DEEPLY FAKE, DEEPLY DISTURBING, DEEPLY CONSTITUTIONAL: WHY THE FIRST AMENDMENT LIKELY PROTECTS THE CREATION OF PORNOGRAPHIC DEEPFAKES

Bradley Waldstreicher†

TABLE OF CONTENTS

† Associate Editor, *Cardozo Law Review*, Volume 42. J.D. Candidate (June 2021), Benjamin N. Cardozo School of Law; B.S., New York University, 2018. I would like to thank my wife, Annie, for her everlasting support and for being my first law school friend, and my parents, Sandi and Stuart, for always having complete faith in me. I would also like to thank Professor Alexander Reinert for his invaluable insight and thoughtful comments throughout the writing process. In memory of my grandfathers, Elliot Waldstreicher and Murray Ehrlich.

INTRODUCTION

In June of 2018, investigative journalist Rana Ayyub told the BBC, "The last few weeks I think I've witnessed hell because every morning I wake up and I see this stream of tweets with screenshots of a pornographic video with my image morphed on it."[1] The episode began when Ayyub, an Indian and Muslim investigative journalist, accepted an invitation from the BBC and Al Jazeera to discuss India's protection of child sex abusers.[2] At the time, an eight-year-old girl had just been raped, and the Bharatiya Janata Party (BJP), an Indian nationalist party, had marched in support of the accused rapist.[3] The day after the interview, Ayyub experienced harassment and abuse on social media.[4] But the next day, the abuse escalated when someone from the BJP texted Ayyub a link to a video that appeared to show Ayyub in a pornographic video.[5] Ayyub watched the first few seconds of the video and froze.[6] She

1 BBC, *Rana Ayyub: 'They Stuck My Face On to a Porn Clip'—BBC News*, YOUTUBE (June 19, 2018), https://www.youtube.com/watch?v=2Mjr6_mhAxg [https://perma.cc/4ALL-QEV6].

2 Rana Ayyub, *I Was the Victim of a Deepfake Porn Plot Intended to Silence Me*, HUFFPOST (Nov. 21, 2018, 8:11 AM), https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316 [https://perma.cc/352B-8CR4].

3 *See id.*

4 The harassment on Twitter included fake tweets that appeared to come from Ayyub's account, and one tweet read, "I hate India and Indians!" *See id.*

5 *See id.*

6 *Internet 'Deepfakes' Threaten Truth and Reality*, PRI (June 13, 2019, 5:15 PM), https://www.pri.org/stories/2019-06-13/internet-deepfakes-threaten-truth-and-reality [https://perma.cc/TY3N-FAYB].

quickly realized the video was fake, since she has curly hair, but she claimed that an average viewer would actually think it was real.[7] Before long, the video was shared all across the internet and social media, and people even reached out to Ayyub in an attempt to pay her for sex.[8] Ayyub suffered immediate effects from this episode, and she checked into a hospital because of horrible reactions from the stress.[9] Now, Ayyub recognizes that she was the victim of a deepfake sex video.[10]

Deepfakes, products of artificial intelligence's ability to alter video content,[11] have become so popular on the internet that the total number of deepfake videos nearly doubled in the first half of 2019.[12] This rise in popularity has not been ignored by state and federal legislators, however, as bills targeting deepfakes have been introduced in the House and the Senate,[13] as well as in several states.[14]

Part I of this Note begins by examining what exactly a deepfake is, how the technology was developed, and how it became popular on the internet. Part I then surveys the current legal landscape of deepfake legislation, which includes enacted state and federal legislation. Part I continues by introducing the problems that Section 230 of the Communications Decency Act[15] poses to the effectiveness of any deepfake legislation. Part I advances by analyzing how deepfake legislation would be reviewed under the First Amendment and explains how pornographic deepfakes may appropriately fit into various categorical exceptions to the First Amendment. Part I concludes by explaining the development of the categorical exceptions of obscenity and child pornography.

---

7 *See id.*

8 *See* Ayyub, *supra* note 2.

9 *See id.*

10 *See* Drew Harwell, *Fake-Porn Videos Are Being Weaponized to Harass and Humiliate Women: 'Everybody Is a Potential Target,'* WASH. POST (Dec. 30, 2018, 10:00 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/VS6J-RF6S].

11 *See* Margaret Rouse, *Definition: Deepfake (Deep Fake AI)*, WHATIS.COM, https://whatis.techtarget.com/definition/deepfake [https://perma.cc/TRZ9-KPUN].

12 *See* Amrita Khalid, *Deepfake Videos Are a Far, Far Bigger Problem for Women*, QUARTZ (Oct. 9, 2019), https://qz.com/1723476/deepfake-videos-feature-mostly-porn-according-to-new-study-from-deeptrace-labs [https://perma.cc/54ZY-3KK3].

13 *See infra* Section I.B.1.

14 *See infra* Section I.B.2.a.

15 Communications Decency Act of 1996, 47 U.S.C. § 230 (2018).

Part II begins by examining whether pornographic deepfakes could be deemed obscene. Part II concludes by analyzing whether pornographic deepfakes should be treated like child pornography, which is exempted from protection as a separate category under the First Amendment. Part III explores how the analysis of pornographic deepfakes under obscenity and child pornography fits with amending Section 230 of the Communications Decency Act, and what ultimately may happen with regard to any deepfake legislation.

I. BACKGROUND

A. *What is a Deepfake?*

Deepfakes are distorted yet highly convincing artificial intelligence-created video, audio, and text that can make it look like something that did not occur actually transpired.[16] The technology behind deepfakes is believed to have been created by Ian Goodfellow, who is currently a Director of Machine Learning in the Special Groups Project at Apple, Inc.[17] This technology learns peoples' facial expressions and movements by extracting information from millions of data points.[18] Then the algorithm seamlessly positions that person's expressions onto somebody else's body, making it look like a person said or did something that they did not actually do.[19]

---

16 Matthew F. Ferraro, *Deepfake Legislation: A Nationwide Survey—State and Federal Lawmakers Consider Legislation to Regulate Manipulated Media*, WILMERHALE (Sept. 25, 2019), https://www.wilmerhale.com/en/insights/client-alerts/20190925-deepfake-legislation-a-nationwide-survey [https://perma.cc/XY9Z-CSGU]; *see* Rouse, *supra* note 11.

17 *See* Martin Giles, *The GANfather: The Man Who's Given Machines the Gift of Imagination*, MIT TECH. REV. (Feb. 21, 2018), https://www.technologyreview.com/2018/02/21/145289/the-ganfather-the-man-whos-given-machines-the-gift-of-imagination [https://perma.cc/7GG5-NG5Q]; Ian Goodfellow, LINKEDIN, https://www.linkedin.com/in/ian-goodfellow-b7187213 [https://perma.cc/VB2Z-D6JD].

18 *See* Karen Hao, *Deepfakes Could Anonymize People in Videos While Keeping Their Personality*, MIT TECH. REV. (Sept. 17, 2019), https://www.technologyreview.com/f/614323/ai-deepfakes-anonymizes-faces-in-videos-photos [https://perma.cc/R6YV-SBHU].

19 *See id.*

The term deepfake is a combination of the words "deep-learning" and "fake,"[20] and the word itself first emerged on Reddit, a social media platform, when a user with the account name "deepfakes" began creating fake pornography videos of celebrities.[21] Almost immediately, others utilized the technology by writing software that allowed anybody to realistically plaster someone's face on another's body.[22] Even though Reddit put an end to this community of people who wrote software and posted pornographic deepfakes, the damage was already done: deepfake technology became highly publicized, and it found a home in fake pornographic videos.[23] Celebrities such as Emma Watson, Natalie Portman, Michelle Obama, Kate Middleton, and Gal Gadot instantly became popular targets of deepfake creators ("Deepfakers").[24] Pornographic deepfakes have become so popular that deepfakes of South Korean pop artists and American and British actresses have accumulated millions of views.[25]

The number of deepfake videos, and particularly pornographic deepfakes, drastically increased from 2018 to 2019, according to a report from a cybersecurity lab.[26] The report found a total of 14,698 deepfake videos on YouTube and other websites as of mid-year 2019, compared to roughly 7,964 videos as of December 2018.[27] Furthermore, the report found that pornographic deepfakes, compared to non-pornographic ones, constituted ninety-six percent of all deepfake videos.[28] Moreover, the total number of views from the top four deepfake pornographic

---

20 India McKinney, Hayley Tsukayama, & Jamie Williams, *Congress Should Not Rush to Regulate Deepfakes*, ELEC. FRONTIER FOUND. (June 24, 2019), https://www.eff.org/deeplinks/2019/06/congress-should-not-rush-regulate-deepfakes [http://perma.cc/RU87-MWKS].

21 *What Is a Deepfake?*, ECONOMIST (Aug. 7, 2019), https://www.economist.com/the-economist-explains/2019/08/07/what-is-a-deepfake [https://perma.cc/AF75-2Y7F].

22 *See id.*

23 *See id.*

24 Dave Lee, *Deepfakes Porn Has Serious Consequences*, BBC (Feb. 3, 2018), https://www.bbc.com/news/technology-42912529 [https://perma.cc/K4Z7-5EJM].

25 Tom Simonite, *Most Deepfakes Are Porn, and They're Multiplying Fast*, WIRED (Oct. 7, 2019, 10:00 AM), https://www.wired.com/story/most-deepfakes-porn-multiplying-fast [https://perma.cc/V2F5-QXQY].

26 *See* Khalid, *supra* note 12.

27 *See id.*

28 HENRY ADJER, FRANCESCO CAVALLI, LAURENCE CULLEN, & GIORGIO PATRINI, DEEPTRACE, THE STATE OF DEEPFAKES: LANDSCAPE, THREATS, AND IMPACT v (2019).

websites was 134,364,438.[29] While there are nine websites that exclusively post deepfake pornography, eight of the top ten pornography websites also make deepfake content available.[30]

Importantly, the report found that pornographic deepfakes exclusively target women, while females were only targeted in thirty-nine percent of non-pornographic deepfakes.[31] Even though celebrities are overwhelmingly targeted in pornographic deepfakes,[32] non-celebrity women are increasingly becoming potential targets.[33] Now, users on deepfake-dedicated forums can even exchange money for custom deepfakes, as long as they possess some images of the target.[34]

B. *Federal and State Deepfake Legislation*

Politicians and scholars have consistently expressed concern over both pornographic and non-pornographic deepfakes since the technology's inception.[35] The first federal deepfake bill was proposed by Nebraska Senator Ben Sasse in 2018.[36] While Senator Sasse's proposed legislation expired, other members of Congress have also introduced legislation.[37] The most sweeping of these proposals was authored by

---

29 *Id.* at 1.

30 *Id.* at 6 (noting that deepfake-dedicated websites host ninety-four percent of all pornographic deepfake videos).

31 *Id.* at 2.

32 *Id.* at 2 (noting that women from the entertainment industry constitute ninety-nine percent of pornographic deepfakes, while news and media constitute the other one percent).

33 *See* Khalid, *supra* note 12.

34 One user said, "Hello, I'd like a high-quality video of a woman friend of mine. Can pay however!" *Id.* One user even requested a deepfake video of his "high school sweetheart." *See id.*

35 *See* Rachel del Guidice, *'Deep Fake' Technology Is a Threat to National Security, Politics, and the Media, Marco Rubio Says*, DAILY SIGNAL (July 19, 2018), https://www.dailysignal.com/2018/07/19/deep-fake-technology-is-a-threat-to-national-security-politics-and-the-media-rubio-says [https://perma.cc/75JM-GVFC]; Sara Rimer, *Q&A: LAW's Danielle Citron Warns that Deepfake Videos Could Undermine the 2020 Election*, B.U. TODAY (Sept. 11, 2019), http://www.bu.edu/articles/2019/qa-laws-danielle-citron-warns-that-deepfake-videos-could-undermine-the-2020-election [https://perma.cc/X9BE-KDCK].

36 *See* Kaveh Waddell, *Lawmakers Plunge into "Deepfake" War*, AXIOS (Jan. 31, 2019), https://www.axios.com/deepfake-laws-fb5de200-1bfe-4aaf-9c93-19c0ba16d744.html [https://perma.cc/6GS9-2ZM4].

37 Devin Coldewey, *DEEPFAKES Accountability Act Would Impose Unenforceable Rules—But It's a Start*, TECHCRUNCH (June 13, 2019, 3:25 PM), https://techcrunch.com/2019/06/13/

Congresswoman Yvette Clarke.[38] However, only on December 20, 2019 was the first federal legislation related to deepfakes signed into law.[39] The legislation, which does not include any substantive law, such as civil or criminal penalties, is part of the National Defense Authorization Act for Fiscal Year 2020 (NDAA).[40]

Several states have also enacted legislation to combat deepfakes.[41] The deepfake statutes passed in Virginia, Texas, and California are discussed below as examples of how states have addressed the problems posed by deepfakes.

1. Federal Deepfake Legislation

a. Senator Sasse's Bill

Nebraska Senator Ben Sasse was the first U.S. politician to target deepfakes with federal legislation,[42] as he introduced the "Malicious Deep Fake Prohibition Act of 2018" on December 21, 2018.[43] Even though the bill expired at the conclusion of 2018, Senator Sasse anticipated that he would reintroduce it in the future but has yet to do so thus far.[44] Senator Sasse's bill aimed at two groups: individual deepfake creators and distributors.[45] Individuals would have been penalized if they created a deepfake with the intention of committing an illegal act, and distributors, such as Twitter, could be penalized only if they knew that they were distributing a deepfake.[46] Punishment under this proposal would have included a possible fine and up to ten years of imprisonment if the deepfake had the potential to disturb an election or

---

deepfakes-accountability-act-would-impose-unenforceable-rules-but-its-a-start [https://perma.cc/GJS4-YZHN]; Waddell, *supra* note 36.

38 See H.R. 3230, 116th Cong. (2019); *see also* Ferraro, *supra* note 16.

39 *See* Jason Chipman, Matthew Ferraro, & Stephen Preston, WilmerHale, *First Federal Legislation on Deepfakes Signed into Law*, JDSUPRA (Dec. 24, 2019), https://www.jdsupra.com/legalnews/first-federal-legislation-on-deepfakes-42346 [https://perma.cc/W26L-9RW2].

40 *See id.*

41 *See infra* Section I.B.2.a.

42 *See* Waddell, *supra* note 36.

43 S. 3805, 115th Cong. (2018).

44 *See* Waddell, *supra* note 36.

45 *See id.*

46 *See id.*

provoke violence.[47] Notably, however, scholar Danielle Citron criticized the breadth of Senator Sasse's proposal, as she claimed that distributors might be inclined to remove even more content than necessary in fear of potential liability.[48]

b. Congresswoman Clarke's Bill

Additionally, on June 12, 2019, New York Congresswoman Yvette Clarke introduced a bill targeting deepfakes called the "Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019" or "DEEP FAKES Accountability Act" (DFAA).[49] Clarke's bill refers to a deepfake as an "advanced technological false personation record."[50] The legislation requires that all deepfakes contain watermarks, which indicate that the video has been altered.[51] DFAA details criminal and civil penalties for anybody who violates this watermark requirement, which is possible either by creating a deepfake without a watermark, or removing the watermark disclosure from a deepfake.[52] DFAA specifically lists the types of deepfakes that are required to be watermarked.[53] These include deepfakes containing sexual content and deepfakes that could interfere with an election.[54] Criminal penalties for malicious deepfakes include fines, imprisonment of up to five years, or both, while civil penalties include a $150,000 fine per record, as well as appropriate injunctive relief.[55]

---

47 *See id.*

48 *See id.*

49 H.R. 3230, 116th Cong. (2019).

50 *Id.*

51 *See* Coldewey, *supra* note 37.

52 H.R. 3230.

53 *Id.*

54 The exact language for the pornographic deepfakes is: "sexual content of a visual nature and appears to feature such person engaging in such sexual acts or in a state of nudity." *Id.* Other categories include deepfakes that can "cause violence or physical harm, incite armed or diplomatic conflict, or interfere in an official proceeding, including an election," deepfakes relating to securities and wire fraud, and lastly, deepfakes by a foreign power with the intent of "interfering in a Federal, State, local, or territorial election." *Id.*

55 *Id.*

Critics have identified several issues with Congresswoman Clarke's bill. First, identifying the creator of the deepfake could be nearly impossible, since it is relatively easy to stay anonymous on the internet.[56] Simply put, the people who are creating deepfakes and who are willing to be identified as the author are likely not the people who are creating malicious deepfakes.[57] Meanwhile, demanding that non-harmful Deepfakers watermark their videos creates more labor for people who do not have dangerous intentions, which could result in a chilling effect on the production of deepfakes.[58] Moreover, it is not difficult to anonymously remove a watermark and distribute that un-watermarked deepfake.[59] While it is unknown whether Congress will pass DFAA, it is clear that it contains major flaws.

c. National Defense Authorization Act for Fiscal Year 2020 (NDAA)

The NDAA addresses deepfakes in two sections.[60] The first section, titled "Report on deepfake technology, foreign weaponization of deepfakes, and related notifications," requires, among other things, that the Director of National Intelligence (DNI) submit a report to the congressional intelligence committees about "the potential national security impacts of machine-manipulated media" and "the actual or potential use of machine-manipulated media by foreign governments to spread disinformation or engage in other malign activities."[61] Notably, the report must include an assessment of the capabilities of both China and Russia to produce and detect deepfakes.[62] Moreover, this section requires the DNI to notify the congressional intelligence

---

56 *See* Roger A. Grimes & Preston Gralla, *17 Steps to Being Completely Anonymous Online*, IDG CONNECT (Jan. 1, 2018, 3:22 AM), https://www.idgconnect.com/idgconnect/news/1007112/steps-completely-anonymous-online [https://perma.cc/UGJ2-6FV4].

57 "The law here is akin to asking bootleggers to mark their barrels with their contact information. No malicious actor will even attempt to mark their work as an 'official' fake." *See* Coldewey, *supra* note 37.

58 *See id.*

59 *See id.*

60 *See* Chipman et al., *supra* note 39.

61 National Defense Authorization Act for Fiscal Year 2020 § 5709, 50 U.S.C. § 3369a (2019).

62 *See id.*

committees each time they encounter any intelligence that suggests that a foreign entity has attempted, or will attempt, to "deploy machine-manipulated media or machine-generated text aimed at the elections or domestic political processes of the United States."[63] The second deepfake-related section of the NDAA details a program in which the DNI can award up to $5,000,000 in prizes for the research of technologies related to deepfakes.[64] While this legislation does not address any criminal or civil consequences for creating deepfakes, the federal government's interest in understanding the potential harms of deepfakes is significant.

2. State Deepfake Legislation

a. Virginia

Several states have passed legislation targeting deepfakes. Virginia amended its revenge porn laws so that nonconsensual deepfakes are now included.[65] Under this law, anyone who creates or shares a pornographic deepfake of someone, without permission, is subject to a misdemeanor that could possibly result in a twelve-month jail sentence and a $2,500 fine.[66] Notably, this law creates a carve out that overlaps with Section 230 of the Communications Decency Act by specifically

---

63 *See id.*

64 National Defense Authorization Act for Fiscal Year 2020 § 5724, 50 U.S.C. § 3024 (2019).

65 Michael Grothaus, *Virginia Updates Its Revenge Porn Laws to Include Deepfakes*, FAST CO. (July 2, 2019), https://www.fastcompany.com/90372079/virginia-updates-its-revenge-porn-laws-to-include-deepfakes [https://perma.cc/U38E-EMAX].

66 *See id.*:

> Any person who, with the intent to coerce, harass, or intimidate, maliciously disseminates or sells any videographic or still image created by any means whatsoever that depicts another person who is totally nude, or in a state of undress so as to expose the genitals, pubic area, buttocks, or female breast, where such person knows or has reason to know that he is not licensed or authorized to disseminate or sell such videographic or still image is guilty of a Class 1 misdemeanor. For purposes of this subsection, "another person" includes 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.

H.B. 2678, 2019 Gen. Assemb., 2019 Reg. Sess. (Va. 2019).

exempting internet service providers from liability for deepfakes that users post to their websites.[67]

b. Texas

Texas is another state that has criminalized deepfakes.[68] Its law, in comparison to Virginia's, does not target pornographic deepfakes, but rather only deepfakes created to influence elections.[69] The punishment for violating Texas's law includes a misdemeanor and a possible jail sentence of a year, plus a $4,000 fine.[70]

c. California

California's two-part attack on deepfakes targets both pornographic deepfakes and deepfakes that could influence elections.[71] California legislators introduced the two deepfake bills after a deepfake that appeared to show Nancy Pelosi slurring words went viral.[72] President Trump even tweeted a version of the Pelosi deepfake video on Twitter, with the caption, "PELOSI STAMMERS THROUGH NEWS CONFERENCE."[73] The first bill prevents the manipulation of audio or video involving a candidate within sixty days of an election, unless there

---

67 *See* Va. H.B. 2678 ("If a person uses services of an Internet service provider, an electronic mail service provider, or any other information service, system, or access software provider that provides or enables computer access by multiple users to a computer server in committing acts prohibited under this section, such provider shall not be held responsible for violating this section for content provided by another person."); *see also infra* Section I.C.

68 Kenneth Artz, *Texas Outlaws 'Deepfakes'—but the Legal System May Not Be Able to Stop Them*, LAW.COM (Oct. 11, 2019, 1:20 PM), https://www.law.com/texaslawyer/2019/10/11/texas-outlaws-deepfakes-but-the-legal-system-may-not-be-able-to-stop-them [https://perma.cc/NJ9X-L2XE].

69 S.B. 751, 86th Leg., Reg. Sess. (Tex. 2019) ("A person commits an offense if the person, with intent to injure a candidate or influence the result of an election: (1) creates a deep fake video; and (2) causes the deep fake video to be published or distributed within 30 days of an election.").

70 *See* Artz, *supra* note 68.

71 *See* Steve Dent, *California Cracks Down on Political and Pornographic Deepfakes*, ENGADGET (Oct. 7, 2019), https://www.engadget.com/2019/10/07/california-deepfake-pornography-politics [https://perma.cc/3PBD-X6HV].

72 Nick Cahill, *California Senate Approves Anti-Deepfake Bill Despite Free Speech Concerns*, COURTHOUSE NEWS SERV. (Sept. 13, 2019), https://www.courthousenews.com/california-senate-approves-anti-deepfake-bill-despite-free-speech-opposition [https://perma.cc/2YLT-JRB7].

73 *Doctored Nancy Pelosi Video Highlights Threat of "Deepfake" Tech*, CBS NEWS (May 25, 2019, 12:39 PM), https://www.cbsnews.com/news/doctored-nancy-pelosi-video-highlights-threat-of-deepfake-tech-2019-05-25[https://perma.cc/N9G6-8Y43].

is a disclosure stating that the material is fake.[74] The second bill, targeting pornographic deepfakes, permits California residents to sue Deepfakers.[75] Previously, a right of action only existed against someone who distributed a nude image.[76] Important exceptions exist in regard to when pornographic Deepfakers are protected, such as when the material is a political work or has newsworthy value.[77]

C. *Problems Posed by Section 230 of the Communications Decency Act*

As noted above, Deepfakers are likely to remain anonymous.[78] But assume a Deepfaker reveals their identity or is easily traceable, and a victim wants to bring a civil suit seeking thousands of dollars in damages. It is possible that a Deepfaker does not have enough money to compensate the victim.[79] A victim in this scenario may pursue relief by suing the publisher of the website where the deepfake can be viewed.[80] However, the publisher may assert a defense under Section 230 of the Communications Decency Act (CDA).[81]

Section 230 of the CDA was passed as a part of the Communications Decency Act of 1996.[82] While many sections of the CDA were struck down as unconstitutional, Section 230 survived, and has been credited with practically creating the internet.[83] The key words from Section 230 are, "No provider or user of an interactive computer

---

74 A.B. 730, 2019 Reg. Sess. (Cal. 2019); *see* Dent, *supra* note 71.

75 A.B. 602, 2019 Reg. Sess. (Cal. 2019); *see* Dent, *supra* note 71.

76 *See* Cal. A.B. 602.

77 *See id.* ("The bill would specify exceptions to those provisions, including if the material is a matter of legitimate public concern or a work of political or newsworthy value.").

78 *See* Grimes & Gralla, *supra* note 56 and accompanying text.

79 Russell Spivak, "*Deepfakes*": *The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 387 (2019).

80 *See id.*

81 Communications Decency Act of 1996, 47 U.S.C. § 230 (2018); *see also* Spivak, *supra* note 79, at 387.

82 47 U.S.C. § 230 (1996); *see also Section 230 Protections*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/bloggers/legal/liability/230 [https://perma.cc/YE52-BJZV].

83 *See* Adi Robertson, *Why the Internet's Most Important Law Exists and How People Are Still Getting It Wrong*, VERGE (June 21, 2019, 1:02 PM), https://www.theverge.com/2019/6/21/18700605/section-230-internet-law-twenty-six-words-that-created-the-internet-jeff-kosseff-interview [https://perma.cc/5YJQ-6TDZ]; *see also Section 230 Protections*, *supra* note 82.

service shall be treated as the publisher or speaker of any information provided by another information content provider."[84] Thus, Section 230 acts as an immunity clause for any online service that publishes third-party content.[85] However, Section 230 does not protect providers or users who violate any federal criminal statute, federal obscenity law, or federal law relating to the sexual exploitation of children.[86] Moreover, as the Virginia deepfake law recognized, Section 230 explicitly preempts any state or local law.[87] As a result of this immunity, Section 230 has allowed the internet to prosper, since without it, many publishers would censor materials posted to their websites in fear of potential liability.[88]

Zeran v. American Online, Inc. is one case that demonstrates the power of Section 230.[89] In that case, the plaintiff brought suit against American Online for not removing defamatory statements posted by a third-party about the plaintiff.[90] The court first stated that Congress's intent in passing Section 230—to block the threat of tort-based lawsuits that would interfere with freedom of speech on the internet—was evident from the conferral of immunity to internet publishers under the statute.[91] Then the court noted that websites like American Online have millions of users, and it would be impossible to filter out every piece of information that could potentially lead to a lawsuit.[92] The court said that if publishers like American Online could be liable for torts by third-parties, it would lead to "an obvious chilling effect" on speech, since a provider like American Online might choose to rigorously restrict the types of posts allowed.[93] Therefore, the court ruled in favor of American Online, citing protection under Section 230 of the CDA.[94]

---

84 *See* 47 U.S.C. § 230(c)(1) (2018).

85 *Section 230 of the Communications Decency Act*, ELEC. FRONTIER FOUND., https://www.eff.org/issues/cda230 [https://perma.cc/M4RY-6EZU] ("This legal and policy framework has allowed for YouTube and Vimeo users to upload their own videos, Amazon and Yelp to offer countless user reviews, craigslist to host classified ads, and Facebook and Twitter to offer social networking to hundreds of millions of Internet users.").

86 *See* 47 U.S.C. § 230(e)(1) (2018).

87 *See id.*

88 *See* Robertson, *supra* note 83.

89 Zeran v. America Online, Inc., 129 F.3d 327 (4th Cir. 1997).

90 *See id.* at 328.

91 *See id.* at 330–31.

92 *See id.*

93 *See id.*

94 *See id.* at 335.

To demonstrate how Section 230 might protect an online publisher from a deepfake posted by a third-party, imagine a hypothetical lawsuit where John Doe's friend posts an embarrassing deepfake of John Doe on a social media platform. John Doe thinks the deepfake is defamatory and requests that his friend remove it. After his friend declines to remove it, John Doe asks the social media platform to remove it, and it also declines. John Doe is so upset about the deepfake that he sues the social media platform for not removing it. The social media platform would likely invoke the protection of Section 230, and a court would be very likely to grant the social media platform summary judgment, meaning that John Doe would lose. Thus, it seems probable that Section 230 would protect online platforms that host deepfakes in the same way that it protects online platforms that host other types of content.

1. Amendments to Section 230

Section 230 of the CDA has already been amended in a significant way, as demonstrated by the "Allow States and Victims to Fight Online Sex Trafficking Act of 2017" (FOSTA) and the Senate bill, the "Stop Enabling Sex Traffickers Act" (SESTA). FOSTA and SESTA's aim is to remove the immunity usually granted to publishers who host materials that facilitate prostitution, thereby eliminating illegal sex trafficking on the internet.[95] One of the main targets of the acts was Backpage.com, a website infamous for its sex worker advertisements.[96] Backpage survived previous lawsuits brought by plaintiffs, mainly because of protection from Section 230.[97] In 2016, Kamala Harris, then serving as California's Attorney General, brought charges against Backpage's

---

95 *See* Anna Windemuth, *The First Challenge to FOSTA Was Dismissed—Along with the First Amendment's Unique Standing Doctrine*, YALE L. SCH.: MFIA (Dec. 27, 2018), https://law.yale.edu/mfia/case-disclosed/first-challenge-fosta-was-dismissed-along-first-amendments-unique-standing-doctrine [https://perma.cc/7NJ9-KJ2L]; *see also* Aja Romano, *A New Law Intended to Curb Sex Trafficking Threatens the Future of the Internet as We Know It*, VOX, https://www.vox.com/culture/2018/4/13/17172762/fosta-sesta-backpage-230-internet-freedom [https://perma.cc/A3X5-TABQ] (last updated July 2, 2018, 1:08 PM).

96 *See* Romano, *supra* note 95 (noting that Backpage has been involved in a number of controversies, including the arrest of individuals caught using the website to pay for sex).

97 *See id.*

founders and CEO, while calling it "the world's top online brothel."[98] The judge acknowledged that California has a strong interest in deterring human trafficking but dismissed the case, citing Section 230's protection and constraints under the First Amendment.[99] Eventually, a 2017 Senate investigation found that Backpage was involved in ads for child trafficking, which led to the passage of FOSTA and SESTA.[100] After these bills were passed, Craigslist shut down its "personals" sections,[101] fearing potential liability.[102] The Electronic Frontier Foundation commenced a suit challenging FOSTA on the grounds that it is unconstitutional under the First and Fifth Amendments, and that it would impede online freedoms.[103] While the benefits and effects of FOSTA and SESTA have been questioned,[104] the important takeaway is that some members of Congress are willing to amend the CDA so that publishers are not guaranteed complete immunity.

2. Other Remedies Available to Deepfake Victims

The question of whether there should be a law that bans deepfakes, or specifically pornographic deepfakes, is complicated because of First

---

98 Max Kutner, *After Backpage.Com Case Dismissed, Anti-Trafficking Advocates Look To Next Battles*, NEWSWEEK (Dec. 12, 2016, 4:46 PM), https://www.newsweek.com/backpage-sex-trafficking-case-ferrer-harris-531187 [https://perma.cc/LY4U-5XXK] (California prosecutors alleged that the site enables sex trafficking, even though "[t]he site's regulations require that people be at least 18 to post adult ads, and ads cannot involve 'any illegal service exchanging sexual favors for money' or exploit minors or involve human trafficking.").

99 *See id.*

100 *See* Romano, *supra* note 95.

101 *See* Merrit Kennedy, *Craigslist Shuts Down Personals Section After Congress Passes Bill on Trafficking*, NPR (March 23, 2017, 3:52 PM), https://www.npr.org/sections/thetwo-way/2018/03/23/596460672/craigslist-shuts-down-personals-section-after-congress-passes-bill-on-traffickin [https://perma.cc/9445-9ENK] (explaining that Craigslist's personals section, which includes ads seeking romance or sexual connections, is no longer going to be available).

102 *Id.* (Craigslist said, "Any tool or service can be misused . . . . We can't take such risk without jeopardizing all our other services, so we are regretfully taking craigslist personals offline. Hopefully we can bring them back some day.").

103 David Greene, *EFF Sues to Invalidate FOSTA, an Unconstitutional Internet Censorship Law*, ELEC. FRONTIER FOUND. (June 28, 2018), https://www.eff.org/deeplinks/2018/06/eff-sues-invalidate-fosta-unconstitutional-internet-censorship-law [https://perma.cc/R8UD-R3GG].

104 *See, e.g.*, Romano, *supra* note 95 (noting that some of the websites that FOSTA and SESTA seek to remove actually give "sex workers a way to advertise, vet, and choose clients online [that] makes them much safer than they are without an online system").

Amendment issues.[105] Some argue that there is no need to create a separate deepfake law, because other laws can provide a remedy to any deepfake victim.[106] For example, if someone creates a deepfake to extort or harass a victim, laws covering those areas will apply.[107] Another remedy a deepfake victim may have is suing someone under the tort of false light, which addresses activities such as photo manipulation.[108] Moreover, rights of publicity claims could arise if the Deepfaker benefits or profits from the sale of a deepfake.[109] Finally, copyright infringement may be asserted because deepfakes, in many instances, modify copyrighted videos.[110] Of course, even with all of these remedies, the issue of Deepfakers remaining anonymous can severely limit many victims' claims. Thus, real change for pornographic deepfake victims could potentially only result from an amendment to Section 230 of the CDA.[111] Such an amendment could allow these victims to sue the online publishers of pornographic deepfakes rather than the Deepfaker, which could drastically decrease the dissemination of pornographic deepfakes.[112] However, as mentioned previously, First Amendment issues will surface if Section 230 is amended to ban or limit deepfakes, or if any separate bill attempts to do so.

---

105 Mathew Ingram, *Legislation Aimed at Stopping Deepfakes Is a Bad Idea*, COLUM. JOURNALISM REV. (July 1, 2019), https://www.cjr.org/analysis/legislation-deepfakes.php [https://perma.cc/A7LC-P9W3].

106 *See* David Greene, *We Don't Need New Laws for Faked Videos, We Already Have Them*, ELEC. FRONTIER FOUND. (Feb. 13, 2018), https:/www.eff.org/deeplinks/2018/02/we-dont-need-new-laws-faked-videos-we-already-have-them [https://perma.cc/3BND-EYPR].

107 *See id.*

108 *See id.* ("To win a false light lawsuit, a plaintiff—the person harmed by the deepfake, for example—must typically prove that the defendant—the person who uploaded the deepfake, for example—published something that gives a false or misleading impression of the plaintiff in such a way to damage the plaintiff's reputation or cause them great offense, in such a way that would be highly offensive to a reasonable person, and caused the plaintiff mental anguish or suffering. It seems that in many situations the placement of someone in a deepfake without their consent would be the type of 'highly offensive' conduct that the false light tort covers.").

109 *See id.*

110 *See* Spivak, *supra* note 79, at 391.

111 This Note only discusses the constitutionality of pornographic deepfakes, and not non-pornographic deepfakes.

112 *See* Spivak, *supra* note 79, at 399–400.

D. *First Amendment Constraints*

The First Amendment states, "Congress shall make no law . . . abridging the freedom of speech."[113] Importantly, freedom of speech does not strictly concern words that are actually spoken.[114] Some other forms of expression that are considered speech include written works, online posts, and video games.[115] Furthermore, the Supreme Court has affirmed that the government may not regulate forms of speech based on content.[116] Thus, one of the first steps a court may take when deciding whether a law violates the First Amendment is determining whether it is content-based or content-neutral.[117] Content-based restrictions apply to speech depending on the message of the speech.[118] In contrast, content-neutral constraints restrict speech without regard to the content of the message.[119] This distinction is relevant because the standard of review applied by a court will differ depending on whether the law is content-based or content-neutral.[120] A court will likely subject content-based laws to review based on strict scrutiny, while content-neutral regulation will be reviewed based on intermediate scrutiny.[121] One reason why content-based laws are subject to a higher standard of review is because it is every person's right, and not the government's, to decide which ideas are worthy of expression.[122]

---

113 U.S. CONST. amend. I.

114 Lata Nott, *Is Your Speech Protected by The First Amendment?*, FREEDOM FORUM INST., https://www.freedomforuminstitute.org/first-amendment-center/primers/basics [https://perma.cc/84FT-SB49].

115 *See id.*

116 *See* David L. Hudson Jr., *Content Based*, FIRST AMEND. ENCYCLOPEDIA, https://www.mtsu.edu/first-amendment/article/935/content-based [https://perma.cc/XN83-BDKT].

117 *See id.*

118 *See* Erwin Chemerinsky, *Content Neutrality as a Central Problem of Freedom of Speech: Problems in the Supreme Court's Application*, 74 S. CAL. L. REV 49, 51 (2000); Geoffrey R. Stone, *Content-Neutral Restrictions*, 54 U. CHI. L. REV. 46, 47 (1987).

119 *See* Stone, *supra* note 118, at 48 (Examples include "[l]aws that restrict noisy speeches near a hospital, ban billboards in residential communities, limit campaign contributions, or prohibit the mutilation of draft cards are examples of content-neutral restrictions.").

120 *See* Hudson, *supra* note 116.

121 *See id.* (Strict scrutiny is "the highest form of judicial review.").

122 *See* Turner Broad. Sys. v. FCC, 512 U.S. 622, 641 (1994) ("At the heart of the First Amendment lies the principle that each person should decide for him or herself the ideas and

Since the standards of review are different, determining whether a law banning pornographic deepfakes would be content-based or content-neutral is critical. *United States v. Playboy Entertainment Group, Inc.*,[123] offers insight into this matter. In that case, the regulation at issue was Section 505 of the Telecommunications Act of 1996,[124] which required cable television providers who exclusively showed sexually-oriented programming to scramble or fully block their channels during hours in which children may be likely to watch.[125] Playboy Entertainment Group, Inc., which provided adult television programming, challenged this regulation as "unnecessarily restrictive content-based legislation."[126] The Court held that the statute was clearly a content-based restriction, since the statute applied only to channels like Playboy's, which primarily showed indecent adult programming.[127]

United States v. Playboy Entertainment Group, Inc. demonstrates that targeting a specific type of message or communication is likely a violation of the First Amendment.[128] Likewise, legislation targeting pornographic deepfakes, rather than all deepfakes, would be content-based, since it would restrict speech based on the content of what it portrays.[129] Furthermore, legislation that targets all deepfakes, including pornographic ones, may be considered content-based in comparison to other videos, or other pornographic videos.

However, there are certain categories of speech that are unprotected by the First Amendment.[130] In fact, categorical exceptions

---

beliefs deserving of expression, consideration, and adherence. Our political system and cultural life rest upon this ideal. Government action that stifles speech on account of its message, or that requires the utterance of a particular message favored by the Government, contravenes this essential right. Laws of this sort pose the inherent risk that the Government seeks not to advance a legitimate regulatory goal, but to suppress unpopular ideas or information or manipulate the public debate through coercion rather than persuasion." (citations omitted)).

123 United States v. Playboy Ent. Group, Inc., 529 U.S. 803 (2000).

124 47 U.S.C. § 561 (1996).

125 *See Playboy*, 529 U.S. at 806.

126 *See id.* at 807.

127 *See id.* at 811–12.

128 *See id.* at 812.

129 *See* Spivak, *supra* note 79 ("[A]ny law banning, or even regulating, deepfakes would be presumptively invalid, given that such a law would fall squarely into content-based or message-based regulation.").

130 *See* Nott, *supra* note 114 (noting that the exceptions include obscenity, fighting words, defamation (including libel and slander), child pornography, perjury, blackmail, incitement of imminent lawless action, true threats, and solicitations to commit crimes).

to First Amendment free speech have been recognized since as early as 1791.[131] Two of those categories, obscenity and child pornography, will help determine whether a law targeting pornographic deepfakes could be constitutional.

E. *Obscenity*

1. Development of the Obscenity Standard

The modern theory of obscenity was established in *Roth v. United States*.[132] At issue in *Roth* was whether a federal obscenity statute infringed on the First Amendment.[133] A jury convicted one of the defendants, who managed a business in which he mailed obscene materials, under this federal obscenity statute.[134] While the Court acknowledged that it had always assumed obscenity to be a categorical exception to the First Amendment, this was the first time the Court was presented with this issue directly.[135] The Court concluded that obscenity was not a category of speech that should be protected by the First Amendment, since it is "utterly without redeeming social importance."[136] In defining obscenity, the Court held that obscene material portrays sex in a way that appeals to prurient interests.[137] Thus, the Court stated that the appropriate test to be used in judging whether material is obscene is "whether to the average person, applying contemporary community standards, the dominant theme of the material taken as a whole appeals to prurient interest."[138] Under this

131 R.A.V. v. St. Paul, 505 U.S. 377, 382–83 (1992) ("From 1791 to the present, however, our society, like other free but civilized societies, has permitted restrictions upon the content of speech in a few limited areas, which are 'of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.'" (citing Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942))).

132 Roth v. United States, 354 U.S. 476 (1957).

133 *See id.* at 479.

134 *See id.* at 480.

135 *See id.* at 481.

136 *See id.* at 484.

137 *Id.* at 487 n.20 (The Court said that prurient refers to "material having a tendency to excite lustful thoughts.").

138 *See id.* at 489–90 (noting that contemporary community standards mean that a jury should "judge the circulars, pictures and publications which have been put in evidence by present-day

standard, the federal obscenity statute at issue here was not unconstitutional, and the defendants' convictions were upheld.[139]

In *Jacobellis v. Ohio*,[140] the Court affirmed that *Roth*'s obscenity test should be applied, even though there was agreement among the Justices that the test was not perfect.[141] Moreover, the Court acknowledged that a balancing test should not be administered when determining whether material is obscene, since only material which is "utterly" without redeeming social importance should be proscribed.[142] Furthermore, the Court agreed that instead of each community having the power to determine whether materials are obscene, a national standard should be used.[143] However, some Justices were still not satisfied with what precisely constituted obscenity.[144] In reference to this challenge, Justice Potter Stewart said, "I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it . . . ."[145] Other judges shared Justice Potter's uncertainty about what exactly constituted obscenity.[146] The Court's lack of clarity on this topic led to the pioneering 1973 decision, *Miller v. California*.[147]

Miller is significant because it was the first time since *Roth* that a majority of the Supreme Court Justices agreed on a formulation for

standards of the community. [A jury] may ask . . . does it offend the common conscience of the community by present-day standards.").

139 *See id.* at 492–94.

140 Jacobellis v. Ohio, 378 U.S. 184 (1964).

141 *Id.* at 191 ("[A]ny substitute would raise equally difficult problems . . . ").

142 *See id.*

143 *See id.* at 193–95 (noting that a standard based on a particular local community would have "the intolerable consequence of denying some sections of the country access to material, there deemed acceptable, which in others might be considered offensive to prevailing community standards of decency." (citing Manual Enters., Inc. v. Day, 370 U.S. 478, 488 (1962))).

144 *See Jacobellis*, 378 U.S. at 197 (Stewart, J., concurring).

145 *See id.*

146 *See* Mishkin v. New York, 383 U.S. 502, 516–17 (1966) (Black, J., dissenting) ("I wish once more to express my objections to saddling this Court with the irksome and inevitably unpopular and unwholesome task of finally deciding by a case-by-case, sight-by-sight personal judgment of the members of this Court what pornography (whatever that means) is too hard core for people to see or read.").

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

what constitutes obscenity.[148] Marvin Miller managed a mass mailing campaign in which he mailed adult material to others.[149] Miller was convicted for a misdemeanor under a California statute since he knowingly mailed obscene material.[150] In recognizing the need for a new formulation, the Court acknowledged that the standards previously adopted were not feasible.[151] Thus, the new formulation the Justices agreed to stated:

> (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.[152]

The first element of this test restates part of the test from *Roth*.[153] The main, and trivial, difference is that *Roth* required that the "dominant theme" of the material appeal to the prurient interest.[154]

The second element requires that three separate hurdles be cleared in order to find that certain materials are obscene.[155] The first is that the material must be "patently offensive."[156] Second, the work must depict or describe sexual conduct that could be considered "hard core."[157] Examples of materials that would satisfy the criteria of what the Court considers hard core sexual conduct under this element include

---

148 Chris Hunt, *Community Standards in Obscenity Adjudication*, 66 CALIF. L. REV. 1277, 1283 (1978).

149 *See Miller*, 413 U.S. at 16.

150 *See id.* at 16–17.

151 *See id.* at 23 n.4.

152 *See id.* at 24.

153 Roth v. United States, 354 U.S. 476, 489–90 (1957).

154 That difference can be attributed to *Miller* recognizing that *Roth*'s standard did not make it clear how to separate obscene materials from valuable materials. *See* Hunt, *supra* note 148, at 1284.

155 *See id.*

156 Patently offensive refers to materials that are "so offensive on their face as to affront current community standards of decency . . . ." *See* Manual Enters., Inc. v. Day, 370 U.S. 478, 482 (1962).

157 *See* Hunt, *supra* note 148, at 1284; *see also Miller*, 413 U.S. at 27 ("Under the holdings announced today, no one will be subject to prosecution for the sale or exposure of obscene materials unless these materials depict or describe patently offensive 'hard core' sexual conduct specifically defined by the regulating state law, as written or construed.").

"[p]atently offensive representations or descriptions of ultimate sexual acts, normal or perverted, actual or simulated" and "[p]atently offensive representations or descriptions of masturbation, excretory functions, and lewd exhibition of the genitals."[158] Lastly, to give fair warning to potential offenders, the state statute utilized must specifically proscribe whichever hardcore materials are being used.[159]

The third element explicitly rejects the "*utterly* without redeeming social value" test by changing the phrasing to "lacks serious . . . value."[160] The issue that the *Miller* Court recognized was that in the old test, it was extremely difficult to prove a negative (that material does not contain *any* redeeming social value).[161] Thus, by changing this standard to "lacks serious . . . value," the determination for the trier of fact of whether material is obscene is made more practicable, since even materials with some value may be deemed obscene.[162]

The *Miller* Court also clarified that each community has the right to decide which materials appeal to the prurient interest, as opposed to implementing a national standard, which was advocated for in *Jacobellis*.[163] The primary reasoning for this decision was that demanding all juries apply a national standard would be a futile exercise since a single formulation could not possibly address the diversity within all fifty states.[164] Thus, *Miller* clearly explains what constitutes obscenity.

---

158 *See Miller*, 413 U.S. at 25.

159 *See* Hunt, *supra* note 148, at 1284.

160 *See Miller*, 413 U.S. at 24–25 (emphasis added); Hunt, *supra* note 148, at 1284–85.

161 *See* Hunt, *supra* note 148, at 1284 n.52.

162 *See id.*

163 *See Miller*, 413 U.S. at 30–35; *see also supra* note 140 (noting the difficulty of formulating a community standard for obscenity when such a standard may vary from one community to the next).

164 *See Miller*, 413 U.S. at 30 ("[O]ur Nation is simply too big and too diverse for this Court to reasonably expect that such standards could be articulated for all 50 States in a single formulation, even assuming the prerequisite consensus exists. When triers of fact are asked to decide whether 'the average person, applying contemporary community standards' would consider certain materials 'prurient,' it would be unrealistic to require that the answer be based on some abstract formulation.").

F. *Separate Categories like Child Pornography*

1. Ferber and Ashcroft

When Justice White, in his majority opinion in *New York v. Ferber*,[165] deemed that the distribution, sale, or exhibition of child pornography does not warrant First Amendment protection, he stressed that "the test for child pornography is separate from the obscenity standard enunciated in *Miller*."[166] Thus, by not analyzing child pornography through the lens of obscenity, the Court created a new categorical exception of materials that are not protected by the First Amendment.

Ferber emphasized that child pornography does not deserve First Amendment protection for a number of reasons. First and foremost, preventing the abuse and sexual exploitation of children is a government objective that is exceptionally important.[167] Second, the Court expressed sensitivity to the fact that child pornography serves as everlasting documentation of a child's involvement in such activities.[168] Moreover, child pornography can be easily circulated and is "intrinsically related to the sexual abuse of children."[169] Furthermore, the Court determined that the value of permitting the distribution of child pornography is de minimis since it is unlikely that the display of children in sexual acts is important in a literary, scientific, or educational sense.[170] *Ferber* led legislators to test the boundaries of what warrants a First Amendment exception. A situation in which this boundary was tested, and rejected, arose in *Ashcroft v. Free Speech Coalition*.[171]

---

165 New York v. Ferber, 458 U.S. 747 (1982).

166 *Id.* at 764.

167 *See id.* at 757–58.

168 *See id.* at 759.

169 *See id.* at 759 n.10 ("[P]ornography poses an even greater threat to the child victim than does sexual abuse or prostitution. Because the child's actions are reduced to a recording, the pornography may haunt him in future years, long after the original misdeed took place. A child who has posed for a camera must go through life knowing that the recording is circulating within the mass distribution system for child pornography." (citation omitted)).

170 *See id.* at 762–63.

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

The Court in *Ashcroft* addressed whether the Child Pornography Prevention Act of 1996 (CPPA) violated the First Amendment.[172] The CPPA prevented, for example, the distribution and possession of pornography in which adults were used to depict minors, or when virtual computer imaging technology could make it appear as if minors participated in such pornography.[173] The Court ultimately held that certain sections of the CPPA abridged freedom of speech and therefore were unconstitutional.[174]

The Court determined that the framework established by *Ferber* was overextended since the CPPA proscribed "child pornography that does not depict an actual child."[175] The Court reiterated that one of the primary reasons why First Amendment protection was not granted to child pornography in *Ferber* was because of the interest in protecting the child manipulated during the production process of the materials.[176] In contrast to *Ferber*, the Court relied on the fact that the CPPA bans materials that do not involve using children in the production process.[177] The Court rejected Congress's claims that these materials should nevertheless be proscribed.[178] One congressional theory was that children would perhaps be more inclined to participate in activities with a pedophile if shown that other children had previously participated.[179] Yet, the Court noted that the potential for crime is not enough of a justification to suppress free speech.[180]

---

172 The CPPA extended "the federal prohibition against child pornography to sexually explicit images that appear to depict minors but were produced without using any real children." *See id.* at 239.

173 *See id.* at 239–40; *see also* 18 U.S.C. § 2251 note (1996) (Act Sept. 30, 1996; Congressional Findings) (Congressional findings about this computer imaging technology found that "new photographic and computer imagining technologies make it possible to produce by electronic, mechanical, or other means, visual depictions of what appear to be children engaging in sexually explicit conduct that are virtually indistinguishable to the unsuspecting viewer from unretouched photographic images of actual children engaging in sexually explicit conduct . . . .").

174 *See Ashcroft*, 535 U.S. at 258.

175 *See id.* at 240.

176 *See id.* (citing New York v. Ferber, 458 U.S. 747, 758 (1982)).

177 *See id.* at 241.

178 *See id.* at 241, 244.

179 *See id.* at 241 ("Pedophiles might use the materials to encourage children to participate in sexual activity.").

180 *See id.* at 245.

Additionally, the Court acknowledged that while obscene materials are not protected by the First Amendment, indecent material, such as an adult depicting a child in pornography, are protected.[181] In ruling certain sections of the CPPA unconstitutional, the Court made it clear that Congress cannot proscribe child pornography based on its denunciation of the material.[182] Instead, the focus must be on Congress's interest in the harm a child may suffer in the production process.[183]

Lastly, it is pertinent to call attention to Section 2256(8)(C) of the CPPA, which the Court did not consider since respondents did not challenge it.[184] This section of the CPPA prohibited the use of computer morphing, a technology that could "alter innocent pictures of real children so that the children appear to be engaged in sexual activity."[185] While the Court recognized that banning computer-morphed materials may be unconstitutional because they are somewhat similar to virtual child pornography, it also reflected that these materials may be "closer to the images in *Ferber*" because real children's images are used.[186] Had the Court determined that morphing, even in the context of child pornography, is protected by the First Amendment, then it would seem clear that deepfakes are also protected. However, if the Court concluded that morphing is not protected in the context of child pornography, it would still be unclear whether deepfakes made from images of adults are protected. Since the Court did not make this determination, it is still unsettled whether pornographic deepfake legislation would be constitutional.

---

181 "[T]he fact that protected speech may be offensive to some does not justify its suppression." *See id.* (quoting Carey v. Population Servs. Int'l, 431 U.S. 678, 701 (1977)).

182 *See id.*

183 S*ee id.* at 240–42.

184 S*ee id.* at 242.

185 *See id.*

186 *See id.*

II. ANALYSIS

A. *Applying the Obscenity Standard to Pornographic Deepfakes*

Ultimately, what the *Miller* obscenity standard requires is that the trier of fact conduct five different analyses: (1) whether the material appeals to the prurient interest, (2) whether the material is patently offensive, (3) whether the material is hardcore, (4) whether the material depicts sexual conduct specifically defined by the applicable state law, and (5) whether the material lacks serious value.[187] Using these five criteria, it is possible to analyze whether a pornographic deepfake can be obscene.

The first step requires examining if the average person in a given community would find that pornographic deepfakes appeal to the prurient interest, which *Roth* defined as "material having a tendency to excite lustful thoughts," or a shameful and morbid interest beyond customary limits.[188] Obviously, every community has different standards for what they might consider to be prurient materials;[189] thus, there cannot be a uniform guideline for what all communities consider prurient. That consideration notwithstanding, it is imperative to note that deepfakes only superimpose someone's face onto an already existing video.[190] Therefore, the only way for a pornographic deepfake to be considered prurient is if the underlying video is prurient.[191] As such, a Deepfaker could be prosecuted if the underlying conduct displayed by the deepfake appeals to what a specific community considers prurient. Since community standards are the determining factor, the first prong would not seem to jeopardize the categorization

187 *See* Hunt, *supra* note 148, at 1285.

188 *See* Roth v. United States, 354 U.S. 476, 487 n.20 (1957).

189 For example, in 1972, a Georgia jury convicted a movie theatre manager for distributing obscene materials by showing the R-rated film *Carnal Knowledge*. *See* Warren Weaver Jr., *Court to Review Obscenity Case*, N.Y. TIMES (Dec. 11, 1973), https://www.nytimes.com/1973/12/11/archives/court-to-review-obscenity-case-carnal-knowledge-appeal-is.html [https://perma.cc/8UXN-LS7K]. However, the film was still shown elsewhere, even though there were community objections. *See, e.g.*, *Film Unveiling Delayed For 'Carnal 'Knowledge,'* N.Y. TIMES (Dec. 25, 1971), https://www.nytimes.com/1971/12/25/archives/film-unveiling-delayed-for-carnal-knowledge.html [https://perma.cc/R875-3LV6].

190 *See* Spivak, *supra* note 79, at 361.

191 *See id.*

of pornographic deepfakes as obscene. The same logic applies when examining whether the material is patently offensive or hardcore.

Next, in order for pornographic deepfakes to be considered obscene, the depicted sexual conduct has to be specifically defined by the applicable state law. Thus, in a potential case against a Deepfaker, the relevant question is whether the deepfake's underlying sexual conduct is specifically defined by the obscenity law of the state where the prosecution is taking place. The best evidence that some states do not believe their obscenity laws specifically define the sexual conduct portrayed in pornographic deepfakes is that states, like Virginia, have enacted separate deepfake statutes.[192] Put another way, if Virginia legislators believed that Deepfakers could be prosecuted under their current obscenity statute, there would be no need to enact any separate statute. However, just because Virginia legislators do not think that pornographic deepfakes fall under their obscenity statute does not mean that other states' legislators feel the same way. Therefore, as long as the state in question specifically defines the deepfake's underlying sexual conduct in its obscenity statute, this prong also does not jeopardize a court potentially ruling that a pornographic deepfake is obscene.

Lastly, pornographic deepfakes must be found to lack serious literary, artistic, political, or scientific value. Pornographic deepfake victims and critics would claim that there is absolutely no value, let alone *serious* value, in allowing people to create videos falsely depicting someone in pornography.[193] Even people who create pornographic deepfakes acknowledge that what they do is "derogatory, vulgar, and blindsiding to the women that deepfakes works on."[194] Those who would argue that pornographic deepfakes do have serious value might explain the benefits of deepfakes' underlying technology, generative

---

192 *See* discussion *supra* Section I.B.2.a.

193 *See* Mutale Nkonde, *Congress Must Act on Regulating Deepfakes*, MEDIUM: ONEZERO (June 17, 2019), https://onezero.medium.com/congress-must-act-on-regulating-deepfakes-1e7e94783be8 [https://perma.cc/9TG4-3RXT] ("If someone were to make deepfake pornographic content of me, it would undermine public trust and derail my career. I do not have the resources to salvage my reputation, to place my attack in a larger social context and to recoup the lost income. This is why I need the government to regulate the spread of deepfakes—and so do the rest of us.").

194 *See* Megan Farokhmanesh, *Deepfakes Are Disappearing from Parts of the Web, but They're Not Going Away*, VERGE (Feb. 9, 2018, 9:00 AM), https://www.theverge.com/2018/2/9/16986602/deepfakes-banned-reddit-ai-faceswap-porn [https://perma.cc/D9EU-9M5M].

adversarial networks (GANs).[195] GANs have advanced what is referred to as "unsupervised learning,"[196] which could drastically improve technologies such as automated-driving technology and voice-activated systems like Siri.[197] Professional artists are also interested in the capabilities that deepfake technology could provide to face and body swapping, citing its accuracy and cost-efficiency.[198] Critics might argue that the technology could still be advanced and developed without the production of pornographic deepfakes; proponents would likely counter that any ban, on any type of deepfake, would have a chilling effect on deepfake technology, since people would be less likely to advance deepfakes' algorithm with the possibility of facing a lawsuit.[199] Ultimately, the determination of whether pornographic deepfakes lack serious value could be construed either way.

B. *Analyzing Pornographic Deepfakes like a Separate Category*

Analyzing pornographic deepfakes under the tests and criteria set out for child pornography in *Ferber* and *Ashcroft* can help determine

---

195 *See* J.M. Porup, *How and Why Deepfake Videos Work—And What Is at Risk*, CSO (Apr. 10, 2019, 3:00 AM), https://www.csoonline.com/article/3293002/deepfake-videos-how-and-why-they-work.html [https://perma.cc/D5SP-F2GC] (What GANs do is set up "two machine learning (ML) models [to] duke it out. One ML model trains on a data set and then creates video forgeries, while the other attempts to detect the forgeries. The forger creates fakes until the other ML model can't detect the forgery. The larger the set of training data, the easier it is for the forger to create a believable deepfake.").

196 *See Supervised and Unsupervised Learning*, GEEKS FOR GEEKS (Jan. 10, 2019), https://www.geeksforgeeks.org/supervised-unsupervised-learning [https://perma.cc/CQX5-BV28] ("Unsupervised learning is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data.").

197 *See* Porup, *supra* note 195.

198 *See* Farokhmanesh, *supra* note 194 ("Before [deepfakes], you needed a team of artists working around the clock to do even a slightly convincing job of a face/body swap . . . . This algorithm is breaking the barriers of uncanny valley, providing a scarily accurate faceswap over a *single* gaming computer, possibly in as short as 24 hours. No team, no render farm, no money.").

199 Jonathon W. Penney, *Internet Surveillance, Regulation, and Chilling Effects Online: A Comparative Case Study*, 6 INTERNET POL'Y REV. 1, 12 (2017) (One study showed that thirty-eight percent of respondents would be much less likely "to share content on the internet that [the respondent] personally created, authored, or made," when faced with a possible legal threat.).

whether such deepfakes should be a separate category that Congress would be able to proscribe under a federal statute.

The Court's concern in *Ashcroft* was about the harm and abuse children suffer in the production process of child pornography.[200] Therefore, in order for pornographic deepfakes to be proscribed as a separate category like child pornography, victims would have to prove harm in the production process of deepfakes. Victims of pornographic deepfakes, however, do not suffer any harm in the production process of a deepfake.[201] Clearly, deepfakes are created using existing images of the person, and the victim might not even know that a pornographic deepfake was created.[202] For example, someone informed Rana Ayyub that a pornographic deepfake depicting her was circulating on the internet.[203] If Ayyub had been harmed in the production process, she would have been aware of the creation of the deepfake. Thus, deepfakes fail to satisfy the fundamental concern set out in *Ashcroft*.

Next, *Ferber* was concerned with the permanent record of the child's participation in such activities.[204] Even though pornographic deepfakes are theoretically permanent, and will likely be available on the internet for an infinite amount of time, such videos do not serve as a permanent record of the *victim's* participation. While viewing a pornographic deepfake can very well be a horrifying experience for the victim, they ultimately know that what they are viewing is fake. Pornographic deepfakes are obviously not like child pornography in this regard, as a child viewing material of themselves can conjure up memories of abuse from participating in the production of such materials.

The intrinsic correlation between the production of child pornography and child sexual abuse was extremely important to the Court in *Ferber*.[205] Thus, if a strong correlation can be shown between pornographic deepfakes and the sexual abuse of those portrayed in the deepfake, then there might be stronger reasons to make pornographic deepfakes an unprotected and separate category like child pornography.

---

200 *See supra* note 176 and accompanying text.

201 *Cf. supra* note 176 and accompanying text.

202 *See supra* note 5 and accompanying text.

203 *See supra* note 5 and accompanying text.

204 *See supra* note 168 and accompanying text.

205 *See supra* note 169 and accompanying text.

There may be direct evidence of someone being physically abused because of their portrayal in a deepfake. Of course, there is a possibility of such harm occurring, as people reached out to Rana Ayyub inquiring about her rate for a potential meetup.[206] However, proponents of pornography argue that, in general, pornography is not correlated with sexual abuse.[207] These proponents point to a fifty-five percent decrease in sexual abuse over the last twenty years, even as the availability of pornography has increased.[208] One journalist argues that if pornography really does create harm, then as a society we would expect a "substantial increase[] in sexual irresponsibility, divorce, and rape," which has not occurred, according to their research.[209] On the other side of the spectrum, there are those who argue that pornography and rape are positively correlated.[210] One study conducted between 1980 and 1982 demonstrated that the correlation between rape and the circulation of sex magazines was as high as .64.[211] One doctor even conducted a study in which he found that people who become addicted to pornography desire even more materials, eventually pushing them to "act out what they've seen."[212]

It remains to be seen whether pornographic deepfakes will increase the rate at which deepfake victims are harmed. While there may be instances where someone views a pornographic deepfake and decides to commit a crime against the person depicted, *Ashcroft* explicitly says that the potential for crime is not a justification for the suppression of

---

206 *See* Ayyub, *supra* note 2.

207 Milton Diamond, the director of the Pacific Center for Sex and Society at the University of Hawaii at Manoa, said, "There's absolutely no evidence that pornography does anything negative." *See* Tim Rymel, *Does Pornography Lead to Sexual Assault?*, HUFFPOST (Aug. 26, 2016, 2:28 PM), https://www.huffpost.com/entry/does-pornography-lead-to-sexual-assault_b_57c0876ae4b0b01630de8c93 [https://perma.cc/V9W4-GQDC].

208 *See id.*

209 *See id.* (noting that instead "sexual irresponsibility has declined, with abortion rates dropping by 41%, and syphilis by a whopping 74%. The teen birth rate dropped by 33% and divorce has decreased by 23%.").

210 *See* Dana A. Fraytak, *The Influence of Pornography on Rape and Violence Against Women: A Social Science Approach*, 9 BUFF. WOMEN'S L.J. 263, 283 (2001).

211 *See id.*

212 *The Concerning Connection Between Sex Crimes and Porn*, FIGHT THE NEW DRUG (Apr. 2, 2018), https://fightthenewdrug.org/the-disturbing-link-between-porn-and-sex-crimes [https://perma.cc/JJV6-YT29]; *see also* VICTOR B. CLINE, PORNOGRAPHY'S EFFECTS ON ADULTS AND CHILDREN (2009).

free speech.[213] There needs to be a much more direct possibility of crime against the victim of a deepfake video in order to warrant the suppression of the First Amendment. Lastly, while many would agree that pornographic deepfakes are indecent, it has been established that indecency alone is not a valid foundational basis to ban certain materials.[214] Thus, a ban on pornographic deepfakes will likely not survive a constitutional challenge under the First Amendment if viewed from the perspective that it should be proscribed as a separate category like child pornography.

III. IMPLICATIONS AND UNCLEAR SOLUTIONS FOR PORNOGRAPHIC DEEPFAKES

Clearly pornographic deepfakes do not fit neatly into the categorical exceptions for obscenity or a separate category like child pornography. However, if pornographic deepfakes fit exclusively into one category, but not the other, how they would be legislated and monitored would be completely different. One reason for this difference is that obscenity depends on a state's obscenity statute,[215] while child pornography is regulated strictly by state and federal authorities.[216] There is even a federal task force, through programs like the FBI's Crimes Against Children, that oversees the trafficking of child pornography.[217] Another task force, the Internet Crimes Against Children Task Force Program, received over $36,000,000 in funding in 2019 and conducted 81,000 investigations.[218] While the George W. Bush administration did create an obscenity task force, it does not appear that

213 *See supra* note 180 and accompanying text.

214 *See supra* note 181 and accompanying text.

215 *See supra* note 187 and accompanying text.

216 *See* Mani Dabiri, *Child-Pornography Possession in State and Federal Court*, THINK DEFENSE APLC (Mar. 31, 2015), https://www.thinkdefense.com/2015/03/2813 [https://perma.cc/2JNX-ED3K].

217 *Crimes Against Children/Online Predators*, FBI, https://www.fbi.gov/investigate/violent-crime/cac [https://perma.cc/MP86-DQJU].

218 *Internet Crimes Against Children Task Force Program*, OJJDP, https://ojjdp.ojp.gov/programs/internet-crimes-against-children-task-force-program [https://perma.cc/AZT7-E5V3].

any such unit currently exists because there is a lack of interest in prosecuting obscenity cases.[219]

It would be easy to imagine how pornographic deepfakes could be monitored federally if they are deemed to be a separate category like child pornography: there would be a federally funded task force whose goal it would be to deny the dissemination of pornographic deepfakes on the internet. Identifying which videos may be deepfakes would be inherently difficult, since the videos are meant to be convincingly deceitful. But there are currently technologies being developed to help detect deepfakes, and the Pentagon's research wing, Defense Advanced Research Projects Agency (DARPA), has contributed funding.[220] Even if these detection algorithms become reliable, surely any video with the words "deepfake" would be deleted, and deepfake-dedicated pornography websites would be shut down. Federally banning pornographic deepfakes would almost immediately cure pornographic deepfake victims' harm, as the availability and distribution of those videos would practically disappear, just as child pornography has disappeared on the internet.[221]

If certain states determine that deepfakes are obscene, but do not fit into a separate category like child pornography, then the only real change for pornographic deepfake victims may occur if Section 230 is amended. Such an amendment could provide that if the applicable state obscenity statute bans the underlying conduct displayed, then internet content publishers that allow pornographic deepfakes to be viewed in those states could be liable, similarly to how sex trafficking was proscribed with the exception created for FOSTA and SESTA. If such an amendment to Section 230 is put in place, the reaction of internet publishers might be drastic, and potentially chilling: publishers may remove more content than necessary in order to evade potential

---

219 *See* Josh Gerstein, *Holder Accused of Neglecting Porn*, POLITICO (Apr. 16, 2011), https://www.politico.com/story/2011/04/holder-accused-of-neglecting-porn-053314 [https://perma.cc/5LAN-KDHK]

220 One technology includes a deepfake-detection system that is ninety-two percent accurate. *See* Will Knight, *A New Deepfake Detection Tool Should Keep World Leaders Safe—For Now*, MIT TECH. REV. (June 21, 2019), https://www.technologyreview.com/s/613846/a-new-deepfake-detection-tool-should-keep-world-leaders-safefor-now [https://perma.cc/CG86-VYB5].

221 Officials even shut down a dark-web child pornography marketplace. *See* Cyrus Farivar & Andrew Blankstein, *Feds Take Down the World's 'Largest Dark Web Child Porn Marketplace'*, NBC NEWS (Oct. 16, 2019, 10:00 AM), https://www.nbcnews.com/news/crime-courts/feds-take-down-world-s-largest-dark-web-child-porn-n1066511 [https://perma.cc/U2D5-73UE].

liability. Therefore, there is currently no solution available that would properly remedy pornographic deepfake victims while protecting freedom of expression under the First Amendment.

CONCLUSION

As evidenced by Rana Ayyub's story, pornographic deepfakes can be devastating for those who are portrayed.[222] States have already begun to pass legislation targeting pornographic deepfakes, but it is unclear how effective those statutes will be if Section 230 is not amended. While Congresswoman Clarke's federal deepfake watermarking legislation may be passed, its effectiveness is questionable.[223] At the very least, more studies need to be conducted to see if Congresswoman Clarke's proposal will actually provide relief to those harmed.[224]

If there is federal deepfake legislation passed that specifically targets pornographic deepfakes, then, in order to be constitutional, pornographic deepfakes may need to fit into the First Amendment categorical exception of being a separate category like child pornography. This Note demonstrates that there are issues with positioning pornographic deepfakes into a separate category like child pornography; the main issues are that the deepfake victim is not involved in the production process of the video and that it is unknown whether there is an intrinsic connection between someone's appearance in a pornographic deepfake and their suffering immediate harm. In determining whether pornographic deepfakes can be obscene under a state's statute by applying the test from *Miller*, it is unclear if deepfakes truly do lack serious scientific value.[225] If pornographic deepfakes fit into either category, a chilling effect may occur, as internet publishers would likely take down more content than necessary to evade liability. One possible solution would be amending Section 230 of the CDA, but doing so could have a drastic impact on the internet's growth, which Section 230 has fostered since its implementation. Furthermore, the implicit deceitfulness of deepfakes conjures up more issues for a

222 *See* Ayyub, *supra* note 2.

223 *See* discussion *supra* Section I.B.1.b.

224 See discussion *supra* Section I.B.1.b.

225 *See* discussion *supra* Part II.

potential lawsuit, such as proving that the content is a deepfake and not just, for example, an unedited video.

Looking to the future, the technology behind deepfakes will likely become more advanced, and it may become more difficult to identify when a video is truly a deepfake. Whether First Amendment freedoms should be restricted in relation to pornographic deepfakes is a difficult issue. Ultimately, as this Note explains, there is no clear solution. At the very least, before substantive federal legislation is passed, researchers need to conduct more studies to learn about the impact that pornographic deepfakes have on victims and internet users.