Deepfakes, Real Enforcement Challenges

David S. Louk*

* J.D., Yale Law School; Ph.D., Jurisprudence & Social Policy, UC Berkeley; Deputy City Attorney, San Francisco City Attorney's Office; Former Academic Fellow, Columbia Law School. The author wishes to thank Professor Jane Ginsburg, the faculty and staff of the Kernochan Center, and the student editors of the *Columbia Journal of Law & the Arts* for hosting the excellent symposium out of which this Article formed, as well as for their skilled and diligent edits to the Article. No non-public information was relied upon in researching and writing this Article. All views professed in this piece are this author's alone and do not reflect the position of any other individuals or institutions.

INTRODUCTION

The advent and rapid advancement of generative AI ("gen AI") technology over the past several years has been accompanied by a proliferation of AI-generated deepfake pornography, which depicts real people in intimate ways that they neither participated in, nor consented to. The ability to cheaply and easily generate realistic-looking, AI-generated audio, video, or images that convincingly mimic real individuals depicted in intimate contexts without their consent, known as non-consensual intimate depictions ("NCID"), presents a unique challenge in technology regulation. This Article focuses on a subset of those depictions, non-consensual intimate images (or "NCII").[1] In contrast to the contentious and heavily debated free-speech and intellectual property issues raised by artificial intelligence and deepfakes in other contexts, NCII features a

1. *See Nonconsensual Distribution of Intimate Images: What to Know*, FED. TRADE COMM'N: CONSUMER ADVICE (Nov. 2024), https://consumer.ftc.gov/articles/nonconsensual-distribution-intimate-images-what-know [https://web.archive.org/web/20260206180803/https://consumer.ftc.gov/articles/nonconsensual-distribution-intimate-images-what-know].

comparatively resolute moral and legal consensus. Indeed, NCII is one of the rare contemporary political issues where there exists bipartisan and near-universal agreement that the practice must be stopped. Despite this, effective enforcement mechanisms to curtail the practice have remained somewhat elusive; the gap between recognizing the harm and implementing effective remedies has yet to be closed.

This Article examines the technological underpinnings of AI-generated NCII deepfakes and the legal apparatuses that have sprung up recently to address it, catalogs the practical and legal obstacles to successful enforcement, and proposes strategic approaches for more effective regulation. While the rapid development and deployment of gen AI technology means that no perfect solution is likely to emerge overnight. Instead, state and national legislatures, cross-border government enforcement offices, and leading multinational technology companies will each need to play critical roles to mitigate the worst of the harms inflicted by NCII deepfakes. This need is also urgent, because the technology to mass produce deepfake nonconsensual intimate video content is rapidly developing, which will surely only multiply the problem.

I. DEFINING THE PROBLEM: NCII DEEPFAKES

Non-consensual intimate images convincingly depict real individuals in false but intimate contexts—and without their consent. To date, existing technology has primarily supported the easy and widespread dissemination of static images, and so this Article's focus is on NCII, though widespread audio and video depictions are sure to follow. NCII differs from the emergence of so-called "revenge porn" cases in the early 2010s, which typically featured *real* intimate content but that was distributed without the consent of the depicted.[2] By contrast, contemporary NCII deepfakes involve wholly fabricated depictions of real individuals who neither consented to their creation nor participated in any underlying intimate activity. Worse, an even more troubling subset of NCII involves Child Sexual Abuse Material (CSAM), which depict minors in sexual contexts, and which can be created from any available image of a clothed minor.[3]

A. OPEN SOURCE MODELS AND THE EROSION OF GUARDRAILS

Given widespread recognition of the social harms and illegality of NCII deepfakes, how have they proliferated so rapidly? At present, most major gen AI platforms, such as ChatGPT's Dall-E and Sora image- and video-generating tools, have implemented guardrails that seek to prevent NCII and CSAM generation.[4] Nevertheless, these

2. *See, e.g.*, Mary Anne Franks, *"Revenge Porn" Reform: A View from the Front Lines*, 69 FLA. L. REV. 1251, 1257–61 (2017).

3. *What Is NCII?*, INHOPE (Feb. 17, 2023), https://web.archive.org/web/20250124021021/https://inhope.org/EN/articles/what-is-ncii.

4. *See, e.g.*, *Child Safety: Adopting Safety by Design Principles*, OPENAI (Apr. 23, 2024) https://openai.com/index/child-safety-adopting-sbd-principles/ [https://web.archive.org/web/20260210060944/https://openai.com/index/child-safety-adopting-sbd-

protections can be circumvented through prompt manipulation,[5] and numerous web forums are dedicated to guiding users in how to override such guardrails.[6] Indeed, Grok, xAI's gen AI image generating software, experienced widespread issues with NCII creation notwithstanding claims it was not designed to produce such content.[7] (More on this below.) What all closed-source models share in common, however, is the ability to implement new safeguards, alter input sanitation methods, and, in extreme circumstances, limit access to the model altogether.

More significant, however, are the systemic vulnerabilities presented by open-source models, a veritable Pandora's box that, once opened, cannot be closed down. For example, when Stability AI released a version of its Stable Diffusion model in 2022, version 1.5, the open-source text-to-image generator inadvertently contained training

principles/] ("strong guardrails and safety measures" in ChatGPT and DALL-E); *Sora System Card*, OPENAI (Dec. 9, 2024) https://openai.com/index/sora-system-card/ [https://web.archive.org/web/20260303044058/https://openai.com/index/sora-system-card/] ("multi-tiered moderation strategy" including classifiers and blocklists); *Building Safeguards for Claude*, ANTHROPIC (Aug. 12, 2025) https://www.anthropic.com/news/building-safeguards-for-claude [https://web.archive.org/web/20260206195902/https://www.anthropic.com/news/building-safeguards-for-claude] (CSAM detection through hash comparison); *Gemini for Safety Filtering and Content Moderation*, GOOGLE CLOUD (updated Mar. 5, 2026) https://docs.cloud.google.com/vertex-ai/generative-ai/docs/multimodal/gemini-for-filtering-and-moderation [https://web.archive.org/web/20260206195940/https://docs.cloud.google.com/vertex-ai/generative-ai/docs/multimodal/gemini-for-filtering-and-moderation] (CSAM safety filters).

5. *See, e.g.*, Emilia Napolano, *Sora: Inappropriate and Harmful Content Creation Easily Bypassed Through Simple Prompt Engineering*, ZENODO (Apr. 26, 2025), https://zenodo.org/records/15295087 [https://perma.cc/7T5G-QFZA]. Users are also able to use image-generating platforms to create "fetish content" of real individuals. Katie Notopoulos, *Sora Might Have a "Pervert" Problem on Its Hands*, BUS. INSIDER (Oct. 24, 2025), https://www.businessinsider.com/sora-video-openai-fetish-content-my-face-problem-2025-10 [https://web.archive.org/web/20260102103649/https://www.businessinsider.com/sora-video-openai-fetish-content-my-face-problem-2025-10].

6. Prompt manipulation techniques—commonly called "jailbreaking"—can bypass AI safety guardrails through various methods, including roleplay scenarios, encoded instructions, multi-turn desensitization, and contextual manipulation. *See* Yi Liu et al., *Jailbreaking ChatGPT via Prompt Engineering: An Empirical Study*, ARXIV (May 23, 2023) https://arxiv.org/abs/2305.13860 [https://web.archive.org/web/20260204204720/https://arxiv.org/abs/2305.13860] (empirical study of seventy-eight jailbreak prompts across ten distinct patterns including pretending, attention shifting, and privilege escalation); *see also* Danny Bradbury, *Researchers Break OpenAI Guardrails*, MALWAREBYTES (Oct. 13, 2025), https://www.malwarebytes.com/blog/news/2025/10/researchers-break-openai-guardrails [https://web.archive.org/web/20251109062710/https://www.malwarebytes.com/blog/news/2025/10/researchers-break-openai-guardrails] (describing "Policy Puppetry" technique that bypassed safety measures across ChatGPT, Claude, and Gemini by disguising prompts as configuration files); Victor Tangermann, *Researchers Find Easy Way to Jailbreak Every Major AI, from ChatGPT to Claude*, FUTURISM (Apr. 25, 2025), https://futurism.com/easy-jailbreak-every-major-ai-chatgpt [https://web.archive.org/web/20251125080933/https://futurism.com/easy-jailbreak-every-major-ai-chatgpt] (reporting HiddenLayer exploit bypassed safety guardrails across all major frontier AI models using prompt injection combined with roleplaying). Unfortunately, research indicates these attacks succeed approximately 20% of the time, requiring an average of forty-two seconds and as few as five interactions to bypass safety guardrails—some succeed in under four seconds. *What Is AI Jailbreaking?*, IRONSCALES (Nov. 2025), https://ironscales.com/glossary/what-is-ai-jailbreaking [https://web.archive.org/web/20251125161610/https://ironscales.com/glossary/what-is-ai-jailbreaking] (citing IBM Research 2024 data).

7. *See infra* section I.C.

data that included intimate and CSAM content.[8] Because open-source models can be freely downloaded and retrained on user-selected datasets, bad actors quickly repurposed this technology.[9] The result was a proliferation of "nudification" and "undressing" applications and websites starting in 2023,[10] which promoted their services with taglines like "[i]magine wasting time taking her out on dates, when you can just use [the website] to get her nudes."[11] These platforms leveraged retrained models specifically optimized for generating non-consensual intimate content.[12]

Thus, by September 2023, an estimated 24 million users per month visited nudification websites—a figure likely significantly higher today.[13] Analysis of deepfake content reveals that approximately 96–99% of deepfake pornography targets women,[14] and a 2024 survey by Thorn found that among 13–20 year-olds, one in eight personally know someone who has been a victim of an NCII deepfake.[15] Additionally, the Center

8. David Evan Harris & Dave Willner, *Was an AI Image Generator Taken Down for Making Child Porn?*, IEEE SPECTRUM (Aug. 30, 2024), https://spectrum.ieee.org/stable-diffusion [https://web.archive.org/web/20251129101851/https://spectrum.ieee.org/stable-diffusion].

9. William Hawkins, Chris Russell & Brent Mittelstadt, *Deepfakes on Demand: The Rise of Accessible Non-Consensual Deepfake Image Generators*, FACCT '25: PROCS. OF THE 2025 ACM CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 1, 1 (2025).

10. *See id.* at 1, 2 (finding that "both Stable Diffusion and Flux models are used for the creation of deepfake models, with 96% of these targeting women," and identifying over 34,000 downloadable deepfake model variants intended to generate images of identifiable individuals); Kyle Wiggers, *Deepfakes for All: Uncensored AI Art Model Prompts Ethics Questions*, TECHCRUCNCH (Aug. 26, 2022), https://techcrunch.com/2022/08/24/deepfakes-for-all-uncensored-ai-art-model-prompts-ethics-questions/
[https://web.archive.org/web/20251113085158/https://techcrunch.com/2022/08/24/deepfakes-for-all-uncensored-ai-art-model-prompts-ethics-questions/] ("Creative and malicious users can abuse the capabilities [of Stable Diffusion] to generate subjectively objectionable content at scale, using minimal resources to run inference —which is cheaper than training the entire model— and then publish them in venues like 4chan.").

11. First Amended Complaint ¶ 6, People v. Sol Ecom, Inc., No. CGC-24-617237 (Cal. Super. Ct. filed Mar. 10, 2025).

12. *See* Santiago Lakatos, *A Revealing Picture*, GRAPHIKA (Dec. 8, 2023), https://graphika.com/reports/a-revealing-picture
[https://web.archive.org/web/20260201082539/https://graphika.com/reports/a-revealing-picture].

13. *Id.*; *see also* Margi Murphy, *"Nudify" Apps That Use AI to "Undress" Women in Photos Are Soaring in Popularity*, TIME (Dec. 9, 2023), https://time.com/6344068/nudify-apps-undress-photos-women-artificial-intelligence/ [https://web.archive.org/web/20251223195706/https://time.com/6344068/nudify-apps-undress-photos-women-artificial-intelligence/].

14. Hailey Reissman, *What Is Deepfake Porn and Why Is It Thriving in the Age of AI?*, ANNENBERG SCH. COMMC'N UNIV. PA. (July 13, 2023), https://www.asc.upenn.edu/news-events/news/what-deepfake-porn-and-why-it-thriving-age-ai
[https://web.archive.org/web/20251128224820/https://www.asc.upenn.edu/news-events/news/what-deepfake-porn-and-why-it-thriving-age-ai]; Manjeevan Singh Seera & Ridoan Karim, *Digital Child Abuse: Deepfakes and the Rising Danger of AI-Generated Exploitation*, LENS (MONASH UNIV.) (Feb. 25, 2025), https://lens.monash.edu/@politics-society/2025/02/25/1387341/digital-child-abuse-deepfakes-and-the-rising-danger-of-ai-generated-exploitation
[https://web.archive.org/web/20251023085721/https://lens.monash.edu/@politics-society/2025/02/25/1387341/digital-child-abuse-deepfakes-and-the-rising-danger-of-ai-generated-exploitation].

15. Olina Banerji, *More Teens Than You Think Have Been "Deepfake" Targets*, EDUC. WEEK (Mar. 3, 2025), https://www.edweek.org/technology/more-teens-than-you-think-have-been-deepfake-

for Democracy & Technology's 2024 survey of high school students found that 15% (representing approximately 2.3 million students) had heard of an NCII deepfake depicting someone at their school during the 2023–24 school year.[16] (These numbers are almost certainly higher today.) Celebrity NCII deepfake content has proliferated through dedicated forums where users solicit custom content for cryptocurrency payments, creating a shadow economy around non-consensual sexual imagery.[17]

B. TECHNICAL EVASION OF DETECTION SYSTEMS

In the face of this profound diffusion of morally troublesome content, the easiest solution would be to implement sound guardrails on the technology itself, but as noted, no guardrails can guarantee prevention of the production of NCII. Closed-source image-generating models like Grok—which permit generation of "spicy" content—have had documented failures in preventing NCII generation that has recently resulted in several prominent lawsuits from victims, even as the platform disclaimed any intent to allow the creation of such content. Open-source platforms fare even worse: Once an open-source model has been released that can be trained to generate deepfake NCII content, in the absence of a kill-switch embedded by the model's creator, there is no way to "shut off" the technology. In such circumstances, the next best approach would be to develop technologies to identify and remove deepfake NCII content, but current detection and attribution technologies remain inadequate. Studies have shown that human detection of deepfake content generally is barely above chance, with overall accuracy of only 55.54%, and odds ratios indicating detection accuracy as low as 39%.[18] In theory, watermarks embedded in deepfake content can provide notice to viewers that content is AI-generated, potentially mitigating deception.[19] However, while some platforms embed watermarks or metadata in generated content, tools can readily remove these markers: Indeed, users of deepfake NCII creation tools will pay premium

---

targets/2025/03 [https://web.archive.org/web/20260123083556/https://www.edweek.org/technology/more-teens-than-you-think-have-been-deepfake-targets/2025/03].

16. Elizabeth Laird, Maddy Dwyer & Kristin Woelfel, *Report—In Deep Trouble: Surfacing Tech-Powered Sexual Harassment in K-12 Schools*, CTR. FOR DEMOCRACY & TECH. (Sep. 26, 2024), https://cdt.org/insights/report-in-deep-trouble-surfacing-tech-powered-sexual-harassment-in-k-12-schools/ [https://web.archive.org/web/20260126002252/https://cdt.org/insights/report-in-deep-trouble-surfacing-tech-powered-sexual-harassment-in-k-12-schools/]; Kara Arundel, *Schools Lack Supports for Victims of Sexually Explicit Deepfake and Real Images*, K-12 DIVE (Sep. 26, 2024), https://www.k12dive.com/news/schools-deepfake-images-student-supports/728107/ [https://web.archive.org/web/20260126002252/https://cdt.org/insights/report-in-deep-trouble-surfacing-tech-powered-sexual-harassment-in-k-12-schools/].

17. Lakatos, *supra* note 12.

18. Alexander Diel et al., *Human Performance in Detecting Deepfakes: A Systematic Review and Meta-Analysis of 56 Papers*, 16 COMPUT. IN HUM. BEHAV. REPS. 1, 4, 6 (2024).

19. Nicola Henry, *AI "Nudify" Sites Are Being Sued for Victimising People. How Can We Battle Deepfake Abuse?*, CONVERSATION (Aug. 21, 2024), https://theconversation.com/ai-nudify-sites-are-being-sued-for-victimising-people-how-can-we-battle-deepfake-abuse-237043 [https://web.archive.org/web/20251204115841/https://theconversation.com/ai-nudify-sites-are-being-sued-for-victimising-people-how-can-we-battle-deepfake-abuse-237043].

fees specifically to remove watermarks that identify images as fake, and watermarks can also be easily cropped using simple graphic editing software.[20] For example, in one survey of such apps, seven of twenty analyzed AI nudification applications watermarked their output images and sold removal of those watermarks at the highest subscription tier, suggesting that watermark removal is a built-in feature of the NCII creation ecosystem.[21]

Gen AI content can itself be detected by AI software, though, to date, with limited efficacy. Current deepfake detection technologies have limited effectiveness in real-world scenarios, with accuracy reduced when lighting conditions, facial expressions, or video quality differ from training data, and future advances in deepfake generation are expected to eliminate current detection hallmarks.[22] Metadata, though more difficult to remove, offers no obvious warning to consumers who view the images.[23] Neither approach provides sufficient fail-safe protection for victims. While transparency practices like watermarking and labeling may aid in accountability and content moderation, doing so does not render noticeably "fake" content unharmful; even noticeably fake NCII can still result in mental and physical harm, reputational damage, and financial costs to the victims depicted.[24]

C. DISTRIBUTION

A third challenge with preventing the spread of NCII deepfakes is the speed and scope with which they can be almost instantaneously disseminated worldwide. As discussed below, the legal distinctions between the websites and apps that generate the images, and the platforms (like Facebook, Snapchat, and X) that host them, have generally limited accountability against the platforms as hosts.

One rare instance where this has not been the case is Grok, the xAI gen AI tool that is integrated directly into X.com's social media website. Grok has recently faced widespread media attention—and subsequent lawsuits—after journalists, regulators, and advocacy groups documented in early 2026 that Grok could be used to generate deepfake NCII by taking photos of real individuals and "undressing" them

20. Marco Viola & Cristina Voto, *Designed to Abuse? Deepfakes and the Non-Consensual Diffusion of Intimate Images*, 201 SYNTHESE 1, 9 (Jan. 13, 2023).

21. Cassidy Gibson et al., *Analyzing the AI Nudification Application Ecosystem*, SEC '25: PROCS. OF THE 34TH USENIX CONF. ON SEC. SYMP. 1, 11 (Nov. 14, 2024).

22. *Science & Tech Spotlight: Combating Deepfakes*, U.S. GOV'T ACCOUNTABILITY OFF. (Mar. 11, 2024), https://www.gao.gov/products/gao-24-107292 [https://web.archive.org/web/20260129175027/https://www.gao.gov/products/gao-24-107292].

23. *See* Michelle L. Ding & Harini Suresh, *The Malicious Technical Ecosystem: Exposing Limitations in Technical Governance of AI-Generated Non-Consensual Intimate Images of Adults*, 2025 CONF. ON HUM. FACTORS IN COMPUTING SYS. SOCIOTECH. AI GOVERNANCE WORKSHOP (Apr. 24, 2025); Barry Collins, *AI or Not? How to Detect if an Image Is AI-Generated*, FORBES (Oct. 14, 2023), https://www.forbes.com/sites/barrycollins/2023/10/14/ai-or-not-how-to-detect-if-an-image-is-ai-generated/ [https://web.archive.org/web/20260401162430/https://www.forbes.com/sites/barrycollins/2023/10/14/ai-or-not-how-to-detect-if-an-image-is-ai-generated/].

24. Ding & Suresh, *supra* note 23.

without consent. Targets included adult women, public figures, and in some cases minors or apparent minors.[25]

What set Grok apart from the cascade of other nudifying websites that have proliferated in recent years is that it is not simply a gen AI model that can, perhaps inadvertently, generate deepfake NCII for an individual user; rather, Grok is integrated with X in a way that lets users easily feed in platform images and generate altered versions at scale, then circulate them on the same network. Grok was thus described as a "one-click harassment machine,"[26] with California Attorney General Rob Bonta accusing xAI of appearing to facilitate the large-scale production of deepfake NCII used to harass women and girls across the internet, including on X.[27]

In response to public backlash in mid-January 2026, *Wired* reported that after days of outrage, X had started limiting Grok image generation on X to paying subscribers, but that this did not solve the problem: Sexualized "undressing" images were still being produced, and the standalone Grok app and website could still generate harmful content.[28] Critics called that change inadequate insofar as it only reduced access on one surface while leaving abuse possible elsewhere—in their view, putting abuse behind a paywall rather than stopping it. Later that month, Attorney General Bonta sent xAI a cease-and-desist letter demanding that it stop the creation and distribution of deepfake NCII and child sexual abuse material, citing possible violations of California civil and criminal law.[29] As of the publication of this article, the Attorney General had not yet announced the results of this investigation.

In addition to the unusually integrated combination of gen AI model and social media platform, another key source of the problem was the gap between xAI's written rules and Grok's actual behavior—another theme discussed below. xAI's Acceptable Use Policy already said users must not violate statutory privacy rights, depict people's likenesses "in a pornographic manner," or sexualize children.[30] Yet despite such written prohibitions, Grok was still reported to be generating exactly that sort of material in practice, and appeared to lack sufficient guardrails to prevent it.

---

25. *See* Press Release, Rob Bonta, Att'y General (Cal.), Attorney General Bonta Launches Investigation into xAI, Grok over Undressed, Sexual AI Images of Women and Children (Jan. 14, 2026), https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-investigation-xai-grok-over-undressed-sexual-ai [https://web.archive.org/web/20260331030724/https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-investigation-xai-grok-over-undressed-sexual-ai].

26. Nilly Patel, *Why Nobody's Stopping Grok*, THE VERGE (Jan. 22, 2026), https://www.theverge.com/podcast/865275/grok-deepfake-undressing-elon-musk-content-moderation [https://web.archive.org/web/20260326132124/https://www.theverge.com/podcast/865275/grok-deepfake-undressing-elon-musk-content-moderation].

27. *See* Press Release, Bonta, *supra* note 25.

28. Matt Burgess, *X Didn't Fix Grok's "Undressing" Problem. It Just Makes People Pay for It*, WIRED (Jan. 9, 2026), https://www.wired.com/story/x-didnt-fix-groks-undressing-problem-it-just-makes-people-pay-for-it/ [https://web.archive.org/web/20260319224509/https://www.wired.com/story/x-didnt-fix-groks-undressing-problem-it-just-makes-people-pay-for-it/].

29. *See* Press Release, Bonta, *supra* note 25.

30. *xAI Acceptable Use Policy*, XAI (effective Jan. 2, 2025), https://x.ai/legal/acceptable-use-policy [https://web.archive.org/web/20260401152721/https://x.ai/legal/acceptable-use-policy].

The controversy with Grok and xAI thus encapsulates many of the challenges with deepfake NCII: the technological hurdles to prevent the creation of any such images, the speed with which they can be distributed, and the challenges with identifying the individuals who created the images (and, Grok notwithstanding, the websites that generated them).

II. THE EXISTING LEGAL FRAMEWORK

In response to the rise of NCII deepfakes, a broad, bipartisan consensus quickly emerged that legal, not just technological, solutions were necessary to address this problem. To date, a patchwork of state and federal laws—as well as overseas laws and regulations—seeks to address the problem.

A. FEDERAL PROHIBITIONS AND SECTION 230

Given the widespread emergence of NCII deepfakes and the largely bipartisan reaction against them, it should not be surprising that even a historically unproductive[31] Congress has recently enacted federal legislation to address aspects of the NCII deepfake problem. Until recently, while multiple federal laws have addressed *aspects* of deepfake NCII content, they did so by reaching deepfake NCII content only incidentally. Moreover, because none of these laws was designed to specifically address the problem, most focus largely on individual content creators or distributors. Such laws include, for example, longstanding state and federal prohibitions on the possession, distribution, and creation of CSAM, even of content not depicting real children.[32] Another federal law, enacted in 2016, prohibits non-consensual intimate disclosure, but it was originally designed to target so-called "revenge pornography," and was enacted against a backdrop where there was an expected relationship between the victim and the perpetrator.[33] In addition, general obscenity prohibitions in the U.S. Code have also provided additional grounds for prosecution of the *distribution* of AI-generated CSAM—if not the possession.[34]

31. Minho Kim & Ashley Wu, *How the House Slumped to Historic Lows of Productivity in 2025*, N.Y. TIMES (Jan. 16, 2026), https://www.nytimes.com/interactive/2026/01/17/us/politics/house-republicans-majority-productivity.html.

32. At the federal level, CSAM is criminalized under multiple statutes. 18 U.S.C. § 2251 prohibits sexual exploitation of children and production of CSAM, with mandatory minimum sentences of fifteen to thirty years for first-time offenders. Sections 2252 and 2252A criminalize transportation, distribution, receipt, and possession of CSAM, with mandatory minimums of five years for receipt and distribution offenses, and up to ten years for simple possession. 18 U.S.C. §§ 2252, 2252A. These statutes apply regardless of whether actual children were depicted.

33. See Consolidated Appropriations Act, 2022, Pub. L. No. 117-103, 136 Stat. 49, which amended 15 U.S.C. § 6851.

34. *See* Rianna Pfefferkorn, *Court Rules That Constitution Protects Private Possession of AI Generated CSAM*, TECH POLICY PRESS (Mar. 20, 2025), https://www.techpolicy.press/court-rules-that-constitution-protects-private-possession-of-aigenerated-csam/ [https://web.archive.org/web/20260306114005/https://www.techpolicy.press/court-rules-that-constitution-protects-private-possession-of-aigenerated-csam/] (discussing U.S. v. Anderegg, Case No. 24-

Moreover, with the exception of CSAM, section 230 of the Communications Decency Act has also, until recently, presented a significant obstacle to platform accountability for deepfake NCII.[35] Under section 230(c)(1), online platforms are not treated as the "publisher or speaker" of information provided by third-party users, effectively immunizing them from liability for user-generated content.[36] Courts have interpreted this provision broadly since the landmark 1997 decision in *Zeran v. America Online, Inc.*, which held that platforms cannot be held liable for third-party content even when they have knowledge of its illegal nature.[37] This sweeping immunity has historically prevented victims of deepfake NCII from pursuing claims against the platforms that host and distribute such images, as platforms successfully invoke section 230 to obtain early dismissal of suits that would treat them as publishers of harmful content.[38]

Given the widespread, bipartisan consensus that deepfake NCII content must be addressed, section 230's broad grant of immunity has recently been reconsidered by Congress. One early result is the]landmark federal law addressing deepfake NCII content, the Take It Down Act, which was signed into law on May 19, 2025, and extends liability beyond individual creators to platforms hosting, generating, or disseminating such content.[39] The Act criminalizes the knowing publication of NCII (both authentic and AI-generated) with penalties of up to two years imprisonment for content involving adults and up to three years for content involving minors.[40] Critically, it also requires covered platforms to establish notice-and-removal mechanisms and to remove reported NCII within forty-eight hours of receiving valid requests.[41] Enforcement authority resides with the Federal Trade Commission, which may treat violations as unfair or deceptive trade practices.[42]

---

cr-50-jdp (W.D. Wisc. Feb. 13, 2025) and noting that the court distinguished between the defendant's possession and distribution of virtual CSAM and denying defendant's motion to dismiss the *distribution* count).

35. 47 U.S.C. § 230 (2018).

36. *Id.* § 230(c)(1) ("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.").

37. Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997) (holding that section 230 creates "a federal immunity to any cause of action that would make service providers liable for information originating with a third-party user of the service").

38. *See, e.g.*, Kaitlin O'Donnell, *Have We No Decency? Section 230 and the Liability of Social Media Companies for Deepfake Videos*, 2021 U. ILL. L. REV. 701, 704 (2021) ("The limitations of federal solutions are due, in part, to Section 230 of the Communication Decency Act which provides immunity to social media companies for content posted on their sites.").

39. Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act (TAKE IT DOWN Act), Pub. L. No. 119-12, 139 Stat. 55 (2025) (codified at 47 U.S.C. § 223) [hereinafter the "Take It Down Act"].

40. *Id.* § 2(a).

41. *Id.* § 3(a)(3).

42. *Id.* § 3(b); *see also* Billee Elliott McAuliffe & Clare H. Nowogrocki, *New Federal TAKE IT DOWN Act Gives Victims an Avenue for the Removal of Deepfake Images*, LEXOLOGY (May 19, 2025), https://www.lexology.com/library/detail.aspx?g=2fed5f0c-e9fa-4a45-ba3c-075866c92f49 [https://web.archive.org/web/20260207034708/https://www.lexology.com/library/detail.aspx?g=2fed5f0c-e9fa-4a45-ba3c-075866c92f49%2A%2A].

While the Take It Down Act's notice-and-takedown mechanisms for NCII represents a seeming carve-out from section 230 immunity,[43] the precise contours of the Take It Down Act's carveout for section 230 immunity remains unresolved, with commentators advancing mixed views about whether (and how) the Act overrides traditional platform immunity. Perhaps the majority view is that the Act eliminates section 230 protection, a position supported by a fair reading of its text. The Act is codified at section 223 of the Communications Act, which section 230(e)(1) explicitly exempts from immunity coverage.[44] Under this view, the Take It Down Act amends section 223 to impose notice-and-takedown obligations on platforms. Since section 230 immunity does not extend to section 223 offenses , platforms cannot invoke section 230 to defend against FTC enforcement actions for failing to comply with takedown requirements.[45] Moreover, proponents of this position argue that the Act "contradicts the basic immunity" of section 230(c)(1) by treating platforms *as publishers or speakers* when it imposes liability for their decisions regarding hosting and removing third-party content—precisely what section 230 was designed to prevent.[46] The Act creates asymmetric liability by providing a safe harbor only for good-faith removals, not for refusing removal requests, thereby incentivizing platforms to remove content without investigation to avoid penalties that can exceed $50,000 per violation.[47]

However, some questions remain as to whether section 230 immunity survives in circumstances outside FTC enforcement actions. This interpretation suggests that

43. *See* Jeffrey D. Neuburger & Jonathan Mollod, *Take It Down Act Signed into Law, Offering Tools to Fight Non-Consensual Intimate Images and Creating a New Image Takedown Mechanism*, PROSKAUER: NEW MEDIA & TECH. L. BLOG (May 29, 2025), https://newmedialaw.proskauer.com/2025/05/29/take-it-down-act-signed-into-law-offering-tools-to-fight-non-consensual-intimate-images-and-creating-a-new-image-takedown-mechanism/
[https://web.archive.org/web/20260207035255/https://newmedialaw.proskauer.com/2025/05/29/take-it-down-act-signed-into-law-offering-tools-to-fight-non-consensual-intimate-images-and-creating-a-new-image-takedown-mechanism/].

44. 47 U.S.C. §§ 230(e)(1), (e)(3) (2018) ("Nothing in this section shall be construed to impair the enforcement of . . . any . . . Federal criminal statute . . . or any . . . State law that is consistent with this section"); Take It Down Act § 2 (to be codified at 47 U.S.C. § 223(h)).

45. *See e.g.*, Neuburger & Mollod, *supra* note 43, ("Since the Take It Down Act states that it will be codified at section 223 of the Communications Act of 1934 (i.e., 47 U.S.C. 223(h)), it appears that platforms would not enjoy CDA protection from FTC civil enforcement actions based on the agency's authority to enforce the Act's requirements that covered platforms 'reasonably comply' with the new Take It Down Act notice-and-takedown obligations.").

46. *See, e.g.*, Thomas J. Cunningham & Michael J. McMorrow, *Platforms Face Section 230 Shift From Take It Down Act*, TROUTMAN PEPPER (June 9, 2025), https://www.troutman.com/insights/platforms-face-section-230-shift-from-take-it-down-act/
[http://web.archive.org/web/20260207041556/https://www.troutman.com/insights/platforms-face-section-230-shift-from-take-it-down-act/] ("[T]he act contradicts the basic immunity provided by [section 230] . . . , that '[n]o 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.' The act treats the providers or users of interactive computer services in exactly that way, by imposing liability on them as the publisher or speaker of 'information provided by another information content provider.'").

47. *Id.* (noting that failure to comply "shall be treated as a violation of a rule" under the FTC Act, with penalties currently at $53,088 per violation, and that the Act "provides no safe harbor . . . for rejecting or refusing to honor a request for removal"); *see also* Take It Down Act § 3(b)(2) .

while platforms likely cannot invoke section 230 against FTC civil enforcement for failure to implement compliant takedown procedures, they *would* retain immunity in other significant respects. First, it is arguable whether platforms maintain section 230 protection from claims related to hosting or publishing NCII that has not been the subject of a prior, valid Take It Down Act notice, since the Act's removal requirements are triggered only upon receiving a compliant takedown request.[48] Second, it may be the case that platforms could assert section 230 immunity against private lawsuits by individuals (as opposed to FTC enforcement) alleging harm from a platform's failure to comply with takedown notices, particularly since the Act creates no private right of action and platforms could argue such suits impermissibly treat them as publishers of third-party content.[49] Third, the Good Samaritan provision of section 230(c)(2) arguably shields platforms from liability for proactively filtering or removing NCII, which would be supplemented by the Act's own safe harbor for good-faith removals.[50] Under this narrower view, the Take It Down Act represents not a wholesale elimination of section 230 immunity but rather a targeted carve-out limited to federal enforcement of notice-and-takedown compliance obligations. Such a view is arguably further supported by the existence of additional proposed legislation that would more clearly strip all section 230 immunity from platforms, such as the Intimate Privacy Protection Act, which would condition section 230 immunity on platforms implementing a "duty of care" to prevent and remove deepfake NCII.[51] Because the Take It Down Act's takedown provisions do not go into effect until May 2026, the precise legal contours of the relationship between the Act and section 230 remain to be seen as of the publication of this article.

B. STATE-LEVEL RESPONSES

Given the severity of the problem posed by NCII deepfakes—and the broad bipartisan consensus against the practice—all fifty states have enacted some version of a non-consensual intimate image law, and at least forty-five states have subsequently amended or added specific prohibitions to target NCII deepfakes.[52] The broad,

48. Neuburger & Mollod, *supra* note 43 ("[T]he Act's requirements for removal of NCII by platforms would not be implicated without a valid removal request.").

49. *Id.* ("Similarly, a platform could make a strong argument that it retains CDA immunity from any claims brought by an individual (rather than the FTC) for failing to reasonably comply with a Take It Down Act notice.").

50. *Id.*

51. Press Release, Rep. Jake Auchincloss, Auchincloss Introduces Bipartisan Bill to Tackle Rise in Non-Consensual Deepfakes on Social Media Platforms (July 30, 2025), https://auchincloss.house.gov/media/press-releases/release-auchincloss-introduces-bipartisan-bill-to-tackle-rise-in-non-consensualdeepfakes-on-social-media-platforms [http://web.archive.org/web/20260207042913/https://auchincloss.house.gov/media/press-releases/release-auchincloss-introduces-bipartisan-bill-to-tackle-rise-in-non-consensualdeepfakes-on-social-media-platforms].

52. *See* Kaylee Williams, *Minors Are on the Frontlines of the Sexual Deepfake Epidemic—Here's Why That's a Problem*, TECH POL'Y PRESS (Oct. 10, 2024), https://www.techpolicy.press/minors-are-on-the-frontlines-of-the-sexual-deepfake-epidemic-heres-why-thats-a-problem/

bipartisan legislative action is reflective of widespread agreement that the problem demands intervention.

California's approach is illustrative of the evolving state-level response. Assembly Bill 602, enacted in 2019, created a civil cause of action for non-consensual image distribution (originally targeting "revenge pornography") and provided for disgorgement and up to $150,000 in statutory damages.[53] The statute applies to persons who create and intentionally disclose sexually explicit material when they know or reasonably should know the depicted individual did not consent, or who intentionally disclose such material they did not create knowing the depicted individual did not consent.[54] In 2025, the California legislature amended this statute to expressly enumerate liability for nudification websites, extending responsibility to platforms facilitating content creation and establishing specific statutory enforcement provisions for public prosecutors, including the attorney general and city and county attorneys.[55] Thus, just as the Federal Trade Commission (FTC) has been delegated primary enforcement authority over the Take It Down Act, California law anticipates the attorney general and city and county attorneys will have a central role to play in enforcement at the state level.

Finally, it is worth noting that the abovementioned state and federal prohibitions on NCII deepfakes also exist within a broader ecosystem of identity rights laws that also apply to NCII deepfakes. While not all states denominate these laws as "right of publicity" statutes, many have expressly adopted common-law privacy rights to protect against unauthorized appropriation of identity or recognized a "right of publicity" either by common law or by statute.[56] These laws have traditionally protected against unauthorized commercial exploitation of a person's name, image, and likeness (as well as, increasingly, voice), though their scope varies significantly from jurisdiction to jurisdiction. Additionally, federal and state trademark laws, unfair competition statutes, and consumer protection laws also restrict unauthorized uses of identity that

---

[http://web.archive.org/web/20260207043301/https://www.techpolicy.press/minors-are-on-the-frontlines-of-the-sexual-deepfake-epidemic-heres-why-thats-a-problem/]; *Revenge Porn Laws: State by State*, C.A. GOLDBERG, https://www.cagoldberglaw.com/resources/states-with-revenge-porn-laws/ [https://web.archive.org/web/20260226051618/https://www.cagoldberglaw.com/resources/states-with-revenge-porn-laws/] (last visited Apr. 1, 2026); *Tracker: State Legislation on Intimate Deepfakes*, Pub. Citizen (updated Mar. 31, 2026), https://www.citizen.org/article/tracker-intimate-deepfakes-state-legislation/ [https://web.archive.org/web/20260310101025/https://www.citizen.org/article/tracker-intimate-deepfakes-state-legislation/].

53. A.B. 602, 2021–22 Leg., Reg. Sess. (Cal. 2021), codified at CAL. CIV. CODE § 1708.86 (West 2021); Douglas E. Mirell & Joshua Geller, *AB 602 and AB 730: Curbing "Deepfakes" in Pornography and Elections*, DAILY J. (Jan. 8, 2020), https://www.dailyjournal.com/articles/355794-ab-602-and-ab-730-curbing-deepfakes-in-pornography-and-elections [http://web.archive.org/web/20260207044110/https://www.dailyjournal.com/articles/355794-ab-602-and-ab-730-curbing-deepfakes-in-pornography-and-elections].

54. CAL. CIV. CODE § 1708.86(b) (West 2025).

55. CAL. CIV. CODE § 1708.86 (West 2025) (as amended).

56. JENNIFER E. ROTHMAN, UNIFORM LAW COMMISSION PROTECTION OF NAME, IMAGE, [VOICE], AND LIKENESS STUDY COMMITTEE: REPORTER'S WELCOME MEMO 10–11 (2025); Jennifer E. Rothman & Robert Post, *The First Amendment and the Right(s) of Publicity*, 130 Yale L.J. 86, 94-95, 88 (2020).

create confusion or falsely indicates an endorsement.[57] As discussed below, these longstanding legal rights can also serve as a tool for enforcement against NCII deepfake content creation and distribution.[58]

III. PRACTICAL ENFORCEMENT CHALLENGES

A. ATTRIBUTION DIFFICULTIES

Identifying whether NCII content is a "deepfake" is challenging enough, but tracing content creation presents even greater difficulties.[59] Identifying both the individual prompting image generation and the specific application or website used proves exceptionally challenging.[60] Users typically operate behind VPNs or burner accounts, uploading content that rapidly proliferates across multiple platforms.[61] Within minutes of initial posting—particularly for celebrity content—attribution becomes functionally impossible as thousands of users redistribute images. For example, AI-generated sexually explicit images of musician and global celebrity Taylor Swift "gained over 45 million views, along with hundreds of thousands of likes, bookmarks, and reposts over a seventeen-hour period" before being taken down.[62] This viral spread outpaces effective legal response, creating a "whack-a-mole" problem where content removal efforts fail to contain dissemination.[63] (The lawsuits against Grok and xAI prove a notable exception, but Grok is also the rare image generator tied to a social media platform that reposts those same images, making attribution unusually straightforward.)[64]

57. *See* Jennifer E. Rothman, *Navigating the Identity Thicket: Trademark's Lost Theory of Personality, the Right of Publicity, and Preemption*, 135 HARV. L. REV. 1271, 1278–79 (2022).

58. *See infra* Part VI.A.

59. *See* Irene Amerini et al., *Deepfake Media Forensics: Status and Future Challenges*, 11 J. IMAGING 1, 29–30 (2025).

60. Gueltom Bendiab et al., *Deepfakes in Digital Media Forensics: Generation, AI-Based Detection and Challenges*, 88 J. INFO. SEC. & APPLICATIONS 1 (2025).

61. Catherine Han et al., *Characterizing the MrDeepFakes Sexual Deepfake Marketplace, in* SEC '25: PROCS. OF THE 34TH USENIX CONF. ON SEC. SYMP. 5169 (2025).

62. Halle Nelson, *Taylor Swift and the Dangers of Deepfake Pornography*, NAT'L SEXUAL VIOLENCE RES. CTR. (Feb. 7, 2024), https://www.nsvrc.org/blogs/feminism/taylor-swift-and-dangers-deepfake-pornography [https://web.archive.org/web/20260211200149/https://www.nsvrc.org/blog_post/taylor-swift-and-dangers-deepfake-pornography/].

63. Kat Tenbarge, *Nude Deepfakes of Taylor Swift Went Viral on X, Evading Moderation and Sparking Outrage*, NBC NEWS (Jan. 25, 2024), https://www.nbcnews.com/tech/misinformation/taylor-swift-nude-deepfake-goes-viral-x-platform-rules-rcna135669 [https://web.archive.org/web/20260211200554/https://www.nbcnews.com/tech/misinformation/taylor-swift-nude-deepfake-goes-viral-x-platform-rules-rcna135669].

64. *See supra* Part I.C.

B. LOW BARRIERS TO ENTRY FOR VIOLATIVE PLATFORMS

The availability of free open-source models and detailed instructional guides on web forums means nudification websites can be easily created and recreated.[65] The substantial revenue generated—driven by millions of monthly visitors—incentivizes continued operation despite legal risks.[66] It is easy to see why: a recent 2025 analysis of eighty-five such websites found that they averaged 18.5 million visitors monthly over a six-month period, with revenue estimates reaching as high as $36 million per website annually.[67] Nor is it especially easy to uncover the identities of the owners and operators: Registry information for these websites is frequently falsified or outdated, and many operate from overseas jurisdictions, complicating enforcement when content affects U.S. residents but originates from servers in Croatia, Estonia, China, or other countries.[68]

IV. LEGAL ENFORCEMENT OBSTACLES

In addition to the practical challenges of reining in deepfake NCII, certain legal obstacles—namely, section 230 and First Amendment protection—may also, at the margins, discourage enforcement. But while the invocation of section 230 and First Amendment defenses frequently looms large over efforts at legal regulation of many forms of deepfakes, such concerns are more remote in the context of NCII content.

A. SECTION 230 IMMUNITY CONCERNS

As noted above, section 230 has long played a central role in limiting enforcement of prohibitions on content creation and distribution online, and while the Take It Down Act has at least partially shifted this landscape, the precise contours remain to be determined. While section 230 immunity may not shield websites explicitly advertising nudification services—as such platforms clearly function as content contributors rather than neutral hosts—ambiguity persists regarding distribution platforms. The Take It Down Act attempts to address these concerns by creating specific takedown obligations that platforms must meet, with FTC enforcement authority treating violations as

65. *See* Hawkins et al., *supra* note 9.

66. *See* Lakatos, *supra* note 12.

67. Matt Burgess, *AI "Nudify" Websites Are Raking in Millions of Dollars*, WIRED (July 14, 2025), https://www.wired.com/story/ai-nudify-websites-are-raking-in-millions-of-dollars/ [https://web.archive.org/web/20260211200950/https://www.wired.com/story/ai-nudify-websites-are-raking-in-millions-of-dollars/].

68. Kolina Koltai, *Behind a Secretive Global Network of Non-Consensual Deepfake Pornography*, BELLINGCAT (Feb. 23, 2024), https://www.bellingcat.com/news/2024/02/23/behind-a-secretive-global-network-of-non-consensual-deepfake-pornography/ [https://web.archive.org/web/20260211201407/https://www.bellingcat.com/news/2024/02/23/behind-a-secretive-global-network-of-non-consensual-deepfake-pornography/] (finding that deepfake NCII websites are "incorporated in ways to hide the identity of their operators" and use virtual office services, with multiple companies using fake business addresses and falsely claiming partnerships with legitimate companies like Microsoft and G2A).

unfair or deceptive practices.[69] As noted, questions remain about whether and how section 230 protections interact with the Act's enforcement mechanisms.[70]

Uncertainty also persists regarding whether state-law identity rights claims—including right of publicity and NCII claims—fall within section 230's exception for intellectual property laws. Federal courts have reached conflicting conclusions on this question, with some holding that section 230 bars state publicity claims while others permit them to proceed.[71] This uncertainty complicates platform liability strategies and creates inconsistent protection for victims depending on jurisdiction.

B. FIRST AMENDMENT CONSIDERATIONS

Although First Amendment concerns frequently loom large in the context of regulating expressive conduct, free-speech challenges appear less formidable in this context than for other kinds of deepfake content. Obscenity and defamation have long constituted recognized exceptions to First Amendment protection,[72] and CSAM receives no constitutional protection at all. Obscenity is unprotected because it lacks serious literary, artistic, political, or scientific value and appeals to prurient interests in a patently offensive manner.[73] Similarly, defamation, which involves false statements that harm an individual's reputation, is generally excluded from First Amendment protection where the individual is a private figure (or, in the case of a public figure, where actual malice is proven).[74] CSAM, on the other hand, is categorically unprotected due to its intrinsic connection to the exploitation and abuse of children.[75] These exceptions reflect the U.S. Supreme Court's longstanding principle that certain types of speech, due to their harmful nature and lack of societal value, do not warrant constitutional safeguards.

The Supreme Court has yet to decide whether NCII—whether real or deepfake—is obscene, and therefore outside the protections of the First Amendment. But challenges to so-called "revenge porn" statutes have failed even where courts have declined to find such content categorically obscene. For example, in *State v. VanBuren*, the Vermont Supreme Court declined to label all nonconsensual pornography "obscene," but nevertheless, applying strict scrutiny, upheld a Vermont statute criminalizing its

69. Take It Down Act §§ 3(a)(3), 3(b).

70. *See* James Grimmelmann, Deconstructing *the Take It Down Act*, COMMC'NS OF THE ACM (July 30, 2025), https://cacm.acm.org/opinion/deconstructing-the-take-it-down-act/ [https://web.archive.org/web/20260211201709/https://cacm.acm.org/opinion/deconstructing-the-take-it-down-act/].

71. *Compare* Hepp v. Facebook, 14 F.4th 204 (3d Cir. 2021) (holding that Section 230 does not bar state right of publicity claim) *with* Perfect 10, Inc. v. CCBill LLC, 488 F.3d 1102 (9th Cir. 2007) (holding that Section 230 immunity barred right of publicity claim).

72. *See* Brown v. Ent. Merchs. Ass'n, 564 U.S. 786, 791 (2011).

73. *See* Miller v. California, 413 U.S. 15, 39 (1973).

74. *See, e.g.*, New York Times Co. v. Sullivan, 376 U.S. 254, 279–280 (1964); Gertz v. Robert Welch, Inc., 418 U.S. 323, 345–50 (1974).

75. *See* New York v. Ferber, 458 U.S. 747, 759–60 (1982); United States v. Williams, 553 U.S. 285, 288 (2008).

dissemination as narrowly tailored to serve a compelling State interest.[76] By contrast, the Illinois Supreme Court went even further, upholding Illinois's law under *intermediate scrutiny*, on the grounds that the prohibition was content-neutral and the statute regulated purely private matters and not speech on matters of public concern.[77]

More problematic are statutory prohibitions that sweep up protected content alongside that which falls outside the First Amendment's safeguards, because statutes that prohibit a substantial amount of protected expression may be deemed unconstitutionally "overbroad," failing even facial review.[78] Because most deepfake NCII statutory prohibitions have been enacted just within the past several years, the precise First Amendment boundaries remain unsettled and will be defined over time. Nevertheless, it is likely most will survive for the same reason nonconsensual pornography bans have: The state has a compelling interest in prohibiting the content and possesses few other available means to do so. The more interesting question is likely potential edge cases swept up by such bans: say, involving an artistic depiction of adults or newsworthy content—such as a journalist photographing a topless protester—that may technically constitute NCII but serve legitimate societal purposes or concern matters of public interest. Here, the risk would come not from whether banning NCII content *itself* is impermissibly overbroad, but whether if doing so inadvertently sweeps up enough protected speech alongside it that could succeed on an as-applied basis.

C. JURISDICTIONAL LIMITATIONS

Another significant enforcement challenge comes from the international scope of deepfake NCII content generation. Cross-border enforcement remains persistently problematic when websites or applications operate from foreign jurisdictions but generate content for users in the United States. The cross-border problem is not just that deepfake NCII content is unlawful in some places but under-regulated in others. Rather, the full abuse chain is typically transnational: The model or service has been developed and operated in one country, hosted in another, prompted by a user in a third, distributed via globally available platforms, and viewed wherever the victim may live. That fragmentation creates several recurring legal and practical obstacles.

V. STRATEGIC APPROACHES FOR ENHANCED ENFORCEMENT

Between the Take It Down Act and the kaleidoscope of recently enacted state laws, avenues for potential enforcement are theoretically vast. Notwithstanding the abundance of laws, however, there remains a deficit of enforcement.

A. PRIORITIZING GOVERNMENT AND PLATFORM ENFORCEMENT OVER INDIVIDUAL

76. State v. VanBuren, 214 A.3d 791, 800 (2019), as supplemented (June 7, 2019).
77. People v. Austin, 155 N.E.3d 439, 458 (2019).
78. Ashcroft v. Free Speech Coal., 535 U.S. 234, 244 (2002).

ACTIONS

At present, individual victim enforcement faces significant obstacles, even in seemingly straightforward cases. The experience of "Jane Doe," a New Jersey teenager, illustrates these challenges vividly. In October 2023, Doe discovered that a classmate, "K.G.," had used an AI nudification website, Clothoff.io, to generate deepfake nude images from her fully clothed Instagram photos and distributed them via Snapchat.[79] Though Doe successfully identified her perpetrator—an exceptional circumstance given that most victims never discover who created their images—enforcement barriers persisted. Criminal charges were not pursued because information gathered by school officials could not be used in the criminal investigation, and the defendant and witnesses refused to cooperate with law enforcement or provide access to their electronic devices.[80]

In February 2024, Doe filed a federal lawsuit against K.G. seeking $150,000 per image, injunctive relief, and destruction of all copies.[81] The psychological toll for Doe was devastating: Doe experienced "enormous distress" that "disrupted her high school education," living with "hopelessness and perpetual fear" knowing the images will "almost inevitably make their way onto the Internet."[82] Even this partial success—obtaining civil recourse against an identified perpetrator in close proximity—left the underlying platform problem unaddressed.

Thus, in October 2025, Doe filed a second lawsuit, this time targeting Clothoff.io itself, which Doe alleged is operated by AI/Robotics Venture Strategy 3 Ltd. in the British Virgin Islands.[83] Doe's second complaint includes both state and federal CSAM claims as well as state invasion of privacy claims, clearly illustrating the overlapping patchwork quilt of legal prohibitions against creating and disseminating NCII deepfake content. However, Doe's second complaint also illustrates the difference between legal rules on the books and effective enforcement on the ground: The complaint reveals the profound difficulties of pursuing overseas operators who use "pseudonyms, fake names and addresses, and third-party payment options to avoid detection and legal accountability," allegedly operating through Belarus-based individuals.[84] The procedural history of Doe's multi-year legal battle to seek justice demonstrates that even when victims overcome initial identification hurdles, cross-border jurisdiction and platform anonymity render individual enforcement anything but straightforward.

In this author's view, these challenges counsel toward systemic government enforcement mechanisms to support and complement individual victim lawsuits, an

79. Complaint at 4–5, Doe v. K.G., No. 2:24-cv-00634 (D.N.J. Feb. 2, 2024).

80. *Id.* at 11–12.

81. *Id.* at 28–29.

82. *Clinics File Suit Against Website that Generates Nonconsensual Nude Images*, YALE L. SCH. (Nov. 4, 2025), https://law.yale.edu/yls-today/news/clinics-file-suit-against-website-generates-nonconsensual-nude-images [https://web.archive.org/web/20260211202247/https://law.yale.edu/yls-today/news/clinics-file-suit-against-website-generates-nonconsensual-nude-images]; *id.* at 13.

83. Complaint at 5, Doe v. AI/Robotics Venture Strategy 3 Ltd., No. 2:25-cv-16671 (D.N.J. Oct. 16, 2025).

84. *Id.* at 30–42; *Clinics File Suit*, *supra* note 82.

approach recently-enacted laws like the Take It Down Act and California's amended NCII laws anticipate. One notable example of this approach is the August 2024 lawsuit filed by San Francisco City Attorney David Chiu against sixteen nudification websites, which collectively received over 200 million visits in the first half of 2024.[85] The suit, brought on behalf of the People of the State of California under the state's unfair competition law, seeks to shut down these websites and permanently enjoin their operation based on violations of state and federal revenge pornography, child pornography, and deepfake laws.[86] One targeted site, Clothoff.io—the same platform used against Jane Doe—had received approximately 26.9 million visits alone.[87] Since the filing of the People's suit, a majority of the sites are no longer accessible,[88] and two have entered into settlements with the People agreeing to permanently shut down the websites.[89]

As noted, the pending enforceability of the Take It Down Act's takedown provisions leaves open the broader question of platform liability—targeting not just content-creation platforms like nudification websites, but also platforms like Facebook and Snapchat. While such an approach has the potential to do much more to address the broader problem of systemic generation and viral dissemination, rather than isolated incidents with individual victims, much remains depending on how the FTC chooses to enforce the Act, how platforms respond, and how courts assess potential availability of section 230 defenses.

B. ENHANCING MODEL DISTRIBUTION REGULATIONS

While potentially controversial among open-source advocates, the emergence of NCII deepfakes also presents a cautionary tale for model distribution practices and raises important questions about the dangers of releasing open-source gen AI models. A proverbial Pandora's box was opened when Stable Diffusion 1.5 provided a sufficiently sophisticated open-source model without adequate retraining guardrails.[90] Since November 2022, the Civitai platform alone has hosted over 34,000 downloadable deepfake model variants, which collectively have been downloaded more than 15

85. First Amended Complaint, *supra* note 11, at 2, 4–6.

86. *Id.* at 20–24; *see* Heather Knight, *San Francisco Moves to Lead Fight Against Deepfake Nudes*, N.Y. TIMES (Aug. 15, 2024), https://www.nytimes.com/2024/08/15/us/deepfake-pornography-lawsuit-san-francisco.html.

87. Complaint at ¶ 136, People v. Sol Ecom, Inc., No. CGC-24-617237, WL 3833798 (Cal. Super. Ct. filed Aug. 14, 2024).

88. Video posted by David Chiu, FACEBOOK, *It has been one year since my Office filed our first-of-its-kind lawsuit against websites that use AI to create deepfake pornography.* (Aug. 28, 2025), https://www.facebook.com/davidchiu.sf/videos/it-has-been-one-year-since-my-office-filed-our-first-of-its-kind-lawsuit-against/745058168296841/ [https://web.archive.org/web/20260207004341/https://www.facebook.com/davidchiu.sf/videos/it-has-been-one-year-since-my-office-filed-our-first-of-its-kind-lawsuit-against/745058168296841/].

89. *See* Stipulated Judgments of May 30, 2025 and Dec. 19, 2025, People v. Sol Ecom, Inc., No. CGC-24-617237 (Cal. Super. Ct. filed Mar. 10, 2025).

90. Harris & Willner, *supra* note 8.

million times.[91] Video technology will rapidly and inevitably reach similar sophistication, and without protective measures, the current image-based crisis will extend to video content.

The challenge also extends beyond NCII to broader questions about identity rights and commercialization. Open-source models that enable unauthorized replication of voices and likenesses threaten not only privacy and dignity interests, but also the commercial value of identity for performers, athletes, and public figures. Policies addressing these concerns must carefully balance innovation benefits against potential harms, recognizing that overly restrictive approaches may stifle legitimate creative and educational uses, while overly permissive rules will fail to prevent determined bad actors from accessing or creating harmful tools.

Policymakers should consider several approaches to reducing the risk of open-source generative AI models contributing to NCII deepfakes. First, all platforms should commit to mandatory safety testing and red-teaming[92] before model release to minimize the risk of unanticipated usage or consequences and to ensure model guardrails cannot be broken.[93] Second, models should include built-in technical safeguards against malicious fine-tuning (although particularly with open-source models, current techniques have proven circumventable, and questions remain about the durability of open-source safeguards).[94] Especially critical are provisions allowing

91. Hawkins, *supra* note 9, at 1603.

92. Red-teaming is a process for evaluating and testing gen AI models to discover vulnerabilities, flaws, and unexpected outputs. To do so, the red-teamers often attempt to generate precisely the kind of content the model was designed to prohibit, either through manual prompt manipulation or use of specialized software that can engage with that model directly." Evelyn Yee, AI Red-Teaming Design: Threat Models and Tools, Oct. 24, 2025, https://cset.georgetown.edu/article/ai-red-teaming-design-threat-models-and-tools/ [http://web.archive.org/web/20260226184716/https://cset.georgetown.edu/article/ai-red-teaming-design-threat-models-and-tools/].

93. *See* Exec. Order No. 14,110, 88 Fed. Reg. 75191 (Oct. 30, 2023) (mandating red-teaming for high-risk AI systems); Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 [hereinafter the "EU AI Act"], arts. 15, 17, 2024 O.J. (L 1689) 1, 61 (requiring providers of high-risk AI systems to conduct testing and validation before deployment); *see also The Future of AI Red Teaming: Challenges, Trends, and What's Next*, AYA DATA (Nov. 13, 2025), https://www.ayadata.ai/the-future-of-ai-red-teaming-challenges-trends-and-whats-next/ [https://web.archive.org/web/20260207015645/https://www.ayadata.ai/the-future-of-ai-red-teaming-challenges-trends-and-whats-next/] (explaining that EU AI Act requires operators of high-risk AI systems to demonstrate accuracy and robustness through rigorous testing, and U.S. Executive Order mandates red teaming with safety test results shared with government agencies before deployment).

94. *See Second Key Update: Technical Safeguards and Risk Management*, INT'L AI SAFETY REP. (Nov. 25, 2025), https://internationalaisafetyreport.org/publication/second-key-update-technical-safeguards-and-risk-management/ [https://web.archive.org/web/20260207022050/https://internationalaisafetyreport.org/publication/second-key-update-technical-safeguards-and-risk-management] (noting that research explores "unlearning" techniques and methods to modify how models process harmful concepts, but such modifications "can often be reversed by actors with the technical skill to fine-tune models"); Xiangyu Qi et al., *On Evaluating the Durability of Safeguards for Open-Weight LLMs,* INT'L CONF. ON LEARNING REPRESENTATIONS 2025 1, 1 (2025) https://proceedings.iclr.cc/paper_files/paper/2025/file/9d3a4cdf6f70559e8c6fe02170fba568-Paper-Conference.pdf [https://perma.cc/QR48-6TWP] (demonstrating through case studies that evaluating

platforms to disable models found to facilitate widespread abuse, potentially backed by liability safe harbors for good-faith moderation efforts.[95] This is especially critical because once open-source models without remote disabling technology become available, it becomes nearly impossible to put the cat back in the bag.[96] Though current open-source models like Stable Diffusion primarily generate static images and lack the video generation capabilities of proprietary systems, more advanced open-source video platforms are quickly emerging, and will pose even greater risks unless preventive frameworks are established. These measures should be incorporated into risk-based regulatory frameworks that distinguish between different capability levels and use cases, as contemplated in the EU AI Act. Such an approach, for example, might distinguish between models than can generate only text-based outputs and those that can also generate images, video, or audio.[97]

C. CROSS-BORDER COOPERATION

Given practical and legal hurdles to worldwide enforcement and persistent extraterritoriality issues, enhanced international cooperation remains essential for meaningful progress. Comparative international approaches reveal both promising approaches and persistent gaps. The United Kingdom's Online Safety Act 2023, implemented over the following two years, criminalizes both the sharing and creation

technical safeguards is "exceedingly difficult" and defenses may mislead audiences about durability); Alex Petropoulos, Bengüsu Özcan & Max Reddel, *Can Open-Weight Models Ever Be Safe?*, CTR. FOR FUTURE GENERATIONS (Sep. 18, 2025), https://cfg.eu/can-open-weight-models-ever-be-safe/ [https://web.archive.org/web/20260207023543/https://cfg.eu/can-open-weight-models-ever-be-safe/] (discussing tamper-resistant architectures and machine unlearning techniques, but noting researchers have criticized these approaches for "either weakening useful capabilities or lacking resilience to circumvention"); Wiggers, *supra* note 10 ("Safety Classifier—while on by default—can be disabled.").

95. *See, e.g.*, Robert Gorwa & Michael Veale, *Moderating Model Marketplaces: Platform Governance Puzzles for AI Intermediaries*, 16 L., INNOVATION & TECH. 341, 341 (2024) (examining how Hugging Face, GitHub, and Civitai moderate models, including use of licensing, access restrictions, and automated content moderation).

96. *See* Martin Anderson, *CivitAI Tightens Deepfake Rules Under Pressure From Mastercard and Visa*, UNITE.AI (May 20, 2025), https://www.unite.ai/civitai-tightens-deepfake-rules-under-pressure-from-mastercard-and-visa/ [https://perma.cc/TT6M-AYD2] (reporting that Civitai banned models designed to replicate real people and added mandatory 50% noise alteration for uploaded images after payment processor pressure); *but see* Emanuel Maiberg, *Hugging Face Is Hosting 5,000 Nonconsensual AI Models of Real People*, 404 MEDIA (July 15, 2025), https://www.404media.co/hugging-face-is-hosting-5-000-nonconsensual-ai-models-of-real-people/ [https://web.archive.org/web/20260207030425/https://www.404media.co/hugging-face-is-hosting-5-000-nonconsensual-ai-models-of-real-people/] (documenting that over 5,000 models designed to create nonconsensual sexual content were reuploaded to Hugging Face after Civitai banned them).

97. *See* EU AI Act, *supra* note 93, arts. 5–6 (establishing prohibited AI practices and classification rules for high-risk AI systems based on their purpose and potential impact); *see also* Haiman Wong, *Mapping the Open-Source AI Debate: Cybersecurity Implications and Policy Priorities*, R ST. INST. (Apr. 17, 2025), https://www.rstreet.org/?post_type=research&p=85817 [https://perma.cc/JMU9-UT2Q] (discussing hybrid approaches that balance transparency with rigorous oversight, noting Meta's Llama model "requires users to apply for access and enforces a license that explicitly prohibits high-risk applications"); Petropoulos, *supra* note 94 (arguing for "tiered approach to evaluating the risks of an open-weight model at a given capability level and deciding whether it can be released safely under current conditions").

of sexually explicit deepfakes without consent, with penalties including unlimited fines and up to two years imprisonment.[98] The Act designates intimate image abuse as "priority illegal content," requiring covered platforms to implement systems for removal and empowering Ofcom—the UK communications regulator—to seek court orders compelling internet service providers to withdraw services from non-compliant sites.[99] Similarly, France amended its Code pénal in 2024 to criminalize non-consensual sexual deepfakes, imposing penalties of up to two years imprisonment and €60,000 fines.[100]

The European Union's AI Act, which entered into force in 2024, mandates transparency for AI-generated content and outlaws the most egregious forms of AI-based identity manipulation.[101] And the General Data Protection Regulation (GDPR)[102] continues to provide protection for depicted individuals' personal data: For instance, a Dutch court, citing the GDPR, recently ordered X/Grok to stop producing AI-generated non-consensual sexualized imagery, which extends even beyond a ban on NCII to include "sexualized" but non-nude depictions.[103] However, the EU approach focuses primarily on disclosure requirements and platform obligations rather than criminal penalties for individual creators. The European Commission's proposed regulation on child sexual abuse explicitly addresses deepfake CSAM, though implementation remains pending.[104]

These varied approaches highlight both the international consensus that NCII deepfakes require intervention and the lack of harmonized enforcement strategies. Effective cross-border cooperation requires more than parallel national laws—it

98. *Criminalising Deepfakes—The UK's New Offences Following the Online Safety Act*, HERBERT SMITH FREEHILLS KRAMER (May 21, 2024), https://www.herbertsmithfreehills.com/notes/tmt/2024-05/criminalising-deepfakes-the-uks-new-offences-following-the-online-safety-act [http://web.archive.org/web/20240619083113/https://www.herbertsmithfreehills.com/notes/tmt/2024-05/criminalising-deepfakes-the-uks-new-offences-following-the-online-safety-act]; Press Release, Sarah Sackman, MP, Ministry of Just. (UK), Better Protection for Victims Thanks to New Law on Sexually Explicit Deepfakes (Jan. 22, 2025), https://www.gov.uk/government/news/better-protection-for-victims-thanks-to-new-law-on-sexually-explicit-deepfakes [https://web.archive.org/web/20260207040234/https://www.gov.uk/government/news/better-protection-for-victims-thanks-to-new-law-on-sexually-explicit-deepfakes].

99. WOMEN AND EQUALITIES COMMITTEE, TACKLING NON-CONSENSUAL INTIMATE IMAGE ABUSE: GOVERNMENT RESPONSE, 2024-5, HC 911, at 4, 20 (UK).

100. Henry Patishman, *Global Legal Actions Against AI Deepfakes: Five Laws of 2025*, REGULA (Aug. 12, 2025), https://regulaforensics.com/blog/deepfake-regulations/ [https://web.archive.org/web/20260207041909/https://regulaforensics.com/blog/deepfake-regulations/].

101. EU AI Act, *supra* note 93, art. 50, (requiring transparency obligations for AI-generated content).

102. Commission Regulation 2016/679, of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

103. *See* Ramsha Jahangir, *Dutch Court Orders X, Grok to Stop AI-Generated Sexual Abuse Content*, TECH POLICY PRESS (Mar. 26, 2026) https://www.techpolicy.press/dutch-court-orders-x-grok-to-stop-aigenerated-sexual-abuse-content/ [https://web.archive.org/web/20260330131413/https://www.techpolicy.press/dutch-court-orders-x-grok-to-stop-aigenerated-sexual-abuse-content/].

104. HERBERT SMITH FREEHILLS KRAMER, *supra* note 98.

demands mutual legal assistance, treaties adapted to digital evidence, coordinated platform takedown procedures, and mechanisms for pursuing operators who deliberately forum-shop across jurisdictions. The *Jane Doe v. Clothoff* litigation exemplifies these challenges: An operator allegedly based in Belarus, incorporated in the British Virgin Islands, using fake addresses in Argentina, and inflicting harm on U.S. victims represents precisely the enforcement quagmire that demands international coordination.[105]

D. THE LIMITS OF TECHNICAL DETECTION SOLUTIONS

Although a technological solution would theoretically be the easiest remedy, technical approaches to identifying deepfakes have proven inadequate as standalone solutions. Digital watermarking—embedding invisible signatures into generated content—faces fundamental vulnerabilities. While platforms like Google's SynthID and Meta's StableSignature embed watermarks intended to survive common image manipulations, researchers have demonstrated that these protections can be systematically defeated. The "UnMarker" system developed by University of Waterloo researchers successfully removed watermarks from seven leading schemes, reducing detection accuracy below random-guess levels without access to training data or internal system details.[106]

Watermarking also suffers from inherent limitations beyond removal efforts. For watermarks to function, they must be universally adopted—yet open-source models and bad-faith actors presently face no obligation, legal or otherwise, to implement them. Watermarks embedded during generation cannot identify content created using models that predate watermarking requirements or that deliberately omit such features. Even when present, watermarks provide attribution rather than prevention; they identify synthetic content after dissemination—potentially mitigating the deceptive features of deepfake content—but do nothing to stop initial creation or distribution.[107]

Perceptual hashing and metadata embedding face similar constraints. While metadata can theoretically track content provenance, it remains easily stripped,[108]

105. Complaint, *supra* note 79, at 3, 5–6, 21.

106. Irfan Ahmad, *Researchers Break Industry Watermarks, Undermining Key Deepfake Detection Methods*, DIGIT. INFO. WORLD (July 24, 2025), https://www.digitalinformationworld.com/2025/07/researchers-break-industry-watermarks.html [https://web.archive.org/web/20260207045148/https://www.digitalinformationworld.com/2025/07/researchers-break-industry-watermarks.html].

107. *See* Nick Gaubitch, *Does Watermarking Protect Against Deepfake Attacks?*, PINDROP (Oct. 30, 2025), https://www.pindrop.com/article/does-watermarking-protect-against-deepfake-attacks/ [https://web.archive.org/web/20260207045515/https://www.pindrop.com/article/does-watermarking-protect-against-deepfake-attacks/].

108. *C2PA in ChatGPT Images*, OPENAI, https://help.openai.com/en/articles/8912793-c2pa-in-chatgpt-images [https://web.archive.org/web/20260228054757/https://help.openai.com/en/articles/8912793-c2pa-in-chatgpt-images] (last visited Feb. 21, 2026).

including by using the same AI tools that generated the content.[109] Perceptual hashing—creating digital "fingerprints" of images to identify duplicates—requires access to databases of known NCII images, operates reactively rather than preventively, and struggles with slight variations in manipulated content.[110] These technical measures may assist in enforcement by providing evidence of synthetic origins, but they cannot substitute for legal frameworks that impose liability on creators and platforms regardless of whether content bears detectable markers.

To date, the rapid evolution of generative AI has consistently outpaced technical countermeasures. Detection algorithms trained on current deepfake artifacts become obsolete as new generation techniques emerge. This arms race dynamic suggests that legal and economic interventions targeting creation and distribution infrastructure offer complementary—and potentially more durable—solutions than technical detection alone.

E. VOLUNTARY COOPERATION

Although important, neither government enforcement nor technical detection alone is likely to eliminate the threat of deepfake NCII content. Voluntary cooperation from key stakeholders also holds great promise as a part of a holistic approach to mitigating harm. Domain registrars, web hosts, and app stores each maintain their own terms of service, which almost always prohibit hosting CSAM content, frequently prohibit hosting NCII content, and often prohibit content and conduct that constitutes cyber-bulling, abuse, and harassment. While not traditional government enforcement mechanisms, private, quasi-contract-based remedies deserve exploration as well. Platforms have historically demonstrated acute sensitivity to CSAM liability concerns, making terms of service violations a potentially effective supplementary enforcement avenue for NCII content more broadly.

Economic pressure points offer additional leverage. Research reveals that just a subset of identified nudification websites collectively generated at least $36 million in annual revenue, supported by mainstream technology infrastructure.[111] Analysis of eighty-five such websites found that sixty-two utilize Amazon Web Services or Cloudflare for hosting and content delivery, while fifty-four employ Google's sign-on system.[112] These sites monetize through tiered subscription models, typically offering limited free trials before requiring payment—commonly accepting cryptocurrency to

109. Jacob Hoffman-Andrews, *AI Watermarking Won't Curb Disinformation*, ELEC. FRONTIER FOUND. (Jan. 5, 2024), https://www.eff.org/deeplinks/2024/01/ai-watermarking-wont-curb-disinformation [https://web.archive.org/web/20260207052221/https://www.eff.org/deeplinks/2024/01/ai-watermarking-wont-curb-disinformation].

110. *See* Hannes Mareen et al., *Fast Fallback Watermark Detection Using Perceptual Hashes*, 10 ELECS. 1155 (2021).

111. *See* Burgess, *supra* note 67.

112. *See id.*

evade financial oversight, but also processing payments through PayPal, Apple Pay, Cash App, Venmo, and traditional credit cards.[113]

This financial ecosystem supporting deepfake NCII content creation thus presents additional opportunities for intervention. Payment processors and financial services companies can refuse to process transactions for known nudification services, as they have done for other categories of harmful content.[114] Cloud infrastructure providers can terminate services to platforms violating their acceptable use policies.[115] While such private enforcement raises concerns about unchecked corporate power and potential overreach,[116] targeted actions against platforms whose core business model centers on facilitating illegal content production present compelling cases for intervention. Cryptocurrency's role in enabling these services—offering anonymity and circumventing traditional financial controls—highlights the need for regulatory attention to payment processing infrastructure supporting illegal content generation.

---

113. Complaint, *supra* note 79, at ¶¶s 41, 50, 59, 74, 82, 141 (detailing payment methods accepted by various nudification websites including PayPal, Apple Pay, Cash App, Venmo, cryptocurrency, and traditional credit cards).

114. *See e.g.*, Michelle Price, *Mastercard, Visa Suspend Ties with Ad Arm of Pornhub Owner MindGeek*, REUTERS (Aug. 4, 2022), https://www.reuters.com/business/finance/mastercard-visa-suspend-ties-with-ad-arm-pornhub-owner-mindgeek-2022-08-04/ [https://web.archive.org/web/20260206184939/https://www.reuters.com/business/finance/mastercard-visa-suspend-ties-with-ad-arm-pornhub-owner-mindgeek-2022-08-04/].

115. Major cloud infrastructure providers maintain acceptable use policies that authorize service termination for policy violations. *See AWS Acceptable Use Policy*, AMAZON WEB SERVS., https://aws.amazon.com/aup/ [https://web.archive.org/web/20260206191410/https://aws.amazon.com/aup/] (last visited Feb. 8, 2026) ("We may . . . remove or disable access to any content…that violates this Policy."); *Google Cloud Platform Terms of Service* § 4.1, GOOGLE CLOUD, https://cloud.google.com/terms [https://web.archive.org/web/20260206192644/https://cloud.google.com/terms] (last visited Feb. 8, 2026) ("Google may Suspend all or part of Customer's use of the Services until the violation is corrected."); *For Online Services*, MICROSOFT, https://www.microsoft.com/licensing/terms/product/ForOnlineServices/all [https://perma.cc/XRK9-T5G6] (last visited Feb. 8, 2026) ("[V]iolations of the Acceptable Use Policy . . . may result in suspension of the Online Service."); *Cloudflare Self-Serve Subscription Agreement*, CLOUDFLARE, https://www.cloudflare.com/terms/ [https://web.archive.org/web/20260206193559/https://www.cloudflare.com/terms/] (last visited Feb. 8, 2026) ("We may at our sole discretion terminate your user account or Suspend or terminate your use or access to the Service at any time, with or without notice for any reason or no reason at all."); *Acceptable Use Policy*, IDENTITY DIGIT., https://www.identity.digital/policies/acceptable-use-policy [http://web.archive.org/web/20260206193749/https://www.identity.digital/policies/acceptable-use-policy] (last visited Feb. 8, 2026) ("Identity Digital reserves the right . . . to deny, suspend, cancel, redirect, or transfer any registration or transaction . . . as it determines necessary . . . to comply with the terms of the applicable registration agreement and Identity Digital' policies [or where] . . . domain name use is abusive or violates this AUP.").

116. *See, e.g.*, *FIRE Statement on Free Speech and Online Payment Processors*, FOUND. FOR INDIVIDUAL RTS. & EXPRESSION (FIRE), https://www.thefire.org/research-learn/fire-statement-free-speech-and-online-payment-processors [http://web.archive.org/web/20260206194300/https://www.thefire.org/research-learn/fire-statement-free-speech-and-online-payment-processors] (last visited Feb. 8, 2026) ("When these companies appoint themselves the arbiters of what speech and views are acceptable, shutting people and organizations out of the online financial ecosystem for wrongthink, they seriously undermine our culture of free expression.").

VI. CONCLUSION

The recent explosion of AI-generated deepfake non-consensual intimate images demonstrates that even when a widespread societal consensus had concluded that a problem demands intervention, implementing enforcement to address it remain profoundly challenging. Practical obstacles abound: Technology makes the creation and distribution of NCII all too easy but the source attribution incredibly challenging. Cheap, readily replicable open-source models create a "whack-a-mole" problem at the platform level. These challenges, coupled with the jurisdictional limitations in enforcing an inherently cross-border phenomenon, make it likely that no single solution will prove sufficient.

Fortunately, a patchwork quilt of solutions stands at the ready, if political and technical know-how are properly employed. Effective responses require coordinated efforts across multiple domains: enhanced platform accountability, reconsideration of open-source model distribution practices, strengthened cross-border cooperation, and creative use of both public enforcement mechanisms and private contractual remedies. As generative AI capabilities continue advancing, the window for establishing protective frameworks before video-based NCID becomes as prevalent as current image-based content continues to narrow. Policymakers, technology companies, and civil society must act decisively while acknowledging that addressing this crisis will require sustained, adaptive efforts rather than singular legislative fixes.