. *Id.*

185. TENN. CODE ANN. § 39-17-318 (2019).

186. *See* Rebecca A. Delfino, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, 88 FORDHAM L. REV. 887, 897 (2019):

> Revenge porn . . . depicts private individuals engaged in intimate acts that were intended to remain private and were not recorded for mass dissemination or entertainment. Indeed, scholars' arguments in support of imposing civil and criminal liability for acts of revenge porn have centered on the violation of the victim's right to sexual privacy.

victim herself has posted online.[187] As a privacy violation, revenge porn does not technically encompass deepfake videos because deepfakes exist in a halfway point between real and fake—"[y]ou can't sue someone for exposing the intimate details of your life when it's not your life they're exposing."[188]

Using revenge porn statutes to provide legal redress for victims is also cabined by the fact that these statutes vary by state.[189] Moreover, the number of victims of revenge pornography is limited to the people who have either taken or had others take nude images of themselves, consensually or nonconsensually. Conversely, the number of potential victims of deepfake pornography is effectively unlimited. This number includes "anyone whose image has been captured digitally" and posted on the internet.[190] This applies to virtually every woman in the country—if not the world—and, therefore, poses an exponentially larger risk.

But the line between revenge pornography and deepfake pornography will continue to blur as deepfake technology improves. There is no functional difference between the effect of a genuine nonconsensual pornographic video and one that is fake but looks real. While a deepfake may not be covered by the vast majority of states' revenge porn laws, the effects of its dissemination will not differ from that of revenge porn. Virginia has amended its revenge pornography statute to include "a person whose image was used in creating, adapting, or modifying a videographic or still image with the intent to depict an actual person and who is recognizable as an actual person by the person's face, likeness, or other distinguishing characteristic," thus

187. *See* DEEPTRACE, *supra* note 11, at 5; Chesney & Citron, *supra* note 104, at 1794 ("[U]sing a person's face in a deep-fake video does not amount to the disclosure of *private* information if the source image was publicly available."); Kristen Dold, *Face-Swapping Porn: How a Creepy Internet Trend Could Threaten Democracy*, ROLLING STONE (Apr. 17, 2018, 8:47 PM), https://www.rollingstone.com/culture/culture-features/face-swapping-porn-how-a-creepy-internet-trend-could-threaten-democracy-629275/ [https://perma.cc/6E4E-DSF9]:

> "The basis for nonconsensual porn laws is that it's private, true information being disclosed without your consent, and you can regulate that. But if it's created – false information – it's no longer considered a privacy violation," says [Mary Anne] Franks, [a technology law professor at the University of Miami and an advisor for the Cyber Civil Rights Initiative]. In other words, despite the fact that your face stitched onto a body of a random porn star doing something explicit is horrific, it's not exactly "true." And that's hard to fight.

188. Ellis, *supra* note 69.

189. *See, e.g.*, Delfino, *supra* note 186, at 909–18 (comparing nonconsensual pornography statutes in California, Texas, Florida, and New York); *see also 46 States + DC + One Territory Now Have Revenge Porn Laws*, *supra* note 183 (detailing state revenge porn laws).

190. Delfino, *supra* note 186, at 898.

sweeping deepfake videos under the statute's coverage.[191] Similarly, in October 2019, California Governor Gavin Newsom approved an addition to the state's civil code that provides a claim for victims to sue the producers of deepfake pornography.[192] In order for state revenge pornography laws to apply to victims of deepfake pornography, the state must affirmatively amend the law, like in Virginia and California.

b. Expired and Pending Federal Legislation

Although Congressmen have introduced bills that criminalize the creation and dissemination of deepfake pornography, none have passed. Other legislation has focused more narrowly on the issue of deepfakes in the election context, while ignoring the ramifications of deepfake pornography.[193] The National Defense Authorization Act for Fiscal Year 2020, for example, which polices the foreign weaponization of deepfakes to spread misinformation and interfere in American elections, became law in December 2019, but it does not address deepfake pornography.[194]

Senator Ben Sasse, a Nebraska Republican, introduced the Malicious Deep Fake Prohibition Act of 2018 in December of 2018.[195] The bill imposed civil and/or criminal penalties on producers who "create, with the intent to distribute, a deep fake with the intent that the distribution of the deep fake would facilitate criminal or tortious conduct . . . [or to] distribute an audiovisual record" with the same

---

191. VA. CODE. ANN. § 18.2–386.2(A) (2019). Virginia is so far the only state to have amended its revenge pornography statutes to include deepfake pornography, although other states have implemented laws that regulate deepfakes, as well. David Ruiz, *Deepfakes Laws and Proposals Flood US*, MALWAREBYTES BLOG (Jan. 23, 2020), https://blog.malwarebytes.com/artificial-intelligence/2020/01/deepfakes-laws-and-proposals-flood-us/ [https://perma.cc/L7J5-HTCR]. However, other states' laws focus in the context of elections. *Id.*

192. CAL. CIV. CODE § 1708.86 (West 2019); *see also* Carrie Mihalcik, *California Laws Seek to Crack Down on Deepfakes in Politics and Porn*, CNET (Oct. 7, 2019, 8:32 AM), https://www.cnet.com/news/california-laws-seek-to-crack-down-on-deepfakes-in-politics-and-porn/ [https://perma.cc/63C6-CW2E] ("[Governor Newsom] also signed AB 602, which gives Californians the right to sue someone who creates deepfakes that place them in pornographic material without consent.").

193. Other potential bills are still pending. *See, e.g.*, Identifying Outputs of Generative Adversarial Networks Act, H.R. 4355, 116th Cong. (2019) (referred to the S. Comm. on Commerce, Sci. & Transp. on Dec. 10, 2019); Deepfake Report Act of 2019, S. 2065, 116th Cong. (2019) (requiring the Department of Homeland Security to produce an annual report on deepfake technology) (referred to the Subcomm. on Consumer Prot. & Commerce on Oct. 29, 2019); A Bill to Require the Secretary of Defense to Conduct a Study on Cyberexploitation of Members of the Armed Forces and Their Families, and for Other Purposes, S. 1348, 116th Cong. (2019) (introduced May 7, 2019).

194. National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116-92, sec. 5709, § 3369(a), 133 Stat. 1198, 2168–70 (2019).

195. Malicious Deep Fake Prohibition Act of 2018, S. 3805, 115th Cong. (2018).

intent and knowing that it is a deepfake.[196] Nonetheless, after it was sent to the Senate Judiciary Committee, the bill expired at the end of 2018 without any cosponsors.[197] The bill's failure may have, in part, been due to the fact that Senator Sasse introduced it the day before the government shutdown.[198]

In the House of Representatives, Representative Yvette Clarke, a New York Democrat, introduced the DEEP FAKES Accountability Act in June 2019.[199] The DEEP FAKES Accountability Act proposes requiring any deepfake to include both a digital watermark and a verbal statement and criminalizes failing to include or removing those elements "with the intent to humiliate or otherwise harass the person falsely exhibited."[200] The bill also imposes civil penalties for failure to disclose and altering disclosures, and directs the Attorney General to assign "a coordinator in each United States Attorney's Office to receive reports from the public regarding potential violations of section 1041 . . . and coordinate prosecutions for any violation of such section."[201] The DEEP FAKES Accountability Act was referred to the Subcommittee on Crime, Terrorism, and Homeland Security in June 2019 and has twenty-eight cosponsors but has not received any further attention.[202] Regardless, this Act penalizes only producers of deepfakes with the requisite intent, not the platforms on which they are hosted.[203] As previously highlighted, penalizing solely the producer of a deepfake will not provide adequate redress for victims.[204]

---

196. *Id.* § 1041(b)(1)-(2).

197. Delfino, *supra* note 186, at 909.

198. Kaveh Waddell, *3. The Newest Front in the Deepfakes War*, AXIOS (Jan. 31, 2019), https://www.axios.com/the-newest-front-in-the-deepfakes-war-1548941120-975f4124-1c66-476b-9a32-973b95de0c5a.html [https://perma.cc/4SU6-JATR].

199. DEEP FAKES Accountability Act, H.R. 3230, 116th Cong. (introduced 2019).

200. *Id.* § 1041(b)-(c), (e), (f)(1)(A)(i), (f)(1)(B)(i).

201. *Id.* §§ 1041(f)(2)(A)-(f)(2)(B), 1042(b).

202. *H.R. 3230 – Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019*, CONGRESS.GOV, https://www.congress.gov/bill/116th-congress/house-bill/3230/actions?KWICView=false (last visited July 16, 2020) [https://perma.cc/GWV6-MYHH].

203. *See* H.R. 3230 § 1041(a).

204. *See* Chesney & Citron, *supra* note 104, at 1792:

> The attribution problem arises in the first instance because the metadata relevant for ascertaining a deep fake's provenance might be insufficient to identify the person who generated it. . . . A careful distributor of a deep fake may take pains to be anonymous, including but not limited to using technologies like Tor. When these technologies are employed, the IP addresses connected to posts may be impossible to find and trace back to the responsible parties.

(citation omitted).

4. The Evolution of Corporate Social Governance

One obvious solution is for websites themselves to take the reins and self-police for deepfake pornography. Indeed, they are in the best position to quickly take down the offending posts. Facebook wields enormous power and sets the standard for other social media websites.[205] If Facebook implemented stringent regulations on deepfake pornography, it is likely that other websites would follow suit.[206] But the same issue that arises with any self-policing system also presents itself here: misalignment of interests. These websites are businesses—Facebook is a publicly traded corporation—and earn the lion's share of their total revenue from advertisements.[207] While victims of deepfake pornography are primarily concerned with immediately removing a post and preventing future posts, these websites are fundamentally moneymaking enterprises—divergent goals that may not always converge into a mutually agreeable solution.[208] That potential outcome is too precarious to suffice as a panacea to this issue.

---

205. *See, e.g.*, Rana Foroohar, *Facebook's Self-Policing Needs an Update*, FIN. TIMES (Sept. 10, 2017), https://www.ft.com/content/f5d04d7e-9481-11e7-a9e6-11d2f0ebb7f0 [https://perma.cc/6R6N-UNVJ] (" 'Commercial entities, political campaigns, governments - and indeed, anyone aspiring to monitor, monetise, control and predict human behavior - are all eager to work with the mega-platforms to achieve their economic and political goals,' says Frank Pasquale, a University of Maryland law professor and noted critic of Big Tech.").

206. *See id.* (noting that technology companies like Facebook have massive influence). It is also worth noting that Facebook, along with Microsoft and academics from several universities, launched the Deepfake Detection Challenge in 2019, which aimed to develop technology to better detect deepfakes and "to spur the industry to create new ways of detecting and preventing media manipulated via AI from being used to mislead others." Mike Schroepfer, *Creating a Data Set and a Challenge for Deepfakes*, FACEBOOK: AI (Sept. 5, 2019), https://ai.facebook.com/blog/deepfake-detection-challenge/ [https://perma.cc/U749-47LL]. Following the close of the challenge, the Partnership on AI released a report in 2020 detailing the findings from the challenge. Claire Leibowicz, *A Report on the Deepfake Detection Challenge*, PARTNERSHIP ON AI (Mar. 12, 2020), https://www.partnershiponai.org/a-report-on-the-deepfake-detection-challenge/ [https://perma.cc/DBV6-TMBC]. Google also announced plans to develop technology to detect deepfakes using a dataset of deepfakes. Nick Dufour & Andrew Gully, *Contributing Data to Deepfake Detection Research*, GOOGLE: AI BLOG (Sept. 24, 2019), https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html [https://perma.cc/923F-TLNK]. The company publicly released its dataset, which it constructed by working with actors to create "hundreds of videos," to enlist researchers to aid in "developing synthetic video detection methods." *Id.* No findings have been released as of this writing.

207. *Facebook Reports First Quarter 2020 Results*, FACEBOOK: INVESTOR RELATIONS (Apr. 29, 2020), https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-First-Quarter-2020-Results/default.aspx [https://perma.cc/S9NN-GJE5] (listing its total revenue for the first quarter of 2020 as $17.737 billion with advertising comprising $17.44 billion—or 98% of its total revenue).

208. *See id.*

Many major video-hosting and social media websites like Facebook,[209] Pornhub,[210] Twitter,[211] Gfycat,[212] and Tumblr[213] have banned deepfakes from their sites. The efficacy of these websites' policies is limited, however, since the policies rely on user reports, and most websites have not yet invested in solutions to detect deepfakes as soon as they are posted.[214] For those websites that have some type of screening technology, any deepfakes posted to the site before the ban are unaffected and are still policed only by user reports.[215]

Facebook's policy, in particular, was panned by media critics for its failure to include deepfakes that are edited by cutting out context or reordering words.[216] This exclusion leaves many deepfakes out of the policy's jurisdiction.[217] One critic, Subbarao Kambhampati, a computer science professor at Arizona State University, worried that Facebook's struggle to properly identify deepfakes within its policy was a "moving

209. *See* Monika Bickert, *Enforcing Against Manipulated Media Population*, FACEBOOK: NEWSROOM (Jan. 6, 2020), https://about.fb.com/news/2020/01/enforcing-against-manipulated-media/ [https://perma.cc/4LTZ-N6WN] (prohibiting videos that have been "edited or synthesized – beyond adjustments for clarity or quality – in ways that aren't apparent to an average person and would likely mislead someone into thinking that a subject of the video said words that they did not actually say" and that are "the product of artificial intelligence or machine learning that merges, replaces or superimposes content onto a video, making it appear to be authentic").

210. Cole, *Pornhub*, *supra* note 48.

211. Cole, *Twitter*, *supra* note 48.

212. Cole, *Gyfcat*, *supra* note 48.

213. *See Our Community Guidelines Are Changing to Keep Tumblr the Constructive, Empowering Place It Should Be*, TUMBLR: STAFF, https://staff.tumblr.com/post/177449083750/new-community-guidelines (last visited July 16, 2020) [https://perma.cc/D8RC-5RTN] ("Lastly, we're eliminating any ambiguity in our zero-tolerance policy on non-consensual sexual images. . . . [I]f we determine a post or blog is . . . engaging in the unwanted sexualization of another person, it will be taken down.").

214. For example, BuzzFeed reported that there were still deepfakes on Pornhub post-ban. Charlie Warzel, *Pornhub Banned Deepfake Celebrity Sex Videos, but the Site Is Still Full of Them*, BUZZFEED NEWS (Apr. 18, 2018, 6:13 PM), https://www.buzzfeednews.com/article/charliewarzel/pornhub-banned-deepfake-celebrity-sex-videos-but-the-site [https://perma.cc/NWA6-5QCJ].

215. For example, Gfycat developed an AI-assisted solution to combat deepfakes, but it does not address pre-existing deepfakes on the site. Cole, *Gyfcat*, *supra* note 48.

216. *See, e.g*, 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?smid=nytcore-ios-share [https://perma.cc/J6LX-XAEJ] (discussing how Facebook's new policy would still allow some altered videos, such as one "edited to make it appear that Speaker Nancy Pelosi was slurring her words" and another that "was heavily edited to wrongly suggest that [former Vice President Joe Biden] made racist remarks").

217. Deepfake pornography could still be covered by Facebook's prohibition of nudity. *See Community Standards: 14. Adult Nudity and Sexual Activity*, FACEBOOK, https://www.facebook.com/communitystandards/adult_nudity_sexual_activity (last visited July 16, 2020) [https://perma.cc/Q8YH-DKTJ] (indicating Facebook's default policy of removing sexual imagery). Facebook's policy does, however, allow for "digitally created content" featuring nudity if "it is posted for . . . satirical purposes." *Id*. This exception could allow for some deepfakes to remain on the site even after they have been reported. *See supra* Section II.A.1 (describing the First Amendment implications of deepfake pornography restrictions).

target" because its system "would have limited reach" and would incentivize producers to work around Facebook's filters.[218] Renee DiResta, a technical research manager with the Stanford Internet Observatory, also noted another consideration Facebook's new policy fails to address: that deepfakes will generally "have already gone viral prior to any takedown or fact check [sic],"[219] nullifying the new policy's post-hoc power.

* * *

This Note's exhaustive review of current remedies and dismissal of each testifies to the desperate need for new legislation that addresses the weaknesses in present solutions and accounts for the strictures of the First Amendment and section 230. If deepfake pornography is not deemed per se obscene, then the First Amendment protects its creation and creators. Moreover, section 230 precludes victims from suing the websites on which deepfake pornography appears. This combination effectively prevents victims from suing anyone at all. And even if a victim can locate the producer of a deepfake, none of the available claims—whether in tort or copyright law or in state revenge porn statutes—sufficiently address the unique challenges posed by deepfake pornography. This Note's proposed solution narrowly attacks the weaknesses in our current legal regime without encroaching on First Amendment rights or inhibiting deepfake technology's ability to flourish in general.

III. OPENING UP PLATFORM LIABILITY AND PENALIZING BAD ACTORS

In order to permit victims to pursue claims against websites that host nonconsensual deepfake pornography, this Note first suggests modifying section 230 of the CDA to return it to its intended purpose of punishing bad actors while immunizing websites that genuinely attempt to protect its users. In Section B, it advocates for legislation that equips victims with a cause of action against both producers and disseminators of deepfake pornography. Finally, in Section C, it proposes to delegate the authority to jointly administer this statute to the Federal Trade Commission ("FTC") and the Federal Communications Commission ("FCC") and establishes a two-track method of obtaining relief.

218. McCabe & Alba, *supra* note 216.

219. *Id.*

A. Amending Section 230

A solution that allows victims to go after producers but neglects to include platforms is woefully incomplete. Producers can vanish from the internet, but platforms should not be categorically immunized from liability when they fail to use a screening mechanism or when they employ an inadequate, half-hearted solution.[220] Therefore, the first step should be to amend section 230 of the CDA to allow victims to sue platforms that refuse to take down videos or that engage in activities that would otherwise be illegal. The brick-and-mortar equivalent of a website that hosts deepfake pornography—say, an adult video store—would not be allowed to sell deepfakes of noncelebrity women in its store.[221] So why should online publishers enjoy more immunity?[222]

Amending section 230 is long overdue.[223] The nature of the internet has dramatically shifted since its adoption in 1996. While the bulletin boards that Prodigy hosted nearly thirty years ago were text-based[224]— more clearly implicating the First Amendment—the internet has exploded into multimedia-based content far beyond what could have been contemplated by section 230's authors.[225] Concurrently, the scale of social media sites has facilitated "the rapid spread of destructive abuse,"[226] and section 230 removes the incentive for websites to properly moderate this content and protect victimized groups.[227]

---

220. *See supra* note 204 and accompanying text.

221. This commodification would be prohibited by wrongful appropriation and right-of-publicity claims. *See supra* notes 147–148 and accompanying text.

222. *See* Citron & Wittes, *supra* note 107, at 421 ("Yes, online platforms facilitate expression, along with other key life opportunities, but no more and no less so than do workplaces, schools, and coffee shops, which are all also zones of conversations and are not categorically exempted from legal responsibility for operating safely.").

223. Modifying section 230 is not beyond the realm of possibility. Hearings have been held in the House of Representatives regarding possible ways to amend the provision. *See infra* note 227 and accompanying text.

224. Stratton Oakmont, Inc. v. Prodigy Services Co., No. 31063/94, 1995 WL 323710, at *1 (N.Y. Sup. Ct. May 24, 1995).

225. *See* Citron & Wittes, *supra* note 107, at 411 ("Twenty years ago, commercial service providers had twelve million subscribers. Now billions of individuals are online in ways that would have been unimaginable when Congress passed the CDA." (footnote omitted)).

226. *Id.* at 411–12 ("If someone posts something defamatory, privacy invasive, or threatening about another person, or even about a nonuser of a given service, and thousands or tens of thousands of people share it, there can be devastating consequences whether or not the targeted individual used the service in question.").

227. *See Fostering a Healthier Internet to Protect Consumers*, *supra* note 106, at 7 ("More often, targeted individuals [of online harassment] are women, women of color, lesbian and trans women, and other sexual minorities."); Citron & Wittes, *supra* note 107, at 413 ("[Websites] have no duty of care to respond to users or larger societal goals. They have no accountability for destructive uses of their services, even when they encourage those uses.").

Of course, any amendment to section 230 must be narrowly tailored to prevent overinclusivity.[228] Modifying section 230 to return it to its roots and to truly protect the "Good Samaritans" while retaining accountability for bad actors is a modest approach that will avoid overbroad liability.[229] Professor Danielle Keats Citron and Benjamin Wittes propose revising section 230(c)(1) to include the following language (additions in italics):

> No provider or user of an interactive computer service that *takes reasonable steps to prevent or address unlawful uses of its services* shall be treated as the publisher or speaker of any information provided by another information content provider *in any action arising out of the publication of content provided by that information content provider*.[230]

Modifying the CDA to mitigate section 230's near-absolute immunization allows websites like Facebook—which has taken some affirmative steps, but not enough, to limit deepfakes on its site—to retain liability.[231] It also clears the way for victims to sue websites specializing in deepfake pornography and websites that fail to take affirmative, protective steps.[232] Amending the CDA deters video-hosting websites like Pornhub from merely updating their terms of service and throwing up their hands at any further moderation—and provides a much more enhanced level of protection for victims.

B. Legislation Prohibiting Nonconsensual Deepfake Pornography

After amending section 230 to allow victims to sue platforms for unlawfully hosting deepfake pornography, Congress should enact legislation that imposes liability on both producers and platforms. This Note's proposed legislation would impose liability on producers for creating and disseminating deepfake pornography without the consent of the depicted individual.[233] But this statute would extend further than

---

228. *See, e.g.*, *supra* note 85 and accompanying text.

229. Professor Citron originally proposed this solution and has advocated for this amendment to the CDA. *Fostering a Healthier Internet to Protect Consumers*, *supra* note 106.

230. *See* Citron & Wittes, *supra* note 107, at 419.

231. *See* Bickert, *supra* note 209.

232. Deeptrace Labs found that eight of the top ten pornography websites host deepfakes, while there are at least nine websites exclusively dedicated to deepfake pornography. This amendment would sweep these websites under its purview. DEEPTRACE, *supra* note 11, at 6.

233. *See, e.g.*, CAL. CIV. CODE § 1708.86(b)(1)-(2) (West 2019):

> (b) A depicted individual has a cause of action against a person who does either of the following:
>
> (1) Creates and intentionally discloses sexually explicit material and the person knows or reasonably should have known the depicted individual in that material did not consent to its creation or disclosure.

any other pending federal or state statutes by also imposing liability on websites that fail to implement methods[234] to affirmatively prevent the disclosure of deepfake pornography, solicit or implicitly condone deepfake pornography, or intentionally disclose deepfake pornography. This prerogative would impel platforms like Facebook and Pornhub to develop techniques that detect deepfake pornography on their sites.[235]

Importantly, the statute would not require the producer or platform to intend or reasonably know that the creation of deepfake pornography could cause harm, emotional or otherwise. By foregoing the intent element, the statute would altogether evade the issues with using tort claims like intentional infliction of emotional distress and defamation, which hinge upon the tortfeasor *intending* to cause harm.[236] Given the inherently personal nature of the harm deepfake pornography poses to victims, the mere action of creating, disseminating, and/or hosting deepfake pornography is what should be punished.[237] The harm caused by the action is not what should be punished. Similarly, acknowledging the fact that many deepfakes are created without the desire to make money, the statute will not require the producer to have derived a financial benefit in order to be punished.[238]

In addition, the legislation must be narrowly tailored to address solely deepfake pornography in order to avoid impinging on First Amendment rights.[239] By explicitly carving out works protected by the First Amendment, this proposed legislation would curtail the scope of covered material and thus pass constitutional muster.[240] Given the

---

> (2) Intentionally discloses sexually explicit material that the person did not create and the person knows the depicted individual in that material did not consent to the creation of the sexually explicit material.

234. This law would define "methods" to include efforts that websites such as Facebook and Gfycat have incorporated to screen for deepfakes on their sites. *See supra* note 48 and accompanying text. However, this definition of "methods" would also include a duty for platforms to tinker with technology-driven approaches to account for new ways that producers could circumvent blocking software. *See supra* note 54 and accompanying text.

235. Technology-forcing statutes such as this one have already been accepted in other contexts. *See, e.g.*, Clean Air Act, 42 U.S.C. § 7521(a) (2012) (granting the EPA Administrator authority to promulgate emissions standards for new motor vehicles, thus forcing manufacturers to produce motor vehicles able to fit the prescribed standards).

236. *See supra* notes 160–167 and accompanying text.

237. This differentiation exposes the tension with revenge porn statutes, in which the violation of privacy in disclosure is what is punished. *See supra* Section II.B.3.a. Here, wrongfully using photos that are already posted publicly would nevertheless establish a violation of the proposed statute.

238. *See supra* notes 147–150 and accompanying text.

239. *See* Waddell, *supra* note 198 (noting that an overly broad law would run the risk of "scar[ing] platforms into immediately taking down everything that's reported as a deepfake – potentially deleting legitimate posts in the process"); *supra* Section II.A.1.

240. *See, e.g.*, CAL. CIV. CODE § 1708.86(c)(1)(B)(i)-(iii) (West 2019):

nature of the harm posed by deepfake pornography, this narrow scope would pass First Amendment scrutiny.[241]

Finally, disclaimers that explicitly state a deepfake's falsity can easily be removed, leading viewers to believe that the video depicts real events. The proposed statute, therefore, will not provide for the usage of disclaimers as a defense.[242]

C. Regulatory Administration

The proposed statute would be jointly administered by the FTC and FCC.[243] Although regulating activity like deepfake pornography does not explicitly fit within the jurisdiction of these two agencies, both possess adjacent authority that could permit joint administration.[244]

The FTC protects consumers by regulating "unfair and deceptive practices in the marketplace."[245] Given that the FTC's express purpose is to safeguard consumers from these threats—and, to be sure, deepfakes are nothing if not unfair and deceptive—its Division of Privacy and Identity Protection within the Bureau of Consumer Protection would serve as a touchpoint through which victims could

---

(c) A person is not liable under this section . . . [if] [t]he material is any of the following:

(i) A matter of legitimate public concern.

(ii) A work of political or newsworthy value or similar work.

(iii) Commentary, criticism, or disclosure that is otherwise protected by . . . the United States Constitution.

241. *See supra* Section I.A; Section II.A.1.

242. *See, e.g.*, CIV. § 1708.86(d):

It shall not be a defense to an action under this section that there is a disclaimer included in the sexually explicit material that communicates that the inclusion of the depicted individual in the sexually explicit material was unauthorized or that the depicted individual did not participate in the creation or development of the material.

243. There is precedent for the FTC and the FCC to partner in regulation concerning the Internet. For example, the two agencies signed a Memorandum of Understanding ("MOU") to delineate the roles each would play in administering *Restoring Internet Freedom*, an Order promulgated by the FCC. This MOU demonstrates that the FCC and the FTC are comfortable sharing authority in a regulatory space. *See FCC Releases Restoring Internet Freedom Order*, FED. COMMC'NS COMM'N (Jan. 4, 2018), https://www.fcc.gov/document/fcc-releases-restoring-internet-freedom-order [https://perma.cc/23SM-XJ5U]; *Restoring Internet Freedom: FCC-FTC Memorandum of Understanding*, FED. TRADE COMM'N (Dec. 14, 2017), https://www.ftc.gov/policy/cooperation-agreements/restoring-internet-freedom-fcc-ftc-memorandum-understanding [https://perma.cc/9FRE-SL7M].

244. Senator Kirsten Gillibrand has proposed the creation of a new federal agency charged with protecting data privacy. If this agency is created, this Note's proposed legislation would fit comfortably within its jurisdiction. *See, e.g.*, Zack Whittaker, *A New Senate Bill Would Create a US Data Protection Agency*, TECHCRUNCH (Feb. 13, 2020, 4:00 AM), https://techcrunch.com/2020/02/13/gilliband-law-data-agency/ [https://perma.cc/RBV7-T69B].

245. *See About the FTC*, FED. TRADE COMM'N, https://www.ftc.gov/about-ftc (last visited July 16, 2020) [https://perma.cc/B9ZG-KVCW].

report violations of this Note's proposed statute,[246] with an officer and support staff in each regional office.[247] Indeed, the FTC already has an online complaint system in place to report violations of fraud, identity theft, and other online scams.[248]

Similarly, while the FCC's Media Bureau is currently charged with regulating broadcast radio and television,[249] it would not be a stretch to include social media platforms within its jurisdiction. The roles radio and television played in disseminating information and providing entertainment in the pre-internet era can easily be analogized to the roles the internet and social media play today.[250] Given that neither agency is explicitly charged with regulating social media nor the internet as a whole, their combined jurisdiction could encompass deepfake pornography.

Adjudicating claims under the proposed statute would follow a two-track system in order to ensure speedy recovery, mirrored largely after the fast-track administrative procedures in place for private parties to challenge patents before the Patent Trial and Appeal Board ("PTAB") in the United States Patent Office. Within the PTAB's two-track regime, victims are able to simultaneously pursue intra-agency adjudication and litigation in Article III federal courts.[251]

The *inter partes* fast-track adjudication system within the PTAB involves three different options[252] and was adopted as a speedier and

246. *See Bureau of Consumer Protection*, FED. TRADE COMM'N, https://www.ftc.gov/about-ftc/bureaus-offices/bureau-consumer-protection (last visited July 16, 2020) [https://perma.cc/FM6Y-MHPJ]; *Division of Privacy and Identity Protection*, FED. TRADE COMM'N, https://www.ftc.gov/about-ftc/bureaus-offices/bureau-consumer-protection/our-divisions/division-privacy-and-identity (last visited July 7, 2020) [https://perma.cc/SB3R-RXFH].

247. The FTC has eight regional offices, located in Seattle, San Francisco, Los Angeles, Dallas, Chicago, Cleveland, Atlanta, and New York City. *See Regional Offices*, FED. TRADE COMM'N, https://www.ftc.gov/about-ftc/bureaus-offices/regional-offices (last visited June 28, 2020) [https://perma.cc/6NRE-GTT8]. This aspect is similar to the proposed DEEP FAKES Accountability Act. *See supra* note 199 and accompanying text.

248. *See Submit a Consumer Complaint to the FTC*, FED. TRADE COMM'N, https://www.ftc.gov/faq/consumer-protection/submit-consumer-complaint-ftc (last visited July 7, 2020) [https://perma.cc/ZX4A-BHCB].

249. *See Offices and Bureaus*, FED. COMMC'NS COMM'N, https://www.fcc.gov/offices-bureaus (last visited July 7, 2020) [https://perma.cc/B3WQ-JKTX].

250. *See What We Do*, FED. COMMC'NS COMM'N, https://www.fcc.gov/about-fcc/what-we-do (last visited July 7, 2020) [https://perma.cc/7BFV-T5XW] (describing the Commission's role as "[s]upporting the nation's economy by ensuring an appropriate competitive framework for the unfolding of the communications revolution" and "[r]evising media regulations so that new technologies flourish alongside diversity and localism").

251. *See* Christopher J. Walker & Melissa F. Wasserman, *The New World of Agency Adjudication*, 107 CALIF. L. REV. 141, 157–58, 170–71 (2019) (describing the PTAB proceedings implemented as part of the Leah-Smith America Invents Act).

252. *See* 37 C.F.R. § 42 (2020). These challenges include *inter partes* reviews, post-grant reviews, and business method disputes. *See* MICHAEL ASIMOW, ADMIN. CONFERENCE OF THE U.S., FEDERAL ADMINISTRATIVE ADJUDICATION OUTSIDE THE ADMINISTRATIVE PROCEDURE ACT 165–67

less expensive option to litigation in federal district court.[253] In order to qualify for this expedited review, the challenger must demonstrate that there is a "reasonable likelihood" of succeeding on at least one of the claims[254] or that it is "more likely than not that at least 1 of the claims challenged . . . is unpatentable."[255] Although expedited, these proceedings are still adversarial, with oral argument and discovery, thus complying with the strictures the Administrative Procedure Act places on formal agency adjudication.[256] The fast-track procedure this Note recommends would correlate with that of the *inter partes* system. In this Note's proposed scheme, victims would be able to submit an online complaint with the FTC (as described above), which would trigger the inner-agency adjudicatory process. A link to the offending video or recording would suffice as authentication to initiate this process.[257]

While the proceeding is pending within the agency, victims would also be able to pursue a temporary restraining order or preliminary injunction in federal district court to enjoin the online platform from keeping the offending post online. Victims should only have to pursue litigation in federal district courts in extreme situations, like if a website refuses to take down the video. Most social media websites have takedown-request features, and deepfake pornography should already violate their policies.[258] Regardless, the statute should explicitly confine this judicial takedown power to deepfake pornography and should not be construed as a broader power for courts to order takedowns of other content.

This proposed inner-agency regime would differ from that of the PTAB in significant respects to better fit the context of deepfake

(2019). Note that the third option, Covered Business Method Review, expired on Sept. 16, 2020. *Transitional Program for Covered Business Method Patents*, USPTO, https://www.uspto.gov/patents-application-process/appealing-patent-decisions/trials/transitional-program-covered-business (last visited July 7, 2020) [https://perma.cc/692P-MVPB].

253. *See* Walker & Wasserman, *supra* note 251, at 158. For a more extensive discussion of the PTAB regime, see ASIMOW, *supra* note 252, at 163–68.

254. 35 U.S.C. § 314(a) (2012) (establishing the criteria for initiating *inter partes* review).

255. 35 U.S.C. § 324(a) (2012) (establishing the criteria for initiating post-grant review).

256. *See* 5 U.S.C. §§ 554, 556, 557 (2012).

257. Of course, this situation could be complicated if the poster or the website takes the video down, thus rendering the link inactive. Victims would be encouraged to immediately screen-grab the video to avoid this possibility, or the websites themselves could provide authentication.

258. *See, e.g.*, *Nudity and Sexual Content Policies*, YOUTUBE: HELP, https://support.google.com/youtube/answer/2802002?hl=en (last visited July 7, 2020) [https://perma.cc/LJL5-25KJ] (banning nudity and pornography); *Terms of Service*, PORNHUB, https://www.pornhub.com/information#terms (last visited July 7, 2020) [https://perma.cc/Z4PA-ZT7H] (prohibiting "Content that contains falsehoods or misrepresentations that could damage . . . any third party" and "any Content depicting . . . non-consensual sexual activity [or] revenge porn"); *supra* note 217.

pornography. For example, the PTAB requires would-be plaintiffs to pay a filing fee[259] in order to qualify for expedited review. This proposed legislation would not require a filing fee to gain entrance to the intra-agency adjudicatory process; any woman, no matter her economic means, can fall prey to deepfake pornography,[260] so access to justice is a critical goal for this Note's proposed statute. Financial benefits underlie many patent disputes,[261] while the dignitary concerns for victims of deepfake pornography significantly outweigh patentholder proprietary dignitary interests.[262] Moreover, the PTAB regime is unique in that it adjudicates claims between two third parties; the PTAB itself is not a party to the case. Here, in order to expand the pool of victims who are able to bring claims, the agency would intervene on the victim's behalf to pursue legal recourse against the platform or the individual.

CONCLUSION

While we are trained to doubt images[263] because the proliferation of technology like Photoshop[264] and Facetune[265] make it easy to seamlessly manipulate photographs with just a few clicks or taps, video remains the last bastion of believability.[266] It is therefore imperative to regulate deepfake pornography when it is actively weaponized against victims without inhibiting its ability to innovate and legally flourish.

An amendment to the CDA that opens the ability to punish bad actors, while still immunizing providers that make genuine attempts to protect victims from harassment, would allow platforms to innovate and experiment with moderation tools. Further, combined with the modification to section 230, this Note's proposed legislation is narrow

259. Walker & Wasserman, *supra* note 251, at 171.

260. *See supra* note 190 and accompanying text.

261. *See* Walker & Wasserman, *supra* note 251, at n.87 (stating that "average costs through trial were $3.5 million" for ten to twenty-five-million-dollar patent controversies in 2015).

262. *See supra* Section I.A.

263. *See, e.g.*, Ajani Bazile, *18 Celeb Photoshop Fails from the 2010s that You Can't Unsee Once You've Seen Them*, BUZZFEED (Nov. 16, 2019), https://www.buzzfeed.com/ajanibazile/celeb-photoshop-fails-2010s [https://perma.cc/46MJ-9E2Q] (discussing various images celebrities posted to social media that were clearly edited).

264. Photoshop is a desktop application that allows users to edit images. *Explore the Photoshop Family of Apps*, ADOBE, https://www.adobe.com/products/photoshopfamily.html (last visited July 7, 2020) [https://perma.cc/PZD3-6BAR].

265. Facetune is a cellphone app that allows users to edit images. FACETUNE2, https://www.facetuneapp.com (last visited July 7, 2020) [https://perma.cc/X2D6-2UDJ].

266. *See, e.g.*, Scott v. Harris, 550 U.S. 372, 380 (2007) ("When opposing parties tell two different stories, one of which is blatantly contradicted by [a videotape on] the record, so that no reasonable jury could believe it, a court should not adopt that version of the facts for purposes of ruling on a motion for summary judgment.").

enough to directly attack deepfake pornography while still allowing other legal uses of deepfake technology to thrive. Providing victims with two avenues of legal redress—one of which is an expedited agency review—allows an offending video to be taken down quickly and bad actors to be punished accordingly.

While allowing the internet to innovate and flourish and protecting First Amendment rights are valid and laudable goals, there is little reason why those goals should trump women's right to sexual privacy and women's ability to participate meaningfully in contemporary society by using the internet. This Note's proposed statutory amendment and legislation safeguard the right to speech and still provide for experimentation with deepfake technology, while penalizing the producers of deepfake pornography and the websites that passively (or explicitly) facilitate this type of online harassment. The tools are here for us to protect the next woman from unwarranted exposure and devastating consequences.

*Anne Pechenik Gieseke**

* J.D. Candidate, 2021, Vanderbilt University Law School; B.J., 2014, University of Missouri. To Drew Gieseke, for his unwavering support, love, and top-notch editing skills. To Tracy and Barry Pechenik, for teaching me the values of hard work and tenacity. To the incredible editors and staff of the *Vanderbilt Law Review*, for providing thorough and diligent feedback. And, finally, to my friends at Vanderbilt, for making me laugh too loudly in the law library.