TAYLOR SWIFT, DEEPFAKES, AND THE FIRST AMENDMENT: CHANGING THE LEGAL LANDSCAPE FOR VICTIMS OF NON-CONSENSUAL ARTIFICIAL PORNOGRAPHY

BY: JULIA STURGES*

INTRODUCTION

In January 2024, artificially generated pornographic images (also known as "deepfakes") of pop superstar Taylor Swift circulated the social media platform, X (formerly Twitter), at an alarmingly quick rate.[1] Within hours, some images were seen more than 45 million times and accrued thousands of shares and likes before eventually being taken down.[2] The incident, which gained media attention in part due to Swift's mega-star status and passionate fan base, brought up fascinating First Amendment questions about the role of social media platforms in regulating obscene speech and protecting victims, especially minors, from these types of attacks. Legal advocates also pondered possible remedies available for victims under current legal standards. Part I of this Article will walk through the history of "deepfakes"[3] and the role of artificial intelligence in the development and circulation of fake pornographic images. Part II will discuss the First Amendment legal standard for obscene speech and how social media platforms may be able to regulate these harmful images and videos on their platforms to mitigate harm. Part III will consider the *New York Times* libel standard for public figures and officials and relevant immunity granted by the Communications and Decency Act Section 230. Part IV will highlight the disproportionate impact and harm that unregulated and widely available deepfakes have on women and girls. And finally, Part V will discuss proposed legislation at the state and federal level and the ways these bills could support victims of cruel deepfakes and prevent future images and videos from wide circulation.

I. HISTORY OF DEEPFAKES

The concept of "deepfakes" is a fairly new phenomenon, but as one source notes, "[t]he manipulation of data is not new. Ancient Romans chiseled names and portraits off stone, permanently deleting a person's identity and history."[4] Fast forward to 1997 when a research paper developed the "Video Rewrite Program" which essentially "automated what some movie studios could do," including "interpret[ing] faces, synthesiz[ing] audio from text, and model[ing] lips in 3D space."[5] In the early 2000's, facial recognition technology began to take over the artificial intelligence space, propelled by a paper authored by Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor, who used "a thorough statistical model to match a shape to an image . . . "[6] The attacks on September 11, 2001 also contributed to the growing need for and

[1] Jess Weatherberd, *Trolls have flooded X with graphic Taylor Swift AI fakes*, THE VERGE (Jan 25, 2024), https://perma.cc/HT2E-2YKP.

[2] *Id.*

[3] *Deepfake*, MERRIAM-WEBSTER.COM, https://perma.cc/5GDE-QXEN.

[4] *See* Meredith Somers, *Deepfakes, Explained*, MIT MGMT. (July 21, 2020), https://perma.cc/D2PW-6PGS.

[5] *See* David Song, *A Short History of Deepfakes*, MEDIUM (Sept. 23, 2019), https://perma.cc/RVN5-U5P2; Christoph Bregler, Michele Covell, & Malcolm Slaney, *Video Rewrite: Driving Visual Speech with Audio*, INTERVAL RSCH. CORP. (1997).

[6] *A Short History of Deepfakes, supra* note 5*;* T. F. Cootes, G. J. Edwards and C. J. Taylor, *Active Appearance Models,* 23 IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 681 (2001).

investment in facial recognition software as a method of crime prevention.[7] In 2016 and 2017, two additional projects, the Face2Face project out of the Technical University of Munich and the Synthesizing Obama project out of the University of Washington[8] significantly contributed to the accessibility of artificial intelligence software, "establish[ing] deepfakes as achievable with consumer grade hardware."[9] After these two projects circulated, it became possible, even easy, to splice together a person's image and voice to get them to do or say just about anything.[10] As artificial intelligence researcher Alex Champandard said about the rise in deepfake technology, "this is no longer rocket science."[11]

It will not come as a shock that the rise in accessibility of artificial intelligence technology and the global use of social media was, and continues to be, the perfect combination for mass distribution of deepfakes. In 2017, a subreddit thread titled r/deepfakes that has since been deleted had, at its peak, nearly 90,000 subscribers[12] and featured many celebrities including Gal Gadot and Scarlett Johansson.[13] Sensity, a "visual threat detection company" based in Amsterdam,[14] released a 2019 report that "detected 14,679 deepfakes online [in 2019] and, in 2020, found that the number rose to 49,081."[15] This shows an almost 300% increase in just a year. Most recently, as stated in the Introduction, artificially generated pornographic images of Taylor Swift circulated X and were seen more than 45 million times.[16] The quick and widespread distribution of these images demonstrates the enormous reach and dangerous impact such images can have on an image's subject(s), its viewers, and pop culture generally.

As artificial intelligence becomes increasingly normalized and accessible, conversations about both the dangers and benefits of its use are happening across industries, continents, and generations. One scholar has noted several schools of thought when it comes to artificial intelligence. Those who are pessimistic about artificial intelligence worry that machines will become smarter than humans, rendering many jobs obsolete and widening the wealth gap between those with access to AI and those without.[17] On the flip side, AI optimists point to the possibility of a utopian future where AI revolutionizes just about everything, from science and

---

[7] Facial Recognition Comes of Age, POLICE, (NOV. 7, 2016) https://perma.cc/7TTU-SPCC.
[8] *See* Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, & Matthias Nießner3, *Face2Face: Real-time Face Capture and Reenactment of RGB Videos*, STANFORD COMPUTER GRAPHICS (2020) https://perma.cc/Q9N7-HN4C; Supasorn Suwajanakorn, Steven M. Seitz, & Ira Kemelmacher-Shlizerman, *Synthesizing Obama: Learning Lip Sync from Audio*, ACM TRANS. ON GRAPHICS (July 2017), https://perma.cc/R9YL-T4F8.
[9] *A Short History of Deepfakes, supra* note 5.
[10] *See* Kate Coleman, *How Deepfakes are Impacting Culture, Privacy, and Reputation*, STATUS LABS https://perma.cc/CYR8-QZRB.
[11] Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All Fucked*, VICE (Dec. 11, 2017), https://perma.cc/M9W6-BURU.
[12] *Id.*
[13] Kavyasri Nagumotu, *Deepfakes Are Taking Over Social Media: Can The Law Keep Up?*, 62 IDEA: THE INTELLECTUAL PROP. L. REV 102 (2022).
[14] SENSITY.COM, https://perma.cc/RLS2-3AYS.
[15] See *Deepfakes Are Taking Over Social Media, supra* note 13.
[16] Weatherberd, *supra* note 1.
[17] *See* Spyros Makridakis, *The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms*, 90 FUTURES 46, 59 (2017).

medicine to technology, employment, and business.[18] While the future of artificial intelligence likely lies somewhere between utopian and dystopian, "[e]xisting AI systems raise real concerns about bias [and] privacy," among others.[19] With the above history in mind, this Article seeks to analyze those concerns, highlighting how young women are disproportionately impacted by deepfakes and calling for heightened regulation of content in the digital space.

II. THE STANDARD FOR OBSCENE SPEECH

Whether or not deepfakes can legally be regulated turns in large part on whether they are considered "obscene" speech. Under *Roth v. United States*, obscene speech is not protected by the First Amendment.[20] In *Roth*, the defendant, a bookstore owner, challenged his conviction under a federal obscenity statute that prohibited mailing certain lewd images, publications, and pamphlets.[21] The Supreme Court held that "obscenity is not within the area of constitutionally protected speech or press" and reasoned that "lewd and obscene . . . utterances are no essential part of any exposition of ideas, and are of such slight social value as a step to truth that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality."[22] Then, in *Miller v. California*, the Court refined *Roth* and provided a three-part framework for determining whether material qualified as obscene.[23] Under *Miller*, the three guiding principles are: whether the average person, applying contemporary community standards would find that the work, taken as a whole, appeals to the prurient interest; whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law; and whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value."[24] Interestingly, the presence of a "community standards" prong in the *Miller* framework asks courts and businesses to consider cultural and community norms when making obscenity determinations, meaning more conservative or religious communities, for example, may end up having different legal standards for obscenity compared to bigger, more progressive towns. Understandingly, the internet has complicated the "community standards" framework because "everything is simultaneously available everywhere."[25] "Indeed, the Third Circuit Court of Appeals upheld an injunction against enforcement of the Child Online Protection Act of 1998 (COPA) on the ground that a community standards test 'would essentially require every Web communication to abide by the most restrictive community's standards.'"[26]And importantly, because obscene speech is not protected by the First Amendment, states have vast leeway to regulate certain lewd speech so long as it does not cross the line of censorship.[27]

---

[18] *Id.* at 50
[19] *How to Worry Wisely about Artificial Intelligence*, THE ECONOMIST (Apr. 23, 2023), https://perma.cc/G85G-58LL.
[20] *See* Roth v. United States, 354 U.S. 476, 485 (1957).
[21] *Id.*
[22] *Id.* (quoting Chaplinsky v. New Hampshire, 315 U.S. 568, 571–72 (1942)).
[23] *See* Miller v. California, 413 U.S. 15 (1973).
[24] *Id.*
[25] Ronald Steiner, *Community Standards*, FREE SPEECH CTR (August 10, 2023), https://perma.cc/S3UP-CFQC.
[26] *Id.*
[27] Jessica Ice, *Defamatory Political Deepfakes and the First Amendment*, 70 CASE W. RES. L. REV. 417 (2019).

These standards are important when it comes to regulating artificially generated pornographic material and determining what, if anything, states and social media platforms can do to restrict these images on their websites and provide causes of action for victims. While some deepfakes under the *Miller* standard for obscenity may *not* receive First Amendment Protection, privacy attorney Evan Enzer said "[p]hony images and videos can qualify as protected speech under the First Amendment depending on the situation and the people involved."[28] However, some states have had success enacting laws to protect victims, many of whom are women, from "revenge porn."[29] In Indiana, such laws have survived First Amendment challenges[30] – a promising outcome for law makers and advocates seeking to protect individuals from harmful online harassment. Other states like California and Illinois are following suit by providing protection from deepfakes and enacting penalties for distributing altered sexual images.[31] There is also a growing push to have "nonconsensual deepfake pornography treated 'under the First Amendment as prohibitions on traditional nonconsensual pornography rather than being dealt with under the less-protective law of defamation.'"[32]

Social media platforms also play a critical role in regulating and protecting victims from harmful deepfakes, especially if the images meet the *Miller* "obscenity" definition. As the Taylor Swift deepfakes demonstrate, millions of users can view, like, share, and save non-consensual artificially generated pornographic images on social media platforms in a matter of seconds. The Constitution provides flexibility for social media platforms to decide their own policies and regulations as they relate to removing harmful, false, or otherwise problematic content from their respective sites.[33] However, Section 230 of the Communications Decency Act, discussed *infra* Part III, does provide some immunity for social media platforms and other online sites for the content their users post and share.

The role of social media in controlling widespread misinformation became a hot topic amid the 2016 and 2020 elections, during which enormous amounts of misleading and completely inaccurate political information spread across social media platforms.[34] Alarmingly, but unsurprisingly, "[r]ecent evidence shows that [in the aftermath of the 2016 election]: 1) 62 percent of US adults get news on social media; 2) the most popular fake news stories were more widely shared on Facebook than the most popular mainstream news stories; 3) many people who see fake news stories report that they believe them; and 4) the most discussed fake news stories tended to favor Donald Trump over Hillary Clinton."[35] This evidence demonstrates the

---

[28] *See* Zach Williams, *States Target AI Deepfakes in Taylor Swift Aftermath*, BLOOMBERG LAW (Feb. 5, 2024), https://perma.cc/MS7R-ND3Z.
[29] *Id.*
[30] State v. Katz, 179 N.E.3d 431 (Ind. 2022).
[31] Titus Wu, California Looks to Boost Deepfake Protections Before Elections, BLOOMBERG LAW (Dec. 15, 2023), https://perma.cc/HUW2-YVRX.
[32] Coleman, *supra* note 10.
[33] Nagumotu, supra note 13 at 118–19.
[34] *See* Nina I. Brown & Jonathan Peters, *Say This, Not That: Government Regulation and Control of Social Media*, 68 SYRACUSE L. REV. 521 (2018); Christine Fernando, Election disinformation campaigns targeted voters of color in 2020. Experts expect 2024 to be worse, AP NEWS (July 29, 2023), https://perma.cc/9UQY-R38U; Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, 31 J. OF ECON. PERSPECTIVES 211 (2017).
[35] Alcott & Gentzkow, *supra* note 32 at 212.

tremendous role that social media platforms play in influencing culture, politics, and everything in between.

As a response to the above, many social media platforms announced new policies and regulations to protect against deepfakes, sexual or otherwise, amid rampant misinformation during the 2016 election. Facebook, for example, released a statement in 2020 announcing new policies to help combat deepfakes, saying "[g]oing forward, we will remove misleading manipulated media if it meets the following criteria: It has been edited or synthesized – beyond adjustments for clarity or quality – in ways that aren't apparent to an average person and would likely mislead someone into thinking that a subject of the video said words that they did not actually say. And: It is the product of artificial intelligence or machine learning that merges, replaces or superimposes content onto a video, making it appear to be authentic."[36] X (formerly Twitter) released a similar statement in February 2020, which read "[y]ou may not deceptively share synthetic or manipulated media that are likely to cause harm. In addition, we may label Tweets containing synthetic and manipulated media to help people understand the media's authenticity and to provide additional context."[37] Furthermore, Reddit, which had previously been known for rampant distribution of deepfakes on its r/deepfakes thread, announced in February 2018 ". . . we have made two updates to our site-wide policy regarding involuntary pornography and sexual or suggestive content involving minors. These policies were previously combined in a single rule; they will now be broken out into two distinct ones. Communities focused on this content and users who post such content will be banned from the site."[38] It is unclear exactly how effective these updated guidelines have been.

Taken together, it is evident that social media platforms can and have taken steps to limit the spread of misinformation, harmful artificially generated content, and pornographic deepfakes from circulating their sites. However, as the Taylor Swift incident indicates, deepfakes are still able to circulate expansively across platforms, indicating a need for stronger surveillance capabilities and standards. With their increased ability to regulate speech under the First Amendment and the *Miller* standard for obscene speech, social media platforms have a unique opportunity, and arguably a responsibility, to regulate the distribution of artificial, non-consesual pornographic images to ensure their users are abiding by community standards and limiting harm. The next question becomes, when technology giants like Facebook and X *don't* regulate these images, or don't regulate them quickly or efficiently enough, can celebrities like Taylor Swift sue?

III. THE *NEW YORK TIMES* STANDARD FOR PUBLIC FIGURES AND SECTION 230 OF THE COMMUNICATIONS AND DECENCY ACT

With almost every resource at their fingertips, one would assume that a celebrity who is the victim of a widespread deepfake attack would be able to take swift action against the creator of the media and/or social media platform on which the image(s)/video(s) circulated.

---

[36] Monika Bickert, Enforcing Against Manipulated Media, META (Jan. 6, 2020), https://perma.cc/Q584-8SAP.
[37] Yoel Roth & Ashita Achuthan, *Building Rules in Public: Our Approach to Synthetic & Manipulated Media*, X BLOG (Feb. 4, 2020), https://perma.cc/G5GG-HFDU.
[38] *Update on Site-wide Rules Regarding Involuntary Pornography and the Sexualization of Minors*, REDDIT https://perma.cc/DL3R-56UD.

Unfortunately, because AI technology is so widely accessible, it would likely be difficult to sue individual deepfake creators given how many people can and do use the technology. As for legal action that someone like Taylor Swift could take against a social media platform for its role in enabling the distribution of the images, there are a few relevant considerations.

First, it is often more difficult for public figures, like Swift, to respond to such images because of a standard for defamation set forth in *New York Times v. Sullivan*. In *New York Times*, an elected commissioner from Alabama sued the New York Times for libel after they published allegedly false and defamatory statements about him.[39] The Supreme Court held that evidence of actual malice, a higher standard not applicable to private individuals,[40] is required in order for public officials to recover punitive damages for defamation and libel.[41] The scope of the *New York Times* holding was extended in *Curtis Publishing Co. v. Butts* and *Associated Press v. Walker*, where the Court applied the actual malice rule to libel actions "instituted by persons who are not public officials, but who are 'public figures' and involved in issues in which the public has justified and important interest."[42] With the above in mind, the question becomes whether Swift would be able to prove that the deepfakes were created and distributed with actual malice – meaning the statement was false or the statement was made with reckless disregard for the truth.[43] I argue that the answer to that question would likely be "yes" given the content of the images, but that is not the end of the equation.

A second consideration is the impact of the Communications and Decency Act of 1996, Section 230. Section 230 provides "limited federal immunity to providers and users of interactive computer services,"[44] and courts have previously applied the law to social media platforms to shield them from liability for their role in circulating harmful third-party content on their platforms.[45] This section of the CDA can and has previously limited celebrities' and private victims' ability to take legal action against social media platforms for their role in circulating lewd content.

There are two primary provisions of Section 230.[46] First, Section 230(c)(1) articulates that service providers and users may not "be treated as the publisher or speaker of any information provided by another information content provider."[47] In *Zeran v. America Online, Inc.*, a 1997 case interpreting this provision, the Fourth Circuit said that Section 230(c)(1) bars "lawsuits seeking to hold a service provider liable for its exercise of a publisher's traditional editorial functions—such as deciding whether to publish, withdraw, postpone or alter content."[48] Second, Section 230(c)(2) says that service providers and their users may not be held liable for voluntarily acting in good faith to restrict access to "obscene, lewd, lascivious, filthy, excessively

[39] New York Times Co. v. Sullivan, 376 U.S. 254 (1964).
[40] Gertz v. Robert Welch, Inc., 418 U.S. 323 (1974).
[41] *Id.* at 283
[42] Curtis Pub. Co. v. Butts, 388 U.S. 130, 134 (1967).
[43] New York Times Co. v. Sullivan, 376 U.S. 254, 279–280 (1964).
[44] CONG. RSH. SERV., R46751, SECTION 230: AN OVERVIEW (2024).
[45] *Id.*
[46] *Id.*
[47] *Id.*; 47 U.S.C. § 230(c)(1).
[48] Zeran v. Am. Online, Inc.,129 F.3d 327, 330 (4th Cir. 1997).

violent, harassing, or otherwise objectionable" material.[49] As one TikTok user aptly points out, most legal and industry experts will tell you that, "in theory, [section 230] is a very good thing, and is vital to both the First Amendment and to the internet in general."[50] From small blogs and nonprofit websites to huge tech giants, Section 230 is meant to protect all of us, and has often been referred to as a "good samaritan law."[51] Furthermore, without Section 230, some online platforms would "overly filter and censor speech" and many others would likely be forced to shut down because the prospect of unimaginable liability to the tune of thousands of lawsuits a day would be too great of a challenge to overcome.[52] However, even with the history of Section 230 as a good samaritan law for the internet aimed to encourage good faith interventions by users and platforms alike, courts have interpreted Section 230 in a way that encourages tech giants to cause harm, with little incentive to reduce harm.[53] As one court noted, "[i]n keeping with this expansive view of the publisher's role, judicial decisions in the area consistently stress that decisions as to whether existing content should be removed from a website fall within the editorial prerogative."[54] Virality drives profit, so without an incentive to remove harmful content primed for virality, and because "attempts to impose liability for the mere refusal to remove content (typically, a refusal to honor a take-down notice), without more, have not yet been met with success," platforms are unlikely to take immediate and sweeping action to remove viral deepfake content.[55] With the above in mind, liability for tech giants like X and Facebook becomes increasingly tricky, limiting available legal recourse for Swift or any other victim.

For celebrities and private individuals alike, understanding the legal and judicial history surrounding the *New York Times Standard* and Section 230 of the CDA is crucial to analyzing the role of social media and possible avenues of legal recourse for victims of non-consensual pornographic content, or any devastating viral posts.

IV. DEEPFAKES DISPROPORTIONATELY IMPACT WOMEN AND GIRLS

The widespread distribution of deepfakes depicting Taylor Swift also demonstrates a grim reality that far too many women know to be true: no matter how much money or social capital you have, no woman, not even Taylor Swift, is immune from gender based digital violence. From Alexandria Ocasio Cortez to Greta Thurnburg, women from every walk of life can be the targets of hateful, non-consensual artificial pornography.[56] As access to artificial intelligence technology continues to increase, more and more people will be able to create and distribute pictures and videos of anyone saying or doing anything – a frightening thought for many people, especially women, who are often targets of online hate.[57] A 2019 study estimated

---

[49] *See* SECTION 230: AN OVERVIEW, *supra* note 41; 47 U.S.C. § 230(c)(2).
[50] @dadchats, TIKTOK, *If we think she's still just a musician, we're not paying attention* (Jan. 30, 2024), https://perma.cc/EFQ4-8F5V.
[51] *Id.*
[52] *Id.*
[53] *Id.*
[54] Cohen v. Facebook, Inc., 252 F. Supp. 3d 140, 156–57 (E.D. N.Y. 2017).
[55] RODNEY A. SMOLLA, SMOLLA & NIMMER ON FREEDOM OF SPEECH, § 23:12 (2023).
[56] Peter Bradshaw, *Another Body review – Terrifying Dive into the World of Deepfake Porn*, THE GUARDIAN (Nov. 22, 2023), https://perma.cc/33N4-JKAL.
[57] *Most online hate targets women, says EU report*, REUTERS (Nov. 29, 2023), https://perma.cc/3PUV-PG8U.

that porn made up around 96% of deepfake videos[58] and in 2021, Sensity, an artificial intelligence firm, estimated that "Of the 85,000 [deepfakes] circulating online, 90 percent depict non-consensual porn featuring women."[59] Furthermore, a vast majority of deepfake creators are men.[60] And as mentioned earlier, celebrities and public figures are not the only targets of these malicious images and videos – plenty of lesser known women deal with this type of online hate all the time.[61] From vengeful ex-boyfriends and hateful friend groups at school to fanatics online,[62] there are a number of reasons why someone would be the target of a deepfake. What is, to some, a funny prank or a harmless joke, can be at best, humiliating and at worst, life altering for those depicted without (and even with) their consent.

Furthermore, Amnesty International has been investigating the abuse of women on X, calling out the site for "failing in its responsibility to respect women's rights online by inadequately investigating and responding to reports of violence and abuse in a transparent manner."[63] Additionally, Amnesty also claims that the lack of action from X is further pushing women into a culture of silence where women are discouraged and scared to share their voices online.[64] Although Amnesty's investigation did not look at deepfakes specifically, one would assume that the presence of and apathy about deepfakes circulating on social media has the same chilling effect for women interested in participating online.[65]

With no obvious legal recourse for celebrities and private individuals, digital violence continues to negatively impact women and girls. At a time when more women are in the classroom, and the board room, than ever before, women should feel safe to use their voices online and participate freely in their schools and communities without fear that they could be the next victim of a malicious cyber attack that uses their likeness without their consent. We cannot have a robust conversation about the benefits and dangers of artificial intelligence, the First Amendment, social media, and privacy without acknowledging the unique ways in which women are targeted in these types of attacks. The lack of legal avenues or legislative response to new artificial intelligence technology that makes it possible for anyone to create a deepfake is disappointing, and allows rape culture, racism, and misogyny to all continue to compound in a way that makes women more vulnerable to hateful, sexualized online rhetoric.

V. PROPOSED SOLUTIONS

At first glance, Taylor Swift appears to have simply been the most recent victim in a long line of celebrities and women who have been the target of a harmful deepfake scheme. However, this most recent scandal may have forced long overdue conversations about liability for big tech

---

[58] Oceane Duboust, Thomas Duthois, Matthew Ashe, & Estelle Nilsson-Julien, *'Violating and dehumanising': How AI deepfakes are being used to target women*, EURONEWS (Nov. 11, 2023), https://perma.cc/DD9B-8YA3.
[59] Sophie Compton, *More and More Women Are Facing the Scary Reality of Deepfakes*, VOGUE (Mar. 16, 2021), https://perma.cc/29WP-4262.
[60] *Id.*
[61] *Id.*
[62] Bradshaw, *supra* note 35.
[63] *Toxic Twitter – A Toxic Place for Women*, AMNESTY INT'L. (Mar. 21, 2018), https://perma.cc/6KTB-4HCX.
[64] *Id.*; Compton, *supra* note 56.
[65] Compton, *supra* note 56.

companies and possible legal remedies for women and girls who are victims of artificial pornography and other online hate. And because Taylor Swift is, well, Taylor Swift, she may be the perfect, and perhaps only, person who could possibly bring about much needed change in this space.

A first possible solution ties back to the prior discussion about the purpose and limits of the CDA Section 230. As discussed, CDA Section 230 is meant to serve as a limit on liability for a platform's third-party posts and content.[66] However, there is still some legal gray area surrounding a site's responsibility to take action as quickly as possible to remove an image or video if they know, or have reason to know, that the image is fake or harmful.[67] A case like Swift's could be the perfect opportunity to challenge this discrepancy and ensure that platforms are not motivated to keep artificial pornography on their sites for longer than necessary because they stand to profit off of the virality of provocative posts.[68] Potentially, such a challenge could call for an exception to Section 230 immunity and allow victims to sue social media platforms.[69] The same Tiktok video mentioned in Section III suggests there should be a standard under Section 230 that "creates a reasonable time frame under which a provider would have to remove content in order to escape triggering liability."

Furthermore, the incident with Taylor Swift also spurred legislative efforts to protect individuals from deepfakes at both the federal and state level.[70] In February, a Federal bill titled The DEFIANCE Act, was introduced as a response to Swift's deepfakes circulating. The DEFIANCE Act would "allow victims to finally defend their reputations and take civil action against individuals who produced, distributed, or received digital forgeries."[71] Previously, Rep. Joe Morelle introduced the Preventing Deepfakes of Intimate Images Act last May, but the bill never got off the ground.[72] Due to recent events, however, perhaps there will finally be movement at the Federal level to allow victims of deepfakes to seek justice, mirroring recent data that shows 85% of Americans support legislation that "would make non-consensual deepfake porn illegal."[73] At the state level, at least 10 states have enacted legislation related to deepfakes, with others in the works.[74] Additionally, Georgia, Hawaii, Texas and Virginia "have laws on the books that criminalize nonconsensual deepfake porn" while "California and Illinois have given victims the right to sue those who create images using their likenesses."[75]

---

[66] @dadchats, *supra* note 47.
[67] *Id.*
[68] *Id.*
[69] *Id.*
[70] Solcyré Burga, *How a New Bill Could Protect Against Deepfakes*, TIME (Jan. 31, 2024), https://perma.cc/C8VV-YKW8.
[71] *See* Press Release, Alexandria Ocasio Cortez, House of Representatives, Rep. Ocasio-Cortez Leads Bipartisan, Bicameral Introduction of DEFIANCE Act to Combat Use of Non-Consensual, Sexually-Explicit "Deepfake" Media (Mar. 7, 2023), https://perma.cc/9R28-6D46.
[72] Burga, *supra* note 67.
[73] *Id.*
[74] Geoff Mulvihill, *What to know about how lawmakers are addressing deepfakes like the ones that victimized Taylor Swift*, AP NEWS (Jan. 31, 2024) https://perma.cc/RJQ8-3MDM.
[75] *Id.*

Other solutions include utilizing technology to monitor, flag, and remove deepfakes from platforms.[76] A computer science professor from University of Buffalo has discussed three approaches, although he notes none of them are perfect, that include deepfake detection algorithms, embedding codes in content people upload that would signal if they are reused in AI creation, and requiring companies offering AI tools to include digital watermarks to identify content generated with their applications.[77]

While we do not currently know how all of this will play out, we do know that these conversations are happening, due in large part to Taylor Swift's current status as the most famous woman in the world and her ability to make the whole world listen. Whether there is a Section 230 loophole that a case like Swift's could challenge, new state or federal legislation, or technological advancements that could better regulate deepfakes, more needs to be done to protect people, especially women and girls, from harmful deepfakes and provide legal recourse for victims. The above solutions are possible first steps at ensuring those who maliciously generate artificial pornography or are complacent in its online circulation are held accountable.

CONCLUSION

Artificial intelligence has made it easier than ever before to create, upload, and circulate digitally altered images and videos that appear realistic. While artificial intelligence does have plenty of benefits, it can also be a dangerous tool when used maliciously and circulated widely without regulation. The three step *Miller* framework for obscene speech arguably puts deepfake pornography out of reach of First Amendment protection, but the heightened libel and defamation standard articulated in *New York Times,* coupled with CDA Section 230's limited liability for websites and social media platforms, means that victims of deepfakes, whether private individuals, public officials, or even Taylor Swift, have limited options for seeking justice. However, Taylor Swift's recent involvement in a widespread deepfake attack is finally kick-starting conversations about how we can protect individuals from deepfake pornography, including arguing for an exception to CDA Section 230 immunity, bringing new state and federal legislation, and using innovative technology to regulate artificially created images and videos. Women and girls continue to be disproportionately impacted by digital violence that can alter the course of their entire lives, and they deserve assurance that if their image and likeness is used inappropriately and without their consent, there is legal recourse available.

In the year of Taylor Swift, when she has added more than 5 billion dollars to the economy,[78] boosted the GDP of the country by nearly half a percentage point,[79] drove 35,000 people to register to vote through a single Instagram story,[80] and boosted female NFL viewership 54% for teen girls and 24% for those between 18-24,[81] her political, social, and economic

---

[76] *Id.*
[77] *Id.*
[78] Abha Bhattarai, Rachel Lerman, & Emily Sabens, *The Economy (Taylor's Version)*, WASH. POST (Oct. 13, 2023), https://perma.cc/P89F-LFKZ.
[79] @dadchats, *supra* note 47.
[80] Becky Sullivan, *A Taylor Swift Instagram Post Helped Drive a Surge in Voter Registration*, NPR (Sept. 22, 2023), https://perma.cc/GFQ2-7QQC.
[81] *See How has the Taylor Swift effect boosted American football?*, AL JAZEERA (Feb. 11, 2024), https://perma.cc/4F52-23XF.

influence cannot be understated. Of course this change is long overdue, and women have been dealing with privacy leaks and digital sexual violence since the internet's earliest days, but the rise in popularity and accessibility of AI technology makes the ramifications of complacency more severe than ever. If anyone can challenge laws and standards that insulate individuals and tech giants from their role in causing harm, that person is Taylor Swift, and the time is now.