Lisa Battaglia

Law 386 Digital Technologies & The Constitution

Final Memo

April 20, 2023

Going Too Deep: How The Law Fails Victims of Nonconsensual Deepfake Pornography

British Twitch influencer "Sweet Anita" learned that "a trove of fake, sexually-explicit videos featuring the faces of Twitch streamers was circulating online."[1] After hearing this, Sweet Anita discovered several videos of her face edited onto another's body in a pornographic video. Confident that she never made a pornographic video in her life, Sweet Anita was horrified to find out that someone created this video using her image without her consent. Her fans, however, believed it was real. Sweet Anita and other Twitch influencers became victims of nonconsensual deepfake pornography.

Although this story made headlines in January 2023, the problem is not exclusive to these influencers. Emma Watson, Gal Gadot, and even Michelle Obama are a few high profile names that have appeared in a deepfake. Deepfakes are not exclusive to public figures either. If a picture or a video of someone exists online, anyone could make a deepfake using their image. This memo analyzes the legal options for a victim of nonconsensual deepfake pornography and how the law protects the deepfake creators and internet service providers.

ISSUE

Does a victim of nonconsensual deepfake pornography have legal recourse?

SHORT ANSWER

[1] Bianca Britton, *They appeared in deepfake porn videos without their consent. Few laws protect them.*, NBC News (Feb. 14, 2023), https://www.nbcnews.com/tech/internet/deepfake-twitch-porn-atrioc-qtcinderella-maya-higa-pokimane-rcna69372

As deepfake technology improves, making videos harder to detect, the law is limited in protecting the person in the video. This memo will a) provide background on deepfakes, b) conclude that state law is limited, c) demonstrate that Section 230 protects internet service providers, d) assess whether deepfakes are protected by the First Amendment, and e) analyze whether victims can claim copyright infringement.

a) What Is A Deepfake?

A deepfake is "a video of a person in which their face or body has been digitally altered so that they appear to be someone else, typically used maliciously or to spread false information," according to the Oxford Dictionary. The Department of Homeland Security qualifies deepfakes as an "emergent type of threat falling under the greater and more pervasive umbrella of synthetic media."[2]

This technology dates back to 1997 when Christoph Bregler, Michele Covell, and Malcolm Slaney created Video Rewrite, a technology that "uses existing footage to create automatically new video of a person mouthing words that she did not speak in the original footage. This technique is useful in movie dubbing, for example, where the movie sequence can be modified to sync the actors' lip motions to the new soundtrack."[3]

Today, deepfakes rely on two core technologies: autoencoders and generative adversarial networks (GANs) which layer the existing image with the inputted image to create a combined, realistic video.[4] Despite the complexity of the technology, free apps and websites make the

---

[2] Increasing Threat of Deepfake Identities, *Department of Homeland Security* (January 2019).

[3] Christoph Bregler, Michele Covell, and Malcolm Slaney, *Video Rewrite: Driving Visual Speech with Audio*, Interval Research Corporation (1997).

[4] Kate Kobriger, Janet Zhang, Andrew Quijano, Joyce Guo, *OUT OF OUR DEPTH WITH DEEP FAKES: HOW THE LAW FAILS VICTIMS OF DEEP FAKE NONCONSENSUAL PORNOGRAPHY*, 28 RMDJLT 204 Richmond J.L. & Tech. (2021)

technology accessible to anyone with internet access, requiring very little technical skill. According to a 2019 report by Sensity, "nonconsensual deepfake pornography accounted for 96% of a sample study of more than 14,000 deepfake videos posted online."[5] To address the concerns with deepfakes is to address the concerns with deepfake pornography.

With easy access and the ability to spread false information quickly, neither the law nor improved algorithms can detect or interfere with this content, making the creators difficult to track down. MrDeepFakes, one of the most prominent deepfake porn websites, advertises jobs for deepfake creators. Offering to compensate employees in cryptocurrency, MrDeepFakes keeps the exchange anonymous and untraceable.[6] The technology itself also protects the deepfake creators: "perpetrators go to great lengths to initiate such attacks at such an anonymous level that neither law enforcement nor platforms can identify them."[7]

Despite the technological manipulations to conceal deepfake creators and the legal protections for the internet service providers hosting the content, the law should have options for victims of nonconsensual, deepfake pornography. Victims, however have very little recourse. Suing to receive an injunction or monetary damages seems unlikely, since the victims do not know who is liable in most cases. In fact, they are likely to lose on most of the options they do

---

[5] Shane Raymond, *Deepfake anyone? AI synthetic media tech enters perilous phase*, Reuters (December 2021), https://www.reuters.com/technology/deepfake-anyone-ai-synthetic-media-tech-enters-perilous-phase-2021-12-13/

[6] Kat Tenbarge, *Found through Google, bought with Visa and Mastercard: Inside the deepfake porn economy*, NBC News (March 2023) https://www.nbcnews.com/tech/internet/deepfake-porn-ai-mr-deep-fake-economy-google-visa-mastercard-download-rcna75071

[7] Shane Raymond, *Deepfake anyone? AI synthetic media tech enters perilous phase*, Reuters (December 2021), https://www.reuters.com/technology/deepfake-anyone-ai-synthetic-media-tech-enters-perilous-phase-2021-12-13/

have. As with many new technologies, "the law is unequipped to handle these impending issues."[8]

b) State Law Is Limited

Although 48 states and Washington D.C. passed laws prohibiting the distribution or production of nonconsensual pornography, only California, Virginia, and Texas enacted laws focusing on deepfakes. Texas became the first to enact a law outlawing political deepfakes, while Virginia and California specify deepfake pornography.[9] State statutes provide little protection since anyone around the world can make and distribute deepfakes on the Internet, emphasizing the need for stronger federal laws.

c) Section 230 Protection for Internet Service Providers

Hosting the content, internet service providers also receive protection, primarily from Section 230 of the Communications Decency Act. The 47 U.S. Code § 230(c)(1) and (c)(2) of the Communications Decency Act of 1996, commonly referred to as Section 230, has been no stranger to the news in the last few years. 47 U.S. Code § 230(c)(1) states: "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider."

Section 230(c)(2) adds "(2) No provider or user of an interactive computer service shall be held liable on account of— **(A)** any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is

---

[8] Douglas Harris, *DEEPFAKES: FALSE PORNOGRAPHY IS HERE AND THE LAW CANNOT PROTECT YOU*, 17 DUKELTR 99, Duke Law & Technology Review (January 2019)

[9] Prajakta Pradhan, *AI Deepfakes The Goose Is Cooked?* University of Illinois Law Review, October 2020

constitutionally protected; or **(B)** any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph." Section 230(c)(2) shields interactive computer services from liability for removing or restricting content. Section 230 provides a liability shield to internet service providers whether they leave content up or take content down. A nonconsensual, pornographic deepfake does not fall under the child pornography or sex trafficking exceptions to Section 230. If a pornographic deepfake circulated online, internet service providers are protected by Section 230. Other parts of the Communications Decency Act that were struck down in 1997, however, might have provided deepfake victims some protection had they remained.

The original Telecommunications Act made it a crime "for anyone to engage in online speech that is 'indecent' or 'patently offensive' if the speech could be viewed by a minor." 47 U.S.C. §231. In *Reno v. American Civil Liberties Union,* 521 U.S. 844 (1997), the ACLU challenged 47 U.S.C. §231. Although the act intended to protect minors from "obscene or indecent material," its broad application infringed on the protected speech between consenting adults. *Reno* concluded that the act "threatened to torch a large segment of the Internet community" and unconstitutionally restricted free speech.[10] Because of this, the rest of the Communications Decency Act was struck down, but Section 230 remained. Prioritizing free speech, the Court could not foresee the implications of striking down this law in 1997. This debate begs the question of whether deepfake pornography is protected speech.

d) <u>Are Deepfakes Protected Speech?</u>

[10] *Id.* at 882.

Even if the victim tracked down the creator, the creator might argue First Amendment protection. The legality of publishing deepfakes "implicates the First Amendment in two contexts: false speech and obscenity."[11] In regards to false speech, the Court suggested that "under the First Amendment there is no such thing as a false idea. However pernicious an opinion may seem, we depend for its correction not on the conscience of judges and juries but on the competition of other ideas."[12] This suggestion leaves the correction of false information up to the free marketplace of ideas, not the legal system.

Deepfake pornography, however, might be unprotected speech and fall under obscenity. This claim is a stretch because courts have "struggled to define pornography and obscenity."[13] While pornography has generally been used to describe sexually explicit material, the courts have shifted the meaning "obscenity" since the U.S. adopted a test from a British case in 1868 in *Regina v. Hicklin*, L.R. 3 Q.B. 360 (1868). Material that is deemed obscene is not constitutionally protected. The *Hicklin* rule provided the following test for obscenity: "whether the tendency of the matter . . . is to deprave and corrupt those whose minds are open to such immoral influences, and into whose hands a publication of this sort may fall."[14]

The Supreme Court then examined the constitutionality of criminal obscenity statutes in *Roth v. U.S.*, 354 U.S. 476 (1957) and determined that "the statutes, applied according to the proper standard for judging obscenity, do not offend constitutional safeguards against convictions based upon protected material, or fail to give adequate notice of what is

---

[11] Harris, *supra* at 107.

[12] *Gertz v. Robert Welch, Inc.*, 418 U.S. 323, 339-40 (1974)

[13] 2 Margaret C Jasper, *The Law of Obscenity and Pornography (Legal Almanac Series)* § 1.1 (2012).

[14] *THE FIRST AMENDMENT ENCYCLOPEDIA*, William Crawford Green, (2009).

prohibited."[15] The Court held that the test to determine obscenity was "whether to the average person, applying contemporary community standards, the dominant theme of the material taken as a whole appeals to prurient interest." The Court then approached obscenity again in *Memoirs v. Massachusetts,* 383 U.S. 413 (1966), and articulated a new three-part test: "(a) the dominant theme of the material taken as a whole appeals to a prurient interest in sex; (b) the material is patently offensive because it affronts contemporary community standards relating to the description or representation of sexual matters; and (c) the material is utterly without redeeming social value."[16]

Now, *Miller v. California*, 413 U.S. 15 (1973) is the leading test for obscenity cases in which it deemed obscenity as unprotected speech. The test states: "The basic guidelines for the trier of fact must be: "(a) whether 'the average person applying contemporary community standards' would find that the work, taken as a whole, appeals to the prurient interest, (b) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law, and (c) whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value."[17] Since *Miller* was decided in 1973, the questions of "average contemporary community standards" and "lacks value" has been left to juries. Leaving these questions to the juries leads to inconsistencies depending on geographic area, religious or political background, upbringing, and so many other factors. There is no one contemporary community standard.

---

[15] *Id* at 492.

[16] *Id.* at 418.

[17] *Id.* at. 15.

This same concern was addressed in *Ashcroft v. American Civil Liberties Union*, 535 U.S. 564 (2002). After *Reno*, Congress attempted to address the concern of exposing children to indecent material by enacting the 47 U.S.C. § 231 Child Online Protection Act (COPA), which drew upon the *Miller* test.[18] Internet content providers and civil liberties groups sued the Attorney General, alleging that COPA violated the First Amendment because relying on "community standards" to identify what material "is harmful to minors" rendered the statute overly broad.[19] The majority determined that COPA's use of "community standards" to identify "material that is harmful to minors" did not violate the First Amendment. In a 8-1 decision, the concurring justices in *Ashcroft* questioned the legitimacy of the *Miller* test. Justice Breyer, in his concurrence, stated "to read the statute as adopting the community standards of every locality in the United States would provide the most puritan of communities with a heckler's veto affecting the rest of the Nation."[20] Justice O'Connor, in her concurrence, stated: "I write separately to express my views on the constitutionality and desirability of adopting a national standard for obscenity for regulation of the Internet."[21] Justice O' Connor wrote that "adoption of a national standard is necessary in my view for any reasonable regulation of Internet obscenity."[22] A national standard, however, would be difficult to determine because of varying backgrounds and opinions on what obscenity means.

---

[18] *Id.* at 564.

[19] *Id*. at 564.

[20] *Id.* at 591.

[21] *Id.* at 586.

[22] *Id.* at 588.

After a remand, the Court heard *Ashcroft v. American Civil Liberties Union* 542 U.S. 656 (2004) again. In the second hearing, the Court discussed the *Miller* test with more precision and reaffirmed the Third Circuit's preliminary injunction of COPA on the grounds "that the statute was not narrowly tailored to serve a compelling Government interest."[23] The Court ultimately concluded that COPA was too restrictive and did not pass the strict scrutiny test for speech regulation.[24] With an intention to protect children from indecent material, broad legislation often regulates content made for consenting adults and therefore infringes on adults' First Amendment rights. While a jury would likely find nonconsensual deepfake pornography obscene, victims of the content cannot rely on its certainty.

The Supreme Court evaluated a similar question in *Ashcroft v. Free Speech Coalition,* 535 U.S. 234 (2002). The Free Speech Coalition challenged the Child Pornography Prevention Act of 1996 (CPPA)[25] which expanded the prohibition on child pornography to include not only pornographic images made using actual children, but also "any visual depiction, including any photograph, film, video, picture, or computer or computer-generated image or picture," that "is, or appears to be, of a minor engaging in sexually explicit conduct,"[26] and any sexually explicit image that is "advertised, promoted, presented, described, or distributed in such a manner that conveys the impression" or depicts "a minor engaging in sexually explicit conduct."[27] The CPPA "bans a range of sexually explicit images, sometimes called 'virtual child pornography,'

---

[23] *Id.* at 664.

[24] *Id.* at 670.

[25] 18 U.S.C. § 2256(8)(A)

[26] § 2256(8)(B)

[27] § 2256(8)(D)

that appear to depict minors but were produced by means other than using real children, such as through the use of youthful-looking adults or computer-imaging technology."[28]

The Supreme Court concluded that computer-generated child pornography not involving actual children was considered protected speech. Relying on the *Miller* standard, the Court concluded that the CPPA's restrictions against what it considered to be virtual child pornography violated the First Amendment. If "virtual child pornography" is considered protected speech under *Ashcroft v. Free Speech Coalition,* adult victims might struggle to fight against the First Amendment protections for deepfake creators.

e) Can Victims Claim Copyright Infringement?

Although a deepfake creator uses the victim's image (and possibly several images) to create the video, the victim's "potential copyright infringement claims are likely to fail despite retaining copyright protection in all of her photographs uploaded online."[29] Copyright protection under 17 U.S.C. § 102(a) protects "original works of authorship fixed in any tangible medium of expression." Unless the victim took a self-portrait, this copyright claim protects the photographer. Additionally, the victim contributes little to the creation of the deepfake, weakening the argument that they are a joint author.

The Ninth Circuit came close to addressing a deepfake case in *Garcia v. Google*, 786 F.3d 733 (9th Cir. 2015). Actress Garcia acted in an action-adventure film in which she had two lines: "Is George crazy? Our daughter is but a child?"[30] Following her debut, Garcia discovered that the movie's writer and director "had a different film in mind: an anti-Islam polemic

[28] *Id.* at 234.

[29] Harris, *supra* at 107.

[30] *Id.* at 737.

renamed *Innocence of Muslims.* The film, featuring a crude production, depicts the Prophet Mohammed as, among other things, a murderer, pedophile, and homosexual. Film producers dubbed over Garcia's lines and replaced them with a voice asking, 'Is your Mohammed a child molester?' Garcia appears on screen for only five seconds."[31]

After receiving death threats, and an Egyptian cleric issuing a fatwa against her, Garcia asked Google to remove the film "asserting it was hate speech and violated her state law rights to privacy and to control her likeness."[32] Garcia also issued takedown requests through the Digital Millenium Copyright Act.[33] Google declined to remove the film, prompting Garcia to sue Google for copyright infringement and intentional infliction of emotional distress.[34]

The Ninth Circuit held that "the district court did not abuse its discretion when it denied Garcia's motion for a preliminary injunction under the copyright laws."[35] Under the Copyright Act 17 U.S.C. § 102(a), the fixation "in any tangible medium of expression" must be done "by or under the authority of the author."[36] The Ninth Circuit asserts that "Garcia is the author of none of this and makes no copyright claim to the film or to the script. Instead, Garcia claims that her five-second performance itself merits copyright protection."[37] Because "she never fixed her acting performance in a tangible medium, as required by 17 U.S.C. § 101," Garcia lost on her copyright claim despite the director using her image in a way she did not approve of that led to

---

[31] *Id.* at 737.

[32] *Id.* at 738.

[33] 17 U.S.C. § 512

[34] *Id.* at 738.

[35] *Id.* at 747.

[36] 17 U.S.C. § 101.

[37] *Id.* at 741.

death threats against her.[38] Like Garcia, a deepfake victim has no involvement in the creative process of the video. Deepfake creators obtain someone's image from the Internet to create a new work, and the Ninth Circuit decision in *Garcia* says the victim does not have a copyright claim.

Under 17 U.S.C. § 107, an exception to the copyright claim may be fair use, further protecting the deepfake creator. In determining whether a use is fair, "the factors to be considered shall include—**(1)** the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; **(2)** the nature of the copyrighted work; **(3)** the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and **(4)** the effect of the use upon the potential market for or value of the copyrighted work." In consideration of the first factor, courts emphasized that the copied work must be transformative, or "add something new, with a further purpose or different character, altering the first with new expression, meaning, or message."[39] A deepfake creator may argue that the work is transformative and not for commercial use, especially if the creator does not make a profit. The more transformative the work, "the less will be the significance of other factors, like commerciality, that may weigh against a finding of fair use."[40] Even with an unlikely strong copyright claim by the victim, the deepfake creator may have a valid fair use defense.

Conclusion: Can We Trust The Internet?

Leaving victims of nonconsensual deepfake pornography with little to no legal remedies asks internet service providers and internet users to choose to protect deepfake victims. A few leading platforms responsibly implemented policies to consciously denounce the fake videos.

---

[38] *Id.* at 743.

[39] *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, (1994)

[40] *Id.* at 579.

Because Twitch recently made headlines for the Sweet Anita controversy, it published a new policy advocating against this content, offering community resources for victims, consulting with experts, and "updating our Adult Sexual Violence and Exploitation policy to make it more clear that intentionally promoting, creating, or sharing synthetic NCEI [nonconsensual explicit imagery] can result in an indefinite suspension on the first offense."[41]

Meta instated a policy in 2020 as well: "consistent with our existing policies, audio, photos or videos, whether a deepfake or not, will be removed from Facebook if they violate any of our other Community Standards including those governing nudity, graphic violence, voter suppression and hate speech."[42] More recently, Meta said in a statement "the company's policy restricts both AI-generated and non-AI adult content and it has restricted the app's page from advertising on its platforms."[43]

Additionally, "TikTok said last month all deepfakes or manipulated content that show realistic scenes must be labeled to indicate they're fake or altered in some way, and that deepfakes of private figures and young people are no longer allowed. Previously, the company had barred sexually explicit content and deepfakes that mislead viewers about real-world events and cause harm."[44] Leading artificial intelligence platforms like OpenAI and Stability AI use image generating tools to "remove the ability to create explicit images," "filter requests," and "block users from creating AI images of celebrities and prominent politicians."[45]

---

[41] *Addressing Explicit Deepfake Content,* Twitch (March 2023). https://safety.twitch.tv/s/article/Addressing-Explicit-Deepfake-Content?language=en_US

[42] Monika Bickert, *Enforcing Against Manipulated Media*, Facebook Transparency Center (January 2020)

[43] Hadero, *supra* at 1.

[44] Hadero, *supra* at 2.

[45] Hadero, *supra* at 1.

With the legal system failing them, Sweet Anita and others are at the mercy of the wild west of the internet. Because the law is limited, Internet service providers must establish a moral stance against the bad actors and remove the controversial imitations from their platforms. As for the rest of the internet providers, the law protects them, allowing the deepfake creators to use, abuse, and even monetize others' images for pornographic deepfakes without repercussions.