CYBER CIVIL RIGHTS INITIATIVE
Coral Gables, FL
cybercivilrights.org
(305)-284-2547

House Committee on Oversight and Accountability
Cybersecurity, Information Technology, and Government Innovation Subcommittee
Hearing on Addressing Real Harm Done by Deepfakes, March 12, 2024

Written Statement of Dr. Mary Anne Franks
President and Legislative Tech Policy Director, Cyber Civil Rights Initiative
Eugene L. and Barbara A. Bernard Professor in Intellectual Property, Technology, and Civil Rights Law, George Washington Law School

I. Background

I am the Eugene L. and Barbara A. Bernard Professor in Intellectual Property, Technology, and Civil Rights Law at George Washington Law School, where I teach and research in the areas of First Amendment law, cyberlaw, criminal law, and family law. I am also the President and Legislative & Tech Policy Director of the Cyber Civil Rights Initiative (CCRI), the nation's leading nonprofit organization dedicated to combating image-based sexual abuse and other technology-facilitated harms.[1]

CCRI began its work in 2013 with a focus on the non-consensual distribution of private, sexually explicit images, also known as "revenge porn," "nonconsensual pornography," or "non-consensually distributed intimate imagery" (NDII). We pioneered a three-pronged approach to this issue aimed at legislative reform, technological reform, and social reform. We created model legislation, worked with tech companies to develop privacy and harassment policies, and provided support and resources to victim-survivors as well as training for lawyers, judges, and law enforcement. Over the last ten years, we have helped bring the number of states with NDII laws from 3 to 48, assisted in drafting the federal criminal nonconsensual pornography bill now known as the SHIELD Act (S. 412), advocated successfully for every major tech platform to implement policies against NDII, and provided more than 20,000 victim-survivors with support and resources through our 24-hour Image Abuse Helpline.

II. Deepfake Pornography and Other Harmful Digital Forgeries

But even as we were making progress against one form of image-based sexual abuse, another form emerged: so-called "deepfake" pornography. The term "deepfake," a portmanteau of

[1] https://cybercivilrights.org/

"deep-learning" (referring to a method of artificial intelligence) and "fake," entered the public lexicon in 2017, when a user going by that name on the website Reddit used AI technology to insert female celebrities' faces into pornographic videos without their consent.[2] A "deepfake porn" industry soon emerged, which now includes websites and applications that make it possible for anyone to create, solicit, and distribute customized, sexually explicit images and videos of anyone they want: celebrities, politicians, work colleagues, ex-girlfriends, next door neighbors, children.[3] It is no longer even necessary to have multiple pictures or videos of a person to generate these highly realistic digital forgeries. For an ordinary member of the public, a single photo will suffice; in the case of famous individuals, no original photo is needed at all.

While anyone can become the victim of deepfake pornography, women and girls are disproportionately targeted. The people who create, solicit, and distribute deepfake porn of women and girls have many motives—to make money, to channel feelings of inadequacy or rejection, to gain the misguided admiration of their peers—but what they all have in common is a refusal to see women and girls as full and equal persons. Like other forms of sexual exploitation, deepfake porn is used to punish, silence, and humiliate women, pushing them out of the public sphere and away from positions of power and influence. This form of image-based abuse inflicts particularly severe and unique dignitary and expressive harms on those targeted, hijacking their images and identities for entertainment.

The harm caused by artificial nonconsensual pornography is virtually indistinguishable from the harm caused by actual nonconsensual pornography[4]: extreme psychological distress that can lead to self-harm and suicide; physical endangerment that can include in-person stalking and harassment; and financial, professional, and reputational ruin. Fake sexually explicit imagery is also used to extort depicted individuals, including children, into sending actual sexually explicit imagery of themselves to their blackmailers. Law enforcement and victim support services expend precious resources investigating graphic child sexual abuse imagery that turn out not to involve any actual abuse, or any actual children, at all.

Dozens of new "deepfake porn" apps and web services launch every month, producing tens of thousands of images every week that are shared on websites that Google and other search engine providers list prominently in their results.[5] They can also go viral on social media platforms, as was the case with the most recent deepfake porn images of singer Taylor Swift. The digital forgeries of Swift were created using tools including Microsoft's text-to-image AI generator by

---

[2] Mary Anne Franks & Ari Ezra Waldman, *Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions*, 78 Md. L. Rev. 892, 893 (2019)

[3] Kat Tenbarge and Liz Kreutz, A Beverly Hills middle school is investigating students sharing AI-made nude photos of classmates, Feb. 28, 2024, https://www.nbcnews.com/tech/misinformation/beverly-vista-hills-middle-school-ai-images-deepfakes-rcna140775

[4] Mary Anne Franks, *"Revenge Porn" Reform: A View from the Front Lines*, 69 Fla. L. Rev. 1251, 1259 (2017)

[5] Matt Burgess, Deepfake Porn is Out of Control, Wired, https://www.wired.com/story/deepfake-porn-is-out-of-control/, Oct 16, 2023; Kat Tenbarge, *Google and Bing put nonconsensual deepfake porn at the top of some search results,* NBC News, Jan. 11, 2024, https://www.nbcnews.com/tech/internet/google-bing-deepfake-porn-image-celebrity-rcna130445.

users on the website 4chan and the Telegram messaging service, but they went viral on X,[6] formerly known as Twitter. Many mainstream AI apps and services can be used to generate sexually explicit imagery without consent, including of minors,[7] often with obscure terms of service that give the companies that own the services the right to use the material any way they wish.[8]

Non-sexually explicit digital forgeries can also cause tremendous harm. Deepfake technology can be used to falsely depict government officials saying or doing unlawful things they never did, as well as the opposite[9]; to manipulate an image of a peaceful protest to appear violent, or vice versa; to impersonate political candidates giving false election information; to create fake audio recordings that sound exactly like a family member asking for help.

In short, digital forgeries wreak havoc on the lives and reputations of depicted individuals; facilitate blackmail, extortion, and financial scams; and threaten election integrity, national security, news reporting; and society's ability to distinguish truth from falsity.

The same three-pronged approach – legislative, technological, and social reform— that CCRI has used to combat the nonconsensual distribution of actual intimate images is urgently needed to combat AI-facilitated image-based sexual abuse. This means carefully targeted, First Amendment-compliant, comprehensive criminal and civil laws, along with Section 230 reform to reward tech industry recklessness. The adoption of global norms and agreements will also be key. Finally, there needs to be increased support for civil society organizations that offer expertise and support regarding technological abuses and dedicated resources for the development of a quick-response mechanism for educating the public about emerging tech-based threats.

III. Recommendations for Federal Legislation

The ideal law to address digital forgeries is one that both provides resources to existing victims and seeks to prevent the creation of future ones. Federal legislation, providing both criminal and civil penalties, is urgently needed to address the escalating problem of digital forgeries. While a handful of states[10] have enacted legislation against deepfake porn, these laws vary widely in definition and scope and the majority of states have yet to act. As digital forgeries often cross

---

[6] Samantha Cole and Emanuel Maiberg, *The Taylor Swift Deepfakes Disaster Threatens to Change the Internet As We Know It*, 404 Media, Jan. 31, 2024, https://www.404media.co/ai-generated-taylor-swift-porn-twitter/

[7] Olivia Snow, *'Magic Avatar' App Lensa Generated Nudes From My Childhood Photos*, Wired, Dec. 7, 2022, https://www.wired.com/story/lensa-artificial-intelligence-csem/.

[8] Meera Navlakha, *Lensa AI app: What to know about the self portrait generator*, Mashable**,** Dec. 5, 2022, https://mashable.com/article/lensa-ai-app-explainer

[9] See, e.g., Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 Cal. L. Rev. 1753, 1785 (2019)

[10] Elliott Davis Jr, *These States Have Banned the Type of Deepfakes That Targeted Taylor Swift*, US News, Jan. 30, 2024, https://www.usnews.com/news/best-states/articles/2024-01-30/these-states-have-banned-the-type-of-deepfake-porn-that-targeted-taylor-swift

multiple jurisdictions, it is essential that prevention and remedies are consistent across the United States and that victims have a clear path for redress.

Criminal prohibition is essential for multiple reasons. Given the severe and irreversible damage inflicted by sexual digital forgeries, the most effective legal response is one that prioritizes deterring perpetrators from engaging in this abuse to begin with. Criminal penalties provide the most potential for deterrence.[11] Would-be perpetrators are far more likely to be deterred by the prospect of jail time than the remote and abstract possibility of being sued. In CCRI's national survey on the victimization and perpetration of nonconsensual pornography, respondents who admitted to having engaged in nonconsensual pornography were asked what might have stopped them from engaging in the abuse. Out of more than a dozen possible factors, the majority of respondents (60%) indicated that harsh criminal penalties would have been the most effective deterrent.[12]

While civil remedies can provide meaningful opportunities after the fact for content removal, restraining orders, and economic compensation, these can at best mitigate, not prevent, the harm. Even if a victim wins damages or obtains an injunction forcing the poster to take down the image, the grim reality is that a harmful image may never truly be removed once it has been made available online.[13] It is also important to remember that pursuing civil actions require money, time, and resources that many victims simply do not have.

Finally, making the distribution of sexually explicit digital forgeries a federal crime also has the benefit of preventing tech platforms from raising Section 230 as a defense against liability.

To be effective and to survive constitutional challenge, a digital forgery law must clearly identify the elements of the harmful conduct, including the required mental state, and ensure that it fits within a category of established First Amendment exceptions. Fortunately, there is longstanding precedent for regulating false, harmful expression that is perceived by others to be true. While false expression that is clearly not intended or likely to be mistaken for real depictions of individuals, such as parody and satire, enjoy considerable protection under the First Amendment, defamation and fraud have historically been considered exceptions to full First Amendment protection, and criminal prohibitions against impersonation, counterfeiting, and forgery have never raised serious constitutional objection. There is also precedent for regulating harmful false expression regardless of whether it is likely to be perceived as authentic, as demonstrated by the

---

[11] While "solid empirical evidence of the deterrent effect of criminalization is hard to come by--it is always hard to discern why people do not engage in crime…. we do know that certainty, more so than severity, of punishment does factor into deterrence, and it is generally acknowledged that most people fear jail more than lawsuits." Mary Anne Franks, *"Revenge Porn" Reform: A View from the Front Lines*, 69 Fla. L. Rev. 1251, 1304 (2017)

[12] Asia A. Eaton et al., *2017 Nationwide Online Study of Nonconsensual Porn Victimization and Perpetration: A Summary Report*, Cyber C.R. Initiative 11 (June 12, 2017), 22, https://www.cybercivilrights.org/wp-content/uploads/2017/06/CCRI-2017-Research-Report.pdf.

[13] Mary Anne Franks, *"Revenge Porn" Reform: A View from the Front Lines*, 69 Fla. L. Rev. 1251, 1300 (2017)

tort of false light and federal criminal legislation prohibiting "morphed" child pornography that combines the faces of real children with the bodies of adults.[14]

At a minimum, federal criminal and civil penalties should be available for the intentional distribution of sexually explicit, photo-realistic visual material that appears to depict an actual, identifiable individual without that individual's consent.

The law should make clear that disclaimers of inauthenticity or non-consent will not immunize distributors from liability. While disclaimers can potentially mitigate the harm of some digital forgeries, this is not the case for sexually explicit digital forgeries. The names of many deepfake porn sites and services make clear that the images and videos are fake (for example, the notorious website MrDeepFakes[15]), but such disclosures do little to reduce the harm of nonconsensual sexual exploitation. Individuals depicted in deepfakes report feeling exposed, humiliated, and alienated from their bodies regardless of whether the material is labeled as fake.

Additionally, in the digital age, audiovisual material rarely remains confined to its original context, especially if the material goes viral. Disclaimers of inauthenticity that originally accompany a digital forgery are likely to be lost as the scope of distribution increases. It is also important to note that identifying false information as false has limited corrective impact in any event. Studies have shown that repeated exposure to false information, even when presented for the purposes of correction, increases the likelihood that the false information will be remembered as true, a phenomenon known as "the illusory truth effect."[16]

IV. Analysis of HR 3106 (Preventing Deepfakes of Intimate Images Act) and the DEFIANCE Act

As an initial matter, CCRI urges Congress to remember that it has not yet prohibited the unauthorized disclosure of *actual* intimate imagery. While we are fully supportive of the effort to prevent and prohibit non-consensual, digitally manipulated sexual imagery, we hope that Congress will at the same time finally pass legislation, namely the SHIELD Act (S. 412), to prohibit the disclosure of private, authentic, intimate imagery.

A. HR 3106, the Preventing Deepfakes of Intimate Images Act

[14] See Clay Calvert, *Artificial Intelligence and Morphed Sexual Imagery of Real Minors: Pushing the Free-Speech Envelope Too Far, AEIdeas*, Nov. 14, 2023, https://www.aei.org/technology-and-innovation/artificial-intelligence-and-morphed-sexual-imagery-of-real-minors-pushing-the-free-speech-envelope-too-far/

[15] See Kat Tenbarge, *Found through Google, bought with Visa and Mastercard: Inside the deepfake porn economy*, NBC News, March 27, 2023, https://www.nbcnews.com/tech/internet/deepfake-porn-ai-mr-deep-fake-economy-google-visa-mastercard-download-rcna75071

[16] Mary Anne Franks & Ari Ezra Waldman, *Sex, Lies, and Videotape: Deep Fakes and Free Speech Delusions*, 78 Md. L. Rev. 892, 895 (2019)

CCRI strongly supports HR 3106, the Preventing Deepfakes of Intimate Images Act[17], which we worked on with sponsor Congressman Joe Morelle and which we believe is the strongest and most effective bill proposed by Congress so far because it provides for both criminal and civil penalties. The bill makes it illegal to disclose or threaten to disclose a manipulated "intimate digital depiction" with reckless disregard for whether the disclosure "will cause physical, emotional, reputational, or economic harm to the depicted individual" or "with the intent to harass, annoy, threaten, alarm, or cause substantial harm to the finances or reputation of the depicted individual." It also amends the existing federal civil remedy for the disclosure of nonconsensual pornography to include manipulated intimate digital depictions, allowing for restraining orders, injunctions requiring the removal of the content, and economic damages.

B. DEFIANCE Act

CCRI applauds the efforts of the DEFIANCE Act.[18] The bill adopts the term "digital forgery," which CCRI suggested using for this content in 2019. Like HR 3106, this bill would expand existing federal civil law on nonconsensual pornography to extend to inauthentic as well as authentic sexually explicit visual material, and offers similar proposed remedies, including restraining orders or injunctions requiring the removal of the content.

The DEFIANCE Act differs from the House bill in two significant ways. It is broader than HR 3106 in that it applies not only to the nonconsensual *distribution* of digital forgeries, but also to the solicitation, production and possession of this content. In that respect, the Act helpfully captures a broader range of abusive conduct.

The other significant difference is that while HR 3106 contains both a criminal provision and a civil provision, the DEFIANCE Act only provides a civil remedy. As noted above, this would severely limit the effectiveness of a legislative solution, and thus we would recommend that it be combined with a criminal penalty along the lines of HR 3106.

[17] https://www.congress.gov/bill/118th-congress/house-bill/3106?q=%7B%22search%22%3A%22hr+3106+hr+3106%22%7D&s=3&r=3
[18] https://acrobat.adobe.com/id/urn:aaid:sc:us:83d22542-02aa-46d7-befe-b26e23ffa1e7