MARCH 8, 2024

A Swift Response: Call to Action on Deepfake Non-Consensual Pornography

by Tatum Andres , Meghan K. Farmer , Nicole D. Allen , Thomas M. Clyde

In an era where digital privacy is increasingly under threat, the recent deepfake incident involving the globally renowned artist, Taylor Swift, has catapulted the issue of nonconsensual pornography to center stage. This high-profile case, that involved fabricated explicit images of Swift being circulated on social media without her consent, has sparked a wave of public indignation and a clamor for legal reform. Beyond the headlines, this incident exposes a grim reality faced by countless individuals who fall victim to what is commonly known as "revenge porn." Despite the issue's growing prevalence, the legal system doesn't always offer many victims a clearcut path to recourse.

The surge in the use of mainstream artificial intelligence (AI) software has raised heightened concerns, including for social media platforms, given AI's capacity to generate remarkably authentic and potentially harmful images. The challenge lies in developing effective content moderation policies without infringing upon the broader ethos of free expression (such as political satire and other free speech that the First Amendment protects). Despite considerable efforts by social media companies and state law, significant gaps remain in the legal protections offered to victims of nonconsensual pornography. The emergence of deepfake technology presents new legal challenges, as exemplified by the Taylor Swift case, where traditional definitions of nonconsensual pornography may not suffice.

The Swift Incident: A Turning Point

The incident involving Taylor Swift, a globally recognized celebrity, marks a significant turning point in public sentiment over nonconsensual pornography. Swift, known for her immense fan following and influential voice, became an unwitting victim when deepfake technology was maliciously used to create and disseminate explicit images of her without her consent. These images, though fabricated, rapidly spread across social media platforms, attracted more than 45 million views, 24,000 reposts, and hundreds of thousands of likes and bookmarks for nearly 17 hours before being shut down by X (formerly Twitter), igniting a firestorm of media coverage and public outcry. Although X was able to block “Taylor Swift” as a search term, some of the false images of Swift continued to circulate the social media platform, because individuals were able to bypass the search block by manipulating search terms, such as adding words between the pop star's first and last name.

What is Deepfake Technology?

Deepfake Technology refers to a form of artificial intelligence utilized for fabricating persuasive images, audio, and video deceptions. This term encompasses both the underlying technology and the fabricated content, derived from a fusion of deep learning and deception. Deepfakes typically involve altering pre-existing source material by substituting one individual with another. Moreover, they generate entirely new content portraying individuals engaging in actions or uttering statements they never actually performed or said. The primary peril associated with deepfakes lies in their capacity to disseminate misinformation that seemingly originates from reliable outlets.

State Initiatives to Combat Deepfake Technology

Laws addressing non-consensual pornography directly apply in instances where deepfakes are employed to generate explicit material without the subject's consent. Yet, the implementation of these laws can be hindered by jurisdictional complexities and the anonymous nature of online content dissemination. Several states have enacted some form of legislation targeting this conduct. However, these laws are far from uniform, with significant variations in definitions, protected activities, and penalties. For instance, some states focus only on images obtained unlawfully or with an intent to harm, while others include broader protections. The penalties range from misdemeanors to felonies, reflecting differing perceptions of the severity of the crime. For example, some states, such as New York, address only the dissemination of such material, rather than the creation of it.

Laws requiring consent for use of an individual's likeness are complicated by First Amendment freedom of speech protections. Therefore, in most jurisdictions, consent requirements depend on factors such as whether the use is commercial or whether the individual depicted is a celebrity. However, enforcing even these limited consent requirements is challenging with deepfakes as they are normally created anonymously and distributed widely.

In addition to a scattered legal landscape, some technology offerings may increase the potential harms from deepfakes. Internet platforms apply various levels of content moderation, so some platforms facilitate photo and video sharing, where deepfakes can be used for defamation, blackmail, and other malicious purposes. Content moderation can be especially difficult on live steams. That lack of moderation is exacerbated by the anonymity and ease of creating and disseminating deepfakes, making it exceedingly difficult to identify the creators and hold them accountable. These factors converge to create an environment ripe for the misuse of this technology.

Social Media Platform Immunity

Due to the protections granted under Section 230 of the Communications Decency Act, holding social media platforms legally accountable for their users' distribution of deepfakes presents challenges. Section 230 of the Communications Decency Act has been a cornerstone of internet law in the United States since its enactment in 1996. This legislation offers fairly wide immunity to online platforms from liability for content posted by their users, fostering a free and open internet. However, in the context of nonconsensual pornography, especially with the emergence of deepfake technology, the provisions of Section 230 necessitate a nuanced analysis to balance the rights and responsibilities of online platforms.

As it stands, Section 230 provides broad immunity to online service providers regarding third-party content. This immunity has been crucial in allowing platforms such as social media sites, forums, and comment sections to flourish without the constant threat of litigation. However, this protection also means that platforms have limited legal incentive to address the spread of nonconsensual pornography proactively (although many have chosen to do so in the interest of creating a safe platform). Last May, the Judiciary Committee advanced a package of five bills related to online child sexual abuse material (CSAM). One bill, Earn It Act, would roll back Section 230 protections when platforms facilitate content that violates civil and state criminal laws on child sexual exploitation. Another bill, the Stop CSAM Act, would create a new cause of action for victims and their families to sue over such material. Neither of these bills have made it to the Senate floor for a vote.

Additionally, the Supreme Court is currently deliberating a crucial issue regarding the extent of state authority over social media platforms.[1] Specifically, the focus is on whether states such as Florida and Texas have the jurisdiction to mandate these platforms to carry content that they deem hateful or

objectionable. The state laws explicitly allow users to sue tech platforms for alleged censorship that Section 230 shields them from. Although deepfake non-consensual pornography was not directly addressed in oral arguments, altering the scope of Section 230 could potentially broaden the scenarios in which social media platforms could face litigation for that content.

Proposed Amendments

On January 30, 2024, the Senate introduced a bipartisan bill, the Disrupt Explicit Images and Non-Consensual Edits Act of 2024 ("DEFIANCE"), intended to hold accountable those who are responsible for the proliferation of nonconsensual, sexually explicit deepfake images and videos. The civil remedy applies to digital forgeries that depict the victim in the nude or engaged in sexually explicit conduct of sexual scenarios. This remedy is enforceable against individuals who produced or possess the forgery with intent to distribute it; or who produced, distributed, or received the forgery, if the individual knew or recklessly disregarded that the victim did not consent to the conduct.

The one pager of the legislation specifically mentions the Swift incident: "Sexually-explicit deepfake content is often used to exploit and harass women—particularly public figures, politicians, and celebrities. For example, in January 2024, fake, sexually explicit images of Taylor Swift that were generated by artificial intelligence swept across social media platforms. Although the imagery may be fake, the harm to the victims from the distribution of sexually explicit deepfakes is very real. Victims have lost their jobs and may suffer ongoing depression or anxiety."

This incident is not just another case of a celebrity targeted by digital abuse; this incident may represent a watershed moment in highlighting the pernicious and widespread issue of deepfake non-consensual pornography. The Swift case has succeeded in drawing attention to the severe implications of such acts, not just for public figures, but for individuals across all walks of life. It underscores the ease with which technology can be abused to invade privacy and inflict harm, raising questions about the effectiveness of existing legal protections against sophisticated forms of digital abuse such as nonconsensual pornography, including deepfakes, while protecting the virtues of free expression and technological innovation.

The below chart outlines the states with legislation that specifically target deepfake content including what conduct is prohibited, the penalties for violations, and explicit references to Section 230 protections for social media platforms.

STATE	STATUTE	PROHIBITED CONDUCT	PENALTY	PLATFORM LIABILITY	NOTABLE INFORMATION
California	California Civil Code §1708.85(a)	Intentionally distributing an altered photograph, film, videotape, recording, or other reproduction, in which a reasonable person would believe based upon its context and content, is authentic, of another person without that person's consent, that exposes an intimate body part of the other person or shows the	"General damages" meaning damages for loss of reputation, shame, mortification, and hurt feelings. "Special damages" meaning all damages that plaintiff alleges and proves that he or she has suffered in respect to his or her property, business, trade,	Explicitly does not purport to alter protections under Section 230 of the Communications Decency Act.	Private right of action A person who may assert a cause of action under Section 377.60 of the Code of Civil Procedure (This section allows certain heirs to bring a lawsuit seeking damages for wrongful death) may also

		other person engaging in an act of intercourse, oral copulation, sodomy, or other act of sexual penetration.	profession, or occupation, including the amounts of money the plaintiff alleges and proves he or she has expended as a result of the alleged libel, and no other.		assert a cause of action under this section.
Florida	Florida Senate Bill 1798	Willfully and maliciously promoting any altered sexual depiction of an identifiable person, without the consent of the identifiable person, and who knows or reasonably should have known that such visual depiction was an altered sexual depiction.	Violation is a felony in the third degree. Monetary damages to include $10,000 or actual damages incurred whichever is greater.	Explicitly does not purport to alter protections under Section 230 of the Communications Decency Act.	
Georgia	Georgia Code §16-11-90	Transmission of photograph or video depicting nudity of sexually explicit conduct of an adult including a falsely created videographic or still image: • Knowingly and without consent. • Transmitting or posting photos of a person engaged in a sexual act, or in a state of nudity. • The transmission or posting is for harassment or causes financial loss to the depicted person. • And serves no legitimate purpose to the depicted person.	Violations are treated in the following ways: • Aggravated misdemeanor, punishable by up to 12 months jail time, $1,000 fine, or both. • Subsequent offenses become a felony, punishable by one to five years in prison, a fine of up to $100,000, or both. • If the material is posted on a website that advertises, the offense is treated as a felony punishable by one to five years in prison, a fine of up to $100,000, or both.	Explicitly does not purport to alter protections under Section 230 of the Communications Decency Act.	

			• But subsequent offenses are a felony, punishable by two to five years in prison, a fine of up to $100,000, or both.		
Hawaii	Hawaii Revised Statute §711-1110.9	Violation of privacy in the first degree: • Knowing disclosure or threaten to disclose a realistic photographic image or video of a composite fictitious person depicted in the nude, or engaged in sexual conduct, that includes the recognizable physical characteristics of a known person so that a reasonable person would believe the realistic photographic image or video appears to depict the known person and not a composite fictitious person, with intent to substantially harm the depicted person with respect to that person's health, safety, business, calling, career, education, financial condition, reputation, or personal relationships,	Violations in the first-degree are a class C felony, punishable by up to 5 years in prison, $10,000 fine, or both. Court can also order destruction or sealing of the photos or video.	Explicitly does not purport to alter protections under Section 230 of the Communications Decency Act.	

		or as an act of revenge or retribution.			
Illinois	Illinois House Bill 2123	Intentional dissemination or threatened dissemination by a person over the age of 18 of a private or intentionally digitally altered sexual image without the depicted individual's consent.	Economic and noneconomic damages proximately caused by the defendant's dissemination or threatened dissemination, including damages for emotional distress whether or not accompanied by other damages; or statutory damages, not to exceed $10,000, whichever is greater		
Minnesota	Minnesota House Bill 1370	Nonconsensual dissemination of a deep fake exists when: (1) A person disseminated a deep fake with knowledge that the depicted individual did not consent to its public dissemination; (2) The deep fake realistically depicts any of the following: (i) the intimate parts of another individual presented as the intimate parts of the depicted individual; (ii) artificially generated intimate parts presented as the intimate parts of the depicted individual; or (iii) the depicted individual engaging in a sexual act; and	General and special damages, including all finance losses due to the dissemination of the deep fake and damages for mental anguish. An amount equal to any profit made from the dissemination of the deep fake by the person who intentionally disclosed the deep fake; and A civil penalty awarded to the plaintiff of an amount up to $100,000.	Explicitly does not purport to alter protections under Section 230 of the Communications Decency Act.	

		(3) The depicted individual is identifiable: (i) from the deep fake itself, by the depicted individual or by another individual; or (ii) from the personal information displayed in connection with the deep fake.			
New York	New York Senate Bill S1042A	Unlawful dissemination or publication of an intimate image: • With intent to cause harm to emotional, financial or physical welfare of another person, they intentionally disseminate or publish a still or video image depicting such other person with one or more intimate parts exposed or engaging in sexual conduct including an image created or altered by digitization, where such person may reasonably be identified...without such other person's consent.	Up to one year in jail and a fine up to $1000.		Private right of action
South Dakota	South Dakota Codified Law § 22-21-4	Knowingly and intentionally disseminate or sell any image or recording of another person: • That has been intentionally	Class 1 misdemeanor, punishable by up to one year in jail, a fine of up to $2,000, or both. If the victim is 17 years old or younger:		

		manipulated to create a realistic but false image or recording that would cause a reasonable person to mistakenly believe that the image or recording is authentic.	• Class 6 felony, punishable by up to two years in prison, a fine of up to $4,000, or both.		
Texas	Texas Penal Code §21.165	Knowingly produces or distributes by electronic means a deep fake video that appears to depict the person with the person's intimate parts exposed or engaged in sexual conduct without the effective consent of the person appearing to be depicted.	Class A misdemeanor, punishable by up to one year in jail, a fine of up to $4,000, or both.		
Virginia	Virginia Code §18.2-386.2	Unlawful dissemination or sale of images of another including a person whose image was used in creating, adapting, or modifying a videographic or still image with the intent to depict an actual person and who is recognizable as an actual person by the person's face, likeness, or other distinguishing characteristic: • With intent to coerce, harass, or intimidate. • Maliciously disseminate or sell. • Any videographic or still image of another person. • That depicts the other person as totally nude, or in a state of undress	Class 1 misdemeanor, punishable by up to 12 months in jail, a fine of up to $2,500, or both.	Explicitly does not purport to alter protections under Section 230 of the Communications Decency Act.	

		so to expose genitals, pubic area, buttocks, or female breast. • Where offender knows they are not licensed or authorized to disseminate or sell such material.			