

**DEEPLY DEHUMANIZING, DEGRADING, AND
VIOLATING: DEEPFAKE PORNOGRAPHY AND THE
PATH TO LEGAL RECOURSE**

Emily Pascale[†]

TABLE OF CONTENTS

ABSTRACT	336
INTRODUCTION	336
I. DEEPFAKE PORNOGRAPHY	337
A. <i>What Are Deepfakes?</i>	337
B. <i>The Proliferation of Deepfake Pornography</i>	338
C. <i>The Harms Inflicted by Deepfake Pornography</i>	340
II. THE LEGAL LANDSCAPE	341
A. <i>Narrowly Constructed Deepfake Legislation</i>	342
1. <i>Proposed Federal Legislation</i>	342
2. <i>State Legislation</i>	343
B. <i>Existing Content Categories</i>	345
1. <i>Defamation</i>	345
2. <i>Obscenity</i>	348
III. CONSTITUTIONAL ANALYSIS: SHOULD DEEPFAKE PORNOGRAPHY BE PROTECTED SPEECH?	350
A. <i>Supreme Court Precedent Relevant to Morphed Pornography</i>	351
1. <i>New York v. Ferber</i>	351
2. <i>Ashcroft v. Free Speech Coalition</i>	352
B. <i>Circuit Court Decisions Recognizing Morphed Child Pornography as Unprotected Speech</i>	353
1. <i>Second Circuit</i>	353
2. <i>Fifth Circuit</i>	355
3. <i>Sixth Circuit</i>	357
C. <i>Applying the Morphed Pornography Cases to Deepfake Pornography</i>	357
IV. LEGISLATIVE RECOMMENDATIONS	363
CONCLUSION	365

[†] J.D. Candidate, Syracuse University College of Law. I would like to thank Professor Laurie Hobart for her helpful insights and suggestions during the writing process, as well as Ben for his unwavering support.

ABSTRACT

Deepfakes have the ability to make people do and say things they never did—a powerful tool that can be used to exploit, manipulate, and deceive. While most commentary has focused on the threats deepfakes pose in the political context, the vast majority of deepfakes online are used for a different purpose: nonconsensual pornography. Deepfake pornography can have devastating consequences on victims’ lives, invading their sexual privacy and inflicting permanent reputational and emotional harm. But despite this prevalent threat, the current legal regime is wholly inadequate in providing victims an avenue of legal recourse. This Note seeks to address this gap, arguing that Congress can and should criminalize nonconsensual deepfake pornography. Specifically, it argues that such legislation would not run afoul of the First Amendment because nonconsensual deepfake pornography should be considered unprotected speech. By regulating deepfakes as a content classification, Congress can avoid triggering free speech concerns and better ensure the legislation stands up in court.

INTRODUCTION

An email pops up on your phone. “There’s a deepfake of you,” it reads. As panic sets in, you begin Googling your name. Instantly, a link to an adult website pops up. After the video downloads, you recognize your face. You watch as your eyes flash to the camera, as your mouth moves expressively. But you never took the video, and the video is pornographic.

Horrified, you frantically scour the internet for other images, hoping it was only a fluke. But then another video pops up, and another. Below each clip are your name and address. You eventually turn to the police, but there is nothing they can do.

Unfortunately, this story is not the work of fiction but reflects a reality faced by victims of deepfake pornography.¹ Products of machine learning technology, deepfakes are the latest iteration of sex

1. This hypothetical is based on the experience of Noelle Martin, a victim of deepfake pornography. See generally Daniella Scott, *Deepfake Pornography Nearly Ruined My Life*, ELLE (June 2, 2020), <https://www.elle.com/uk/life-and-culture/a30748079/deepfake-porn/> (detailing Noelle Martin’s story).

cybercrimes that convincingly map a person's face onto a pre-existing pornographic image.²

This Note explores the ramifications of deepfake pornography and considers legal solutions to thwart its spread. Section I explains the technology underlying deepfakes and examines the danger associated with deepfake pornography. Section II analyzes whether deepfake pornography is protected speech under the First Amendment and discusses the inadequacy of existing solutions. Finding the existing legal regime inadequate, Section III argues that deepfake pornography should be exempted as a separate category like morphed child pornography. Section IV concludes by proposing legislation within the confines of this new content classification, using the existing morphed pornography statute as a guide.

I. DEEPPAKE PORNOGRAPHY

A. *What Are Deepfakes?*

Deepfakes are fabricated images that make someone appear to say or do something they never did.³ The first widely known application of this technology emerged on a Reddit thread in 2017, when a user posted videos that superimposed female celebrities onto pornographic films.⁴ The Reddit user who posted the videos was known as “u/deepfakes,” coining the term that is a portmanteau of “deep learning” and “fake.”⁵

Though resembling the effects of Photoshop, deepfakes are the product of artificial intelligence, relying on neural networks to generate “realistic impersonations out of digital whole cloth.”⁶ The process involves inputting hours of video footage of a specific individual to train the neural network to “understand” the nuances of that person's face.⁷ Once the network is trained, it can digitally graft

2. See Rebecca A. Delfino, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, 88 *FORDHAM L. REV.* 887, 890–91 (2019).

3. Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 *CALIF. L. REV.* 1753, 1758 (2019).

4. James Vincent, *Why We Need a Better Definition of 'Deepfake'*, *THE VERGE* (May 22, 2018), <https://www.theverge.com/2018/5/22/17380306/deepfake-definition-ai-manipulation-fake-news>.

5. *Id.*

6. Chesney & Citron, *supra* note 3, at 1758.

7. Sally Adey, *What Are Deepfakes and How are They Created?*, *IEEE SPECTRUM* (Apr. 29, 2020), <https://spectrum.ieee.org/what-is-deepfake>.

one person's face onto another person's body.⁸ Because it takes a large data set to render a realistic image, it is easier to create deepfakes of public figures who have been photographed and filmed extensively.⁹

As time has gone on, deepfake technology has grown more sophisticated and hyper-realistic thanks to reliance on generative adversarial networks (GANs).¹⁰ Rather than using a single neural network, GANs involve two networks competing against each other to create a synthesized product that is convincingly refined.¹¹ One network, known as the generator, uses a sample set of data to create a synthetic image mimicking the original data set.¹² The second network, called the discriminator, draws on the same data set to evaluate the authenticity of the image.¹³ Feedback is sent to the generator in an iterative fashion until the discriminator can no longer distinguish between the real and fake image.¹⁴ This technology has evolved to such an extent that viewers often have a hard time recognizing that an image is fabricated.¹⁵

B. The Proliferation of Deepfake Pornography

The first widespread application of deepfake technology, nonconsensual pornography, has proliferated rapidly since its emergence in 2017.¹⁶ According to a report by Sensity, the number of pornographic deepfakes has doubled every six months since 2018, accounting for ninety-six percent of all deepfakes online.¹⁷ By 2021,

8. *Id.*

9. See Matthew Bodi, *The First Amendment Implications of Regulating Political Deepfakes*, 47 RUTGERS COMPUT. & TECH. L. J. 143, 146 (2021).

10. See Will Knight, *The US Military is Funding an Effort to Catch Deepfakes and Other AI Trickery*, MIT TECH. REV. (May 23, 2018), <https://www.technologyreview.com/2018/05/23/142770/the-us-military-is-funding-an-effort-to-catch-deepfakes-and-other-ai-trickery/>.

11. See Mika Westerlund, *The Emergence of Deepfake Technology*, 9 TECH. INN. MNGT. REV. 39, 40–41 (2019).

12. *See id.*

13. *See id.*

14. *See* Knight, *supra* note 10.

15. *See* Shane Raymond, *Deepfake Anyone? AI Synthetic Media Tech Enters Perilous Phase*, REUTERS (Dec. 13, 2021, 5:41 PM), <https://www.reuters.com/technology/deepfake-anyone-ai-synthetic-media-tech-enters-perilous-phase-2021-12-13/>.

16. *See* DEEPTRACE, *THE STATE OF DEEPFAKES: LANDSCAPE, THREATS, AND IMPACT 1* (2019).

17. *See id.*; Nina Schick, *Deepfakes Are Jumping from Porn to Politics. It's Time to Fight Back*, WIRED (Dec. 28, 2020, 6:00 AM), <https://www.wired.co.uk/article/deepfakes-porn-politics>.

2023] **Deeply Dehumanizing, Degrading, and Violating** 339

180,000 deepfakes were posted across the internet.¹⁸ That number rose to 720,000 during the summer of 2022, generating hundreds of millions of views.¹⁹

Driving this proliferation are rapid advancements in machine learning technology that have removed previous technological barriers to entry.²⁰ Certain websites, for instance, are “specifically designed” to help people without technological expertise produce deepfakes.²¹ Instead of requiring thousands of pictures and hours of video footage, these apps enable users to create deepfakes using a single picture, often taken from a victim’s social media account.²² Once the picture is uploaded, users can select among a “library of porn videos” to “generate a preview of the face-swapped result within seconds.”²³ Other apps allow users to virtually strip women of clothing, revealing what appears to be their nude bodies.²⁴

As this technology becomes more streamlined, the apps and websites that specialize in deepfake pornography are cashing in.²⁵

18. Schick, *supra* note 17; DEEPTRACE, *supra* note 16, at 1.

19. Schick, *supra* note 17; DEEPTRACE, *supra* note 16, at 1.

20. See DEEPTRACE *supra* note 16, at 8.

21. Adam Dodge & Erica Johnstone, *Using Deepfake Technology to Perpetuate Intimate Partner Abuse*, WITHOUT MY CONSENT 5 (2018) (available at <https://withoutmyconsent.org/perch/resources/2018-04-25deepfakedomesticviolenceadvisory.pdf>).

22. Karen Hao, *A Horrifying New AI App Swaps Women into Porn Videos with a Click*, MIT TECH. REV. (Sept. 13, 2021), <https://www.technologyreview.com/2021/09/13/1035449/ai-deepfake-app-face-swaps-women-into-porn/>.

23. *Id.*

24. DEEPTRACE, *supra* note 16, at 8. One such app that has recently exploded in popularity is Lensa. Olivia Snow, ‘*Magic Avatar*’ App Lensa Generates Nudes from my Childhood Photos, WIRED (Dec. 7, 2022), <https://www.wired.com/story/lensa-artificial-intelligence-csem/>. This app allows users to generate hyper-realistic avatars of themselves after uploading at least 10 selfies. *Id.* While most of the digital portraits are innocuous, the app also, without warning, generates nude and hypersexualized images of the user. Melissa Heikkilä, *The Viral AI Avatar App Lensa Undressed Me—Without My Consent*, MIT TECH. REV. (Dec. 12, 2022), <https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/>. Lensa has drawn criticism not only for producing unprompted sexual images of the user, but also for providing a platform to create nonconsensual nude images of other people. See *id.*; see Snow, *supra* note 24.

25. See DEEPTRACE, *supra* note 16, at 3. In 2022, Lensa generated \$16.2 million in revenue, \$8.2 million of which was produced in a five-day period. David Curry, *Lensa AI Revenue and Usage Statistics (2023)*, BUS. OF APPS (Jan. 9, 2023), <https://www.businessofapps.com/data/lensa-ai-statistics/#:~:text=Lensa%20AI%20has%20made%20%2416.2%20million%20rev>

After drawing in users with free trial runs, most deepfake websites charge users for continued access, offering weekly and monthly subscriptions.²⁶ Some of these websites have thousands of paying customers and acquire new users through sophisticated referral processes.²⁷ Additionally, the websites that are dedicated to hosting deepfake pornography are also profiting by featuring advertisements.²⁸ Given the financial incentives to produce and disseminate deepfake pornography, deepfake websites will continue to proliferate, exposing anyone with an online presence to virtual sexual exploitation.

C. The Harms Inflicted by Deepfake Pornography

Deepfake pornography can have devastating repercussions on victims' lives.²⁹ By non-consensually displaying an individual in a pornographic setting, deepfake pornography reduces individuals to "sexual objects that can be exploited and exposed."³⁰ As one victim described it, it was "dehumanizing, degrading, [and] violating to just see yourself being misrepresented and being misappropriated in that way."³¹ Others have suffered such pervasive invasions that they were forced to change their names.³²

Sexual privacy intrusions can also lead to severe physical and psychological harm.³³ This was the case for Indian journalist Rana Ayyub, who was the subject of a pornographic deepfake released in

enue%20in,alone%2C%20already%20surpassing%20the%20total%20generated%20in%202021.

26. See Matt Burgess, *The Biggest Deepfake Abuse Site Is Growing in Disturbing Ways*, WIRED (Dec. 15, 2021), <https://www.wired.com/story/deepfake-nude-abuse/>. Lensa charges \$7.99 for a pack of 50 images, and \$35.99 for a year subscription. Curry, *supra* note 25.

27. Burgess, *supra* note 26. In short, the process involves providing users who share a personalized link a reward every time someone clicks. *Id.* These links have been shared widely across Twitter, YouTube, Telegram, and other forums. *Id.*

28. See DEEPTRACE, *supra* note 16, at 6.

29. See Danielle Keats Citron, *Sexual Privacy*, 128 YALE L. J. 1870, 1924–28 (2019).

30. *Id.* at 1925.

31. Justin Sherman, "Completely Horrifying, Dehumanizing, Degrading": One Woman's Fight Against Deepfake Pornography, CBS NEWS (Oct. 14, 2021), <https://www.cbsnews.com/news/deepfake-porn-woman-fights-online-abuse-cbsn-originals/>.

32. Citron, *supra* note 29, at 1925.

33. *Id.* at 1926.

2023] **Deeply Dehumanizing, Degrading, and Violating** 341

retaliation to her reporting on political corruption.³⁴ When the video went viral, Ayyub’s social media was flooded with harassing comments and solicitations.³⁵ The crush of slut-shame and hateful messages soon overwhelmed her, and she was rushed to the hospital with heart palpitations.³⁶ In the months following the video’s release, Ayyub described feeling isolated and broken.³⁷ Like Ayyub, many deepfake victims are left with extreme anxiety and depression, some even contemplating suicide.³⁸

And the pernicious effects do not end there; deepfake pornography can also cause social and employment-related consequences.³⁹ When making decisions about candidates, most employers consult online search results, often declining to hire candidates who have “unsuitable photos” on the internet.⁴⁰ As a result, “companies may refuse to interview or hire” candidates whose “search results include . . . deep-fake sex videos.”⁴¹ Even where victims retain employment, after the revelation of a nonconsensual pornographic image, they may experience difficulties “concentrating, eating, and working.”⁴² Outside of work, victims are prone to slip into self-isolation, and find it more difficult to form intimate relationships.⁴³

II. THE LEGAL LANDSCAPE

The rapid proliferation of deepfake pornography and its attendant harms necessitates an avenue of legal recourse for victims. But in order to determine the appropriate legislative response to deepfake pornography, it is helpful to begin with the First Amendment constraints to regulation.

34. 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>.

35. *Id.*

36. Rana Ayyub, *I Was the Victim of a Deepfake Porn Plot Intended to Silence Me*, HUFFINGTON POST (Nov. 21, 2018, 8:11 AM), https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316.

37. *See id.*

38. Citron, *supra* note 29, at 1926.

39. *Id.* at 1927.

40. *Id.* (“According to a Microsoft study, nearly eighty percent of employers use search results to make decisions about candidates, and in around seventy percent of cases, those results have a negative impact.”).

41. *Id.* at 1928.

42. *Id.* at 1926.

43. *See* Citron, *supra* note 29, at 1924–25.

The First Amendment states that “Congress shall make no law . . . abridging the freedom of speech.”⁴⁴ When considering whether a regulation would offend this mandate, the first step is to determine if the government would be prohibiting speech because of its content, which is presumptively unconstitutional.⁴⁵ As the Supreme Court has previously recognized, content-based restrictions exist where the government regulates speech because of the “topic discussed or the idea or message expressed,” including a false message.⁴⁶ Thus, since the need to proscribe deepfake pornography arises from its false nature, any deepfake regulation would easily qualify as a content-based restriction.

In order to rebut the presumption of unconstitutionality, the government has two options. First, it can construct a narrowly tailored law that fits within the confines of the First Amendment.⁴⁷ This approach has been taken by the few states that passed laws addressing deepfake pornography. The other option is to regulate deepfake pornography under existing categories of unprotected speech, which include obscenity, true threats, defamation, fighting words, and child pornography, among others.⁴⁸ As currently constructed, the categories most applicable to deepfake pornography are obscenity and defamation. Both of these avenues, and their challenges, are explored below.

A. Narrowly Constructed Deepfake Legislation

1. Proposed Federal Legislation

In an effort to comply with the First Amendment, deepfake legislation proposed at the federal level has been quite narrow in scope. The first bill designed to curb the dissemination of deepfakes,

44. U.S. CONST. amend. I.

45. *See Reed v. Town of Gilbert*, 576 U.S. 155, 163 (2015) (“Content-based laws—those that target speech based on its communicative content—are presumptively unconstitutional and may be justified only if the government proves they are narrowly tailored to serve compelling state interests.”).

46. *Id.*; *see also United States v. Alvarez*, 567 U.S. 709, 715 (2012) (concluding that a statute prohibiting speech based on its false nature is a content-based regulation).

47. In other words, the content regulation must be “narrowly tailored to serve compelling state interests.” *Reed*, 576 U.S. at 163.

48. *Alvarez*, 567 U.S. at 717 (first citing *Miller v. California*, 413 U.S. 15 (1973); then citing *Watts v. United States*, 394 U.S. 705 (1969); then citing *New York Times Co. v. Sullivan*, 376 U.S. 254 (1964); then citing *Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942); then citing *New York v. Ferber*, 458 U.S. 747 (1982); and then citing *Brandenburg v. Ohio*, 395 U.S. 444 (1969)).

2023] **Deeply Dehumanizing, Degrading, and Violating** 343

called the Malicious Deep Fake Prohibition Act of 2018, was introduced by Senator Ben Sasse.⁴⁹ The bill would make it a federal crime to create or distribute a deepfake with the actual knowledge the video is a deepfake and intent to facilitate unlawful activity.⁵⁰ In the context of deepfake pornography, the law’s impact would be quite limited since deepfake pornography does not necessarily involve an underlying crime. And yet, despite the proposed law’s narrow reach, some have argued that it risks chilling protected speech.⁵¹ At the end of 2018, the bill expired without leaving committee.⁵²

The following year, Representative Yvette Clarke introduced a slightly more aggressive piece of deepfake legislation, the DEEP FAKES Accountability Act.⁵³ The bill would require manipulated content to comply with digital watermark and disclosure requirements that notify viewers of its falsity.⁵⁴ Critics of the bill, however, argue that this provision could actually benefit wrongdoers.⁵⁵ Since malicious creators are less likely to adhere to disclosure requirements, viewers “primed to believe the authenticity of any media not containing a watermark or disclosure” would therefore be “easily duped.”⁵⁶ Further, even if the creator labels the deepfake as false, these watermarks can be removed, and even done so anonymously.⁵⁷ Nonetheless, like Senator Sasse’s bill, the DEEP FAKES Accountability Act was never passed.

2. State Legislation

Without federal action, states have been left to enact their own legislation to combat the threat of pornographic deepfakes. Despite

49. Malicious Deepfake Prohibition Act of 2018, S. 3805, 115th Cong. (2018).

50. *Id.* § 1041(b).

51. Kaveh Waddell, *Lawmakers Plunge into “Deepfake” War*, AXIOS (Jan. 31, 2019), <https://www.axios.com/2019/01/31/lawmakers-plunge-into-deepfake-war-1548948076> (criticizing the Malicious Deep Fake Prohibition Act as placing “over-broad liability on distributors” that threatens to “scare platforms into immediately taking down everything that’s reported as a deepfake.”).

52. Bodi, *supra* note 9, at 153.

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

54. *Id.* § 1041(b).

55. See Devin Coldeway, *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/deepfakes-accountability-act-would-impose-unenforceable-rules-but-its-a-start/>.

56. Jack Langa, Note, *Deepfakes, Real Consequences: Crafting Legislation to Combat Threats Posed by Deepfakes*, 101 B.U. L. REV. 761, 789 (2021).

57. Coldeway, *supra* note 53 (describing how even a “sophisticated whole-frame watermark” can be automatically stripped by bots).

promising movement in 2019, state efforts have largely stalled, with only a handful of states successfully banning deepfake pornography.⁵⁸

Of the states that decided to act, most have addressed the problem by providing a private right of action. In 2019, California passed legislation allowing victims to sue creators of fake porn.⁵⁹ Specifically, the law prohibits the dissemination of “computer-generated” images that depict an “individual engaging in sexual conduct in which the depicted individual did not engage.”⁶⁰ Shortly thereafter, New York passed a similar law providing a civil penalty for the nonconsensual disclosure of sexually explicit material that is a result of “digitization.”⁶¹ Both bills have been criticized as infringing on the deepfake creator’s free speech rights.⁶²

Other states have decided to criminalize deepfake pornography.⁶³ Virginia, for instance, makes it a Class 1 misdemeanor to disseminate an image made by “creating, adapting, or modifying a videographic or still image with the intent to depict an actual person.”⁶⁴ In an effort to comply with the First Amendment,⁶⁵ the statute requires that the defendant disseminated the material with the “intent to coerce, harass, or intimidate.”⁶⁶ But considering that many creators do not expect that the victim will discover the deepfake, the intent requirement would be difficult to prove.⁶⁷ As Professor Mary Anne Franks points out, most

58. See Korey Clark, ‘Deepfakes’ Emerging Issue in State Legislatures, LEXISNEXIS, <https://www.lexisnexis.com/en-us/products/state-net/news/2021/06/04/Deepfakes-Emerging-Issue-in-State-Legislatures.page> (last visited Dec. 28, 2022).

59. See CAL. CIV. CODE § 1708.86(b) (West 2022).

60. § 1708.86(a)(6)(B)–(C).

61. See N.Y. CIV. RIGHTS LAW §§ 52-c(1)(a)–(b), (2)(a) (McKinney 2022).

62. See Katyanna Quach, *New York State is Trying to Ban ‘Deepfakes’ and Hollywood Isn’t Happy*, THE REG. (June 12, 2018), https://www.theregister.com/2018/06/12/new_york_state_is_trying_to_ban_deepfakes_and_hollywood_isnt_happy/; Will Fischer, *California’s Governor Signed New Deepfake Laws for Politics and Porn, but Experts Say They Threatened Free Speech*, BUS. INSIDER (Oct. 10, 2019, 12:51 PM), <https://www.businessinsider.com/california-deepfake-laws-politics-porn-free-speech-privacy-experts-2019-10>.

63. See VA. CODE ANN. § 18.2-386.2(A) (2022); MD. CODE ANN., CRIM. LAW § 11-208 (LexisNexis 2022).

64. VA. CODE ANN. § 18.2-386.2(A) (2022).

65. See Delfino, *supra* note 2, at 932 (noting that civil liberties organizations such as the ACLU and Electronic Frontier foundation have “historically opposed statutes that do not include an ‘intent to harm’ provision under the theory that failure to require proof of intent may violate the First Amendment.”).

66. § 18.2-386.2(A).

67. See Delfino, *supra* note 2, at 921.

of these videos are “created for the pleasure of their creator rather than the humiliation of the object of their desire.”⁶⁸ Requiring a subjective intent element also “miss[es] the point”; irrespective of the creator’s motivation, the wrongdoing occurs from disclosing what appears to be one’s naked body without consent.⁶⁹

Additionally, while many states criminalize nonconsensual pornography more generally, these laws do not apply to deepfake pornography.⁷⁰ Nonconsensual pornography, also known as revenge porn, involves the disclosure of “private, sexually explicit image[s] to someone other than the intended audience,” and is criminalized as a sexual privacy violation.⁷¹ Though deepfakes involve similar intrusions, the privacy concerns are arguably less severe because they are typically made using images posted online.⁷² Moreover, by morphing two images together to create a person who does not exist, deepfakes do not expose the actual intimate details of someone’s life—however realistic they may appear.⁷³ Thus, to apply to deepfake pornography, revenge porn laws would have to be amended to specifically cover nonconsensual deepfake images.⁷⁴ Regardless, these laws have raised First Amendment concerns of their own.⁷⁵

B. Existing Content Categories

1. Defamation

Instead of crafting a narrowly tailored law, some commentators have suggested regulating deepfakes under existing content categories, such as obscenity and defamation.⁷⁶ While defamation law

68. 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/>.

69. See Delfino, *supra* note 2, at 921 (quoting Diane Bustamante, Comment, *Florida Joins the Fight Against Revenge Porn: Analysis of Florida’s New Anti-Revenge Porn Law*, 12 FIU L. REV. 357, 387 (2017)).

70. Anna Pechenik Gieseke, Note, “*The New Weapon of Choice*”: Law’s Current Inability to Properly Address Deepfake Pornography, 73 VAND. L. REV. 1479, 1501 (2020).

71. Delfino, *supra* note 2, at 896.

72. See *id.* at 897; see also Gieseke, *supra* note 70, at 1501.

73. *Id.* at 897–98.

74. Gieseke, *supra* note 70, at 1503.

75. Delfino, *supra* note 2, at 925.

76. See Nina I. Brown, *Deepfakes and the Weaponization of Disinformation*, 23 VA. J. L. & TECH. 1, 36–40 (2020) (arguing that laws relating to defamation and fraud apply to deepfakes); see also Matthew B. Kugler & Carly Pace, *Deepfake Privacy: Attitudes and Regulation*, 116 NW. U. L. REV. 611, 671 n.257 (2021) (suggesting that

varies across states, the tort typically consists of the elements laid out in the Second Restatement of Torts.⁷⁷ According to the Restatement, defamation requires:

- (a) a false and defamatory statement concerning another;
- (b) an unprivileged publication to a third party;
- (c) fault amounting to at least negligence on the part of the publisher; and
- (d) either actionability of the statement irrespective of special harm or the existence of special harm caused by the publication.⁷⁸

In *New York Times Co. v. Sullivan*, the Supreme Court famously heightened the standard of fault from negligence to actual malice where the plaintiff is a public official.⁷⁹ A statement is made with actual malice if the speaker or publisher knows the statement was false or recklessly disregarded the truth.⁸⁰ This means that merely showing that a statement is false is insufficient; a plaintiff must demonstrate that the speaker acted with a “high degree of awareness” that the statement is not true.⁸¹ Even applying this heightened standard, many deepfake creators likely evince the requisite state of mind. Deepfakes are, of course, inherently false, depicting people saying and doing things they never did. And unlike typical defamation defendants who can claim ignorance of a statement’s falsity, deepfake creators cannot.⁸² Indeed, creating a deepfake requires a person to take “affirmative steps to swap someone’s face onto another video.”⁸³ The mere fact that a deepfake was created is sufficient to show that the creator knew he was telling a lie, satisfying either the negligence or actual malice standards of fault.

it is “possible” to criminalize deepfakes under obscenity laws); *see also* Chesney & Citron, *supra* note 3, at 1792–95 (concluding that in particular contexts, deepfake creators may be subject to specific categories of civil liability).

77. *See* Russell Spivak, “Deepfakes”: *The Newest Way to Commit One of the Oldest Crimes*, 3 *GEO. L. TECH. REV.* 339, 369–70 (2019).

78. *RESTATEMENT (SECOND) OF TORTS* § 558 (AM. L. INST. 1977).

79. *New York Times Co.*, 376 U.S. at 279–80.

80. *Id.* at 280.

81. Cass R. Sunstein, *Falsehoods and the First Amendment*, 33 *HARV. J. L. & TECH.* 387, 410 (2020); *Gertz v. Robert Welch, Inc.* 418 U.S. 323, 332 (1974) (quoting *Saint Amant v. Thompson*, 390 U.S. 727, 731 (1968)).

82. Jessica Ice, Note, *Defamatory Political Deepfakes and the First Amendment*, 70 *CASE W. RESV. L. REV.* 417, 433–34 (2019).

83. *Id.* at 434.

2023] **Deeply Dehumanizing, Degrading, and Violating** 347

In addition to the creator's state of mind, a plaintiff must prove that the false statement was defamatory. Broadly speaking, defamation is defined as "the making of a statement which tends to expose the plaintiff to public contempt, ridicule, aversion or disgrace."⁸⁴ Undoubtedly, depicting someone engaging in lewd sexual acts can create the kind of reputational harm that invites "public contempt, ridicule, aversion or disgrace."⁸⁵ This conclusion is also supported by a federal district court case out of Virginia, where a defendant distributed doctored photographs depicting plaintiff as a porn star.⁸⁶ The court determined that the photos were defamatory because they "impute an unfitness for Plaintiff to perform the duties of a youth soccer coach" and "prejudice Plaintiff in his profession."⁸⁷

That said, deepfakes that do not purport to be authentic are generally not considered defamatory. In *Milkovich v. Lorain Journal Co.*, the Supreme Court held that a defamation claim must fail if the statement at issue "cannot reasonably be interpreted as stating actual facts about an individual."⁸⁸ Thus, deepfakes that are clearly parody, such as those that turned Nicolas Cage into Indiana Jones and James Bond,⁸⁹ are not grounds for a defamation action. More ominously, this also means that creators of deepfake pornography can evade defamation liability by slapping "fake" or "parody" in the video's title or caption.⁹⁰ Similarly, if a video is obviously fake due to poor editing, the creator could argue that it is not reasonably believable.⁹¹ And since deepfake technology is still in its early stages, many of the videos posted online do not totally pass as authentic material.⁹²

84. *Stepanov v. Dow Jones & Co., Inc.*, 987 N.Y.S.2d 37, 41 (N.Y. App. Div. 2014) (quoting *Foster v. Churchill*, 665 N.E.2d 153, 157 (N.Y. 1996)).

85. *Id.* (quoting *Foster*, 665 N.E.2d at 157).

86. *Tharpe v. Lawidjaja*, 8 F. Supp. 3d 743, 786 (W.D. Va. 2014).

87. *Id.*

88. *Milkovich v. Lorain J. Co.*, 497 U.S. 1, 20 (1990) (citing *Hustler Mag. v. Falwell*, 485 U.S. 46, 50 (1988)).

89. Sam Haysom, *People Are Using Face-Swapping Tech to Add Nicolas Cage to Random Movies and What Is 2018*, MASHABLE (Jan. 31, 2018), <https://mashable.com/article/nicolas-cage-face-swapping-deepfakes>.

90. Aasha Shaik, *Deepfake Pornography: Beyond Defamation Law*, YALE CYBER LEADERSHIP F. (July 20, 2021), <https://cyber.forum.yale.edu/blog/2021/7/20/deepfake-pornography-beyond-defamation-law>; see also Spivak, *supra* note 77, at 373 (noting context of a video supplies the defamation analysis).

91. Shaik, *supra* note 90.

92. Gieseke, *supra* note 70, at 1498 (arguing that although the technology will evolve to an extent that "renders this concern moot," many other women will be left without legal redress in the interim).

While a clearly fabricated deepfake seems less likely to cause harm, that is not always the case. In their 2021 study examining people's attitudes toward labeled deepfake pornography, Professor Matthew Kugler and Carly Pace presented two groups of people with either labeled or unlabeled deepfake pornography.⁹³ After surveying the participants' impressions, they found that the perceived wrongfulness of the deepfake did not diminish for the videos that were labeled fake.⁹⁴ This conclusion reflects the understanding that simply disclaiming that a video is fake fails to address the loss of dignity at the center of the deepfake issue.⁹⁵ Irrespective of whether the video discloses its falsity, the deepfake appropriates one's sexual identity, exhibiting it to the world without consent.⁹⁶ Accordingly, the law should protect victims from this unwarranted exposure.⁹⁷

2. *Obscenity*

It has long been recognized that the First Amendment does not protect obscene speech.⁹⁸ Still, the Supreme Court has remained sensitive "to any infringement on genuinely serious literary, artistic, political, or scientific expression."⁹⁹ As a result, the permissible scope of an obscenity regulation is quite narrow, reserved only for those materials that can be described as "hard core" pornography.¹⁰⁰ To determine whether content is considered obscene, the Supreme Court created the *Miller* test, which asks:

93. See Kugler & Pace, *supra* note 76, at 636. Before the participants were presented with the sets of deepfake videos, they were provided vignettes that described how deepfakes videos are made. *Id.*

94. *Id.* at 640 (concluding that there was "no significant effect of labeling on the perceived harmfulness or blameworthiness of the video").

95. See Shaik, *supra* note 90.

96. *Id.*

97. It is also worth noting that even intentionally messy deepfakes have been able to fool the public. See *id.* When a Belgian political party released a deepfake video of Donald Trump, they "assumed that the poor quality of the fake would be enough to alert their followers to its inauthenticity." Oscar Schwartz, *You Thought Fake News Was Bad? Deep Fakes Are Where Truth Goes to Die*, THE GUARDIAN (Nov. 12, 2018, 5:00 AM), <http://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth>. However, many people who viewed the video believed it was real. See *id.*

98. This content category was first recognized in 1942. See *Chaplinsky*, 315 U.S. at 572.

99. *Miller*, 413 U.S. at 23.

100. *Id.* at 24, 27.

2023] **Deeply Dehumanizing, Degrading, and Violating** 349

- (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.¹⁰¹

Given these considerations, not all deepfake pornography will qualify as obscene. For instance, since the obscenity determination is made using “contemporary community standards,” whether a deepfake is obscene will vary from locality to locality. What one community may deem appealing to the prurient interest, another might find palatable.¹⁰² The depicted sexual activity in a deepfake may also be “specifically defined” in the applicable law of some states but not others.¹⁰³ Moreover, since pornography does not automatically constitute obscenity under the *Miller* test, the application of obscenity law to a specific pornographic deepfake depends on the degree of offensiveness of the underlying image or video.¹⁰⁴ Consequently, whether a deepfake qualifies as obscene must be determined on a case-by-case basis. Not only does this handicap deterrence efforts, but it also precludes the legislature from using obscenity law as a basis for categorically regulating deepfake pornography.¹⁰⁵

While the application of obscenity law to deepfake pornography will vary, obscenity would likely fail to cover certain nude images, one of the most common types of deepfake porn. Indeed, in many jurisdictions, nudity alone is not considered obscene; the nude person must be positioned “for the purpose of focusing the camera on their sex organs.”¹⁰⁶ Thus, because apps like DeepNude can strip women of clothing regardless of whether they were posed in a salacious

101. *See id.* at 24 (citing *Kois v. Wisconsin*, 408 U.S. 229, 230 (1972)) (quoting *Roth v. United States*, 354 U.S. 476, 489 (1957)).

102. *See Spivak, supra* note 77, at 360 (describing how different states take different approaches to obscenity law).

103. *See id.*

104. *See id.*

105. *See id.* (explaining that deepfakes are not facially obscene, and thus “any legislation regulating deepfakes would not pass muster under this obscenity exception”).

106. *See, e.g., City of Youngstown v. DeLoreto*, 251 N.E.2d 491, 501 (Ohio Ct. App. 1969).

manner,¹⁰⁷ these deepfakes would fall outside the scope of obscenity law.

Yet even if obscenity law applies, prosecuting the possession of obscene deepfake pornography would not be an option. In *Stanley v. Georgia*, the Court struck down a law banning the possession of obscene material because it infringed on the right to receive information in the privacy of one's home.¹⁰⁸ Consequently, the government is limited to prosecuting the distribution and sale of obscene materials. As is explained *infra*, this would pose a challenge to addressing the issue of deepfake pornography, which may be difficult to prosecute at the distribution level.

III. CONSTITUTIONAL ANALYSIS: SHOULD DEEPPAKE PORNOGRAPHY BE PROTECTED SPEECH?

As the law currently stands, most victims are forced to suffer the consequences of deepfake pornography without the possibility of redress. While defamation and obscenity laws seem like a logical fit, the very artifice of deepfake images will preclude most of these claims. Moreover, despite promising legislation in some states, these efforts are moving slowly and often criticized as infringing on First Amendment rights.

To avoid the existing issues in regulating deepfake pornography, this Note argues that deepfake pornography should be treated as its own content category outside of First Amendment protections. Specifically, deepfake pornography should be classified as unprotected speech in a similar way to how morphed child pornography is classified as unprotected speech by three federal circuits.¹⁰⁹ Much like deepfake pornography, morphed child pornography is created “when an innocent photo of an actual child is edited to make it appear as though the child is engaging in a sexual act.”¹¹⁰ This morphing can be created through various methods, “ranging from rudimentary scissors and glue to sophisticated editing programs” such as Photoshop.¹¹¹ Also like deepfakes, the creators of

107. See DEEPTRACE, *supra* note 16, at 8.

108. *Stanley v. Georgia*, 394 U.S. 557, 568 (1969).

109. See generally *United States v. Hotaling*, 634 F.3d 725 (2d Cir. 2011); *Doe v. Boland*, 698 F.3d 877 (6th Cir. 2012); *United States v. Mecham*, 950 F.3d 257 (5th Cir. 2020) (holding that morphed child pornography is not protected speech).

110. Caleb Beacham, Note, *Metamorphosis: Changing Oklahoma Law to Protect Children from Morphed Child Pornography*, 55 TULSA L. REV. 311, 316 (2020).

111. *Id.*

2023] **Deeply Dehumanizing, Degrading, and Violating** 351

computer-morphed images often rely on pictures of children that are taken from social media or other websites.¹¹² Given these similarities, if morphed child pornography is unprotected speech, deepfake pornography should be unprotected too.

A. Supreme Court Precedent Relevant to Morphed Pornography

To date, morphed child pornography is classified as unprotected speech by all but one circuit court that has addressed the issue.¹¹³ However, because the Supreme Court has not yet ruled on the matter, it is helpful to begin with the precedent on which the lower courts relied. Those cases are *New York v. Ferber*¹¹⁴ and *Ashcroft v. Free Speech Coalition*.¹¹⁵

1. New York v. Ferber

Forty years ago, the Supreme Court held that child pornography was categorically exempt from First Amendment protections in *New York v. Ferber*.¹¹⁶ In that case, a bookstore owner was convicted under a New York state law criminalizing child pornography after selling two films to an undercover officer.¹¹⁷ Challenging his conviction, the proprietor claimed that the law violated the First Amendment because it proscribed the sale of non-obscene child pornography.¹¹⁸ The Court, however, disagreed, holding that states could validly regulate child pornography that falls outside the scope of obscenity law.¹¹⁹

In arriving at that conclusion, the Court provided five reasons why states should be afforded “greater leeway in the regulation of pornographic depictions of children.”¹²⁰ First, using children as subjects of pornographic content is “harmful” to their “physiological, emotional, and mental health.”¹²¹ Second, the distribution of child pornography is “intrinsically related to the sexual abuse of children” because it creates a “permanent record” of the abuse of children.¹²²

112. *Id.* at 317.

113. *Compare Hotaling*, 634 F.3d at 728, *Doe*, 698 F.3d at 885, and *Mecham*, 950 F.3d at 267, with *United States v. Anderson*, 759 F.3d 891, 894–95 (8th Cir. 2014) (holding that morphed child pornography is protected speech).

114. *Ferber*, 458 U.S. at 747.

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

116. *Ferber*, 458 U.S. at 764.

117. *Id.* at 751–52.

118. *Id.* at 752.

119. *Id.* at 756, 760–61.

120. *Id.* at 756.

121. *Ferber*, 458 U.S. at 758.

122. *Id.* at 759.

Third, because the sale of child pornography provides an “economic motive” for the exploitation of children, cutting off the “distribution network” for this material is necessary to dry up the market for child abuse.¹²³ Fourth, the expressive value of child pornography is “exceedingly modest, if not *de minimis*.”¹²⁴ And, finally, exempting child pornography from the First Amendment due to its harmful effects is in line with other content categories that are based on a similar “balance of competing interests.”¹²⁵ Indeed, as the Court had previously recognized, when “the evil to be restricted so overwhelmingly outweighs the expressive interests, if any, at stake, . . . no process of case-by-case adjudication is required.”¹²⁶

2. *Ashcroft v. Free Speech Coalition*

In *Ashcroft v. Free Speech Coalition*, the Court suggested that *Ferber*’s rationales apply to morphed pornography.¹²⁷ That case concerned an overbreadth challenge to the Child Pornography Protection Act of 1996 (CPPA), which extended the federal prohibition on child pornography to “any visual depiction” or “computer-generated visual image or picture” that “is, or appears to be, of a minor engaging in sexually explicit conduct.”¹²⁸ In essence, this provision meant that explicit imagery that purports to show children but does not actually feature real children—also known as “virtual” child pornography—was considered prohibited child pornography.¹²⁹ The statute, in turn, captured a “range” of visual depictions, including pornographic content featuring youthful-looking adults, teenage actors engaged in sexually explicit scenes in Hollywood movies, and computer-generated or animated child pornography.¹³⁰

By prohibiting pornographic material that does not depict an actual child, the Court concluded the CPPA stretched beyond the bounds of *Ferber*.¹³¹ Writing for the majority, Justice Kennedy explained that banning child pornography is only justified because of

123. *Id.* at 759, 761.

124. *Id.* at 762.

125. *Id.* at 764.

126. *Ferber*, 458 U.S. at 763–64.

127. *Ashcroft*, 535 U.S. at 240.

128. Child Pornography Protection Act of 1996, H.R. 4123, 104th Cong. § 3 (1996).

129. *Ashcroft*, 535 U.S. at 239, 241.

130. *Id.* at 241.

131. *Id.* at 239.

the harms it inflicts on actual children.¹³² In contrast, a pornographic depiction that uses youthful-looking adult actors or computer-generation involves “no crime and creates no victims by its production.”¹³³ While the government argued that virtual child pornography is used to “whet[] the appetites” of pedophiles, thereby indirectly harming children, the Court found that this threat was too remote to justify removing First Amendment protections.¹³⁴

Notably, and although not at issue, the Court also drew attention to § 2256(8)(C) of the CPPA, which criminalizes child pornography created by using computer morphing.¹³⁵ As explained *supra*, morphed pornography, like a deepfake, is created by splicing together an innocent image of a child with an explicit image depicting sexual acts. Notwithstanding the absence of sexual abuse in the creation of these images, the Court found that they are in a sense “closer to the images in *Ferber*” because they “implicate the interests of real children.”¹³⁶ Though *dicta*, this passage signifies the Court’s willingness to recognize that morphed pornography—and by extension deepfake pornography—is exempt from the First Amendment.

B. Circuit Court Decisions Recognizing Morphed Child Pornography as Unprotected Speech

In the years following *Ashcroft*, a majority of circuits heeded the Court’s suggestion, concluding that morphed child pornography does not deserve First Amendment protections. In doing so, the lower courts began to lay a path for extending *Ferber* to adult deepfake pornography.

1. Second Circuit

The Second Circuit was the first to address this issue.¹³⁷ In *United States v. Hotaling*, defendant was charged with possession of child pornography after morphed pornographic images were found on his computer.¹³⁸ The images featured the identifiable faces of minors that had been “pasted” onto adult bodies engaged in sexual conduct.¹³⁹ Hotaling challenged the indictment, arguing that the images were

132. *Id.* at 249.

133. *Id.* at 250.

134. *Ashcroft*, 535 U.S. at 241, 250, 253–54.

135. *See id.* at 242.

136. *Id.*

137. *Hotaling*, 634 F.3d at 725.

138. *Id.* at 727.

139. *Id.*

protected speech because they did not harm an “actual minor” but simply “record[ed] his mental fantasies.”¹⁴⁰ The district court rejected these arguments, and Hotaling pleaded guilty to the charges.¹⁴¹

On appeal, the Second Circuit upheld the conviction, holding that *Ferber’s* categorical exemption extended to morphed child pornography.¹⁴² Echoing *Ashcroft*, the court began by noting that the “underlying inquiry is whether an image of child pornography implicates the interests of an actual minor.”¹⁴³ To the court, the fact that the children involved had not engaged in sexual activity during the creation of the photographs was not dispositive.¹⁴⁴ Rather, the children’s interests were implicated because their identifiable faces were used to “make it appear that they are performing sexually explicit acts.”¹⁴⁵ As a result, the children were at “risk of reputational harm and suffered the psychological harm of knowing that their images were exploited.”¹⁴⁶ Prevention of these harms, the court explained, has been a “long recognized” justification for excluding child pornography from First Amendment protections.¹⁴⁷

Having recognizing morphed pornography as unprotected speech, the Second Circuit next determined that the government could ban its possession.¹⁴⁸ In doing so, the court relied on *Osborne v. Ohio*, a case where the Supreme Court held that criminalizing the possession of child pornography did not offend the First Amendment.¹⁴⁹ According to *Osborne*, the need to ban possession arose after the market for child pornography was “driven underground,” making it nearly impossible to prosecute.¹⁵⁰ By penalizing those who possess the material, demand for the material drops, sparing more children from

140. *Id.*

141. *Id.*

142. *Hotaling*, 634 F.3d at 728.

143. *Id.* at 729.

144. *See id.*

145. *Id.* at 729–30. This connection between the child and sexual conduct was further clarified given that the names of the children featured was attached to each photo. *Id.*

146. *Hotaling*, 634 F.3d at 730.

147. *Id.* at 728.

148. *Id.* at 730.

149. *Osborne v. Ohio*, 495 U.S. 103, 110–11 (1990). Distinguishing *Stanley*, *Osborne* explained that unlike prohibiting possession of obscenity, which was based on the illegitimate state interest in “controlling a person’s private thoughts,” criminalizing possession of child pornography is based on the need to “protect the victims of child pornography.” *Id.* at 109.

150. *Id.* at 110.

2023] **Deeply Dehumanizing, Degrading, and Violating** 355

abuse.¹⁵¹ Given the state’s “compelling” interest in protecting children from the “physiological, emotional, and mental” harms associated with child pornography, the majority “[could not] fault” the state for “attempting to stamp out this vice at all levels in the distribution chain.”¹⁵²

In the Second Circuit’s view, the rationales from *Osborne* applied with equal force to morphed pornography.¹⁵³ Like regular child pornography, the distribution of morphed images has been driven underground, circulating primarily from “trafficker to trafficker.”¹⁵⁴ Accordingly, the Second Circuit rejected the defendant’s contention that the morphed photos were mere “records of the defendant’s fantasies,” but instead were images “primed for entry into the distribution chain.”¹⁵⁵ Congress, therefore, could legitimately ban possession of morphed pornographic images.¹⁵⁶

2. Fifth Circuit

In *United States v. Mecham*, the Fifth Circuit similarly determined that morphed child pornography is unprotected speech.¹⁵⁷ In that case, police discovered over 30,000 images of morphed child pornography on the defendant’s computer.¹⁵⁸ Like the Second Circuit, the Fifth Circuit grappled with the issue of whether “underlying criminal conduct” is necessary to exempt an image of child pornography from First Amendment protections.¹⁵⁹

In arguing that such conduct is necessary, the defendant relied on the narrow interpretation of *Ferber* in a case called *United States v. Stevens*.¹⁶⁰ There, the Supreme Court was confronted with the issue of whether videos of animal torture should receive First Amendment protections.¹⁶¹ In a departure from its previous cases, the Court rejected the government’s contention that depictions of animal cruelty should be treated as unprotected speech because the images’ harms

151. *Id.* at 109–10.

152. *Id.* (quoting *Ferber*, 458 U.S. at 756–58).

153. *Hotaling*, 634 F.3d at 729–30.

154. *See id.* at 730.

155. *Id.*

156. *See id.*

157. *Mecham*, 950 F.3d at 265–67.

158. *Id.* at 260.

159. *Id.* at 263.

160. *Id.* at 263–64; *United States v. Stevens*, 559 U.S. 460 (2010).

161. *Stevens*, 559 U.S. at 464.

outweigh the expressive value.¹⁶² When classifying categories of unprotected speech, the Court explained that it does not rely on this “startling and dangerous . . . ad hoc balancing of relative social costs and benefits,” but looks to “historic and traditional categories long familiar to the bar.”¹⁶³ In an attempt to fit *Ferber* into this new doctrinal scheme, the Court recast it as a decision grounded in a “previously recognized, long-established category of unprotected speech”: speech integral to criminal conduct.¹⁶⁴ Accordingly, the analysis did not rest on the “balance of competing interests alone,” but rather child pornography’s relationship to the underlying crime of sexual abuse, a historically-accepted rationale for exempting speech from the First Amendment.¹⁶⁵ By this reasoning, Mecham argued, *Ferber* does not extend to morphed child pornography because it does not involve actual sexual abuse.¹⁶⁶

Although acknowledging that Mecham’s logic was adopted by the Eighth Circuit in *United States v. Anderson*, the Fifth Circuit rejected it.¹⁶⁷ In the Fifth Circuit’s view, a “one-paragraph discussion of child pornography” in *Stevens*, a case dealing with the unrelated topic of animal abuse videos, cannot justify abandoning longstanding precedent.¹⁶⁸ This is particularly so, the Fifth Circuit continued, because *Stevens* “makes no mention of the interest in preventing reputational or emotional harm to children” addressed in *Ferber* and *Ashcroft*.¹⁶⁹ The Fifth Circuit further noted that while *Stevens* classified *Ferber* as a “special case” due to child pornography’s relationship to sexual abuse, it never declared that this was the only rationale.¹⁷⁰ Finally, requiring the display of sexual abuse would limit the reach of regular child pornography laws, which apply to the exhibition of nudity that may “stop short of depicting child abuse.”¹⁷¹ Having determined that the absence of underlying abuse does not bar prosecution, the Fifth Circuit concluded that the state’s interest in

162. *Id.* at 469–70.

163. *Id.* at 468, 470 (quoting *Simon & Schulster, Inc. v. Members of N.Y. State Crime Victims Bd.*, 502 U.S. 105, 127 (1991)).

164. *Id.* at 471 (first citing *Ferber*, 458 U.S. at 761–62; then citing *Osborne*, 495 U.S. at 110).

165. *Stevens*, 559 U.S. at 470 (quoting *Ferber*, 458 U.S. at 763).

166. See *Mecham*, 950 F.3d at 263.

167. *Id.* at 265–66.

168. *Id.* at 265.

169. *Id.* at 266.

170. *Id.*

171. *Mecham*, 950 F.3d at 266.

protecting the emotional wellbeing of the minor is sufficient to justify suppression of morphed pornography.¹⁷²

3. Sixth Circuit

The Sixth Circuit came to the same conclusion in *Doe v. Boland*.¹⁷³ There, defendant edited stock images of children to make it appear as though they were engaging in sexual conduct.¹⁷⁴ In finding that the images fall outside the First Amendment, the Sixth Circuit first focused on the “evils” inflicted by morphed child pornography.¹⁷⁵ Specifically, the court explained that morphed pornography violates “the individual interest in avoiding disclosure of personal matters” and creates “many of the same reputational, emotional, and privacy interests as actual pornography.”¹⁷⁶ In the court’s view, these harms suffered by identifiable children serve as a basis for distinguishing *Ashcroft*.¹⁷⁷ Unlike computer-generated pornography that does not depict actual children, morphed images create “real victims with real injuries,” and thus “bear a closer similarity to actual pornography.”¹⁷⁸

In addition to the harms suffered by victims of morphed pornography, the Second Circuit also pointed to the images’ “relatively weak expressive value.”¹⁷⁹ Given *Ferber*’s recognition that the speech value of child pornography is “exceedingly modest, if not *de minimis*,” morphed child pornography, which depicts identical subject matter, must similarly lack First Amendment value.¹⁸⁰ Importantly, the Sixth Circuit did not find that the sophisticated technology often used to make morphed pornography added anything to its expressive value.¹⁸¹

C. Applying the Morphed Pornography Cases to Deepfake Pornography

Taken together, these decisions provide a strong basis for classifying deepfake pornography as an additional category of unprotected speech. Indeed, in extending *Ferber* to morphed child

172. *Id.* at 267.

173. *Doe*, 698 F.3d at 877.

174. *Id.* at 879.

175. *Id.* at 883.

176. *Id.* at 880–81.

177. *See id.*

178. *Doe*, 698 F.3d at 883.

179. *Id.*

180. *Id.* (quoting *Ferber*, 458 U.S. at 762–63).

181. *See id.* at 883–84.

pornography, the circuit courts elucidated four principles that provide a roadmap for exempting deepfake pornography from the First Amendment.

First, the animating justification for treating morphed pornography as unprotected speech is the prevention of emotional and reputational harm. Although morphed pornography does not display children's actual nude bodies, children are nonetheless harmed by knowing their images are being exploited and viewed by pedophiles. According to the Second, Fifth, and Sixth Circuits, the government has a compelling interest in preventing this harm.

In the context of deepfake pornography, the need to protect victims from harm is equally as strong. Deepfakes, like morphed child pornography, use the likeness of real people, superimposing their faces onto pre-existing pornographic images. As explained *supra*, this not only creates profound sexual privacy invasions, but can lead to an array of psychological and emotional issues similar to those from morphed child pornography. When victims see themselves falsely depicted in a pornographic image, they experience intense shame, and many develop anxiety and depression. Moreover, like morphed pornography, these harms are exacerbated by the circulation of the material. Once the image makes its way online, the victim's reputation is often permanently damaged, causing financial and social repercussions. To make matters worse, the images are often difficult to remove from the distributional chain, creating a sort of "permanent record" of the depiction that retraumatizes the victim again and again.

These harms, which are severe and long-lasting, can also be distinguished from the speculative, remote threats the court rejected in *Ashcroft*. Unlike the government's argument that virtual child pornography should be banned because it "whets the appetites" of pedophiles, the harm caused by deepfake pornography does not depend "upon some unquantified potential for subsequent criminal acts."¹⁸² Rather, by displaying the sexual identity of victims without consent, deepfakes directly cause dignitary and reputational repercussions. Thus, the government would not merely be "controlling a person's private thoughts," as the defendant in *Ashcroft* argued,¹⁸³ but would be preventing real harm to real people.

To be sure, adult deepfake pornography differs from morphed child pornography in that it does not implicate the interests of minors,

182. See *Ashcroft*, 535 U.S. at 250, 253.

183. *Id.* at 253 (quoting *Stanley*, 394 U.S. at 566).

the protection of whom *Ferber* found “compelling.”¹⁸⁴ While the vulnerability of children certainly strengthens the government’s interest in suppressing child pornography, it does not foreclose extending *Ferber* to other areas of nonconsensual pornography where victims suffer similar harms. Further, since morphed pornography can be made using “lower tech means”¹⁸⁵—such as gluing a child’s face onto a pornographic image—there are arguably instances in which adult deepfake pornography has even greater potential to cause reputational harm. Indeed, some deepfake images are so realistic that viewers struggle to discern whether they are fabricated. Consequently, these images are more likely to fool the public than a rudimentary morphed photograph.

Second, while some argue the absence of sexual abuse in deepfake pornography fails to satisfy a “fundamental concern” set out in *Ferber*,¹⁸⁶ the morphed pornography cases make clear that the display of underlying abuse is not necessary. Emphasizing the need to protect victims of child pornography from its downstream repercussions, the circuit courts recognize that *Ferber*’s analysis does not turn on production-related harm alone.¹⁸⁷ Rather, in light of *Ashcroft*, the harm need only be inflicted on an actual child for *Ferber* to apply; whether the harm flowed from creating the image or its aftermath is irrelevant. In support of this argument, the Second and Sixth Circuits also pointed to *Osborne*’s concern that the image’s “continued existence” will cause the victim “continuing harm by haunting the [child] in years to come.”¹⁸⁸ The focus on the threats arising from the image’s circulation shows that the Court was not only concerned with the harms perpetrated by underlying abuse. Finally, while *Stevens* declared that child pornography is only unprotected due to its relationship to sexual abuse, the Fifth Circuit correctly characterized this assertion as *dicta*, carrying little weight for future decisions concerning pornography.¹⁸⁹ In rejecting *Stevens*’ application, the Fifth Circuit also noted that the child pornography statute upheld in *Ferber* extended to nude images of children that do

184. *Ferber*, 458 U.S. at 757 (quoting *Globe Newspaper Co. v. Superior Court*, 457 U.S. 596, 607 (1982)).

185. *Id.* at 242.

186. Waldstreicher, *supra* note 102, at 757–59 (arguing that the *Ferber* exception does not apply because deepfake victims are not sexually abused).

187. *See id.* at 752–53.

188. *Osborne*, 495 U.S. at 111.

189. *See Mecham*, 950 F.3d at 265.

not necessarily involve the display sexual contact.¹⁹⁰ Accordingly, classifying deepfake pornography as unprotected speech is not precluded due to the absence of an underlying sexual abuse crime.

Third, in addition to the interest in preventing dignitary harm, the low speech value of morphed pornography can justify its exclusion from the First Amendment. In *Ferber*, the Court explained that the speech value of child pornography is “exceedingly modest, if not *de minimis*” because it is unlikely to be an “important and necessary part of a literary performance or scientific or educational work.”¹⁹¹ Echoing this conclusion, the Sixth Circuit determined that morphed pornography has “relatively weak expressive value” because “morphed images are never necessary to achieve an artistic goal.”¹⁹² In so reasoning, the court drew a distinction with virtual child pornography, which often overlaps with sexual depictions holding literary, scientific, and artistic value, such as a pictorial in a physiology textbook or a portrayal of “star-crossed teenage lovers.”¹⁹³ By contrast, morphed child pornography would rarely capture these traditionally expressive works.

Nonconsensual deepfakes similarly lack artistic or literary value. Like morphed child pornography, it seems unlikely that inserting someone into a crude sexual depiction is necessary for an artistic or educational aim. In the off chance that it is, the creator could avoid harm to the subject by attempting to gain their consent.¹⁹⁴ At the same time, proponents of deepfakes may argue that the First Amendment value in deepfake pornography lies in the underlying technology. In fact, machine learning technology has been used for a variety of expressive and educational purposes, such as creating virtual museums¹⁹⁵ and delivering an anti-malaria message in nine

190. *Id.* at 264 (citing *Ferber*, 458 U.S. at 765).

191. *Ferber*, 458 U.S. at 762–63.

192. *Doe*, 698 F.3d at 883–84.

193. *Id.* at 884 (citing *Ashcroft*, 535 U.S. at 247).

194. In *Ferber*, the Court suggested that insofar as the creator could avoid harm to the child, a depiction of underage sexual conduct is permissible. *See Ferber*, 458 U.S. at 763 (“[I]f it were necessary for literary or artistic value, a person over the statutory age who perhaps looked younger could be utilized. Simulation outside of the prohibition of the statute could provide another alternative.”).

195. *Yes, Positive Deepfake Examples Exist*, THINK AUTOMATION, <https://www.thinkautomation.com/bots-and-ai/yes-positive-deepfake-examples-exist/> (last visited Dec. 28, 2022).

2023] **Deeply Dehumanizing, Degrading, and Violating** 361

languages.¹⁹⁶ That said, morphed child pornography often relies on similarly sophisticated technology, yet the lower courts have not found that the technology itself adds to the expressive value of what it depicts.

Fourth, and finally, when extending *Ferber* to new areas of exploitative pornography, the Court is not limited to those “historic and traditional” categories of speech “long familiar to the bar.”¹⁹⁷ According to some commentators, *Stevens* restricted the Court’s ability to identify new classes of unprotected speech by insisting that it must point to a “long-settled tradition of subjecting that speech to regulation.”¹⁹⁸ While it can hardly be said that there is a well-established tradition of proscribing child pornography made by computer morphing, the circuit courts rejected *Stevens*’ applicability to pornography cases.¹⁹⁹ As the Fifth Circuit explained, because *Stevens* concerns the unrelated topic of animal abuse imagery, the decision does not carry enough weight to undermine the rationales set out in *Ferber*.²⁰⁰ Moreover, the lower courts rejected *Stevens*’ attempt to reclassify child pornography under a traditional First Amendment category, speech integral to criminal conduct.²⁰¹ To fit this category of speech, pornography must depict the underlying crime of sexual abuse.²⁰² However, as the Fifth Circuit noted in *Mecham*, the federal definition of child pornography does not require the depiction of sexual abuse, but also encompasses nude images of minors.²⁰³ In the

196. *What, How, and Why of Deepfake Technology Explained*, DEEPWORD, <https://www.deepword.co/deepfake-technology-explained/> (last visited Dec. 28, 2022).

197. *See Stevens*, 559 U.S. at 468 (quoting *Simon & Schuster, Inc. v. Members of the N.Y. State Crime Victims Bd.*, 502 U.S. 105, 127 (1991) (Kennedy, J., concurring in the judgment)).

198. *See, e.g., Nadine Strossen, United States v. Stevens: Restricting Two Major Rationales for Content-Based Speech Restrictions*, 2010 CATO SUP. CT. REV. 67, 84 (2009) (arguing that *Stevens* “limit[ed] the categories of unprotected expression to the finite set that it has historically recognized”); *see also* John D. Moore, *The Closed and Shrinking Frontier of Unprotected Speech*, 36 WHITTIER L. REV. 1, 18, 48 (2014) (positing that in light of *Stevens*, the Court has “abandoned the balancing approach to evaluating low-value speech in favor of a historical-categorical analysis” which has “shrunk the realm of categorically unprotected speech”).

199. *See, e.g., Mecham*, 950 F.3d at 267.

200. *See id.* at 265.

201. *See id.* at 266–67; *see also* *United States v. Price*, 775 F.3d 828, 838 (7th Cir. 2014) (rejecting a reading of *Stevens* that would protect images of child pornography unless they depict sexual abuse).

202. *See Mecham*, 950 F.3d at 264.

203. *See id.* at 267 (explaining that the “federal definition of *real* child pornography is not limited to images that depict sexual abuse of a minor”).

court's view, narrowing the definition of child pornography to align with a historically-based content category would be a "significant doctrinal development, and likely not hidden in a case about crush videos."²⁰⁴

Thus, while it would certainly be difficult to ground a proscription of deepfake pornography in history and tradition, the Supreme Court's analysis need not be so confined. By rejecting *Stevens*' historical treatment test, the Supreme Court would be free to use *Ferber*'s rationales to recognize deepfake pornography as an additional category of unprotected speech.

Admittedly, persuading the Court to classify a novel technology like deepfakes as a new content category would not be an easy task. At the same time, some members of the Court have recognized the need to expand content classifications in light of new technology.²⁰⁵ Writing separately in *Ashcroft*, both Justice Thomas and Justice O'Connor forewarned that technological advancements may require the Court to extend *Ferber* to virtual child pornography.²⁰⁶ In a concurring opinion, Justice Thomas raised the concern that "technology may evolve to the point where it becomes impossible to enforce actual child pornography laws because the Government cannot prove that certain pornographic images are of real children."²⁰⁷ Where this occurs, Thomas explained, a "narrowly drawn restriction" on virtual pornography would be permissible.²⁰⁸ Justice O'Connor echoed the same sentiment in a dissent joined by Chief Justice Rehnquist and Justice Scalia. She wrote:

Of . . . concern is the prospect that defendants indicted for the production, distribution, or possession of actual child pornography may evade liability by claiming that the images attributed to them are in fact computer-generated Given the rapid pace of advancements in computer-graphics technology, the Government's concern is reasonable.²⁰⁹

Broadly speaking, these points raised by the Justices suggest a willingness to adapt content classifications in response to shifts in technology.

204. *Id.* at 266 (quoting *Price*, 775 F.3d at 839).

205. See *Ashcroft*, 535 U.S. at 259 (Thomas, J., concurring in the judgment); *id.* at 264 (O'Connor, J., concurring in the judgment in part and dissenting in part).

206. *Id.* at 259, 264.

207. *Id.* at 259 (Thomas, J., concurring).

208. *Id.*

209. *Ashcroft*, 535 U.S. at 263–64 (O'Connor, J., dissenting).

Beyond child pornography, Justice Gorsuch expressed similar sentiments in the context of defamation. In a dissent from the denial of certiorari, Gorsuch argued that defamation law should be expanded due to “technological changes” in the media landscape.²¹⁰ According to Gorsuch, the rise of smartphones and social media are undermining the traditional justifications for protecting false speech directed at public officials.²¹¹ Where it was once necessary to protect these statements to ensure a healthy public debate, now, “in a world in which everyone carries a soapbox in their hands,” the force of this rationale is “less obvious.”²¹² Although couched in terms of defamation, this observation suggests that the Justice may be open to rethinking First Amendment principles in light of technological advancements.

IV. LEGISLATIVE RECOMMENDATIONS

Given the strong arguments for creating a new content category for deepfake pornography, Congress is presented with an opportunity to legislate. By relying on a categorical exemption to First Amendment protections, lawmakers can avoid the challenges of narrow tailoring, and better ensure deepfake pornography legislation stands up in court.

Despite this wider latitude to legislate, Congress must still take measures to avoid definitional ambiguities that would render the law overbroad. To do so, Congress should model the language of the legislation after the existing morphed child pornography statute contained in 18 U.S.C. §§ 2252A(a) and 2256(8)(C), which has been upheld multiple times on appeal. In the relevant part, § 2252A(a) prohibits “knowingly” receiving, distributing, advertising, reproducing, or possessing child pornography.²¹³ Section 2256(8)(C) defines child pornography to include morphed child pornography, which is any “visual depiction [that] has been created, adapted, or modified to appear that an identifiable minor is engaging in sexually explicit conduct.”²¹⁴

This language is instructive in a few respects. First, the *mens rea* from § 2252A(a) does not require an intent to harm.²¹⁵ Proving a

210. *Berisha v. Lawson*, 141 S. Ct. 2424, 2427–28 (2021) (Gorsuch, J., dissenting from denial of certiorari).

211. *See id.* at 2427–28.

212. *Id.* at 2427.

213. 18 U.S.C. § 2252A(a) (2022).

214. § 2256(8)(C).

215. § 2252A(a).

specific aim to harm or harass the victim is not only exceedingly difficult, but has no bearing on whether the harms from *Ferber* exist. Rather, by adopting the *mens rea* from § 2252A(a), the state need only show that the defendant knew the material he possessed or distributed was a nonconsensual deepfake. This state of mind strikes the appropriate balance between ensuring effective prosecution and limiting coverage to the most culpable. Additionally, the statute correctly requires that the subject in the image be identifiable. Like the victims of morphed pornography, victims of deepfake pornography are harmed insofar as they can be identified in the image. Without this qualifying language, the deepfake would be considered “virtual” pornography, which is protected speech under *Ashcroft*.

Another provision to adopt from the morphed pornography statute is the ban on possession. Applying *Osborne*, the Second Circuit determined that possession of morphed child pornography could be criminalized because it is nearly impossible to attack the problem of child pornography at the distribution level.²¹⁶ Indeed, most of the market for child pornography has been driven underground, where the images are exchanged on dark websites, not storefronts.²¹⁷ Likewise, as major online platforms are beginning to crack down on deepfake pornography,²¹⁸ deepfake proprietors are fleeing to the far corners of the internet, reducing the prospect of effective prosecution.²¹⁹ To better protect the public from the evils inflicted by deepfake porn, the government would thus be justified in attempting to root out this pernicious content “at all levels of the distributional chain.”²²⁰ Moreover, proscribing possession of deepfake pornography can also be distinguished from proscribing the possession of obscenity, which is unconstitutional under *Stanley*.²²¹ When the government prohibits private possession of obscene materials, it aims to “control[] a person’s private thoughts,” an illegitimate interest in the eyes of the Court.²²² In contrast, proscribing possession of deepfake pornography

216. *Hotaling*, 634 F.3d at 730.

217. *See id.*

218. Many platforms, including Twitter, Facebook, Reddit, Pornhub, and Tumblr, have instituted some sort of deepfake policy. *See* Gieseke, *supra* note 70, at 1485, 1506. These policies, however, have failed to redress the issue of deepfake pornography more broadly. *See id.*

219. *See* Tim Starks, *Deepfakes Advertised on Underground Markets, Signaling Possible Shift, Recorded Future Says*, CYBERSCOOP (April 29, 2021), <https://www.cyberscoop.com/deepfakes-doctored-video-audio-future/>.

220. *Osborne*, 495 U.S. at 110.

221. *Id.* at 109.

222. *Stanley*, 394 U.S. at 566.

2023] **Deeply Dehumanizing, Degrading, and Violating** 365

is not based on this “paternalistic interest,” but rather the compelling need to protect victims from reputational and emotional harm.²²³

While the morphed pornography statute is helpful in some respects, it must be clarified and refined in others.²²⁴ For one, the language from § 2256(8)(C), which defines morphed pornography, is too sweeping. By prohibiting visual depictions that are “created, adapted, or modified” to appear as though one is engaged in sexual activity, the statute would cover images not necessarily created by deepfake technology, but also more rudimentary means.²²⁵ This language should, therefore, be tightened to capture deepfakes specifically.²²⁶

Further, since adults, unlike children, have the capacity to consent, the law should exclude deepfakes that are created consensually. When the subject of the deepfake consents to its creation, the underlying harms identified by *Ferber* are not present. For example, individuals with certain disabilities may choose to superimpose “their faces and that of consenting partners” onto pornographic videos, “enabling virtual engagement with an aspect of life unavailable to them in a conventional sense.”²²⁷ Here, because the sexual privacy invasion is absent, the resulting harms justifying the First Amendment exemption do not exist. Nonetheless, if one of the partners were to distribute the video without the other’s consent, the law would apply.

CONCLUSION

Deepfake pornography presents a unique and disturbing threat to sexual privacy in the digital age. Although its initial targets were

223. *Osborne*, 495 U.S. at 109 (similarly distinguishing the state’s interest in protecting victims of child pornography).

224. Perhaps obviously, “identifiable minors” should be changed to “persons.” See 18 U.S.C. § 2256 (2022).

225. 18 U.S.C. § 2256(8)(C) (2022).

226. When defining deepfake pornography, it may be helpful to look to the definitions contained in state law. See CAL. CIV. CODE § 1708.86(a)(4) (West 2022) (defining “depicted individual” as “an individual who appears, as a result of digitization, to be giving a performance they did not actually perform or to be performing in an altered depiction.”); see also N.Y. CIV. RIGHTS LAW § 50-f(c) (McKinney 2022) (defining “digital replica” to mean a “newly created, original, computer-generated, electronic performance by an individual in a separate and newly created, original expressive sound recording or audiovisual work in which the individual did not actually perform, that is so realistic that a reasonable observer would believe it is a performance by the individual being portrayed and no other individual.”).

227. Chesney & Citron, *supra* note 3, at 1771.

celebrities, the rapid advance of this technology exposes any person with an online presence to the specter of sexual impersonation. Given the inadequacy of the current legal regime in addressing this emerging threat, action by Congress is required. As this Note demonstrates, Congress can ensure that legislation prohibiting deepfake pornography passes Constitutional muster by arguing that such speech, like morphed child pornography, is categorically exempt from the First Amendment. By regulating deepfakes under the purview of *Ferber*, Congress can avoid infringing on free speech rights and better ensure that the legislation stands up in court. But if no action is taken, virtually nothing is stopping deepfake creators from continuing to exploit, manipulate, and deceive.