DEFAMATION IN THE TIME OF DEEPFAKES

ABIGAIL GEORGE*

*Abstract***

Deepfake technology, powered by artificial intelligence, has enabled the quick and easy creation of hyperrealistic videos that superimpose one person's face onto another's body. While the technology has benign applications, it has also been overwhelmingly used to create nonconsensual pornography. Deepfake pornography is a severe sexual offense that has targeted hundreds of thousands of women. This Note, the first comprehensive analysis of deepfake pornography under defamation law, sketches a framework for advocates and judges to apply defamation to cases of deepfake pornography.

This Note argues that deepfakes—in achieving photorealism and simulating someone's true body and private life—qualify as defamatory false statements of fact. As this Note shows, when alleged defamatory statements strive for (and achieve) hyperrealism, and they purport to reveal a truth about someone's private sex life, they qualify as false statements of fact. Cursory indications that a deepfake is "fake" or even viewers' knowledge that it is "synthetic" refer solely to the manner of creation, not its signified meaning. The photovisual realism of deepfakes collapses the distinction between form and meaning or signified and signifier. As signifiers whose forms perfectly resemble their signified, deepfakes leave no room for the person depicted to disavow their message or for the statements to transform into a parody or commentary protected by the First Amendment. Thus, the knowledge that a deepfake is fake does little to undermine the reputational harm and, consequently, the defamation claim. Finally, this Note addresses defamation law's peculiar and controversial "actual malice" scienter requirement. As actual malice relates to knowledge or reckless disregard for the falsity of the statement and not a defamatory intent, it applies to creator-

distributors who use synthetic processes to make deepfakes, albeit often claiming a benign or parodic purpose.

INTRODUCTION

In April 2018, Rana Ayyub—a well-known investigative journalist from India—received a seemingly innocuous message that a video of her was circulating online.[1] Ayyub had recently made headlines for campaigning for justice following the rape and murder of an eight-year-old Kashmiri girl, so she wondered if the clip was from her recent interviews on BBC News and Al Jazeera in which she condemned the Bharatiya Janata Party's (BJP) support of the accused.[2] She hoped that the video would not exacerbate the backlash she was experiencing.[3] When she clicked on the link, she was shocked.[4] "What he sent me was a porn video, and the woman in it was me."[5]

The video was a nonconsensual deepfake pornography (NCDP) video created using artificial intelligence ("AI") and publicly available photos of Ayyub.[6] After appearing on the internet, the video of Ayyub was shared tens of thousands of times via WhatsApp and Twitter.[7] "I started throwing up. I just didn't know what to do. In a country like India, I knew this was a big deal. I didn't know how to react, I just started crying."[8]

---

1   *See* 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/DCS8-A86K].

2   *See id.*

3   *See id.*

4   *See id.*

5   *Id.*

6   *See id.*

7   Ayyub, *supra* note 1; *see also* Lilian Stolk, *If Deepfakes Are a Threat, This Is It: A Feminist Perspective on the Impact of Deepfake Pornography*, The Hmm (June 5, 2020), https://thehmm.nl/if-deepfakes-are-a-threat-this-is-it/ [https://perma.cc/Z2GS-4NDP] ("Via WhatsApp, the video ended up on almost every phone in India.").

8   Ayyub, *supra* note 1.

As the video continued to circulate, Ayyub faced a wave of online and physical harassment and familial, personal, health, and professional consequences.[9] She was inundated with harassing comments on social media.[10] She was called "Jihadi Jane" and "Isis Sex Slave."[11] After receiving messages asking her rates for sex, Ayyub was hospitalized due to anxiety and heart palpitations.[12] Ayyub always considered herself an outspoken feminist but was silenced due to the abuse: "From the day the video was published, I have not been the same person . . . I've self-censored quite a bit out of necessity."[13]

Unfortunately, Ayyub's story is not uncommon. Deepfake pornography,[14] also known as sexual deepfakes,[15] refers to synthetically created, sexually explicit images or videos of people that are produced without their consent.[16] Deepfake pornography has targeted over 105,000 women, and nearly half of U.S. high school students have heard of deepfakes depicting classmates.[17] Women affected by deepfake pornography have dropped out of

---

9   *See id.*

10   *See id.*

11   *Id.*

12   *Id.* ("I used to be very opinionated, now I'm much more cautious about what I post online.").

13   *Id.* ("I always thought no one could harm me or intimidate me, but this incident really affected me in a way that I would never have anticipated.").

14   "Deepfake" is a portmanteau of "deep" and "fake," referring to the fact that it uses "deep learning," a subset of machine learning that relies on artificial neural networks and is manipulated content. *See* GRAHAM MEIKLE, DEEPFAKES 2 (2022). Deepfakes owe their name to the Reddit user u/xual who created the anonymous Reddit forum r/Deepfakes in 2017 to create and share deepfake pornography. *Id.* at 3.

15   *See* Victoria Rousay, Sexual Deepfakes and Image-Based Sexual Abuse: Victim-Survivor Experiences and Embodied Harm, 12–13 (May 2023) (A.L.M. thesis, Harvard University) (arguing that the term "sexual deepfakes" better captures their violation, abuse, and lack of consent).

16   I am adapting Danielle Citron and Mary Anne Franks' definition of nonconsensual pornography, which they define as "the distribution of sexually graphic images of individuals without their consent." Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345, 346 (2014).

17   *See* MEIKLE, *supra* note 14, at 75 (noting that the app DeepNude had over 500,000 download requests, targeting 104,852 women); Elizabeth Laird, Maddy Dwyer & Kristin Woelfel, *In Deep Trouble: Surfacing Tech-Powered Sexual Harassment in K–12 Schools*, CTR. FOR DEMOCRACY & TECH., 11 (Sept. 2024), https://cdt.org/wp-content/uploads/2024/09/2024-09-26-final-Civic-Tech-Fall-Polling-research-1.pdf [https://perma.cc/EN5E-B5G9]; *see also infra* notes 64–65 and accompanying text.

school,[18] left their jobs,[19] and implemented "permanent" and "significant" changes in their behavior, attitudes, and relationships.[20]

Deepfake pornography is not limited to an obscure corner of the internet. Since first appearing on Reddit in 2017, [21] the accuracy and accessibility of deepfake pornography have each grown at an exponential rate with no signs of slowing. At least 244,625 videos of deepfake pornography circulate online,[22] and the top six deepfake-pornography websites receive over thirty-one million visits per month.[23] Deepfake technology has evolved from requiring hundreds of facial images and advanced computer skills to needing just one image, twenty-five minutes, and basic computer skills.[24] Sexual deepfake images can even be made on user-friendly iPhone apps.[25] In a few years, it will be possible to make realistic

---

18   *See* Noelle Martin, *Image-Based Sexual Abuse and Deepfakes: A Survivor Turned Activist's Perspective*, The Palgrave Handbook of Gendered Violence and Technology 55, 60 (Anastasia Powell, Asher Flynn & Lisa Sugiura eds., 2021).

19   *See* Rousay, *supra* note 15, at 107 ("Unfortunately, for some victim-survivors like Lia, despite the content being fake, she was still fired from her job after her employer saw the video.").

20   *Id.* at 110 (discussing a "state of permanency" in which survivors remain hindered by fear of being recognized or revictimized); *see also* Jayna Nickert, *The Damage Caused by Deepfake Porn*, Healthnews (Nov. 16, 2023), https://healthnews.com/mental-health/anxiety-depression/the-damage-caused-by-deepfake-porn/ [https://perma.cc/36JY-BQEQ] (noting that deepfake pornography leads to anxiety, panic, depression, posttraumatic stress disorder, and dissociation).

21   *See* Meikle *supra* note 14.

22   Matt Burgess, *Deepfake Porn Is Out of Control*, Wired (Oct. 16, 2023), https://www.wired.com/story/deepfake-porn-is-out-of-control/ [https://perma.cc/2EUK-4LVV] (noting that over 113,000 such videos were uploaded in the first nine months of 2023 and more were produced in 2023 than in all other years combined).

23   [John Doe Website Analytics] (data on file with *Columbia Journal of Gender & Law*); *see also* *2023 State of Deepfakes: Realities, Threats, and Impact*, Sec. Hero (2023), https://www.securityhero.io/state-of-deepfakes/#key-findings [https://perma.cc/52V4-NXSZ] (finding that the ten most popular deepfake pornography platforms had over 300 million combined views in 2023).

24   *See 2023 State of Deepfakes*, *supra* note 23 (reporting that it takes less than twenty-five minutes and is free to create a sixty-second deepfake video with just one face image); *see also* Tate Ryan-Mosley, *A High School's Deepfake Porn Scandal Is Pushing US Lawmakers Into Action*, MIT Tech. Rev. (Dec. 1, 2023), https://www.technologyreview.com/2023/12/01/1084164/deepfake-porn-scandal-pushing-us-lawmakers/ [https://perma.cc/NFF3-HNZT] ("Creating a convincing deepfake five years ago required hundreds of images . . . which meant those at greatest risk . . . were celebrities and famous people with lots of publicly accessible photos [but] now, deepfakes can be created with just one image.").

25   For example, the app DeepNude—advertising itself as "the superpower you always wanted"—allowed users to upload photos of any woman for the app to generate naked images of. Meikle, *supra* note 14, at 73.

three-dimensional sex avatars of any person without their consent.[26] Accordingly, while deepfakes initially primarily targeted women in the public sphere,[27] private individuals are now the main target,[28] with the number of deepfake videos increasing 900% each year.[29]

Deepfakes are experiencing a meteoric rise as a form of gender-based violence.[30] Yet, they have largely escaped political, legal, and public scrutiny for at least three reasons. First, deepfake pornography causes a "silencing effect," wherein victims remove themselves from online spaces and do not risk speaking out publicly for fear of provoking retaliation and driving more viewers to their intimate content.[31] Moreover, with few exceptions, the inability of mainstream media and politicians to see deepfakes as gender-based violence is astonishing.[32] Instead, they focus attention on deepfakes as political or security threats,[33]

---

26   *See* Danielle Citron, The Fight for Privacy: Protecting Dignity, Identity and Love in the Digital Age 48 (2022) [hereinafter Citron, Fight for Privacy].

27   In 2019, a report showed that ninety-nine percent of subjects in deepfake pornography were actresses or musicians working in the entertainment sector. *See* Henry Ajder et al., Deeptrace, The State of Deepfakes: Landscape, Threats, and Impact 2 (2019), https://regmedia.co.uk/2019/10/08/deepfake_report.pdf [https://perma.cc/EN2W-75QQ]. Deepfake pornography has targeted almost every woman in the public eye. There are over 1,000 deepfake videos of Emma Watson that collectively have over fifteen million views, rendering it "essentially a separate porn genre on its own." Meikle, *supra* note 14, at 61.

28   *See* Citron, Fight for Privacy, *supra* note 26, at 48 (noting that sixty-three percent of users of a deepfake chatbot uploaded photos of girls or women they knew personally); *see also* Meikle, *supra* note 14, at 72 (referring to the shift toward targeting private individuals as the "domesticat[ion]" of synthetic porn); Stolk, *supra* note 7, at 7–9 (noting that creators accept requests for deepfakes of specific people and frequently receive requests for deepfakes of ex-girlfriends).

29   Matthew Miller, *Deepfakes: Real Threat*, KPMG, 3 (2023), https://kpmg.com/kpmg-us/content/dam/kpmg/pdf/2023/deepfakes-real-threat.pdf [https://perma.cc/BCU4-WP77].

30   *See supra* notes 21–23 and accompanying text.

31   *See* My Image My Choice, https://myimagemychoice.org/ [https://perma.cc/6G2B-WJP5] (describing the "silencing effect" causing victims to "modify their behavior, retreat from online spaces, and [be] shut out from full participation in public discourse – especially online . . . Most people don't want to risk speaking out about their experiences because this might provoke retaliation, or drive more viewers to their intimate content."); Rousay, *supra* note 15, at 110 (performing a qualitative data analysis of fifty-eight survivors of deepfakes to find that all participants "remove[d] or self-censor[ed] their online presence").

32   *See infra* notes 118–136 and accompanying text.

33   *See, e.g.*, NSA, FBI & Cybersecurity & Infrastructure Sec. Agency, Cybersecurity Information Sheet: Contextualizing Deepfake Threats to Organizations 1 (2023), https://media.defense.gov/2023/Sep/12/2003298925/-1/-1/0/CSI-DEEPFAKE-THREATS.PDF [https://perma.cc/EQ9M-C7XC] (listing the ways deepfakes can be "abused" as threatening brands, impersonating leaders, and gaining access to sensitive

eschewing the fact that ninety-eight percent of deepfakes are pornographic.[34] Of those images, ninety-nine percent depict women.[35] To put it simply, deepfakes are—and always have been—a gender-based violence issue.[36] Finally, existing legal mechanisms have largely left the women targeted without recourse.[37] Women who seek legal recourse are often told by law enforcement that the perpetrator did not break any laws.[38] Aside from a few piecemeal convictions and creative lawsuits, there have been no truly effective ways for victims to protect themselves.[39]

The sexualization of women in the public sphere and threats of gender-based violence are hardly new. But nonconsensual deepfake pornography presents a novel and urgent threat because technological advances have rendered the videos effectively indistinguishable

---

information); Dan Robitzski, *Pentagon's AI Director Calls for Stronger Deepfake Protections*, FUTURISM: THE BYTE (Aug. 30, 2019), https://futurism.com/the-byte/pentagon-ai-director-deepfake-protections [https://perma.cc/2KMZ-EVU6] (noting that the Pentagon is allocating vast financial resources to combating this challenge to national security).

34   *See 2023 State of Deepfakes*, *supra* note 23.

35   *See id.* While deepfake pornography primarily affects women, "it won't be long" until homosexual deepfake pornography of men is used to "cost someone their life or liberty in certain parts of the world." NINA SCHICK, DEEPFAKES: THE COMING INFOCALYPSE 159–60 (2020). The most popular deepfake website recently released a "gay version." [JOHN DOE WEBSITE #1] (link on file with *Columbia Journal of Gender & Law*).

36   *See, e.g.*, Citron & Franks, *supra* note 16, at 353 (arguing that image-based sexual abuse is also a form of sex discrimination); Rousay, *supra* note 15 ("Findings from this study suggest that sexual deepfake abuse is a severely gendered phenomenon in which heteronormativity has become the template for enacting sexual violence."); Stolk, *supra* note 7 (arguing that the real danger of deepfakes is to gender equality, not "the truth").

37   *See, e.g.*, Anne Pechenik Gieseke, Note, *"The New Weapon of Choice": Law's Current Inability to Properly Address Deepfake Pornography*, 73 VAND. L. REV. 1479 (2020); Ryan-Mosley, *supra* note 24 ("[T]he dearth of regulation and legal precedent on deepfake pornography means that victims . . . have little to no recourse."); *see also infra* Section I.B.1 (explaining why revenge porn statutes do not apply to victims of NCDP since deepfakes do not expose the woman's "real" body).

38   *See, e.g.*, Ayyub, *supra* note 1; Martin, *supra* note 18, at 58 (explaining that she called the police, went to the police station, and hired a private investigator, but "there was nothing they could do, or did do").

39   *See* Micah Kindred, *Deepfakes: The Effect on Women and Potential Protections*, U. CIN. L. REV. (Aug. 2, 2023), https://uclawreview.org/2023/08/02/deepfakes-the-effect-on-women-and-potential-protections/ [https://perma.cc/QC97-ZPLY]; *see also* Markus Scheiber, *San Francisco Files First-of-its-Kind Lawsuit to Tackle AI Deepfake Nudes*, POLITICO (Aug. 17, 2024), https://www.politico.com/news/2024/08/17/san-francisco-lawsuit-ai-deepfake-nudes-00174487 [https://perma.cc/58GP-V45C] (noting that the 2024 lawsuit targeting deepfakes is "first-of-its-kind"); *infra* Section I.C.1 (describing the current laws targeting deepfakes).

from real videos and undetectable as fake.[40] This Note is the first comprehensive analysis of deepfake pornography under defamation law. Previously, defamation has been overlooked, cursorily dismissed, or only hinted at in passing as a potential legal pathway.[41] This Note argues that defamation law—while by no means a panacea[42]—is the best avenue to pursue legal recourse in the time of deepfakes for at least three reasons. First, at a time when mediation of reputations occurs primarily in cyberspace, courts have shown their willingness to award defamation damages for harm that occurs beyond traditional jurisdictional boundaries.[43] Second, defamation law consistently recognizes noneconomic, dignitary interests implicated by false statements shared publicly, which are precisely the interests infringed upon by deepfake pornography.[44] Third, as this Note shows, when allegedly defamatory statements strive for—and achieve—hyperrealism and the video insinuates that it is revealing a "truth" about someone's private sex life, these images fit squarely into the defamation framework, contrary to what some scholars have argued.[45] Cursory indications that a deepfake is "fake" or viewers' knowledge that it is "synthetic" refer solely to the manner of creation. This knowledge does little to undermine the reputational harm and, consequently, the defamation claim.[46]

---

40   *See*, *e.g.*, Emily van der Nagel, *Verifying Images: Deepfakes, Control, and Consent*, 7 Porn Stud. 424, 424–25 (2020) (arguing that deepfakes "continue a long history of women's images being used to harass, humiliate, and harm them").

41   *See, e.g.*, Gieseke, *supra* note 37, at 1500 ("Defamation . . . fails as an option due to the intent requirement."); Aasha Shaik, *Deepfake Pornography: Beyond Defamation Law*, Yale Cyber Leadership F. (July 20, 2021), https://www.cyber.forum.yale.edu/blog/2021/7/20/deepfake-pornography-beyond-defamation-law [https://perma.cc/G9PJ-S7UD] (arguing that using defamation "would be missing the actual point, which is the violation of consent"); Moncarol Y. Wang, Comment, *Don't Believe Your Eyes: Fighting Deepfaked Nonconsensual Pornography with Tort Law*, 2022 U. Chi. Legal F. 415, 441 (2023) (arguing that the "Achilles heel" of defamation is "that the publication must be to a third party").

42   *See infra* notes 99–105 (noting the downsides of civil liability for cyberviolence).

43   *See infra* Section III.A.3.

44   *See infra* Part II.

45   *See, e.g.*, Jessica Ice, *Defamatory Political Deepfakes and the First Amendment*, 70 Case W. Rsrv. L. Rev. 417, 434 (2019) ("[G]ood-faith deepfake creators will have a rather easy shield against culpability: any indication, either in the video itself or on the location (webpage) where the video was posted, that the video is a fake."); Shaik, *supra* note 41 ("It would be trivially easy for producers and distributors of nonconsensual deepfake pornography to skirt this issue entirely by simply posting 'fake' in the title . . . ."); Stolk, *supra* note 7 ("Since everybody knows [deepfakes of Emma Watson] are fake, they probably won't damage her image and reputation very much."); *see infra* Part III.

46   *See infra* Section III.B.

This Note argues that deepfakes—in achieving photorealism and purporting to be revealing a truth about a person's sex life—qualify as defamatory. Part I establishes deepfake pornography as a novel and egregious sexual offense facilitated by exponential technological advancements. It then discusses and reveals the shortcomings of legislation not specific to deepfakes, legislation specific to deepfakes, and the common law of torts. Part II canvasses the nature of deepfakes' harm, establishing them as a sexual offense infringing upon core rights of autonomy, dignity, and reputation. Finally, Part III discusses how deepfake pornography—regardless of whether it is indicated as or known to be fake—is actionable under defamation.

I. What Deepfakes Are and What Deepfakes Are Not

This Part situates deepfake pornography at the intersection of disruptive new AI technologies and gender-based cyberviolence—both of which suffer from a profound lack of legal attention and remedies. Section I.A describes the history and technology of deepfakes. Section I.B places deepfake pornography on the continuum of image-based sexual abuse and discusses why revenge-pornography legislation does not apply to deepfake pornography. Section I.C addresses potential paths for legal recourse, settling on defamation law.

A. The Landscape of Synthetic Media and Deepfake Technology

Although definitions vary, deepfakes essentially involve taking a small portion of a person's voice or photo and using generative AI to create a hyperrealistic portrayal of that person doing or saying things they never did.[47] Deepfakes are a form of synthetic media, or media doctored or generated by artificial intelligence, which itself is on the spectrum of manipulated media.[48]

---

47   *See* Dean Fido & Craig A. Harper, Non-Consensual Image-Based Sexual Offending: Bridging Legal and Psychological Perspectives 3 (2020) (deepfakes entail "[u]sing visual editing software to superimpose the likeness of another onto sexually explicit material"); Gieseke, *supra* note 37, at 1481 ("Deepfake technology uses artificial intelligence to realistically manipulate videos by splicing one person's face onto another's.").

48   *See* Meikle, *supra* note 14, at 3.

Of course, manipulated media is nothing new.[49] But deepfakes distinguish themselves from previous forms of manipulated media in both scale and kind.[50] Deepfakes are created via deep-learning algorithms, most commonly Generative Adversarial Networks (GAN), or diffusion networks. While the technological specificities are noteworthy,[51] for the purposes of this Note, these networks are quick, accessible, and lead to consistent, eerily photorealistic representations of a woman's intimate life.[52] Deepfakes come in many forms,[53] but the vast majority consist of face-swapping, or taking a woman's face, analyzing her body shape over her clothes, and superimposing her likeness onto the naked body of another.

It is difficult to convey just how realistic deepfakes are to those unfamiliar with recent technological advancements. The realism of today's deepfakes is nothing short of extraordinary. Research indicates that high-quality deepfake videos can "easily" fool the public, with less than a quarter being recognized as fake.[54] To the dismay of investors allocating billions of dollars toward research,[55] mechanical deepfake detectors currently

---

49   Photographs and audio recordings have been manipulated since their advent. Early examples include altered photos of Abraham Lincoln and Joseph Stalin. *See* Schick, *supra* note 35, at 26–27.

50   Ice, *supra* note 45, at 427 ("[T]he method of creation (by deep learning or artificial intelligence) [is] a key way to distinguish deepfakes from other faked videos . . . because the use of deep learning in a video's creation implies that such a video can be created more easily [and look more realistic] than a manually manipulated video.").

51   *See generally* Loveleen Gaur, Deepfakes: Creation, Detection, and Impact 1–6 (2022). *See also* Schick, *supra* note 35, at 44–45 (describing Generative Adversarial Networks as employing a "constant iterative process" until achieving a close-to-perfect depiction).

52   *See* Schick, *supra* note 35, at 44–45.

53   These include face reenactment (manipulating someone's facial features), face generation (creating a new face not based on any real individual), face swapping (replacing one person's face with another's), and speech synthesis (replicating voices). *See* Carolyn Pepper, Peter Raymond & Talia Fiano, *Reputation Management and the Growing Threat of Deepfakes*, Bloomberg L. (July 9, 2021), https://news.bloomberglaw.com/us-law-week/reputation-management-and-the-growing-threat-of-deepfakes [https://perma.cc/4T6D-FLMF].

54   *See* Pavvel Korshunov & Sébastien Marcel, *Deepfake Detection: Humans vs. Machines*, ArXiv, 4 (Sept. 7, 2020), https://arxiv.org/pdf/2009.03155 [https://perma.cc/E42M-ZA9E] (finding a bias toward assuming the authenticity of videos and that participants identified deepfake videos as fake 24.5% of the time); Klair Somoray & Dan J. Miller, *Providing Detection Strategies to Improve Human Detection of Deepfakes*, 149 Comput. Hum. Behav. 1, 8 (2023) (finding that even when participants were told they would be shown deepfakes and instructed on detection techniques, their ability to detect them was "generally poor" and only slightly above chance (60.7%)).

55   *See also* NSA et al., *supra* note 33, at 6–7 (noting that the organizations developing deepfake detectors include Microsoft, Intel, Google, the Air Force Research Lab, and Adobe).

fare no better than human detectors.[56] More importantly, a successful deepfake detector would do little to mitigate the harm to the hundreds of thousands of women. For those victims, their injury is not derived from viewers mistaking the content for being real but from the public exposure of sexualized depictions of their bodies.[57]

Deepfake pornography is not magic, and it is "no longer rocket science."[58] While realistic synthetic video was previously costly and "an exceedingly complex operation for even the most experienced digital artists," it is now "a single button press to create a face-swapped video."[59] Sexual deepfake photographs are even more accessible. For example, the app DeepNude—advertising itself as "the superpower you always wanted"—allows users to upload photos of any woman.[60] Then, the app would generate a naked image of her.[61] The use of gendered pronouns is intentional: The databases used by DeepNude and other deepfake pornography software are typically trained exclusively on cisgender female bodies, meaning they can only generate deepfakes of women.[62]

----

56   As deepfake technology improves, detection methods have tried to catch up, with some detection methods analyzing light, shadows, eye movement, and even blood circulation. But as soon as detection methods advance, deepfake technology responds by addressing the weak point. *See* Jason Haas, *Deepfake Dilemma*, Intell. Prop. Mag., Sept. 2019, at 33. *See also* NSA et al., *supra* note 33, at 6–7 (calling the development of detectors a "cat and mouse game").

57   *See infra* Section II.B.

58   Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All Fucked*, Vice (Dec. 11, 2017), https://www.vice.com/en/article/gydydm/gal-gadot-fake-ai-porn [https://perma.cc/ME6C-28Q9].

59   Erik Gerstner, *Face/Off: "DeepFake" Face Swaps and Privacy Laws*, Def. Couns. J., Jan. 2020, at 2; *see also* NSA et al., *supra* note 33, at 2 ("[T]he market is now flooded with free easily accessible tools . . . that make the creation or manipulation of multimedia essentially plug-and-play."). Creators of deepfakes can exploit these publicly available images as "training data" for their algorithms. These sophisticated software tools meticulously analyze the photographs to accurately identify and replicate facial expressions, mannerisms, and idiosyncratic gestures. *See* Meikle, *supra* note 14, at 72.

60   Meikle, *supra* note 14, at 73. DeepNude had more than 500,000 download requests, targeting more than 104,852 women. Even though the app's creators took it offline after it launched in 2019, versions of the code continue to circulate widely. *See id.* at 73–74.

61   *See id.* at 73.

62   For example, DeepNude was trained on photographs of 10,000 female bodies. *See id.* at 74. Similarly, Porn Star by Face, one of the most popular databases used by creators of deepfake videos, was trained with data from 4,000 women and does not provide "matches" for men. *See id.* at 72. The algorithm crunches data of physical features to find a "match" between a woman they want to make deepfake pornography of and an existing pornography star. The site describes itself as "The First Porn Star face-recognizing search engine based on deep neural networks." *Id.*

Reddit's removal of the original forum for deepfakes and nominal bans by mainstream platforms have done little to quell the extensive and resilient online ecosystem dedicated to the creation and dissemination of deepfake pornography.[63] Over 9,500 websites specialize in nonconsensual sexual imagery.[64] Deepfake creators can use any one of the forty-two open-source, user-friendly machine-learning tools available online and ask questions in any of the fifteen deepfake-creation community websites totaling over 600,000 members.[65] Websites that host deepfake-pornography websites are "participatory culture[s],"[66] providing extensive guides and forums dedicated to answering questions.[67] These sites typically employ a revenue-sharing model encouraging users to become content creators.[68]

As the adage goes, technology is neither good nor bad; nor is it neutral.[69] Deepfakes have beneficial applications, such as protecting the identity of victims when sharing their stories or testifying to Congress.[70] Outside of pornography, deepfakes can be employed

---

63   Reddit's removal of /r/Deepfakes on February 7, 2018, did little to stop the proliferation of online deepfake forums and tools. *See* Meikle, *supra* note 14, at 50–51. NCDP is hosted both on dedicated deepfake pornography websites and mainstream pornography websites. *See id.*; *see* Ajder et al., *supra* note 27, at 3 (demonstrating the exponential growth of GANs since 2007).

64   *See* Citron, Fight for Privacy, *supra* note 26, at 71.

65   *See 2023 State of Deepfakes*, *supra* note 23.

66   The largest website that hosts deepfake pornography had more than 275,000 members as of February 2022, but there are likely to be even more users since most videos can be accessed without an account. Meikle, *supra* note 14, at 56–57. *See also id.* at 60 (calling the largest deepfake website a "participatory culture" analogous to the interactive models of major social media platforms such as TikTok).

67   The website includes extensive training materials for users who want to learn how to make deepfake videos. Over 10,000 forum posts on the website respond to questions and explain the technical aspects of creating deepfakes. *See id.* at 59 ("The site walks the user through how to extract images of their chosen celebrity. . . and . . . explains how to optimize and align images . . . how to train the neural networks, how to merge the resulting images, and how to use basic post-production techniques . . . .").

68   *See id.* at 58.

69   This is Melvin Kranzberg's first law of technology. *See* Melvin Kranzberg, *Technology and History: "Kranzberg's Laws,"* 27 Tech. & Culture 544, 545 (1986).

70   *See* Another Body (Willa 2023) (using deepfake technology to create a documentary in which victims told their stories without being identified).

for comedy,[71] entertainment,[72] malevolent political purposes,[73] and fraud.[74] As stated previously, most scholarship and government resources focus on deepfakes in the political and national security spheres,[75] turning a blind eye to the fact that deepfake technology was both pioneered for pornography and that it is overwhelmingly used to create nonconsensual pornography.[76]

B. Gender-Based Cyberviolence and Pornography

Deepfakes are not just a form of synthetic media; they also exist on the continuum of image-based sexual abuse[77] and cyber gender violence.[78] Other forms of image-based

---

71　　*See, e.g.*, BuzzFeedVideo, *You Won't Believe What Obama Says in This Video*, YouTube (Apr. 17, 2018), https://www.youtube.com/watch?v=cQ54GDm1eL0&ab_channel=BuzzFeedVideo [https://perma.cc/FM7J-M8DE].

72　　*See* Gerstner, *supra* note 59, at 3.

73　　*See* Paul Sonne, *Fake Putin Speech Calling for Martial Law Aired in Russia*, N.Y. Times (June 5, 2023), https://www.nytimes.com/2023/06/05/world/europe/putin-deep-fake-speech-hackers.html [https://perma.cc/3UAC-CTBJ].

74　　*See, e.g.*, Kindred, *supra* note 39 ("Deepfakes have been used for cybercrime, extortion, targeted attacks, misinformation, fraud, getting around authentication methods, and threats to personal, professional, and company reputations.").

75　　*See supra* note 33.

76　　*See* MEIKLE, *supra* note 14, at 51 ("Non-consensual deepfake porn . . . is . . . the predominant use of synthetic video to date."). Notwithstanding the lack of formal legal structures addressing the gendered issues posed by the rapid proliferation of digital media and information, pornography itself has been at the forefront of many of these innovations and the development of new forms of media. In fact, "demand for, and take-up of, new technologies has been consistently driven by the desire of audiences to access pornographic material more easily and more privately." Rebecca Sullivan & Alan McKee, PORNOGRAPHY: STRUCTURES, AGENCY AND PERFORMANCE 49 (2015). Internet pornography spurred the development of webcams, secure online credit card payment systems, banner advertisements and pop-ups, and streaming video technologies. *See* Susanna Paasonen, *Online Pornography*, *in* THE SAGE HANDBOOK OF WEB HISTORY 551, 551 (Niels Brügger & Ian Milligan eds., 2018).

77　　Image-based sexual abuse refers to the umbrella of "offences involving the nonconsensual-generation, taking, and/or distribution of private sexual images." FIDO & HARPER, *supra* note 47, at 7.

78　　*See, e.g.*, Danielle Citron, *The Continued (In)visibility of Cyber Gender Abuse*, 2023 YALE L.J.F. 333, 341 [hereinafter Citron, *Continued (In)visibility*].

sexual abuse include revenge pornography,[79] upskirting,[80] downblousing,[81] and cyber-flashing.[82] Professor Danielle Citron coined the term "cyber gender abuse" to capture the "gendered nature" of cyberviolence.[83] Scholar Emma Jane writes of cyberhate that "[m]isogynists have never had so many opportunities to collectivize and abuse women with so few consequences."[84] One in twelve American adults under thirty have been victims of image-based sexual abuse, and almost two-thirds have been harassed online.[85] But, despite the prevalence of technology-facilitated violence, there is a "never-ending dismissal of cyber gender abuse"[86] due in part to a "tendency to tolerate, trivialize, or dismiss these harms."[87]

1. Deepfake Pornography and Nonconsensual "Revenge" Pornography

Nonconsensual "revenge" pornography (NCP)[88] is a form of image-based sexual abuse with considerable similarities to deepfake pornography. Owing in large part to the work

---

79   See *infra* Section I.B.2.

80   *See* Fido & Harper, *supra* note 47, at 3 (defining upskirting as "[t]he non-consensual and surreptitious capturing of intimate images under an individual's clothing").

81   *See* Citron, Fight for Privacy, *supra* note 26, at 73 (remarking that "A down-blouse thread on a hidden camera site had more than 150,000 videos with titles like 'Very busty white girl spotted on Japan street with jiggling big boobs,' 'Black woman with dreadlocks in bikini,' and 'Sexy Asian Teen.'").

82   *See* Fido & Harper, *supra* note 47, at 3 (defining cyber-flashing as "[s]haring sexually explicit images via digital technologies . . . to unsuspecting or non-consenting recipients").

83   Citron, *Continued (In)visibility*, *supra* note 78, at 337 (coining the term "cyber gender abuse" to refer to the "gendered nature" of cyber abuse).

84   Emma Jane, Misogyny Online: A Short (and Brutish) History 51 (2017).

85   *See* Emily A. Vogels, *The State of Online Harassment*, Pew Rsch. Ctr. (Jan. 13, 2021), https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/ [https://perma.cc/S6MH-UCTG]; My Image My Choice, *supra* note 31.

86   Citron, *Continued (In)visibility*, *supra* note 78, at 340 (connecting the law's historic nonrecognition of harms that disproportionately affect women to the law's current nonresponse to cyber gender abuse).

87   Citron & Franks, *supra* note 16, at 347 (attributing the dearth of effective legal protections to a "lack of understanding about the gravity, scope, and dynamics of the problem; historical indifference and hostility to women's autonomy; inconsistent conceptions of contextual privacy; and misunderstandings of First Amendment doctrine").

88   Nonconsensual "revenge" porn is "intimate media that is created, obtained, or distributed without the subject's consent." Chad D. Post, *101: How to Combat Revenge Porn*, Wis. Law. (Feb. 11, 2020), https://www.

of feminist activists and scholars, forty-six states have now criminalized nonconsensual pornography,[89] and a civil cause of action for it exists at the federal level.[90] But while deepfake pornography and nonconsensual pornography are normatively equivalent, they are legally distinct.[91] Deepfake pornography is synthetically created, whereas nonconsensual pornography consists of non-manipulated, real depictions.[92] In other words, the intimate depictions in deepfakes—even when indistinguishable from physical reality— are not actual photographic captures of the featured woman's body. Indeed, there is an often-overlooked second victim in deepfake pornography: the person whose images and sex work are used nonconsensually to stock the database of bodies.[93] As such, deepfakes are excluded from revenge-porn legislation due to statutory phrases like "person whose

wisbar.org/NewsPublications/WisconsinLawyer/Pages/Article.aspx?Volume=93&Issue=2&ArticleID=27466 [https://perma.cc/6RDP-J3FT]. Most scholars prefer the term "nonconsensual pornography" to revenge porn. *See id.* (showing how "revenge porn" is a misnomer because only eleven percent of perpetrators share intimate images for the purposes of revenge).

89   *See* Citron & Franks, *supra* note 16; *see also* Brooklynn Armesto-Larson, *Nonconsensual Pornography: Criminal Law Solutions to a Worldwide Problem*, 21 Or. Rev. Int'l. L. 177 (2020).

90   In 2022, the reauthorization of the Violence Against Women Act established "a federal civil cause of action for individuals whose intimate visual images are disclosed without their consent." *Fact Sheet: Reauthorization of the Violence Against Women Act (VAWA)*, The White House (Mar. 16, 2022), https://www. whitehouse.gov/briefing-room/statements-releases/2022/03/16/fact-sheet-reauthorization-of-the-violence-against-women-act-vawa/ [https://perma.cc/SL2R-NZRX].

91   *See* Fido & Harper, *supra* note 47, at 17–18 (stating that revenge porn laws do not apply to deepfakes); *see also* Wang, *supra* note 41 (arguing that "seeking relief for DNCP victims via NCP laws—on the theory that both involve nonconsensual acts and reputational damage—is likely insufficient").

92   Most of the photographs or videos in nonconsensual pornography were taken by the victim themselves. *See* Post, *supra* note 88.

93   Professor Rebecca Delfino calls attention to this often-overlooked second victim in a deepfake pornography video: "Although the actor whose body is featured may have consented to the original pornographic video, they likely never agreed to have another person's face superimposed onto their body. They, too, have been victimized." Rebecca A. Delfino, *Pornographic Deepfakes: The Case for Federal Criminalization of Revenge Porn's Next Tragic Act*, 88 Fordham L. Rev. 887, 898 (2019). The President of the Adult Performance Artist Guild (APAG), the largest union of adult actors, stated, "[S]top pretending that the people in porn are not human beings—that we're not being exploited 10 times worse" than mainstream actors. Hallie Lieberman, *Inside the Glitchy and Horny World of AI Porn*, The Daily Beast (Aug. 25, 2023), https://www.thedailybeast. com/inside-the-glitchy-and-horny-world-of-ai-porn. [https://perma.cc/NS8L-SNMP]. Some adult performers are finding creative ways to respond to their changing industry, such as by creating AI images of themselves to increase their revenue stream. *See id.*

intimate parts" or "engaged in sexual conduct."[94] But the feature of deepfakes that closes the door to NCP laws—namely, their falsity—pushes the door open to defamation, which was never a viable path for nonconsensual "real" porn.[95]

It may be argued that deepfake pornography and revenge porn are normatively distinct because viewers of deepfakes presumably are aware that the content is fake,[96] whereas the "allure" of revenge porn is that the images are real.[97] But this distinction turns out to be largely misplaced. Viewers of deepfakes are often misled and unable to discern the falseness of the video or photo, and the "allure" of deepfakes is that they are indistinguishable from a real depiction of the person.[98] Creators go to lengths to use real women's faces with the overarching (and often attained) goal of photorealism.

There are several roadblocks to legal accountability for cyber gender abuse. First, challenges arise with identifying perpetrators of online abuse due to VPNs and software that make their IP addresses unidentifiable.[99] In the world of deepfake pornography, perpetrators are even more likely to try to hide their identity—in fact, the websites actively encourage it. For example, the largest deepfake-pornography website urges users to "always . . . stay[] anonymous online [because] your privacy is important to us."[100] Of course, the privacy of

---

94   *See, e.g.*, Or. Rev. Stat. Ann. § 163.472 (West 2024); Ohio Rev. Code Ann. § 2917.211 (West 2019) (using the statutory language of [the person in the image] is "in a state of nudity or is engaged in a sexual act"); Me. Rev. Stat. Ann. tit. 17-A, § 511-A (West 2016) (using the statutory language of "photograph, videotape, film or digital recording of another person in a state of nudity or engaged in a sexual act"). For a compilation of state revenge porn statutory language, see *States with Revenge Porn Laws*, C.A. Goldberg, PLLC (https://www.cagoldberglaw.com/states-with-revenge-porn-laws/#1558636661661-9e82fe80-c2a5 [https://perma.cc/UC9X-YZQZ].

95   *See infra* Part III.

96   *See* Delfino, *supra* note 93, at 897 (arguing that deepfake viewers are "in on the joke").

97   *See* Citron, *Continued (In)visibility*, *supra* note 78, at 347 ("The draw to these sites is that the women featured have not consented to the posting of their images.").

98   *See* Fido & Harper, *supra* note 47, at 16 (noting that with deepfake pornography, it is "hard to tease apart fact from fiction"); *see infra* Part II.

99   *See* message from Jane Doe to Abigail George (Jan. 7, 2023, 17:36 EST) (on file with *Columbia Journal of Gender & Law*) (expressing that even though she knew who the perpetrator was, she did not pursue legal action because he lived in Canada and she lived in the United States).

100   [John Doe Website] allows anyone who is "verified" to upload videos. The verification process only involves showing links to their datasets, an email address, and a username. The email address does not even have to be "real." *See* [John Doe Website #2] (on file with *Columbia Journal of Gender & Law*).

the women they target is seemingly not a concern—an irony that likely bypasses deepfake creators.

Furthermore, online platforms—"the best-positioned entities to respond to most harmful content"—are immune from liability for distributing deepfake pornography under Section 230 of the Communications Decency Act (CDA).[101] Passed by Congress in 1996,[102] the CDA grants platforms sweeping immunity from liability for user-generated content.[103] Accordingly, these "platforms' power now includes the ability to ignore the propagation of damaging deep fakes."[104] But while Section 230 immunizes platforms, individual creators and distributors can be held legally liable.[105]

C. Three Legal Approaches and Their Shortcomings

The obstacles to legal accountability for cyber gender abuse, coupled with the sheer rate of technological advancement, have allowed deepfakes to proliferate into one of the gravest gender-based violence issues of our time. This section turns to existing or proposed paths to legal recourse. A few nuclear options have been placed on the table, such as banning all deepfake technology or telling all women to avoid posting photos of themselves

---

101   Danielle K. Citron & Robert Chesney, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CAL. L. REV. 1753, 1798 (2019) ("Section 230 has evolved into a super-immunity that, among other things, prevents the best-positioned entities to respond to most harmful content."); *see* 47 U.S.C. § 230(c)(2).

102   *See* 47 U.S.C. § 230(c)(2) ("No provider or user of an interactive computer service shall be held liable on account of . . . any action voluntarily taken in good faith to restrict access to . . . material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable . . . .").

103   *See, e.g.*, Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401 (2017) (detailing cases in which Courts have provided broad immunity for internet platforms); *see also* Herrick v. Grindr, 306 F. Supp. 3d 579, 585–86 (S.D.N.Y. 2018) (dismissing a case against Grindr after it failed to remove an impersonator who posted the plaintiff's nude photos, falsely claimed he had rape fantasies, and shared his home address, causing over 1,000 men to approach his home).

104   Citron & Chesney, *supra* note 101, at 1798.

105   While Congress has held hearings on Section 230, no substantive changes have been made. For an overview of the history and interpretation of Section 230, see JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET (2019).

online.[106] These options are not only impractical but also undesirable.[107] Moreover, an option such as requiring watermarks on all deepfakes—something China implemented[108] and the Biden Administration considered[109]—may be less nuclear but is equally ill-conceived.[110] More tenable legal paths to recourse include (1) legislation not specific to deepfakes, (2) legislation specific to deepfakes, and (3) the common law of torts.

1. Legislation Not Specific to Deepfakes

In response to the urgent crisis of deepfakes, lawyers and victim advocates have attempted to fit square pegs into round holes. Existing crimes such as impersonation,[111] cyberstalking,[112] and harassment are recognized in several states. At the federal level, suggestions that the Anti-Stalking Punishment and Prevention Act or the Video Voyeurism

---

106   *See* Jesselyn Cook, *Here's What It's Like to See Yourself in a Deepfake Porn Video*, HuffPost (June 23, 2019), https://www.huffpost.com/entry/deepfake-porn-heres-what-its-like-to-see-yourself_n_5d0d0faee4b 0a3941861fced?ncid=engmodushpmg00000004/ [https://perma.cc/C96T-NA4L] (reporting that women have been told that they only thing they can do to keep themselves safe is stay offline).

107   As stated previously, there are many beneficial applications of deepfakes. *See supra* notes 71–73.

108   *See* Asha Hemrajani, *China's New Legislation on Deepfakes: Should the Rest of Asia Follow Suit?*, The Diplomat (Mar. 8, 2023), https://thediplomat.com/2023/03/chinas-new-legislation-on-deepfakes-should-the-rest-of-asia-follow-suit/ [https://perma.cc/X8B3-ACKV].

109   *See* Press Release, The White House, Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/ statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ [https://perma.cc/6RS8-KNMW] ("The Department of Commerce will develop guidance for content authentication and watermarking to clearly label AI-generated content.").

110   For an explanation of why watermarking is not effective, see Ben Colman, *The Inadequacy of Deepfake Watermarking*, Reality Defender (May 15, 2023), https://realitydefender.com/insights/the-inadequacy-of-deepfake-watermarking/ [https://perma.cc/CAE4-BFHE]. Moreover, services exist to remove watermarks. *See* Citron, Fight for Privacy, *supra* note 26, at 48.

111   Several states, including California, Hawaii, Louisiana, Mississippi, New York, Rhode Island, and Texas, have impersonation crimes. *See* Cal. Penal Code § 528.5 (West 2011); Haw. Rev. Stat. Ann. § 711-1106.6 (West 2008); N.Y. Penal Law § 190.25 (McKinney 2019); 11 R.I. Gen. Laws Ann. § 11-52-7.1 (West 2019); Tex. Penal Code Ann. § 33.07 (West 2019).

112   Thirty-four states have cyberstalking laws. For an overview of the various state cyberstalking laws, see Ashley N.B. Beagle, *Modern Stalking Laws: A Survey of State Anti-Stalking Statutes Considering Modern Mediums and Constitutional Challenges*, 14 Chap. L. Rev. 457 (2011).

Prevention Act apply to deepfakes have turned out to be wishful thinking.[113] In 2023, the first and only American prosecution for deepfake pornography occurred under New York's aggravated harassment and cyberstalking laws.[114] But overall, legislation not specific to deepfakes is unlikely to ever result in significant prosecutions due to their typically high mens rea requirements, prosecutors' unwillingness to interpret such statutes loosely, and other idiosyncrasies.[115] More broadly, since showing the requisite intent is often a limiting factor for cyber-abuse prosecutions, defamation law's reduced and peculiar scienter requirement is a better fit.[116]

2. Legislation Specific to Deepfakes

Many scholars and advocates have pushed for statutory bans on deepfake pornography.[117] In 2023, the United Kingdom became the first jurisdiction to expressly criminalize deepfake pornography.[118] In the United States—reflecting general trends toward the importance of

---

113   *See* Anti-Stalking Punishment and Prevention Act, 18 U.S.C. § 2261A(2) (2012) (criminalizing the use of an "interactive computer service or electronic communication service . . . to engage in a course of conduct that . . . causes, attempts to cause, or would reasonably be expected to cause substantial emotional distress to a person"). While this law appears promising, it is unlikely to result in any prosecutions because of the focus on repeat offenders and the intent requirement. *See id.*; Video Voyeurism Prevention Act, 18 U.S.C. § 1801 (2004) (penalizing intentional and nonconsensual "capturing" images of a person's private area, particularly when the person reasonably expects privacy). This Act does not extend to deepfakes for the same reason as revenge porn laws. *See id. See also supra* Section I.B.1.

114   In an especially egregious case, a man was prosecuted and convicted in New York for creating and disseminating deepfake pornography of several underage women using photos from their social media. *See* Press Release, District Attorney, County of Nassau, Seaford Man Sentenced to Jail and 10 Years' Probation as Sex Offender for 'Deepfaked' Sexual Images (Apr. 18, 2023), https://www.nassauda.org/CivicAlerts. aspx?AID=1512 [https://perma.cc/VET7-2VX6].

115   *See* Citron & Chesney, *supra* note 101, at 1801 ("Although a wide range of deep fakes might warrant criminal charges, only the most extreme cases are likely to attract the attention of law enforcement.").

116   *See infra* Section III.A.2.

117   *See, e.g.*, Delfino, *supra* note 93 (making the case for criminalizing deepfake pornography); Douglas Harris, *Deepfakes: False Pornography is Here and the Law Cannot Protect You*, 17 Duke L. & Tech. Rev. 99, 102 (2019).

118   *See* Online Safety Act 2023, c. 50, § 66(B)(1)(a) (UK) (criminalizing "intentionally shar[ing] a photograph or film which shows, or *appears to show*, another person . . . in an intimate state") (emphasis added); *see also* Press Release, New Laws to Better Protect Victims From Abuse of Intimate Images, Ministry of Justice & Right Hon. Dominic Raab (Nov. 25, 2022) (UK), https://www.gov.uk/government/news/new-laws-to-better-protect-victims-from-abuse-of-intimate-images [https://perma.cc/5XLN-CVEN] (announcing that the Online Safety Act would "specifically criminalize[]" nonconsensual sexual deepfakes).

state-conferred rights[119] and mirroring the path of revenge-porn laws—states are once again blazing the trail on deepfake legislation. In 2019, a small smattering of states began introducing and promulgating deepfake-related legislation.[120] In 2024, the tide turned, with more deepfake legislation introduced in the first half of that year than in the previous six years combined.[121] As of October 2024, twenty-nine states have enacted legislation dealing with pornographic deepfakes.[122] The approaches to such legislation vary, with some states amending existing revenge-porn statutes to include "digital images,"[123] while others have instituted wholly new crimes or civil penalties.[124] Arizona, California, Colorado, Hawaii, Illinois, Massachusetts, Minnesota, New York, Vermont, and Wyoming have laws addressing the nonconsensual creation and distribution of adult deepfake pornography.[125]

---

119   *See, e.g.*, Jessica Bulman-Pozen & Miriam Seifter, *State Constitutional Rights and Democratic Proportionality*, 123 Colum. L. Rev. 1855, 1856–59 (2023).

120   In 2019, California became the first state to pass legislation giving victims of nonconsensual deepfake pornography a private right of action. *See* Cal. Civ. Code § 1708.86 (West 2021). *See also* Eric Kocsis, *Deepfakes, Shallowfakes, and the Need for a Private Right of Action*, 126 Dick. L. Rev. 621, 638–39 (2022) (analyzing the California law).

121   *See* Ballotpedia, State of Deepfake Legislation 2024 Annual Report 10 (2024) (reporting that forty-seven deepfake-related bills were enacted in 2024 compared to thirty-one between 2019 and 2023).

122   *See id.*

123   *See* Vt. Stat. Ann. tit. 13, § 2606 (West 2024) (effective June 6, 2024) (amending the definition of "visual image" to include "an image created or altered by digitalization"); Wash. Rev. Code Ann. § 9A.86.030 (West 2024) (effective June 6, 2024) (amending several existing laws on disclosing intimate images to include "fabricated depiction" and "digitalization"); H.B. 2678, 2019 Leg., Reg. Sess. (Va. 2019). As of July 2023, Illinois citizens could sue for damages for deepfakes under the state's law on revenge pornography. *See* 740 Ill. Comp. Stat. Ann. 190/5 (West 2024).

124   *See* Wyo. Stat. Ann. § 6-4-306 (West 2021) (effective July 1, 2021) (defining "image" to include a "computer generated image that purports to represent an identifiable person" and establishing the offense as a misdemeanor punishable by up to one year in prison). *See also* Cassandre Coyer, *States Are Targeting Deepfake Pornography— But Not in a Uniform Way*, ALM Law (Aug. 10, 2023), https://www.law.com/legaltechnews/2023/08/10/states-are-targeting-deepfake-pornography-but-not-in-a-uniform-way/?slreturn=20231021183045 [https://perma.cc/F68D-6W4J].

125   *See* H.B. 2394, 56th Leg., 2nd Sess. (Ariz. 2024); Assemb. B. 602, 2019 Leg., Reg. Sess. (Cal. 2019); CO S.B. 24-011, 2024 Leg., Reg. Sess. (Colo. 2024); S.B. 309, 31st Leg., Reg. Sess. (Haw. 2021); H.B. 2123, 103rd Gen. Assemb., (Ill. 2023); H.B. 4744, 193rd Gen. Ct. (Mass. 2024); H.B. 1370, 2023 Leg., 93rd Sess. (Minn. 2023); S.B. 1042, 2023 Leg., Reg. Sess. (N.Y. 2023); H.B. 2678, 2019 Leg., Reg. Sess. (Va. 2019); Wyo. Stat. Ann. § 6-4-306 (West 2021). *See generally* Ballotpedia, AI Deepfake Legislation Tracker, https://legislation.ballotpedia.org/ai-deepfakes/search?status=Enacted&category=Pornographic%20material&orderby=stateAsc&session=2024&session=2023&session=2022&session=2021&session=2020&session=2019&page=1 [https://perma.cc/RC23-NHZ3].

In addition, Alabama, Florida, Georgia, Idaho, Indiana, Iowa, Louisiana, New Hampshire, South Dakota, Texas, Utah, Virginia, and Washington have legislation addressing both adult and child deepfake pornography.[126] Finally, legislation in Kentucky, Mississippi, North Carolina, Oklahoma, Tennessee, and Wisconsin is limited to addressing pornographic deepfakes involving children.[127]

At the federal level, the situation looks bleak. Not surprisingly, the only legislation passed by Congress mentioning deepfakes dealt exclusively with national security.[128] Even so, several other deepfake bills have been introduced in Congress. The Malicious Deepfake Prohibition Act of 2018[129] focused primarily on political interference, and the DEEP FAKES Accountability Act of 2019[130] proposed a watermark requirement. Both bills expired at the end of their terms.[131] At the time of writing, the most promising bill is the Preventing Deepfakes of Intimate Images Act, introduced in May 2023.[132] The Act would establish a criminal and a civil cause of action,[133] entitling victims to up to $150,000 in

---

126 *See* Ala. Code § 13A-6-240 (2024); Fla. Stat. Ann. § 775.0847 (West 2022); Ga. Code Ann. § 16-11-90 (West 2021); Ga. Code Ann. § 39-5-5 (West 2025); Iowa Code Ann. § 708.7 (West 2024); Idaho Code Ann. § 18-6606 (West 2024); Ind. Code Ann. § 32-21.5-2-1 (West 2024); La. Stat. Ann. § 73:14.14 (2024); N.H. Rev. Stat. Ann. § 644:9-a (2024); S.D. Codified Laws § 22-21-4 (2022); Tex. Penal Code Ann. § 21.165 (West 2023); Utah Code Ann. § 76-5b-103 (West 2024); Vt. Stat. Ann. tit. 13, § 2606 (amended 2024); Wash. Rev. Code Ann. § 9A.86.030 (West 2024).

127 *See, e.g.*, H.B. 591, 2024 Gen. Assemb., Reg. Sess. (N.C. 2024).

128 *See* National Defense Authorization Act for Fiscal Year 2020, Pub. L. No. 116-92, § 5709, 133 Stat. 2168 (2019) (requiring the Director of National Intelligence to generate a report on the "national security impacts of machine-manipulated media" and assess China's and Russia's capabilities).

129 S. 3805, 115th Cong. (as introduced to the Senate, Dec. 21, 2018). The Act would have imposed criminal liability on distributors and creators, designating a federal sentence of ten years for deepfakes that disturb elections, but it was stalled and expired due to concerns over First Amendment violations. *See id.*

130 H.R. 3230, 116th Cong. (2019) (allowing for a private right of action to receive statutory damages and injunctive relief). *But see* Fido & Harper, *supra* note 47, at 18 (noting that the Deep Fake Accountability Act "does little to aid victims of deepfake media production who have been depicted in a sexualized manner as it only necessitates the need to include a watermark to indicate that the media is fake").

131 *See* S. 3805, 115th Cong. (2018); H.R. 3230, 116th Cong. (2019).

132 *See* H.R. 3106, 118th Cong. (as introduced to the House, May 5, 2023).

133 *See id.* §§ 1309A(b)(1), 2252D(a)(1) (establishing a civil cause of action provided the discloser knew or "recklessly disregard[ed] whether, the individual has not consented to such disclosure" and a criminal cause of action for defendants who distribute "with the intent to harass, annoy, threaten, alarm or cause substantial harm to the finances or reputation of the depicted individual").

damages and an injunction to remove the images.[134] While the Act would be a "good first step,"[135] as of September 2024, it only has fifty-nine co-sponsors in the House and a "one" percent chance of being enacted.[136]

Moreover, the inherent digital and global nature of deepfake pornography poses significant roadblocks to the effectiveness of a ban in one jurisdiction without banning it in all jurisdictions.[137] Even so, a U.S. federal ban would carry the most weight globally since over half of pornography videos originate from and most deepfake-pornography websites are hosted in the United States.[138] Of course, this Note does not intend for defamation law to displace the importance of new legislation. In fashioning a response to this gender-based violence crisis, the adoption of a federal statute offers a promising—albeit incomplete—solution. This Note argues for a common law path to recourse in addition to any potential statutory one.

---

134   *See id.* § 1309A(d)(1)(A)(i)–(iv) (entitling the plaintiff to recover any of the defendant's profits from the disclosure, damages sustained by the individual, including for emotional distress or up to $150,000 in liquidated damages, punitive damages, and attorney fees). The Act would also allow the plaintiff to remain anonymous. *See id.* § 1309A(d)(2).

135   Ryan-Mosley, *supra* note 24**.**

136   *See* GOVTRACK, H.R. 3106: PREVENTING DEEPFAKES OF INTIMATE IMAGES ACT, https://www.govtrack.us/congress/bills/118/hr3106 [https://perma.cc/F42L-8PNT] (analyzing the Bill's current posture and support).

137   *See* Citron, *Continued (In)visibility*, *supra* note 78, at 348 ("Most sites are hosted in countries like the United States where the risk of liability for privacy invasions is low."); FIDO & HARPER, *supra* note 47, at 20 (arguing with regard to image-based sexual abuse that "there is a clear argument for uniformed punishments and legislation to be applied in a globally connected world"); *see also* Martin, *supra* note 18, at 58 (explaining that the police could not do anything about her deepfakes because the websites were hosted overseas).

138   *See* Marleen J.E. Klaassen & Jochen Peter, *Gender (In)equality in Internet Pornography: A Content Analysis of Popular Pornographic Internet Videos*, 52 J. SEX RSCH. 721, 725 (2015) (finding evidence that 51.8% of pornography videos originated in the United States).

3. Liability via Common Law Torts

In the realm of tort law, torts other than defamation, such as intentional infliction of emotional distress (IIED),[139] false light,[140] and the right of publicity,[141] have been proposed and may provide recourse for deepfake pornography. Although promising, bringing an IIED claim without an accompanying defamation claim is a long shot due to its status as a "disfavored cause of action"[142] and difficulties in demonstrating "severe emotional distress."[143] The right of publicity and false light are limited to nonconsensual commercial uses of identities, so while potentially viable for monetized deepfakes, they do not address deepfake pornography at large as a sexual offense.[144] More broadly, privacy torts are "widely deemed out-of-date" to address the harms experienced in the digital and information age.[145]

139　The Restatement (Second) of Torts establishes IIED as "extreme and outrageous conduct intentionally or recklessly caus[ing] severe emotional distress to another is subject to liability for such emotional distress, and if bodily harm to the other results from it, for such bodily harm." § 46(1) (Am. L. Inst. 1965). Courts have applied IIED to cases of NCP, even awarding up to $6.4 million, suggesting that they might be willing to apply it in cases of NCDP. *See* Christine Hauser, *$6.4 Million Judgment in Revenge Porn Case Is Among Largest Ever*, N.Y. Times (Apr. 11, 2018), https://www.nytimes.com/2018/04/11/us/revenge-porn-california.html [https://perma.cc/8DCX-XA3K] (awarding $6.4 million in a revenge porn case); *see also* Wang, *supra* note 41, at 434 (arguing that IIED "is likely the most powerful tort available for DNCP victims").

140　False-light torts arise when a plaintiff's identifying features are used in connection with a controversial issue such that it creates a false message linking the two. *See* Bruce A. McKenna, *False Light: Invasion of Privacy*, 15 Tulsa L.J. 113 (1979).

141　*See* Michael D. Murray, Right of Publicity in a Nutshell 2 (2d ed. 2022) ("[T]he right of publicity protects a person's name, image, likeness, persona, and often their voice or other distinctive characteristics, from unauthorized commercial exploitation by others.").

142　*See* Russell Fraker, *Reformulating Outrage: A Critical Analysis of the Problematic Tort of IIED*, 61 Vand. L. Rev. 983, 984 (2019) (demonstrating that IIED is a "disfavored cause of action" and "[c]ourts appear wary of holding defendants liable for plaintiffs' emotional injuries").

143　Restatement (Second) of Torts § 46(1) (Am. L. Inst. 1965).

144　*See* Murray, *supra* note 141.

145　Mala Chatterjee, Extending the Legal Person 35 (Oct. 11, 2023) (unpublished manuscript) (on file with *Columbia Journal of Gender & Law*); *see generally* Pavesich v. New England Life Ins. Co., 122 Ga. 190, 205 (1905) (using natural law principles to establish the first common law right to privacy, reasoning that a "right to withdraw from the public gaze … [is] embraced within the right of personal liberty . . . [which] embraces the correlative right of privacy.").

In contrast, defamation is alive and well in the age of the internet and mass communication.[146] A longtime—albeit unlikely—ally in the fight against gender-based violence,[147] defamation once again fits the bill in the time of deepfakes. At its core, defamation assigns liability to people who disseminate false statements, presented as true, leading to reputational damage.[148]

II. The Harm and Proper Remedy of Deepfakes

A cardinal doctrine in tort law states that the remedy is proportional to the harm.[149] So, in the quest for a proper remedy, this Part reframes nonconsensual deepfake pornography from benign internet amusement to a severe sexual offense infringing on core rights of sexual autonomy and reputation. Section II.A shows that by violating a person's sexual autonomy and consent, deepfakes are normatively equivalent to contact-based sexual offenses, thereby calling for heightened scrutiny. Section II.B shows how deepfakes are a harm to reputation as protected by the law of defamation.

As a preliminary matter, much of the harm caused by deepfake pornography is intangible, and the American legal system, save for defamation, has long been unwilling to acknowledge and rectify non-material or non-economic harms. Thus, this Part will be framed by Professor Mala Chatterjee's "extended selves" thesis.[150] Chatterjee argues that since "[mental processes] can extend beyond our bodies into the external world, so too do any interests and rights we might have with respect to our mental processes." [151] Thus, "[t]here is no good reason for the law to distinguish between our bodies and

---

146   *See, e.g.*, Jane E. Kirtley, *Uncommon Law: The Past, Present and Future of Libel Law in a Time of "Fake News" and "Enemies of the American People*,*"* 2020 U. CHIC. LEGAL F. 117, 117 (2020) ("[T]he United States is experiencing a growth in libel suits brought by both public officials and private figures.").

147   *See infra* notes 255–259 and accompanying text.

148   *See* RESTATEMENT (SECOND) OF TORTS § 558 (AM. L. INST. 1965).

149   *See* Abram Chayes, *The Role of the Judge in Public Law Litigation*, 89 HARV. L. REV. 1281, 1282–83 (1976) (identifying one of the defining features of civil litigation as "[t]he scope of the relief is derived more or less logically from the substantive violation under the general theory that the plaintiff will get compensation measured by the harm caused by the defendant's breach of duty . . . in tort by paying the value of the damage caused.").

150   Professor Mala Chatterjee argues that "our extended selves are presently either protected as property or legally unprotected." Chatterjee, *supra* note 145, at 25. She writes that the American legal system "conceptualize[s] information as either (owned) property or (free) speech." *Id.*, at 26.

151   *Id.* at 8.

certain information entities when delineating the boundaries of ourselves."[152] Accordingly, deepfake pornography implicates sexual autonomy and dignity because, as an informational entity carrying one's likeness and identity, deepfakes are extensions of the self.

A. Establishing Deepfake Pornography as a Sexual Offense

By violating a person's sexual autonomy and consent, deepfakes are normatively equivalent to contact-based sexual offenses, thereby calling for heightened scrutiny. This section shows that the protected rights and interests justifying heightened scrutiny for sexual crimes apply squarely to deepfake pornography. But what are those protected rights? Legal systems have a long, disturbing history of conceiving of sexual violence in terms of male and patriarchal interests.[153] In the last few decades—largely in response to calls from scholars and feminist advocates—sexual crimes have been reconceptualized in terms of violations of sexual autonomy.[154] "Sexual autonomy has emerged as something like a fundamental right."[155]

Despite not involving physical contact or force, nonconsensual deepfake pornography is a sexual offense to the extent that it appropriates a sexual identity and obliges sexual conduct onto that identity (an extension of the self) without the identity holder's consent. In doing so publicly, deepfakes reduce the victim's identity to sex in the eyes of the community.

---

152   *Id.* at 31.

153   *See* Mustafa T. Kasubhai, *Destabilizing Power in Rape: Why Consent Theory in Rape Law Is Turned on Its Head*, 11 Wis. Women's L.J. 37, 52 (1996) (noting that ancient legal systems for punishing rapists were premised not on the woman's harm but on harm caused to the victim's father); *see also* Susan Estrich, *Rape*, *in* Feminist Jurisprudence 158, 162 (Patricia Smith ed., 1993) (noting that in rape law "while the focus is on the female victim, the judgment of her actions is entirely male").

154   *See* Jeb Rubenfeld, *The Riddle of Rape-by-Deception and the Myth of Sexual Autonomy*, 122 Yale L.J. 1372, 1382 (2013); Nora Scheidegger, *Balancing Sexual Autonomy, Responsibility, and the Right to Privacy*, 22 Ger. L.J. 769, 770 (2021) ("The right to sexual autonomy has developed into a fundamental human right worthy of state protection."); *see also* Coker v. Georgia, 433 U.S. 584, 597 (1977) (describing the wrongness of rape as a violation of "personal integrity and autonomy" and "the ultimate violation of self"); Stephen Schulhofer, Unwanted Sex 111 (2000) (defining sexual autonomy as "the freedom of every person to decide whether or when to engage in sexual relations"); Joan McGregor, Force, Consent, and the Reasonable Woman, *in* In Harm's Way: Essays in Honor of Joel Feinberg 231, 250 (Jules L. Coleman & Allen Buchanan eds., 1994) (describing the moral wrongness of rape as a violation of one's "sexual self-determination").

155   Rubenfeld, *supra* note 154, at 1382.

Deepfake pornography is not parody, satire, or "research."[156] It is just pornography. Even under Justice Potter Stewart's famous "I know it when I see it" formulation, most deepfakes qualify.[157] Pornography differs from other explicit media in that its primary purpose is sexual arousal.[158] Deepfake pornography—like real pornography—is often "masturbation material" and "[w]hat was words and pictures becomes, through masturbation, sex itself."[159]

Pornography and feminism—to put it mildly—have a vexed relationship. Antipornography feminists—most notably, Catharine MacKinnon and Andrea Dworkin—lambaste pornography as the linchpin of women's inequality and as systematically objectifying women.[160] For philosopher Martha Nussbaum, one of the worst kinds of objectification is denial of autonomy, which happens by "treat[ing] the object as lacking in autonomy and self-determination."[161] Thus, objectification is intrinsically linked to violations of autonomy. While Mill-inspired pro-pornography liberals and antipornography feminists disagree over whether pornography's objectification extends to all women,[162] they can agree that by forcing the likeness of a non-consenting victim to appear in sexual and often violent depictions, the creator treats the depicted woman as lacking in self-determination.[163] If the creators acknowledged her autonomy, they would have asked for

---

156　　*See* [John Doe Website #2] (on file with *Columbia Journal of Gender & Law*) (describing deepfake pornography as "machine-learning research").

157　　Jacobellis v. Ohio, 378 U.S. 184, 197 (1964) (Stewart, J., concurring).

158　　*See* Mari Mikkola, Pornography: A Philosophical Introduction 2 (2019).

159　　Catharine A. MacKinnon, Only Words 25 (1996).

160　　Fundamentally, objectification involves perceiving and treating an individual as a mere object or tool for use. In this regard, anti-pornography feminists extend their critique beyond the objectification of female performers in the production and consumption of pornography. This critique also alleges that men's consumption of pornography fosters a general perception of women as objects, primarily for sexual gratification, thus perpetuating their objectification—which has been a central concept in feminism generally. *See, e.g.*, *id.*; Andrea Dworkin, *Censorship, Pornography and Equality*, Feminist Jurisprudence 449, 454 (Patricia Smith ed., 1993) (arguing that "[pornography] is the subordination of women perfectly achieved").

161　　Martha C. Nussbaum, *Objectification,* 24 Phil. & Pub. Affs. 249, 257 (1995) (elaborating on feminists' objectification critique of pornography by listing seven ways that being treated "as an object" occurs: instrumentality, denial of autonomy, inertness, fungibility, violability, ownership, and denial of subjectivity).

162　　*See* Judith Hill, *Pornography and Degradation*, 2 Hypatia 39, 45 (1987) (explaining that even if pornography treats women as means and instruments for male pleasure, the harm does not extend to women generally).

163　　Whereas objectification via instrumentalization and fungibility occurs in regular pornography, in deepfake pornography, there is also a denial of autonomy, inertness, ownership, and subjectivity. *See* Nussbaum,

consent and confined themselves to how she desired to be depicted sexually or respected her wish to not be depicted at all.

Of course, erotic depictions carry a wide range of cultural and social meanings, and there is a growing movement of feminist pornography that focuses both on ethical production and challenging ideas about desire, beauty, and power.[164] But content-wise, the representations popular on mainstream pornography and now deepfake-pornography websites have been the subject of widespread criticism[165] for disproportionately showing "aggressive acts against women such as gagging, choking, sadomasochism, and gang rape, as well as aggressive anal sex and degrading practices like 'ass-to-mouth' and ejaculation in women's faces and mouths."[166] Quantitative content analyses of the most popular pornography have found acts of physical aggression in forty to eighty-eight percent of scenes.[167] This category of depictions appears to have even higher rates in deepfake pornography.[168] A few examples include "Nikki Haley Loves Sexual Torture," "Greta Thunberg gets a facial at the gym," and "AOC swallows 35 loads."[169] Thus, deepfake pornography not only objectifies the woman depicted, it often does so violently.

---

supra note 161; Citron, Fight for Privacy, *supra* note 26, at 114 ("Unwanted exposure of our naked bodies makes us acutely aware that others see us as objects that can be violated, rather than as human beings deserving respect.").

164   *See generally* The Feminist Porn Book: The Politics of Producing Pleasure (Tristan Taormino et al. eds. 2013).

165   *See* Elizabeth Wolgast, *Pornography and the Tyranny of the Majority*, *in* Feminist Jurisprudence 431, 435 (Patricia Smith ed., 1993) (arguing that women are represented as "inferior or less-than-human beings to be used by others in sexual and sadistic ways"); MacKinnon, *supra* note 159, at 17 ("With pornography, men masturbate to women being exposed, humiliated, violated, degraded, mutilated, dismembered, bound, gagged, tortured and killed . . . men come doing this.").

166   Max Waltman, Pornography: The Politics of Legal Challenges (2021).

167   Ana J. Bridges et al., *Aggression and Sexual Behavior in Best-Selling Pornography Videos: A Content Analysis Update*, 16 Violence Against Women 1065 (2010) (finding acts of physical aggression such as spanking, gagging, hair pulling, choking, and slapping in 88% of scenes, and verbal aggression such as name-calling in 48% of scenes); Marleen J. E. Klaassen & Jochen Peter, *Gender (In)equality in Internet Pornography: A Content Analysis of Popular Pornographic Internet Videos*, 52 J. Sex Rsch. 721, 728 (2015) (finding 40% of videos depicted physically violent acts toward women, most commonly spanking or gagging).

168   *See* [John Doe Website #2] (links on file with *Columbia Journal of Gender & Law*).

169   [John Doe Website #2 & 3] (links on file with *Columbia Journal of Gender & Law*).

For years, cyberfeminists have been screaming that digital and physical spaces and selves are coproduced and inseparable.[170] "Digital dualism" refers to the "systematic bias to see the digital and physical as separate."[171] This bias obscures our perception, constructing a boundary between on- and offline.[172] Cyberfeminist Nathan Jurgenson argues that "people are enmeshing their physical and digital selves to the point where the distinction is becoming increasingly irrelevant."[173] The self is an "augmented self," or a physical body and digital profile "acting in constant dialogue."[174]

Recall that deepfakes aren't about generating random, artificial personas—creators focus on replicating specific people, usually someone known personally. Thus, when deepfake creators appropriate someone's likeness with hyperrealistic accuracy, they usurp one's identity. Deepfakers are not shy about including identifying features of the victim with their uploads. As deepfakers hide behind VPNs, their videos usually include the full name of the subject and often other identifying information such as their college, city, address, phone number, and credit card number.[175] As a usurpation of identity, deepfake pornography is greater than a simple invasion of privacy or identity theft. [176] Sexual

---

170   *See generally* Donna Haraway, A Cyborg Manifesto (1985) (theorizing about a socialist, feminist cyborg that challenges singular identities and control matrices that contain women and other marginalized groups); Legacy Russell, Glitch Feminism: A Manifesto (2020). The term "cyberfeminism" was coined by Sadie Plant in a 1996 essay in which she defines cyberfeminism as "an insurrection on the part of the goods and materials of the patriarchal world, a dispersed, distributed emergence composed of links between women, women and computers, computers and communication links, connections and connectionist nets." Sadie Plant, *On the Matrix: Cyberfeminist Simulations*, The Gendered Cyborg 325, 335 (Fiona Hovenden et al. eds., 1st ed. 2000).

171   Nathan Jurgenson, *Digital Dualism Versus Augmented Reality*, Cyborgology (Feb. 24, 2011), https://thesocietypages.org/cyborgology/2011/02/24/digital-dualism-versus-augmented-reality/ [https://perma.cc/ST6R-Z3KC].

172   *See id.* (noting that IRL ("in real life") should be replaced by AFK ("away from keyboard")).

173   *Id.* ("And our selves are not separated across these two spheres as some dualistic 'first' and 'second' self but is instead an augmented self.").

174   *Id.*

175   *See* Citron, Fight for Privacy, *supra* note 26. For example, in the case of Taylor Klein, the deepfaker included her phone number and the name of her college, prompting many of her classmates to send messages on Instagram and threaten to come to her house; *see also* Another Body, *supra* note 70.

176   Sex law's progress has yet to extend to image-based sexual abuses, which are still usually seen as mere invasions of privacy rather than sexual offenses. *See* Roni Rosenberg & Hadar Dancig-Rosenberg, *Reconceptualizing Revenge Porn*, 63 Ariz. L. Rev. 199, 219 (2021) (arguing that revenge pornography should be classified as a sexual offense and not merely an invasion of privacy).

violations reduce a person's identity to sex alone, whereas other forms of identity theft, such as credit card fraud, invade a person's privacy but do not reduce a person's identity to their wallet.[177]

Many deepfake victims experience an invasion of the self that is incredibly similar to contact-based assault. Rape is a violent, possessory act: "a rape victim's body is taken over, invaded, occupied, taken control of—*taken possession of*—in a fashion and to a degree not present in ordinary acts of theft, robbery, assault and so on."[178] Rape is "such complete and invasive physical control over [your body that it] … is in an elemental sense no longer your own."[179] Likewise, creators of deepfake pornography take full control of their target's identity to assert unwavering possession of their likeness, forcing them to conform to their will. One victim testified: "When it's your own face reacting and moving, there's this panic that you have no control."[180] Another said: "There's something really visceral about seeing an incredibly hyper-realistic image of yourself in somebody's extreme misogynistic fantasy of you."[181] Thus, deepfake pornography not only subjects a person to public objectification, but it also converts their digital body to an entity they no longer can control.

Even though deepfake pornography is not a physical occupation of the body, when someone watches the video of themselves, they experience it as such. "Sexual-privacy invasions are experienced like physical penetrations of the body."[182] The diversity and extent of trauma of contact-based sexual offenses and other kinds of image-based sexual

---

177   For Professor Jed Rubenfeld, the unique harm or "special violation" of rape, making it worse than assault or battery, is a violation of self-possession, not sexual autonomy. Rubenfeld, *supra* note 154, at 1426. Rubenfeld writes: "The right to self-possession implies the freedom not to have another person forcibly take sexual possession of one's body, which in turn implies the freedom not to be forced into sexual service." *Id.* at 1443.

178   *Id.* at 1426.

179   *Id.*

180   Cook, *supra* note 106; *see also, e.g.*, Rousay, *supra* note 15; Message from Jane Doe to Abigail George (Dec. 29, 2023, 16:15 EST) (on file with *Columbia Journal of Gender & Law*) (describing that she had trouble looking at herself in the mirror and taking showers after the incident).

181   Emine Saner, *Inside the Taylor Swift Deepfake Scandal: 'It's Men Telling a Powerful Woman to Get back in Her Box,'* THE GUARDIAN (Jan. 31, 2024), https://www.theguardian.com/technology/2024/jan/31/inside-the-taylor-swift-deepfake-scandal-its-men-telling-a-powerful-woman-to-get-back-in-her-box   [https://perma.cc/QFK8-GLRN].

182   Danielle Citron, *Sexual Privacy*, 128 YALE L.J. 1870, 1925 (2019) [hereinafter Citron, *Sexual Privacy*].

abuse have been well-documented.[183] Deepfakes incur the same indelible trauma as contact-based sexual offenses.[184] Many victims make explicit comparisons to contact-based sexual assault: "In 2018, I was inebriated at a party, and I was used for a man's sexual gratification without my consent. Today, I have been used by hundreds of men for sexual gratification without my consent."[185] Moreover, as a digital record, deepfake pornography is a material trace that haunts the victim for the rest of their life.[186] One survivor explained: "It felt like the image was ammunition that could be used against me for the rest of my life."[187]

1. The Contextual Nature of Consent

Even though there is no physical contact or force involved while making them, deepfakes constitute violations of consent. Most nonconsensual deepfakes are created using publicly available photos of an individual—usually, photographs freely shared on social media or other online directories. Most women are not aware of a deepfake's creation until someone (or some *bot*) brings it to their attention.[188] This means they do not (and cannot) actively object to its creation; as they are not given any opportunity to reject, they have not

---

183   *See generally* Crime Survivors Speak 2022: National Survey of Victims' Views on Safety and Justice, ALL. FOR SAFETY & JUST. (Sept. 2022), https://allianceforsafetyandjustice.org/wp-content/uploads/2022/09/Alliance-for-Safety-and-Justice-Crime-Survivors-Speak-September-2022.pdf   [https://perma.cc/6CKV-KCCZ]; Samantha Bates, *Revenge Porn and Mental Health: A Qualitative Analysis of the Mental Health Effects of Revenge Porn on Female Survivors*, 12 FEMINIST CRIMINOLOGY 22, 22 (2017) (finding that victims suffer from "trust issues, posttraumatic stress disorder (PTSD), anxiety, depression, suicidal thoughts, and several other mental health effects"); Corey Call, *Perceptions of Image-Based Sexual Abuse Among the American Public*, 22 CRIMINOLOGY, CRIM. JUST., L. & SOC.'Y 30, 32 (2021) (finding that image-based sexual abuse causes "powerlessness, shame, humiliation, anxiety, depression, loss of self-esteem, eating disorders, and other psychological issues"); Mudasir Kamal & William J. Newman, *Revenge Pornography: Mental Health Implications and Related Legislation*, 44 J. AM. ACAD. PSYCH. L. 359, 359 (2016).

184   *See* Bates, *supra* note 183 (demonstrating the similar effects of revenge porn and sexual assault); *see also* Nandini Comar, *The Rise of Revenge Porn*, GARBO (Oct. 29, 2021), https://www.garbo.io/blog/revenge-porn [https://perma.cc/V82U-BPMB] ("The repercussions of such abuse are the same as in-person sexual assault.").

185   Maya Elaine Higa (@mayahiga), X (Jan. 31, 2023, 7:55 PM), https://twitter.com/mayahiga/status/1620586546083803136 [https://perma.cc/YRM3-WN8Q].

186   *See* Rousay, *supra* note 15.

187   ANOTHER BODY, *supra* note 70.

188   *See* Message from Jane Doe to Abigail George (Jan. 5, 2024, 17:35 EST) (on file with *Columbia Journal of Gender & Law*); ANOTHER BODY, *supra* note 70.

consented.[189] Even the *Model Penal Code* now acknowledges that someone does not have to say "no" or physically resist to show they do not consent.[190] When a deepfaker makes a pornography video without the consent of the woman featured, they know she is unable to object to it. This is especially true when the deepfaker knows the woman personally and could easily ask for permission but chooses not to, knowing she would almost certainly refuse.[191]

Consent's contextual nature refers to the fact that "sharing information for one purpose is not permission to share for other uses."[192] Thus, it should be glaringly obvious that a woman who made her photographs publicly available online did not consent to her image being used to turn her likeness into pornography. While "passive acquiescence" is a long and enduring indication of consent in rape law,[193] defamation recognizes the contextual nature of consent regarding photographs.[194] When a plaintiff "had no reason to anticipate" a photograph's distortion, a defendant "should have . . . shown [them] before publication."[195] Therefore, as the next section continues to show, deepfakes constitute egregious public sexual offenses implicating fundamental rights that are protected by defamation law.

---

189   Mustafa Kasubhai argues that rather than focusing on force, sexual assault law should focus on a requirement of affirmative consent: "rape law should concentrate on consent rather than force" and "non-consent must be presumed." Kasubhai, *supra* note 153, at 41.

190   MODEL PENAL CODE § 213.0(2) (AM. L. INST. 2021).

191   *See supra* note 28 and accompanying text (showing the growing rate of deepfakers targeting people they know personally).

192   Citron & Franks, *Criminalizing Revenge Porn*, *supra* note 16, at 355. For a discussion about the contextual nature of privacy, see HELEN NISSENBAUM, PRIVACY IN CONTEXT (2009).

193   *See* Susan Estrich, *Rape*, *in* FEMINIST JURISPRUDENCE 158, 177 (Patricia Smith ed., 1993) ([T]he law puts a special burden on the rape victim to prove through her actions her nonconsent . . . .); Perez v. State, 94 S.W. 1036, 1038 (Tex. Crim. App. 1906) (finding that "[m]ere copulation, coupled with passive acquiescence, is not rape . . . . [E]very exertion in her power under the circumstances must be made to prevent the crime, *or* consent *will be presumed*."); Kasubhai, *supra* note 153, at 53 (noting that still today, in the United States, "few courts will consider verbal non-consent sufficient to convict a man of rape").

194   *See, e.g.*, Burton v. Crowell Pub. Co., 82 F.2d 154, 156 (2d Cir. 1936).

195   *Id.*

B. Deepfakes, Dignity, and Reputation

In contrast to the law's recent recognition of a woman's right to sexual autonomy, reputation is a long-established core right—legally protected by the law of defamation.[196] Although it is uncontested that reputation is the sole protected interest underlying the tort of defamation,[197] reputation itself is notoriously elusive.[198] Definitions vary,[199] but a common thread is "the social apprehension that we have of each other."[200] The Second Restatement of Torts defines defamatory communication as that which "tends to harm the reputation of another as to lower [them] in the estimation of the community."[201] Scholar David Rolph argues that "reputation as celebrity" is appropriate in the digital and information era, wherein the community within and by which one's reputation rises or falls is no longer a single social class or group but a global village.[202] Reputation as celebrity recognizes that one interacts with one's community primarily through media, and so "all reputation, and consequently all damage to reputation, is mediated, that is, it occurs through some form of medium of communication."[203] As publicized violations of a sexual nature, pornographic deepfakes inflict reputational harm by tarnishing how the targeted woman is seen by others in her community.

---

196   *See* DAVID ROLPH, REPUTATION, CELEBRITY AND DEFAMATION LAW 1 (2008).

197   *See id.* ("It is clear that reputation is the sole interest directly protected by the law of defamation." (citations omitted)).

198   *See* Robert C. Post, *The Social Foundations of Defamation Law: Reputation and the Constitution*, 74 CAL. L. REV. 691, 692 (1986) ("Reputation . . . is a mysterious thing."); HAROLD LUNTZ & DAVID HAMBLY, TORTS: CASES AND COMMENTARY 1.4.03 (1985) (describing reputation as "nebulous yet much cherished"); ROLPH, *supra* note 196, at 1 (noting that "there has been scant attention given to this crucial concept").

199   *See, e.g.*, Reputation, MERRIAM-WEBSTER, https://www.merriam-webster.com/dictionary/reputation [https://perma.cc/QQ2D-PJWB] (defining reputation as "overall quality or character as seen or judged by people in general" and "a place in public esteem or regard: good name"). "Reputation" comes from the Latin verb *reputare*, meaning "to take into consideration." *Id.* Robert Post famously introduced a three-part typology of legally cognizable types of reputation: reputation as property, reputation as honor, and reputation as dignity. *See* Post, *supra* note 198, at 693.

200   *See* Post, *supra* note 198, at 692.

201   RESTATEMENT (SECOND) OF TORTS § 559 (AM. L. INST. 1965).

202   ROLPH, *supra* note 196, at 172.

203   *Id.* at 172. Reputation as celebrity is premised on "the plaintiff interacting with his or her audience through the media." *Id.*

Deepfakes publicly allege that their sexual depiction is the truth of the woman's sexuality. For example, the app DeepNude claimed to be "[r]evealing truth hidden under clothing."[204] The goal of deepfakes is not to depict a fictional world cut off from reality or to be satirical—they aim to be authentic portrayals of the target's intimate life. Creators focus on making deepfakes as realistic as possible, with many forums dedicated to troubleshooting glitches that expose their synthetic nature. In other words, deepfakes may have a false manner of creation, but they do not purport to depict a false image of the victim's sex life. Instead, deepfakes say: *this is how this person's naked body looks*, and *these are their facial and emotional responses to sex*.

Deepfakes' synthetic creation does not preclude them from alleging truths about their subjects. As a genre, pornography straddles fact and fiction. Sociologist Matthew Ezzell analyzes empirical data to conclude that consumers view pornography as real and not just fantasy.[205] Even if pornography consumers "know" that the subjects are actors, "[t]he majority of consumers consider the performers and the sex to be real."[206] They understand the depictions as a model for real sexual life.[207] This model bleeds off the screen, "directly shaping" real-world social interaction and behaviors.[208] Defamation law recognizes that something can be fictional and still presuppose or imply derogatory and false things.[209] So, regardless of whether viewers know that a deepfake is synthetic, they perceive an alleged truth of the featured person, which often imputes ideas like "Emma Watson enjoys rape" or "[Alexandria Ocasio-Cortez] enjoys choking."[210] This is not an abstract philosophical idea; it is a concrete lived experience for victims. Consider the hundreds of men who sent private

---

204   Meikle, *supra* note 14, at 73.

205   *See* Matthew B. Ezzell, *Pornography Makes the Man*, *in* The Philosophy of Pornography: Contemporary Perspectives 17 (Lindsay Coleman & Jacob Held eds., 2016).

206   *Id. See also* Mikkola, *supra* note 158.

207   *See* Ezzell, *supra* note 205, at 24 (showing a national poll that pornography is leading to habituation, desensitization, increased tendencies to objectify women in and outside of pornography, pushing sexual partners to try positions seen in pornography, and pressuring women to have sex).

208   *Id*. Following a comprehensive national survey of men who consume pornography, journalist Pamela Paul testified that the men "found the way they looked at women in real life warping to fit the pornography fantasies they consumed on screen." Testimony of Pamela Paul, Written Testimony to Congress (Nov. 10, 2005), https://www.judiciary.senate.gov/imo/media/doc/paul_testimony_11_10_05.pdf [https://perma.cc/83VC-P2PQ].

209   *See* Mikkola, *supra* note 158, at 165 ("one can communicate [a] view . . . without explicitly saying so"). Mikkola argues that "pornography may well say harmful things about women and sexuality, even if it does not do so explicitly and even if it purports to be fictional." *Id. See* Section III.2.

210   [John Doe Websites 2 & 3] (links on file with *Columbia Journal of Gender & Law*).

messages to deepfake-pornography victims Taylor Klein and Rana Ayyub asking their rates for sex or asking "to come over to fuck" them.[211]

In addition to aiming at exposing a truth about their subject, deepfakes publicly claim knowledge of the subject.[212] Philosopher Rae Langton proposes that pornography generates "maker's knowledge" or "the special knowledge someone has of something, in virtue of making that thing."[213] Langton compares maker's knowledge to a blueprint that a designer or architect uses, as their beliefs are a blueprint for the real world.[214] Maker's knowledge suggests "an agent who is maker of his own actions" or "somehow a maker of the actions of others."[215] Deepfake creators pick from the library of blueprints of existing pornography videos and fit a real woman onto one, creating a blueprint of that woman's sexual life. The target's sexual life—her desires, fantasies, and pleasure—are no longer her own. They belong to the blueprint creator, who makes them according to his will.[216] Deepfakes that play on the target's professional capacities not only assert knowledge of her sex life but also her professional identity. For example, there are videos titled "[G]lobal [W]arming has [C]reated a Monster—And Greta [Thunberg] Fucks It," "Alexandria Ocasio-Cortez Goes to 'Great Lengths' to Snatch Up New Voters," and "Kate Middleton—Duchess of Anal."[217] These explicitly play on the target's professional capacities, publicly associating her professional identity and achievements with sex.

---

211   Ayyub, *supra* note 1; MY IMAGE MY CHOICE, *supra* note 31.

212   Pornography has been theorized both as a speech act and as a social ontology. The speech-act view of pornography, drawing on J. L. Austin, views pornography as doing something in addition to saying something. Perlocution is the idea that speech acts can cause additional extra-linguistic effects beyond the literal meaning of the words spoken. *See* RAE LANGTON, SPEECH ACTS AND UNSPEAKABLE ACTS (1993). Mari Mikkola wants to shift the conversation about pornography from speech act theory to social ontology. Mikkola suggests an artifactual analysis of pornography: "The term 'pornography' does not pick out an abstract entity but an array of concrete things—something that a proper philosophical understanding of pornography in my view should bear closely in mind." MIKKOLA, *supra* note 158, at 240.

213   Rae Langton, *Speaker's Freedom and Maker's Knowledge*, *in* SEXUAL SOLIPSISM: PHILOSOPHICAL ESSAYS ON PORNOGRAPHY AND OBJECTIFICATION 289, 289–310, 301 (2009).

214   *See id.* at 308. *See also* MIKKOLA, *supra* note 158, at 251–52 (building on Langton's contention to argue that "who the maker is makes a difference").

215   Langton, *supra* note 213, at 308.

216   *See id.* at 216 (The subject of the video "is somehow an extension of himself" and "her autonomy has been excluded from the picture.").

217   [JOHN DOE WEBSITES 2 & 3] (links on file with Abigail George).

Applied to deepfake pornography, maker's knowledge reveals how deepfakes are not just harmless fantasies but claims of knowledge over their subjects.[218] Beliefs can both aim at truth and count as knowledge.[219] Thus, when deepfake creators strive for (and typically achieve) hyperrealistic depictions of women—whether public figures or personal acquaintances—they both aim at truth and assert knowledge about her in a public manner. Nominal indications that a video or image is "fake" do little to change that.[220]

So, deepfakes are false statements asserting a truth, disseminated to a person's community, and affecting how she is seen in the eyes of others. As a result, victims lose autonomy or control over their reputations. For example, one affected woman described: "I founded a non-profit animal sanctuary at 22 years old. I have raised over $1 million for conservation work at 24 years old. I have created zero sexual content in my three years on Twitch. Despite this, my face was stolen so men could make me into a sexual object to use for themselves."[221] As her testimony suggests, deepfakes displace reputations that victims worked to build (founding a non-profit, etc.) with images of sexualization.[222]

There are several objections to characterizing nonconsensual deepfake pornography as a sexual offense incurring reputational harm. First, people may argue that deepfake pornography is closer to a sexual fantasy than a sexual offense.[223] But deepfakes are categorically distinct from sexual fantasies for the simple reason that they are shared with

---

218   *See* Langton, *supra* note 213, at 305 (contending that non-synthetic pornography consists of "justified beliefs, some true and justified beliefs and perhaps some knowledge").

219   *See id.* at 304 ("[The] self-fulfilling nature of a belief about what we are going to do is what enables that belief both to aim at truth, and to count as knowledge.").

220   *See* Saner, *supra* note 181 (describing the impact as being unrelated to whether people know it is not real).

221   Maya Elaine Higa (@mayahiga), X (Jan. 31, 2023, 7:55 PM), https://twitter.com/mayahiga/status/1620586546083803136 [https://perma.cc/YRM3-WN8Q].

222   Noelle Martin described: "They are literally robbing your right to self-determination, effectively, because they are misappropriating you, and your name and your image and violating you permanently." Max Aitchison, *Aussie Student's X-Rated Horror After Innocently Googling her own Name to Discover Someone Had Done the Unthinkable - and Her Life Will Never Be the Same Again*, Daily Mail (Apr. 17, 2023), https://www.dailymail.co.uk/news/article-11981501/Aussie-students-horror-Googling-life-never-again.html [https://perma.cc/9A4W-TT2N]. Moreover, recall that journalist Ayyub was discredited from her work. *Supra* notes 1–13 and accompanying text.

223   *See generally* Carl Öhman, *Introducing the Pervert's Dilemma: A Contribution to the Critique of Deepfake Pornography*, 22 Ethics & Info. Tech. 133 (2020).

others.[224] If someone makes a nonconsensual deepfake and no one else knows about it, the creator is within a *locus poenitentiae*, and little (if any) harm occurs.[225] As previously explained, the depicted person is harmed when they, or others, see or know about it. On a related note, defendants are likely to contend that when deepfakes are labeled as such, no harm ensues since viewers know they are fake.[226] But as previously explained, "fake" refers to deepfakes' synthetic creation, not to the claims they make about the woman depicted.[227]

III. A Defamation Framework

The previous Parts described how deepfake pornography—a sexual offense infringing upon the victim's right to reputation—presents a novel legal issue that has evaded existing legal protections. This Part shows how that harm has a legal remedy in the time-tested principles of defamation law. It starts with the less-contested ways deepfakes fit into the defamation framework and ends with defamation's biggest love-hate relationship: freedom of speech. Section III.A addresses publication, mental state, and damages in defamation claims for deepfakes. Section III.B shows how deepfake pornography, in constituting a fundamental harm to reputation, qualifies as a defamatory statement. Finally, Section III.C examines balancing the plaintiff's right to reputation with the defendant's freedom of speech.

A. Lower-Hanging Fruit

Notwithstanding minor interstate variations, the essential components of a defamation claim are a false and defamatory assertion regarding another, publication to a third party, at

---

224   *See id.*

225   In Latin, "locus poenitentiae" means "place where one can change their mind or attitude." John Kaplan, Robert Weisberg & Guyora Binder, Criminal Law: Cases and Materials (9th ed. 2021). In criminal law, it refers to the concept that crimes occurring solely in the minds or thoughts of defendants are not legally recognizable. *See id.*

226   The fantasy critique claims that pornography or deepfake pornography does not say anything about women because it is fictional. *See* Alan Soble, *Pornography*, 11 Soc. Theory & Prac. 61, 73 ("As fantasy, pornography is a vision of the way things ought to be or could be, regardless of the way things actually happen to be[:] Pornography cannot [therefore] be charged with falsely and maliciously describing women.").

227   "Deepfake" is a misnomer since the videos assert truths of their subjects. So even when the viewer or maker knows that it is a deepfake and synthetically created, they aim at (and assert) a truth about a subject. *See supra* note 204 and accompanying text.

least negligence on the part of the defendant, and harm arising from the publication.[228] This section tackles the last three. Section III.B addresses the first element.

1. Publication

Publication refers to the act whereby the defamatory content is either intentionally or negligently relayed to a third person.[229] Posting deepfakes online amounts to publication.[230] This is consistent with their injury: the gravamen of the harm occurs when deepfakes are shared and when the targeted person sees the content themselves.[231] Even in cases where deepfakes are used for (s)extortion or blackmail, the crux of the threat is that they will be disseminated to a third party.[232] Moreover, creators incur liability if a third party sees the content in a non-purposeful, negligent way, such as by looking at someone's phone[233] or glancing at a left-open tab.[234] Finally, people who disseminate deepfakes they did not personally create may also face liability under the "republication rule."[235]

228　*See* Restatement (Second) of Torts § 558 (Am. L. Inst. 1965).

229　*See id.* § 577 cmt. a.

230　*See* Rolph, *supra* note 196, at 71–72.

231　*See supra* Part II.

232　*See* Citron, *Continued (In)visibility*, *supra* note 78, at 342.

233　*See* Message from Jamie Doe to Abigail George (Jan. 6, 2024, 15:35 EST) (on file with *Columbia Journal of Gender & Law*) (explaining that a friend found deepfakes of her while looking through her boyfriend's photographs).

234　In a now notorious incident, Atrioc, a famous YouTube live streamer, accidentally left open a tab showing deepfake pornography of another famous live streamer. *See* Max Miceli, *Atrioc Apologizes A Second Time With Pledge To Help Women Affected By Deepfake Websites*, Dot eSports (Feb. 1, 2023), https://dotesports.com/streaming/news/atrioc-apologizes-a-second-time-with-pledge-to-help-women-affected-by-deepfake-websites [https://perma.cc/6XS5-SYLC].

235　*See* Restatement (Second) of Torts § 581(1) (Am. L. Inst. 1965) ("[O]ne who only delivers or transmits defamatory matter published by a third person is subject to liability [as if he had originally published it] if, but only if, he knows or has reason to know of its defamatory character."). The republication rule has largely been curtailed for internet postings. *See* Benjamin C. Zipursky, *The Monsanto Lecture: Online Defamation, Legal Concepts, and The Good Samaritan*, 51 Valparaiso U.L. Rev. 1, 5–6 (2016). *But see* Matthew D. Bunker & Clay Calvert, *"Defamation Live": The Confusing Legal Landscape of Republication in Live Broadcasting and a Call for a "Breaking News Doctrine,"* 39 Colum. J.L. & Arts 497 (2016) (discussing the relevance of the republication rule for live broadcasts).

2. Mental State of Perpetrators

In *New York Times v. Sullivan*, the Supreme Court established that defamation of public figures requires showing "actual malice," a term of art meaning "knowledge that [the alleged defamatory statement] was false or [] reckless disregard of whether it was false."[236] Crucially, the reference point in determining fault is the falsity of the statement—not the intended effects on the plaintiff. Thus, defendants in defamation actions involving public figures are subject to strict liability as to whether the statements defame the plaintiff, and their mental state must only be proved vis-à-vis the statement's falsity. In other words, even when the defendant believed the defamatory statements conveyed the victim in a neutral or positive light, they are liable as long as others reasonably understood a defamatory meaning referring to the plaintiff.[237] The fault standard is even lower for defaming private persons. Ten years after *Sullivan*, the Court in *Gertz v. Welch* refused to extend the actual malice standard to cases involving private individuals, giving states the green light to "define for themselves the appropriate standard of liability."[238] While states have adopted varied approaches, most apply a negligence standard.[239]

---

236   New York Times Co. v. Sullivan, 376 U.S. 254, 279–80 (1964); *see also* McCafferty v. Newsweek Media Grp., Ltd., 955 F.3d 352, 359 (3d Cir. 2020) ("'Actual malice' is a term of art that does not connote ill will or improper motivation."). Even "the outer limit of malice," reckless disregard, means that the defendant "entertained serious doubts as to the truth of the statement or . . . had a subjective awareness of probably falsity." Kendall v. Daily News Pub. Co., 716 F.3d 82, 89, 91 (3d Cir. 2013).

237   *See* Prosser and Keeton on the Law of Torts 809 (W. Page Keeton et al. eds., 5th ed. 1984). As long as a defamatory meaning referring to the plaintiff is reasonably understood by others, the publisher has strict liability, regardless of whether they intended it as defamatory. *See id.* ("If a defamatory meaning, which is false, is reasonably understood, the defendant publishes at his peril, and there is no possible defense except the rather narrow one of privilege."). *See* W.S. Holdsworth, *A Chapter of Accidents in the Law of Libel*, 57 L.Q. Rev. 74, 83–84 (1941) (noting that defendants are strictly liable even if they publish statements that are innocuous on their face, could reasonably be construed to be innocuous on its face, or not intended to refer to the plaintiff); Jeremiah Smith, *Jones v. Hulton: Three Conflicting Views as to Defamation*, 60 U. Pa. L. Rev. 365, 461–63 (1912).

238   Gertz v. Welch, 418 U.S. 323, 347, 349 (1974) (noting that showing knowledge of falsity or reckless disregard for the truth is required for awarding of punitive damages).

239   *See* Restatement (Second) of Torts § 580B (Am. L. Inst. 1965) (summarizing the fault standard for private individuals as the defendant "act[ing] negligently in failing to ascertain" whether a statement was false); *see also, e.g.*, Haueter v. Cowles Pub. Co., 811 P.2d 231, 236 (Wash. Ct. App. 1991).

A deepfake creator's mental state as to any intended harm to the plaintiff varies significantly from case to case.[240] But in every case, creators know the content is fake since they use synthetic processes to make it.[241] Accordingly, public figures who bring defamation cases for deepfakes will have no problem showing that creator-distributors acted with actual malice. In cases of private individuals, plaintiffs will have an even easier time meeting the negligence bar and can likely sweep in defendants who distribute deepfakes without having created them.

More broadly, establishing the requisite mens rea for perpetrators has long been a hurdle in cases of gender-based cyberviolence[242] and is likely to be a significant pitfall in criminal convictions for deepfake pornography or under the federal cause of action proposed in the Preventing Deepfakes of Intimate Images Act.[243] At recent oral arguments in a criminal cyberstalking case in which a stalker had sent thousands of threatening messages over the course of years, Chief Justice John Roberts suggested that the message "[s]taying in cyber life is going to kill you. Come out for coffee. You have my number," could be an innocuous and solicitous "invitation to get off the computer."[244]

---

240   *Compare* Danielle Citron, Hate Crimes in Cyberspace 17 (2014) (noting that perpetrators "know that women will be seen as sluts . . . . [and] make them unemployable, undateable, and at risk for sexual assault") *with* Fido & Harper, *supra* note 47, at 41 (claiming that deepfake creators may be motived by several factors such as curiosity, compulsivity, or gratification of sexual fetishes). *See also* [John Doe Website #4] (on file with *Columbia Journal of Gender & Law*) (deleting all his videos, a famous creator acknowledged that he "feel[s] like the total piece of shit I am" and "the best course of action" is to "wipe my part off the internet and help decrease the number of future videos of those involved").

241   Those who disseminate but do not create may not meet the actual malice standard for public figures, but they would likely still meet the negligent standard for private individuals.

242   *See* Citron, *Continued (In)visibility*, *supra* note 78, at 363. Interestingly, seventy-four percent of regular viewers of deepfake pornography report that they do not feel guilty about it. *2023 State of Deepfakes*, *supra* note 23. Yet over two thirds said they would feel "shocked and outraged by the violation of someone's privacy and consent" if someone close to them became a victim of deepfake pornography, and nearly three quarters of respondents would report the incident to the authorities. *Id.* The dissonance between consuming it and thinking about the consequences if someone close to them was affected reveals how cyberspace acts as a moral vacuum.

243   H.R. 3106, 118th Cong. (as introduced to the House, May 5, 2023). *See supra* text accompanying note 132.

244   Transcript of Oral Argument at 53–56, Counterman v. Colorado, 600 U.S. 66 (2023) (No. 22-138). Counsel for Counterman compared the statements to a child saying, "I will kill you," after their sibling took the last brownie. *Id.* at 32. Likewise, Justice Alito considered how such statements could take place in the context of a mystery novel. *See id.* at 32–33. Justice Thomas lamented the growing hypersensitivity of society, saying some of the stalker's statements were not "threatening in and of themselves, and yet someone could be triggered by those statements or hypersensitive about those statements and feel threatened." *Id.* at 73. For an

There is a real threat that judges will interpret deepfakes similarly, especially considering that ninety percent of judges have never heard of deepfake pornography.[245] So while deepfake websites' and creators' claims that the videos are "not meant to harm or humiliate anyone"[246] may shield them from criminal liability, such lip service will not bar defamation claims, which only require showing knowledge of the statement's falsity, not intent to harm the victim.

3. Making Victims Whole Again: Damages for Non-Economic Harms

Defamation holds a peculiar place in the common law for its recognition of non-economic harms in calculating damages and—subject to a few exceptions—not barring suits that do not show pecuniary loss.[247] Plaintiffs who bring defamation cases for deepfake pornography can receive damages for reputational harm, emotional distress, medical expenses for psychiatric help, attorney's fees, and punitive damages.[248] Defamation adheres to a well-established principle—dating back to the Middle Ages and now applying in cyberspace—that the greater the visibility and accessibility of the publication, the greater the harm and corresponding remedy.[249] Modern courts acknowledge the compounding

---

overview of the impact of the decision in this case on cyber gender abuse, see Citron, *Continued (In)visibility*, *supra* note 78, at 359–64.

245   *See* Vanessa Caldwell, *Her Face Was Deepfaked Onto Porn. When Police Wouldn't Help, She Did Her Own Investigation*, CBC Docs (Nov. 22, 2023), https://www.cbc.ca/documentaries/the-passionate-eye/her-face-was-deepfaked-onto-porn-when-police-wouldn-t-help-she-did-her-own-investigation-1.7035523 [https://perma.cc/MX8K-WTSM].

246   *See* [John Doe Website #5] (link on file with *Columbia Journal of Gender & Law*). In fact, the major deepfake pornography websites appear to have cursorily consulted lawyers. They include disclaimers such as "the videos are created for entertainment and learning purposes only." [John Doe Website #2] (link on file with *Columbia Journal of Gender & Law*).

247   Historically, the common law of defamation occurred in the Star Chamber, where defendants could be made to pay damages to a plaintiff as well as receive criminal punishments. *See* Rolph, *supra* note 196, at 52. "There was no necessary relationship with provable damage to reputation and the damages awarded . . . . [It] developed as a pragmatic remedy." *Id.* at 79. In the Middle Ages, slander, because of its transient nature, required proof of "special damages" or economic harm to recover. Leslie Yalof Garfield, *The Death of Slander*, 35 Colum. J. L. & Arts 17, 17 (2012). *See also* Gertz v. Robert Welch, Inc., 418 U.S. 323, 349 (1974) ("The common law of defamation is an *oddity* of tort law, for it allows recovery of purportedly compensatory damages without evidence of actual loss." (emphasis added)).

248   *See* Rolph, *supra* note 196, at 77–83 (providing an overview of defamation damages).

249   *See* Restatement (First) of Torts § 621, cmt. c (Am. L. Inst. 1938) (noting that in determining the amount of damages, the juror or other trier of fact "may consider the area of dissemination and the extent and

harm to reputation as defamatory statements spread in cyberspace.[250] For example, in Rudy Giuliani's 2023 election-fraud defamation case, the plaintiffs were awarded over $32 million to compensate for the reputational harm incurred from statements Giuliani made on Twitter.[251] Thus, in deepfake cases, plaintiffs' damages stand to increase as the deepfake spreads online.[252] Many perpetrators spread deepfakes through bots that send anonymous emails and messages to people in the victim's community.[253] Finally, as the deepfake-pornography industry becomes increasingly monetized, more targeted women will be able to add to their damages by showing pecuniary loss.[254]

Of course, in cases of sexual abuse, it is difficult to ever make plaintiffs "whole again." But defamation damages at least reflect the unique harms inflicted by allegations related to the plaintiff's sex life. In fact, sex is no stranger to defamation law. There is a centuries-old, plentiful (and occasionally troubling) body of case law arising from imputations related

---

duration of the circulation of the publication"); Cantu v. Flanigan, 705 F. Supp. 2d 220, 227–28 (E.D.N.Y. 2010) (listing "the extent to which the statements were circulated" as a factor that the jury may consider when calculating damages); ROBERT D. SACK, SACK ON DEFAMATION: LIBEL, SLANDER, AND RELATED PROBLEMS § 10.5.1 (2008) (discussing the criteria and proof juries may consider).

250   *See* ROLPH, *supra* note 196, at 73.

251   Guiliani tweeted suggesting that two election workers engaged in election fraud in Georgia, leading to them receiving scores of abusive and harassing phone calls, as well as messages and attacks by Trump supporters. *See* Eileen Sullivan, *Jury Orders Giuliani to Pay $148 Million to Election Workers He Defamed*, N.Y. TIMES (Dec. 15, 2023), https://www.nytimes.com/2023/12/15/us/politics/rudy-giuliani-defamation-trial-damages.html [https://perma.cc/H8FN-RVD5]. Moreover, in the recent Jean Carroll-Donald Trump defamation case, the jury awarded Carroll $1.7 million for reputational harm. *See* Rachel Schilke, *Jury Orders Trump to Pay $5 Million in Damages to Carroll for Battery and Defamation*, WASH. EXAM'R (May 9, 2023), https://www.washingtonexaminer.com/news/1638923/jury-orders-trump-to-pay-5-million-in-damages-to-carroll-for-battery-and-defamation/ [https://perma.cc/B8VF-RP5T].

252   *See* ROLPH, *supra* note 196, at 73 ("[P]ublishers who chose to publish defamatory matter online did so in the full knowledge that the greater the coverage, the greater the potential profit and thus the greater the risk of defamation that such publishers ought to consider or bear." (citation omitted)).

253   *See* Message from Jane Doe to Abigail George (Jan. 5, 2024, 17:35 EST) (on file with *Columbia Journal of Gender & Law*).

254   Many videos on deepfake pornography websites, such as MrDeepFakes, are shortened versions of longer videos available for purchase on Fan-Topia, a platform allowing online creators to charge for their content. *See* Kat Tenbarge, *Found Through Google, Bought with Visa and Mastercard: Inside the Deepfake Porn Economy*, NBC NEWS (Mar. 27, 2023), https://www.nbcnews.com/tech/internet/deepfake-porn-ai-mr-deep-fake-economy-google-visa-mastercard-download-rcna75071 [https://perma.cc/MNX7-FQ5A]. Creators on MrDeepFakes also accept requests and advertise for paid positions to help them create content. *See id.*

to sexual behavior.[255] In early slander courts, female plaintiffs commonly brought slander actions against male defendants for allegations of unchastity or adultery, most commonly being called a "whore."[256] These courts "protect[ed] the interests of women rather than men,"[257] and ninety percent of cases were brought by women.[258] In the early nineteenth century, a gendered reform movement led to the passing of the Slander of Women Act in several common law countries and American states, declaring imputations of unchastity or adultery towards women defamatory without needing to show economic loss.[259] While the normative values underlying defamation cases for deepfake pornography would differ from such sexual slander cases, the similarities are striking: women using defamation to fight against sexual violence when other legal mechanisms have failed them.

B. Defining Defamatory: Moral Fault, Moral Standards, and Ridicule

Although the common law varies by state, prevalent formulations establish a statement as defamatory if it tends to "harm the reputation of another as to lower [them] in the estimation of the community" or exposes them "to hatred, contempt, or ridicule."[260] Deepfake pornography plaintiffs can meet this standard in three ways: by showing that the

---

255   *See generally* Laura Gowing, Domestic Dangers: Women, Words, and Sex in Early Modern London (1999) (describing the history of women and defamation in early modern England); *see also* S.M. Waddams, Sexual Slander in Nineteenth-Century England: Defamation in the Ecclesiastical Courts, 1815–1855 xii (2000) (providing a historical analysis of defamation case papers in ecclesiastical courts in the first half of the nineteenth century).

256   Waddams, *supra* note 255, at 66, 121 (noting that the most common insult was "whore," and in some towns almost everyone must have known someone involved in a sexual slander defamation suit).

257   *Id.* at xii ("[It] was a jurisdiction that put a certain amount of power directly into women's hands, even where social stature and financial resources were lacking.").

258   *Id.*; *see also* Jessica Lake, *Whores Aboard and Laws Abroad: English Women and Sexual Slander in Early Colonial New South Wales*, 35 Gender & Hist. 916, 917 (2023) (noting a "pronounced gender pattern" in slander cases in the first half of the nineteenth century in England).

259   North Carolina was the first jurisdiction to pass the Act, in 1808. Lake, *supra* note 258, at 916. *See also* Slander of Women Act, 1891, 54 & 55 Vict., c. 51 (UK) (establishing that imputing "unchastity" or "adultery" to a woman was per se defamation); Alexandros Antoniou & Dimitris Akrivos, *Homosexuality, Defamatory Meaning and Reputational Injury in English Law*, *in* Diverse Voices in Tort Law 175, 189–90 (Kirsty Horsey ed., 2024) (describing how the term "lesbian" came to be defamatory under these acts).

260   Restatement (Second) of Torts § 559 (Am. L. Inst. 1965); Parmiter v. Coupland, 151 ER 340 (1840). *But see* Scott v. Sampson, 8 QBD 491 at 503 (1882) (establishing a more general test for defamation as "a false statement [about a man] to his discredit").

deepfake (1) caused a sizable part of the community to impose moral fault on the plaintiff; (2) generated an immoral association; or (3) exposed the plaintiff to ridicule.

In assessing standards for injury to reputation, defamation is famously flexible due to its intrinsic linkage with subjective moral standards.[261] Societal prejudices have often formed the basis of actionable defamatory claims, including imputations of being homosexual, having mental illness, having HIV/AIDS, and performing abortions.[262] As historical work has shown, defamation acts as a barometer of a community's contingent moral taxonomy.[263] When community morals shift, the law of defamation responds.[264] This is both a hazard and a strength of the common law. Before it is too late and the public exposure of nonconsensual sexual deepfakes becomes far too commonplace to be shunned by the community and collective morality, judges and juries can declare nonconsensual sexual deepfakes defamatory.

Importantly, even when libelous statements are clearly doctored, fictional, or known to be fake, they can still be defamatory. Defamation can often be established through "inference, implication or insinuation."[265] There are always mediations in any medium that acts as a vessel for libelous statements. Thus, in the case of deepfakes, it does not matter whether viewers think the material is synthetic or real. Deepfakes *infer* that the depicted individual is the *type* of person that would engage in the depicted conduct, *insinuate* that the depictions correspond to the plaintiff's real intimate life, or *impute* that the plaintiff's moral worth is equal to their sexuality. In other words, the question is not whether viewers know that the deepfake is synthetically created or not; it is whether they associate the recognizable subject of the video with the conduct depicted in the video. And they unequivocally do.[266]

---

261   Lawrence McNamara, Reputation and Defamation 5 (2007) (noting "deep-seated problems with the legal framework"); Lake, *supra* note 258, at 917 (asserting that English defamation law "is riddled with doctrinal anomalies and unparalleled complexity").

262   *See, e.g.*, Hepburn v. TCN Channel Nine [1983] 2 NSWLR 682 (Austl.); McNamara, *supra* note 261, at 103–04 (noting that the enduring legacy of the "ethical dominance of Christian tradition" has run up against burgeoning hegemony of liberal ethics).

263   *See* Rolph, *supra* note 196, at 12 (arguing that "the concept of reputation changes historically, reflecting social, political, economic, cultural and, most importantly, technological changes.").

264   *See* McNamara, *supra* note 261, at 101.

265   Carwile v. Richmond Newspapers, 82 S.E.2d 588, 592 (Va. 1954).

266   *See supra* notes 238–244 and accompanying text.

The first path to establishing a statement as defamatory asks if the statement injures the plaintiff's reputation in the views of the "right-thinking" person.[267] In contrast to their English and Australian counterparts, American courts tend to apply a sectional standards approach, which instead asks if a substantial part of the population regards the referenced conduct as immoral.[268] Thus, it places no moral judgment on the imputation itself, thereby sidestepping culture wars and prejudice. By extension, declaring sexual deepfakes as injurious to one's reputation does not involve value judgments about the morality of engaging in pornographic sex.[269]

In a strikingly analogous case to deepfake videos referencing the depicted person's employment or school, a federal district court found doctored photographs depicting the plaintiff acting in a sexually explicit matter defamatory; the photographs identified him as a "porn star," and contained references to his employer.[270] The photographs were defamatory despite being undoubtedly faked because they "impute[d] an unfitness for Plaintiff to perform the duties of a youth soccer coach" and "prejudice[d] Plaintiff in his profession or trade."[271] In contrast to the technological prowess behind deepfakes, the photographs at issue were crudely doctored via Photoshop. Thus, deepfakes that impute that the subject should be defined by their sexuality and not their professional accomplishments are defamatory even if viewers know they are fake.

---

267   Kimmerle v. New York Evening Journal, 186 N.E. 217, 218 (N.Y. 1930). *See Developments in the Law: Defamation*, 69 Harv. L. Rev. 875, 885 (1956) ("When an idea would lower the plaintiff in the esteem of a significant segment of the community, the courts have shown no hesitation in finding that its publication will support an action, however small the number of actual recipients, when damage is presumed or special damage is shown." (footnotes omitted)).

268   Such an approach is reflected in the Restatement (Second) of Torts, which establishes the standard as "in the eyes of a substantial and respectable minority[.]" § 559 cmt. e (Am. L. Inst. 1965). The sectional standards approach contrasts with the general standards test. The latter asks what "ordinary, decent folk in the community" or the "average sensible citizen" would consider unethical. McNamara, *supra* note 261, at 117. Thus, the sectional standards approach avoids attempts at discerning collective, unified community moral standards. In so doing, it can sidestep culture wars and prejudice and is better suited to a pluralistic society. *See id.*; Lyrissa Barnett Lidsky, *Defamation, Reputation, and the Myth of Community*, 71 Wash. L. Rev. 1, 7–8 (1996) (arguing that a substantial and respectable minority standard "embodies the traditional liberal values of tolerance and respect for diversity necessary in a multi-cultural, multi-ethnic society.").

269   *See* Shaik, *supra* note 41 (arguing that using a defamatory approach means arguing that pornographic content harms a woman's reputation, thereby "perpetuat[ing] the patriarchal notion that it is wrong for women to be expressly sexual").

270   *See* Tharpe v. Lawidjaja, 8 F. Supp. 3d 743, 778, 786 (W.D. Va. 2014).

271   *Id.* at 786.

Second, statements can be defamatory via immoral association without suggesting the plaintiff bears moral fault. Consider the famous case brought by Princess Youssoupoff after the film *Rasputin, the Mad Monk* depicted Rasputin raping her.[272] The court found the film defamatory even though being raped does not impute any moral discredit on the victim's part.[273] Two relevant principles arise from this case. First, statements can be defamatory even if they do not suggest any fault on the part of the plaintiff.[274] Being raped, even though it did not imply "moral discredit," tarnished Princess Youssoupoff's reputation.[275] Second, defamation can apply to largely fictionalized depictions—even ones that are heavily dramatized and aggrandized. Most people who watch Hollywood blockbusters take the veracity of their historical depictions with a grain of salt. Yet because the film suggested there was a kernel of truth in its depiction of Princess Youssoupoff, it affected her reputation.[276] And since the depiction implied a "diminution of moral worth," it was defamatory.[277]

Finally, defamation cases premised on exposing the plaintiff to "ridicule" can succeed even when there is no negative moral judgment, no imputed moral fault, and an obvious mistake such that no one would think the material was true.[278] These principles have already

272   *See* Youssoupoff v. Metro-Goldwyn Mayer (1934) 50 TLR 581, CA.

273   *See id. See also* MCNAMARA, *supra* note 261, at 142–47.

274   McNamara writes that "*Youssoupoff* shows that the court thought that female sexual purity could be lost either with or without any fault on a woman's part. A woman's reputation could be disparaged by allegations that she has been sexually promiscuous, but it could also be disparaged by imputations that suggested she had been engaged in or subjected to some kind of sexual act, regardless of her own conduct, will, or volition." MCNAMARA, *supra* note 261, at 147.

275   *Id.* at 145.

276   In a similar fashion, in Snyder v. New York Press Co., 121 N.Y.S. 944 (App. Div. 1910), a newspaper story described how a process server entered the plaintiff's home and saw the plaintiff in the bath. *See id.* at 944–45. Again, the article did not "impute immoral conduct or character" but had "a direct tendency to subject the plaintiff[] to unfavorable comment, to diminish her respectability, to abridge her comfort and enjoyment, and to expose her to public ridicule, such as provokes contempt, not merely such as may be sportive and thoughtless." *Id.* at 945–46.

277   MCNAMARA, *supra* note 261, at 148. McNamara distinguishes "moral fault" from "immoral association," in which a plaintiff's reputation is damaged not by their choice to engage in frowned upon behavior but by "someone else's behavior" or some other circumstances. *Id.* He writes: "[I]t was not the absence or moral fault that was determinative of actionability, but the diminution of moral worth." *Id.* (emphasis omitted).

278   For example, a defendant was liable for defamation for publishing an article on evolution and placing a photograph of the plaintiff next to the gorilla, stating, "Stanislaus Zbyszko, the Wrestler, Not Fundamentally Different from the Gorilla in Physique." Zbyszko v. New York American, Inc. 239 N.Y.S. 411, 412 (App. Div.

been applied in photographs exposing the naked bodies of plaintiffs.[279] The landmark case for assessing defamatory statements based on exposing the plaintiff to "ridicule" is *Burton v. Crowell*, written in 1936 by Judge Learned Hand.[280] The plaintiff, a steeplechaser, had posed for photographs that were published in an advertisement for Camel cigarettes.[281] But, by way of a trompe d'œil, a girth hanging over the stirrup appears as the plaintiff's oversized and exposed penis.[282] The plaintiff's name appeared near the photograph with the caption, "Get a lift with a camel."[283] Judge Hand reasoned that statements exposing a plaintiff "to more than trivial ridicule" were actionable provided that the plaintiff was not "thin-skinned" or a "self-important prig," and the incident was not capable of being "laugh[ed] off."[284] In his typical fashion, Learned Hand cryptically analyzed what constituted the reputational harm despite the "obvious mistake," meaning no one would think it exposed the plaintiff's actual naked body.[285] First, the plaintiff suffered harm by "being known indefinitely as the absurd victim of the unhappy mischance."[286] In fact, "[t]he obvious mistake only added to the amusement."[287] Second, Judge Hand elevated the plaintiff's sense of self-worth to a legally protected interest.[288] He stated, "[t]he gravamen

---

1930). The Court reasoned: "[I]t is not necessary that words impute disgraceful conduct to the plaintiff." *Id.* at 413; *see also* McNAMARA, *supra* note 261, at 166 (commenting that the article in *Zbyszko* "did not attribute any flaw for which the plaintiff was somehow responsible, but it still seemed to suggest that the plaintiff was a lesser person than others even though he bore no moral responsibility for carrying that lesser status.").

279   *See* Ettingshausen v. Austl. Consol. Press, [1991] 23 NSWLR 443, 443–44, 447 (finding that a rugby player who had a naked photograph of him taken in the showers after a game was defamed because "as a result of the exposure of his genitals, he has been held up (or exposed) to ridicule"); *see also* Obermann v. ACP Publ'g Pty Ltd, [2001] 1022 NSWSC ¶ 40 (holding that an Olympic water polo player's exposed breasts in a photograph exposed her to ridicule).

280   *See* Burton v. Crowell Publ'g. Co., 82 F.2d 154, 154 (2d Cir. 1936).

281   *See id.*

282   *See id.*

283   *Id.*

284   *Id.* at 155–56.

285   *Id.* at 155 ("Nobody could be fatuous enough to believe any of these things; everybody would at once see that it was the camera, and the camera alone, that had made the unfortunate mistake").

286   *Burton*, 82 F.2d at 155.

287   *Id.* at 156 (reasoning that "it was the mere association on the plaintiff . . . that was thought to lower him in others' esteem").

288   *See id.*

of the wrong in defamation . . . [is] the feelings, that is, the repulsion or the light esteem, which [the] opinions [of others] engender."[289]

While the "mistake" in deepfake pornography is by no means "obvious," even a narrow reading of *Burton* establishes that deepfakes that are crudely made or have watermarks can still expose the plaintiff to ridicule.[290] Moreover, legally recognizable harm occurs even if people merely have heard about the deepfake without seeing it since the deepfake associates the plaintiff with a pornography campaign in the minds of the public.[291] For example, a boss fired one victim after hearing about her deepfakes.[292] *Burton* also shows that reputational harm can include "the feelings" of deepfake victims.[293] More broadly, the *Burton* principle reveals that defamation suits do not provide recourse to the plaintiff by showing the world that the statements are false. Likewise, the harm from deepfake pornography is not primarily derived from people thinking the videos are real and would not dissipate if they were exposed as false.[294]

C. The First Amendment: Distinguishing Deepfakes from Parody

The last—and often hardest—hurdle for defamation claims is the defendant's freedom of speech, which must always be balanced with the plaintiff's right to reputation. Once a defamatory statement crosses the thin and blurry line from "false statement of fact" to "parody," it enjoys First Amendment protection.[295] Thus, plaintiffs in deepfake cases will need to show that the deepfake pornography contains a false statement of fact, which is a claim about the plaintiff that reasonable viewers perceive to be true. Echoing the lip service paid by some major deepfake hosting websites, defendants in defamation actions

---

289   *Id.*

290   *Id.*

291   *See* Ayyub, *supra* note 1 (describing her family's reaction from hearing about the deepfake).

292   *See* Rousay, *supra* note 15, at 107.

293   *Burton*, 82 F.2d at 156.

294   *See supra* text accompanying note 57. "Fake" is in the name of the medium and in the name of the largest website dedicated to deepfake pornography. [John Doe Website #2] (link on file with *Columbia Journal of Gender & Law*).

295   *See* Gertz v. Robert Welch, Inc., 418 U.S. 323, 340 (1974) (holding that in the realm of defamatory statements, "there is no constitutional value in false statements of fact").

for deepfakes are likely to assert the defense of parody.[296] As an entirely new technology that is only just beginning to be litigated, the contours of whether and when deepfakes qualify as parody will certainly need to percolate in lower federal courts. But previous principles suggest that deepfakes are doing something entirely different from parody.

The Supreme Court fleshed out the parody-false statement of fact distinction in *Hustler v. Falwell*.[297] The material in question, a satirical advertisement in *Hustler* magazine, featured a photograph of Reverend Jerry Falwell, a religious right-wing leader known for his public admonitions about sexual morality.[298] Paralleling an advertising campaign in which celebrities recounted their first time drinking Campari, the advertisement depicted Falwell recounting his "first time" having sex with his mother.[299] The advertisement contained exaggerated and implausible content, such as a goat in an outhouse, and included a disclaimer that it was "not to be taken seriously."[300] Chief Justice William Rehnquist, writing for the majority, upheld the dismissal of Falwell's defamation claim because advertisement parody, albeit "patently offensive and [] intended to inflict emotional injury," is not "reasonably [] interpreted as stating actual facts."[301] Lower circuit courts explain: If "a reasonable reader would not accept the statements at face value, the statements do not cause damage to the plaintiff's reputation."[302]

---

296   *See* [John Doe Website #2] (link on file with *Columbia Journal of Gender & Law*) (claiming the videos are for "entertainment and learning purposes only"); [John Doe Website #2] (link on file with *Columbia Journal of Gender & Law*) (requiring users to acknowledge a disclaimer that the videos "are intended as parody of the celebs portrayed"). Of course, these disclaimers are mere lip service. On a forum post about potential legal consequences, one of the top moderators and creators replied, "It's all just machine learning research ;)." [John Doe Website #5] (link on file with *Columbia Journal of Gender & Law*).

297   *See* Hustler Mag. v. Falwell, 485 U.S. 46 (1988). The false statement of fact framework also applies to IIED claims, as it did in Hustler. *See id.* at 52.

298   *See id.* at 47; Susan Dudley Gold, Parody of Public Figures: Hustler Magazine v. Falwell 14 (2014) (describing Falwell as "a preacher with his own radio and television shows" who was "well known for his religious and political activities," called his political organization the "Moral Majority," and had "proclaimed he had tried to turn the public against pornography 'with every breath in my body'").

299   *Hustler*, 485 U.S. at 48.

300   *Id.*

301   *Id.* at 48.

302   Hogan v. Winder, 762 F.3d 1096, 1106 (10th Cir. 2014) (citing Mast v. Overson, 971 P.2d 928, 933 (Utah Ct. App. 1998)).

In a well-publicized case, Kimerli Jayne Pring, who had recently been crowned Miss Wyoming, sued *Penthouse* magazine for publishing an article in which Miss Wyoming—bestowed with magical sexual powers—performs oral sex on a football player, causing him to levitate.[303] In what was the largest libel award to date, a jury awarded Pring twenty-six million dollars.[304] But the Tenth Circuit overturned the award on appeal because "a reader would . . . have understood that the charged portions were pure fantasy and nothing else."[305] The material "described something physically impossible in an impossible setting."[306]

These precedents, *inter alia*, reveal the judicial philosophy behind requiring defamation claims to be false statements of fact.[307] If a reasonable person interprets the statements as hyperbole or parody, they are protected by the First Amendment.[308] Parody, as defined by copyright law, uses some elements of a prior composition to create a new one that

---

303   *See* Pring v. Penthouse Int'l, Ltd., 695 F.2d 438, 440–44 (10th Cir. 1982). For an account by Pring's attorney, see GERRY SPENCE, TRIAL BY FIRE: THE TRUE STORY OF A WOMAN'S ORDEAL AT THE HANDS OF THE LAW (1986).

304   *See Jury Says Penthouse Magazine Libeled a Former Miss Wyoming*, N.Y. TIMES (Feb. 21, 1981), https://www.nytimes.com/1981/02/21/us/jury-says-penthouse-magazine-libeled-a-former-miss-wyoming.html [https://perma.cc/P7ZC-HMH8].

305   *Pring*, 695 F.2d at 443 ("It is impossible to believe that anyone could understand that levitation could be accomplished by oral sex before a national television audience or anywhere else. The incidents charged were impossible. The setting was impossible.").

306   *Id.*

307   Moreover, in Greenbelt v. Bresler, a well-known real estate developer sued a local newspaper for libel. *See* Greenbelt Co-op. Publ'g Ass'n v. Bresler, 398 U.S. 6, 7–8 (1970). The newspaper had reprinted statements in which city council meeting attendees described Bresler's negotiating position as "blackmail." *Id.* at 7. Justice Stewart reasoned that no reader would believe the council members were actually charging Bresler with blackmail and "even the most careless reader must have perceived that the word was no more than rhetorical hyperbole, a vigorous epithet used by those who considered Bresler's negotiating position extremely unreasonable." *Id.* at 14. Similarly, the Superior Court of New Jersey found that a college newspaper's statements claiming that a school administrator was available on a "whoreline" for "good telephone sex" were not assertions of fact, instead being "treated as a protected expression[s] of opinion." Walko v. Kean Coll. of New Jersey, 561 A.2d 680, 683–84 (Law. Div. 1988). The court opined that "[n]o reasonable person, even glancing at the offending ad, could possibly conclude that it was a factual statement of plaintiff's availability for 'good telephone sex.'" *Id.* at 683.

308   *See* Hustler Mag., Inc. v. Falwell, 485 U.S. 46, 57 (1988) ("But for reasons heretofore stated this claim cannot, consistently with the First Amendment, form a basis for the award of damages when the conduct in question is the publication of a caricature such as the ad parody involved here.").

comments on the original work.[309] Across all the cases in which a defamation claim was struck down for not being a false statement of fact, the alleged libelous material depicted unrealistic—usually, physically impossible—situations.[310] A statement that a plaintiff's fellatio made someone levitate is clearly untrue.[311] Claiming that a moral majority religious leader engaged in drunken incestuous relations in an outhouse is so exaggerated that no one would think it was asserting a truth.[312] Even though the *Hustler* ad provides commentary and refers to the plaintiff (thereby affecting public perception of him), readers are aware of a dissonance between the statements and reality.[313] In contrast, the allure of deepfake pornography is precisely that it realistically exposes the plaintiff's body and sex life.

Once again, technological advancements enable deepfake pornography to be identical to the plaintiff's physical body, distinguishing it from parody.[314] For the women whose identities are appropriated, the most harmful part of deepfakes is their realism.[315] Deepfake pornography asserts something about who the victim is: a sexual object. In contrast to the statement that the plaintiff's fellatio made someone levitate,[316] deepfake pornography is not an impossible claim: reasonable viewers see it as realistic, naked depictions of the victim.[317] It should be painfully evident to judges and juries that these videos serve the

---

309   *See* Campbell v. Acuff-Rose Music, 510 U.S. 569, 580 (1994) ("For the purposes of copyright law, the nub of the definitions, and the heart of any parodist's claim to quote from existing material, is the use of some elements of a prior author's composition to create a new one that, at least in part, comments on that author's works.").

310   *See Pring*, 695 F.2d 443 (describing fellatio with the power to cause levitation); New Times, Inc. v. Isaacks, 146 S.W.3d 144 (Tex. 2004) (describing a six-year-old placed in ankle shackles); Walko v. Kean Coll. of New Jersey, 561 A.2d 680, 683–84 (Law. Div. 1988) (describing a school administrator placing an ad in the school magazine for good telephone sex).

311   *See Pring*, 695 F.2d at 443.

312   *See Hustler*, 485 U.S. at 48.

313   *See id.*

314   *See supra* notes 54–56.

315   *See, e.g.*, Cook, *supra* note 106 (quoting a victim as saying, "When it's Photoshop, it's a static picture and can be very obvious that it's not real . . . [b]ut when it's your own face reacting and moving, there's this panic that you have no control.").

316   *See Pring*, 695 F.2d at 443.

317   *See supra* notes 54–56 (listing studies showing the low detection rate of deepfakes).

primary purpose of realistic sexual exposure of the plaintiff.[318] More broadly, the Court has justified certain protections for defamatory statements to give freedom of expression necessary "breathing space."[319] If the Court is concerned about giving freedom of expression enough room to breathe, it should acknowledge that deepfake pornography is suffocating the women it targets.[320]

CONCLUSION

Deepfakes are particularly pernicious public sexual offenses—whose accuracy, accessibility, and popularity only continue to grow.[321] Deepfakes may be a new phenomenon, but they fit into old defamation principles. What is "fake" about deepfakes is not their purported representation or "message" but their manner of creation or "medium." Something can be fake (in the false sense) but still allege a truth about someone—like a lie. In fact, that is precisely what defamation protects: false statements, alleging a truth, causing reputational harm.[322] Thus, a defamation frame reveals deepfake pornography as a core infringement of noneconomic, dignitary, and reputational interests. Moreover, in a time of "(in)visibility of cyber gender abuse," defamation acknowledges the compounding nature of harm as it spreads in cyberspace and the particularly nefarious nature of sexual imputations.[323] It is not necessary to have large-scale publicized defamation suits à la Johnny Depp-Amber Heard for each of the 100,000 women targeted by deepfakes.[324] Common law rulings do more than resolve disputes between the parties in the action; they announce

---

318   There is an even stronger case when deepfakes appear on conventional porn sites rather than dedicated deepfake sites. *See* New Times, Inc. v. Isaacks, 91 S.W.3d 844, 853–54 (Tex. App. 2002) (holding that determining if statements are false statements of fact depends upon a reasonable person's perception of the entirety of the publication, not merely on individual statements).

319   New York Times Co. v. Sullivan, 376 U.S. 254, 271–72 (1964) (justifying certain protections for erroneous statements as necessary "breathing space" for freedom of expression).

320   *See 2023 State of Deepfakes*, *supra* note 23; *see also supra* notes 14–16 and accompanying text. Professor Mary Anne Franks stated, "There's a massive chilling effect that deepfake pornography has on women's speech because the way to make yourself safer is to censor yourself." Cook, *supra* note 106.

321   *See supra* notes 21–29 and accompanying text.

322   *See* RESTATEMENT (SECOND) OF TORTS § 558 (AM. L. INST. 1965).

323   Citron, *Continued (In)visibility*, *supra* note 78; *supra* Section III.A.3.

324   For an account of the Johhny Depp-Amber Heard defamation case, see NICK WALLIS, DEPP V. HEARD: THE UNREAL STORY (2023).

moral and legal norms.[325] Defamation suits can succeed. When they do, they will generate ripple effects, sending a message to those behind their screens that they are not immunized from legal accountability and that sexual abuse remains legally cognizable even when it occurs in cyberspace. Everyone deserves the ability to construct and direct their sexual lives and reputations—on and off the screen.

---

325   For the criticisms of the dispute resolution model of litigation, see generally Owen M. Fiss, *Two Models of Adjudication*, *in* HOW DOES THE CONSTITUTION SECURE RIGHTS? 36 (Robert A. Goldwin & William A Chambra eds. 1985); Susan A. Bandes, *The Idea of a Case*, 42 STAN. L. REV. 227 (1990).