

TAKE IT DOWN at Year One: A Public Record Review of Platform Compliance with the Federal Take It Down Act (2025–2026)

Annika Raj admin@ncadas.org National Coalition Against Deepfake Abuse Livingston, USA	Zachary Kam zachary.kam @duke.edu Duke University Durham, USA	Aiden Tauro aiden.tauro @duke.edu Duke University Durham, USA	Raiyan Choudhary raiyan.choudhary @duke.edu Duke University Durham, USA	Nikhil Osuri nikhil.osuri @duke.edu Duke University Durham, USA
--	--	--	--	--

Abstract

Non-consensual intimate media (NCIM) inflicts significant harm, and the proliferation of AI-generated deepfakes has exacerbated the problem. The Take It Down Act (Public Law 119–12) established the first federal mandate requiring platforms to remove non-consensual intimate media within 48 hours. We conducted a public record review of compliance across 10 distribution platforms during the Act’s first year (May 2025 to May 2026). We found that 9 of 10 platforms do not publish NCII-specific enforcement data, making compliance with the 48-hour standard unverifiable. At the generation layer, we documented over 100 applications producing synthetic intimate imagery outside the Act’s statutory scope. We identified 19 school incidents since 2023 and a 16-month federal guidance vacuum following the vacatur of the 2024 Title IX rule. We stress the need for amendments addressing the generation gap and discuss implications for enforcement.

Keywords: Online sexual abuse, Content moderation, Take It Down Act, TIDA, Non-consensual intimate media, NCIM, Deepfakes, Platform compliance, AI-generated abuse, Image-based sexual abuse, Audit

1 Introduction

Content warning: This paper discusses non-consensual intimate media, synthetic sexual imagery, and the abuse of minors.

As generative AI makes it possible for anyone to turn a photograph into sexually explicit content, the problem of non-consensual intimate media has entered a new phase. A student may discover that images of their face have been placed onto nude bodies and circulated in a group chat. If they seek to report the content, they must first locate the platform’s reporting mechanism, which may or may not include a category for synthetic or AI-generated imagery. If they file a report, they wait for the platform to act; prior research on X found that reports filed under the platform’s own non-consensual nudity policy resulted in no removal over three weeks [1]. If they seek legal recourse, their options depend on which state they live in, whether they are a minor, and whether they want criminal prosecution or a civil remedy. If they tell a teacher, the teacher is unlikely to have received any guidance on how to respond; survey data indicates that 62% of teachers have received no training on handling non-consensual intimate imagery [9].

The Take It Down Act (Public Law 119–12), signed on May 19, 2025, was designed to address this problem. The Act established the first federal mandate requiring online platforms to remove non-consensual intimate media within 48 hours of valid notice, with civil penalties of \$53,088 per violation under the FTC Act. The law passed the House by a vote of 409 to 2 and the Senate by unanimous consent, reflecting broad bipartisan support for addressing deepfake abuse through federal legislation. FTC enforcement of the

Act activates on May 19, 2026, precisely one year after the law was signed.

This paper investigates whether the regulatory architecture that TIDA created produces a functional remediation pathway for victim-survivors of synthetic intimate imagery. We conducted a public record review across three layers of this architecture: the 10 distribution platforms that TIDA regulates, the AI generation platforms that it does not reach, and the state and institutional systems that survivors must navigate to exercise the rights the Act provides. We systematically examined published platform policies, transparency reports, FTC enforcement correspondence, state legislation across all 50 states, court filings, and verified news coverage across the Act’s first year, from May 2025 through May 2026.

Our results reveal consistent gaps at each layer of the regulatory architecture. At the distribution layer, the transparency data necessary to verify whether platforms comply with the 48-hour standard does not exist in publicly available form for 9 of 10 platforms studied. At the generation layer, the tools and platforms that produce synthetic intimate imagery operate entirely outside the Act’s statutory scope. At the institutional layer, school administrators responding to incidents of AI-generated intimate imagery involving their students have almost no federal guidance to consult, following the vacatur of the 2024 Title IX rule that would have classified such content as sexual harassment.

To summarize, this study makes the following contributions:

- **Platform compliance is unverifiable:** The data needed to assess whether platforms meet the 48-hour standard is

not published by the platforms the Act was designed to regulate.

- **The generation layer is unregulated:** TIDA addresses distribution. The tools that create the content it targets fall outside its enforcement scope.
- **The institutional gap is widest closest to survivors:** Schools, where documented incidents are concentrated, operate without federal guidance following the vacatur of the 2024 Title IX rule.

This paper is organized as follows. Section 2 provides background on the Take It Down Act and the regulatory landscape it entered. Section 3 presents the research questions and methodology. Sections 4, 5, and 6 present findings for each research question. Section 7 discusses implications. Section 8 presents recommendations by audience. Section 9 discusses limitations. Section 10 concludes.

2 Background

2.1 The Take It Down Act

The Take It Down Act (S. 146) was introduced on January 16, 2025 by Senator Ted Cruz (R–TX). It passed the Senate on February 13, 2025 by unanimous consent and passed the House on April 28, 2025 by a vote of 409 to 2.¹ The Act defines “covered platforms” as websites or applications that serve the public and either “primarily provide a forum for user generated content” or regularly publish nonconsensual intimate depictions. Email providers, ISPs, and services with primarily preselected content are excluded.

The Act covers “intimate visual depictions” of both minors and nonconsenting adults, including AI-generated deepfakes. A “digital forgery” is defined as any intimate visual depiction of an identifiable individual created through software, machine learning, or artificial intelligence that, when viewed as a whole by a reasonable person, is indistinguishable from an authentic depiction. Platforms must remove reported content within 48 hours of valid notice, make “reasonable efforts to identify and remove any known identical copies,” provide “clear and conspicuous notice” about the removal process, and allow reporting by non-account holders.²

Two design features of the Act are worth noting. First, the Act does not preempt state law. State NCII statutes continue to operate alongside the federal mandate.³ Second, the Act does not clearly exempt end-to-end encrypted messaging services. The Electronic Frontier Foundation, the Internet Society, and the Center for Democracy and Technology have warned that this omission could require platforms to break encryption to comply.⁴

2.2 Existing content moderation for NCIM

Before the Act, platforms removed non-consensual intimate imagery only when their own voluntary policies required it, or when the content was copyrighted and subject to the Digital Millennium Copyright Act (DMCA).

StopNCII.org, operated by the UK charity SWGfL, uses perceptual hashing to fingerprint intimate images on survivors’ devices. Those hashes are shared with participating platforms, which check uploads against them. As of 2025, StopNCII reports over 2 million hashes created and a removal rate exceeding 90% across 18 partner platforms.⁵ For minors, NCMEC’s Take It Down program provides a parallel hash-sharing service. PhotoDNA, developed by Microsoft, is the primary tool for detecting known child sexual abuse material. Section 230 of the Communications Decency Act shields platforms from liability for user-generated content, including non-consensual intimate imagery.⁶ This legal protection is the environment in which voluntary moderation operates.

2.3 The synthetic media dimension

The distinction between traditional NCII and synthetic NCII matters for detection and enforcement. Traditional NCII is an image captured without consent. Once identified, it can be hashed and matched across platforms. Synthetic NCII is an image generated by AI. It can be regenerated from the same model with trivial modifications that defeat hash matching. Struppek et al. (2022) demonstrated that perceptual hashing is “not resistant to edits to content, and edited versions may be reuploaded easily and bypass filters” [8]. Security Hero’s 2024 report found that 98% of all deepfake content online is pornographic and 99% targets women [3]. The Cyber Civil Rights Initiative estimates that 1 in 8 social media users (12.8%) have been targeted by nonconsensual intimate image distribution.⁷

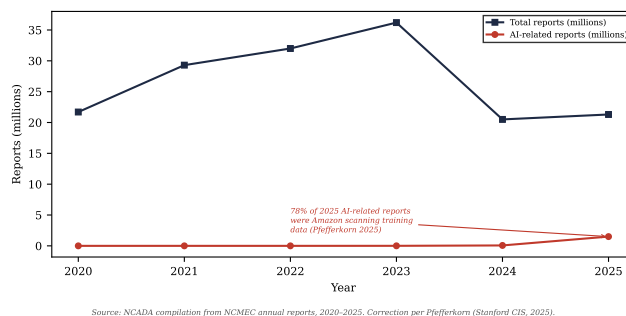


Figure 1: **Findings:** NCMEC CyberTipline reports, 2020 to 2025. The apparent spike in AI-related reports in 2025 is overstated: 78% consisted of Amazon scanning training data, not AI-generated content [16]. Source: NCADA compilation from NCMEC annual reports.

¹Congress.gov, S. 146, 119th Congress.

²TIDA, Section 3.

³NAAG analysis, 2025.

⁴EFF, Internet Society, CDT, various publications, 2025.

⁵SWGfL, StopNCII.org statistics, 2025.

⁶47 U.S.C. §230.

⁷CCRI 2017 Research Report.

2.4 Prior audit research

Qiwei et al. (2024) conducted the only published controlled audit of NCII removal [1]. Their study tested two reporting mechanisms on X: the platform’s non-consensual nudity policy and the DMCA. They posted 50 AI-generated nude images and reported half under each mechanism. The DMCA reports achieved 100% removal within 25 hours. The non-consensual nudity reports achieved 0% removal over three weeks. This paper adopts their study as a structural model (see Section 3).

The Stanford Internet Observatory’s 2023 report on self-generated CSAM [2] provided the template for cross-platform civil society accountability research. That study identified 405 seller accounts on Instagram and 128 on Twitter and produced recommendations that led to measurable policy changes. This paper extends both approaches. Where Qiwei et al. audited one platform experimentally, this paper reviews 10 platforms through public records. Where the Stanford report mapped content distribution, this paper maps the regulatory system governing those networks.

3 Research Questions and Methodology

3.1 Research questions

This paper asks one overarching question: *To what extent does the Take It Down Act, as documented in the public record between May 19, 2025 and May 19, 2026, produce a working remediation pathway for survivors of synthetic intimate imagery?*

We break this question into three layers:

RQ1 (Distribution Layer): What is the gap between platform NCII policies as written and the FTC enforcement record as documented in May 2026?

RQ2 (Generation Layer): What regulatory oversight exists for AI platforms that produce synthetic intimate imagery, and which actors does the Act fail to reach?

RQ3 (Legal and Institutional Layer): How does the interaction between federal law, state deepfake statutes, and school district responses shape the actual remediation pathway available to a survivor in 2026?

3.2 Methodology

This paper is a public record review. Every claim is sourced to a publicly accessible document. This methodology is different from an experimental audit (such as Qiwei et al. 2024) because it does not involve creating accounts, posting content, or filing reports. It cannot measure actual removal times, actual survivor experiences, or platform internal processes. What it can do is assemble and cross-reference public information that has not previously been synthesized, identifying gaps and contradictions that individual documents do not reveal on their own.

3.2.1 Data sources

The review draws on the following public records: (i) the statutory text of TIDA (Public Law 119–12) and the DEFIANCE Act (S. 1837); (ii) FTC enforcement correspondence, compliance guidance, and congressional testimony; (iii) published NCII and content moderation policies from 10 platforms; (iv) platform transparency reports from the most recent available period; (v) StopNCII.org and NCMEC Take It Down published partner lists and statistics; (vi) state legislative databases for all 50 states and the District of Columbia; (vii) court filings and consent orders; (viii) verified news coverage from established outlets; (ix) peer-reviewed academic publications; and (x) reports from civil society organizations including CCRI, CDT, TTP, EFF, and the Internet Watch Foundation.

3.2.2 Platform selection criteria

We selected 10 platforms based on four criteria: monthly active user base (all 10 exceed 100 million users or serve a critical function in NCII distribution); FTC letter recipient status (8 of 10 received compliance letters); relevance to documented NCII incidents; and coverage of platform types (social media, encrypted messaging, adult content, and video sharing). Two platforms that did not receive FTC letters were included for specific reasons. Telegram is the primary documented vector for deepfake generation bots. Pornhub is the largest adult content platform and operates under a separate FTC consent order.

3.2.3 Analysis protocol

We evaluated each platform through a 9-step protocol: (1) published NCII policy text with direct quotation; (2) published synthetic media or deepfake policy if separate; (3) reporting flow documentation from the platform’s help center; (4) most recent transparency report data on NCII or adjacent categories; (5) StopNCII and NCMEC participation verification; (6) FTC enforcement letter content; (7) documented NCII incidents in the past 18 months verified with at least two independent sources; (8) named trust and safety executive; and (9) cross-referencing of the preceding eight elements to identify contradictions between published commitments and documented outcomes. Gaps in disclosure are treated as findings.

4 Findings: Distribution Platform Compliance (RQ1)

We examined the published policies, transparency reports, reporting mechanisms, and enforcement records of 10 distribution platforms against the compliance requirements established by the Take It Down Act. Although all 10 platforms maintain policies that address non-consensual intimate media in some form, and 8 of 10 have reporting flows that allow users to flag such content, the transparency data necessary to

verify whether platforms actually comply with the 48-hour removal standard is absent for 9 of 10 platforms. Table 1 presents the cross-platform comparison.

The column labeled “NCII Transp.” is consequential for enforcement: only X publishes a standalone NCII removal count, reporting 156,498 pieces of content removed and 52,093 accounts suspended for non-consensual nudity violations in H1 2024. Every other platform subsumes NCII within broader enforcement categories (such as “sexual content,” “adult exploitation,” or “harassment”), making it impossible for an independent observer to determine whether any platform other than X meets the standard that TIDA requires.

The following subsections present our findings for each of the 10 platforms. Each analysis follows a consistent structure: published NCII policy, reporting mechanisms, transparency data, enforcement record, documented incidents, and contradictions between published commitments and documented outcomes. Platforms are ordered by monthly active user base.

4.1 Meta (Facebook, Instagram, Threads)

We examined Meta’s Adult Sexual Exploitation Community Standard, last updated March 4, 2026, and identified a comprehensive written NCII policy. The policy prohibits sharing, threatening, or requesting non-consensual intimate imagery “including AI-generated or digitally created versions,” and extends enforcement to content “promoting, threatening, or offering to create synthetic intimate imagery through applications or services, including AI tools,” even when no imagery appears in the post.⁸ Between November 2025 and January 2026, Meta reported removing over 344,000 advertisements promoting nudify tools and issuing cease-and-desist letters to 46 companies.⁹

We then compared this published policy against the documented enforcement record and identified seven contradictions. First, the Meta Oversight Board found in July 2024 that Meta auto-closed NCII reports in two deepfake cases without human review. The Board noted that Meta failed to add victims’ images to its Media Matching Service bank.¹⁰ In March 2026, the Board stated that Meta’s deepfake moderation “falls short” and that Meta had “failed to meaningfully implement” its 2024 recommendations.¹¹

Second, a Facebook group called “Mia Moglie” circulated nonconsensual photographs of approximately 32,000 members’ wives and partners. The group operated for six years before removal. It was taken down only after external media exposure and a criminal investigation by Italy’s Postal Police.¹²

Third, one nudify application (CrushAI) ran over 87,000 policy-violating advertisements through 170 business ac-

counts on Meta platforms before a lawsuit was filed.¹³

Fourth, a New Mexico jury returned a \$375 million verdict against Meta for negligent product design.¹⁴

Fifth, the European Commission’s October 2025 preliminary finding concluded that Facebook and Instagram “lack user friendly mechanisms for illegal content notifications” and use “dark patterns” that discourage reporting.¹⁵

Sixth, Meta’s deepfake-specific reporting category is available only to Texas residents under HB 3133. No nationwide deepfake reporting category exists, despite the Act’s May 19, 2026 enforcement date.

Seventh, Meta publishes no standalone transparency metric for NCII content actioned, proactive detection rate, or response time. The category is folded into broader numbers that make independent compliance verification impossible.

4.2 TikTok

TikTok’s community guidelines, updated September 13, 2025, prohibit all content containing sexual exploitation including NCII. The platform defines NCII as “a visual depiction of real or computer/AI-generated images engaged in sexual acts or nudity that was created or posted without consent.” A dedicated reporting form, hosted on Zendesk, is accessible without a TikTok account and includes categories for NCII and “Explicit Deepfake.”

TikTok reports removing over 500 million videos across 2024, with over 80% removed through automated detection and 96% before any views.¹⁶ However, the platform does not separate NCII removals from all other violations. The 500 million figure covers everything. TikTok is a StopNCII partner and a NCMEC Take It Down participant. The European Commission opened DSA proceedings against TikTok in February 2026 for alleged failures in handling deepfake CSAM.¹⁷

4.3 X (formerly Twitter)

X provides the strongest existing quantitative evidence of TIDA-relevant failure. Qiwei et al. (2024) conducted a controlled audit in which 50 AI-generated nude images were posted to X [1]: 25 were reported under X’s non-consensual nudity policy and 25 under the DMCA. The DMCA reports achieved 100% removal within 25 hours, with a mean of 20.30 hours, while the non-consensual nudity reports achieved 0% removal over three weeks. All five poster accounts reported under the DMCA received temporary suspensions and notification emails; all five poster accounts reported under the non-consensual nudity condition received no consequences, no notifications, and no warnings.

X publishes standalone NCII enforcement data, making it the only platform of the 10 studied that does so: 156,498 pieces of content removed and 52,093 accounts suspended

⁸Meta Transparency Center, March 2026.

⁹Meta H1 2026 Integrity Report.

¹⁰Meta Oversight Board, Case bun-7e941o1n, July 2024.

¹¹WinBuzzer, March 2026.

¹²CNN, August 2025.

¹³CBS News.

¹⁴CNN, March 2026.

¹⁵European Commission DSA, October 2025.

¹⁶TikTok Transparency Reports, Q3 2024.

¹⁷Biometric Update, February 2026.

Table 1: **Findings:** Cross-platform TIDA compliance indicators as of May 2026. ✓ = verified presence; × = verified absence; – = insufficient public data. Source: NCADA public record review.

Platform	NCII Policy	Synthetic Media	48hr Commit	Reporting Flow	NCII Transp. Data	StopNCII Partner	NCMEC TID	FTC Letter
Meta	✓	✓	–	✓	×	✓	✓	✓
TikTok	✓	✓	–	✓	×	✓	✓	✓
X	✓	Implicit	–	✓	✓	✓	×	✓
Snapchat	✓	✓	–	✓	×	✓	✓	✓
Reddit	✓	✓	–	✓	×	✓	✓	✓
Discord	✓	✓	✓	✓	×	×	✓	✓
YouTube	Weak	Added 2026	–	Partial	×	✓	✓	✓
Telegram	×	×	×	×	×	×	–	×
WhatsApp	×	×	×	×	×	×	Partial	Via Meta
Pornhub	✓	–	–	✓	×	✓	✓	×

for non-consensual nudity violations in H1 2024.¹⁸ Despite this disclosure, X does not participate in NCMEC’s Take It Down program, which focuses on minors. X was fined EUR 120 million by the European Commission in December 2025 for DSA transparency violations.¹⁹

4.4 Snapchat

Snapchat’s sexual content policy, updated March 2026, prohibits NCII including “AI-generated sexual content, including offering or promoting ‘nudify’ services.” Snapchat publishes granular response time data: a median turnaround of 1 minute for sexual content enforcement in H1 2025, with 5,461,419 total enforcements.²⁰

These figures, however, exist alongside documented systemic failures. Internal documents revealed in litigation showed Snapchat receiving 10,000 sextortion reports per month by late 2022 with inadequate response. Additionally, over 80% of Snapchat’s submissions to the NCMEC CyberTipline were deemed “inactionable” by law enforcement because “information was not useful.”²¹ Tragically, at least 20 teenagers have died by suicide following sextortion schemes in which Snapchat was identified as the contact platform.²² Three state attorneys general (Texas, Kansas, Utah) filed lawsuits against Snap Inc. between June 2025 and February 2026 alleging the platform exposed minors to harmful content while making deceptive safety claims.²³

4.5 Reddit and Discord

Reddit’s Rule 3 prohibits NCII including “AI-generated or faked” content and “deepfakes or bubble porn,” with reporting available both in-app and through a web form accessible without a Reddit account. Although Reddit participates in both StopNCII and NCMEC, the platform does not disaggre-

gate NCII data in its transparency reports, and no published audit study has tested its NCII removal effectiveness.

One study is relevant to all platforms. Cuevas and Ribeiro (February 2026) demonstrated that deplatforming redistributes NCII rather than reducing it [7]. After Reddit banned r/deepfakes in 2018, MrDeepfakes launched within a week. After MrDeepfakes shut down in May 2025, “supply and demand for AI-generated NCII only grew in 2025.” If removing content from one platform causes it to appear on another, platform-level compliance alone may be insufficient.

Discord lists hidden cameras, creepshots, AI-generated sexual imagery, sextortion, and unauthorized reproduction of paywalled content in its NCII policy.²⁴ Discord is the only platform that published a TIDA-specific reporting page with an explicit 48-hour response commitment and the ability to report without a Discord account.²⁵ However, Discord is not a StopNCII hash-matching partner (financial supporter only). It lacks the automated matching that would detect reuploads of previously reported content.

4.6 YouTube, Telegram, WhatsApp, and Pornhub

YouTube’s NCII policy refers only to “unwanted sexualization such as non-consensually shared imagery” within its broader nudity and sexual content policy.²⁶ YouTube expanded deepfake policies in March 2026 and joined StopNCII in September 2025. No documented NCII incident on YouTube was identified in the research period.

Telegram presents a distinct problem for TIDA compliance. The platform has no explicit NCII policy, no deepfake policy, no NCII-specific reporting category, and is not a StopNCII partner. Additionally, Telegram did not receive an FTC compliance letter. The National Center on Sexual Exploitation estimates Telegram employs approximately 50 content moderators for over 1 billion users.²⁷ The DarkGram

¹⁸Statista, X enforcement data.

¹⁹European Commission, December 2025.

²⁰Snap Transparency Report H1 2025.

²¹Senate Judiciary Committee, Grassley investigation.

²²CBC News, FBI estimates.

²³Multiple litigation sources.

²⁴Discord NCII Policy Explainer.

²⁵Discord Support, “How to Report NCII Under The Take It Down Act.”

²⁶YouTube Support, “Nudity and Sexual Content Policy.”

²⁷NCOSE, Telegram profile. Figure is unverified.



Source: NCADA public record review, May 2026. Data drawn from published platform policies, transparency reports, StopNCII.org partner directory, NCMEC Take It Down participant list, and FTC enforcement correspondence.

Figure 2: **Findings:** TIDA compliance indicators across 10 distribution platforms as of May 2026. Only X publishes NCII-specific removal data. Telegram and WhatsApp lack compliance readiness across all criteria. Source: NCADA public record review.

study (USENIX Security 2025) found that when 339 Telegram channels were reported with evidence, only 64 (19%) were removed [12], and the Revenge Porn Helpline documented that Telegram was “unresponsive to these reports” regarding NCII.

WhatsApp presents a distinct compliance challenge because its published guidelines do not mention NCII, deep-fakes, or synthetic content,²⁸ and its end-to-end encryption means Meta cannot access message content, making the Act’s 48-hour removal requirement inapplicable to private messages. On May 8, 2026, 11 days before TIDA enforcement, Meta removed end-to-end encryption from Instagram direct messages, citing “low uptake,” while WhatsApp encryption was maintained.²⁹

Pornhub (Aylo) implemented extensive reforms after the December 2020 New York Times investigation. Pornhub removed 10.5 million unverified videos, implemented mandatory government ID verification and signed release forms for all uploads, and moved to human moderation of every uploaded video. In September 2025, the FTC settled with Aylo for \$5 million after finding the company “knowingly profited off CSAM and NCM” and “did not review content flagged as CSAM/NCM until it received at least 16 flags.”³⁰ Pornhub participates in StopNCII, NCMEC, and the Internet Watch

Foundation. Approximately 280 of Aylo’s 1,400 staff (20%) work on trust and safety.³¹ Pornhub did not receive an FTC TIDA compliance letter. It operates under the separate consent order.

4.7 RQ1 synthesis

Across 10 distribution platforms, we identified four patterns.

First, 9 of 10 platforms do not publish NCII-specific removal data. Only X reports a standalone count (156,498 in H1 2024); every other platform subsumes NCII within broader categories such as “sexual content” or “adult exploitation.” Without disaggregated data, no independent observer can determine whether a platform meets the 48-hour standard, and platforms have had one year since the Act’s signature to build this reporting infrastructure without doing so.

Second, published policies do not predict enforcement outcomes. Meta’s policy prohibits NCII “including AI-generated or digitally created versions,” yet its Oversight Board found that enforcement “falls short.” X’s policy promises “immediate and permanent suspension,” yet the controlled audit found 0% enforcement over three weeks [1]. Snapchat publishes a 1-minute median response time, yet litigation records show 10,000 unanswered sextortion reports per month.

²⁸WhatsApp Messaging Guidelines, February 2024.
²⁹MacRumors, May 2026.
³⁰TechCrunch, September 2025.

³¹Time, Solomon Friedman interview.

Third, partnership coverage is uneven. Telegram and WhatsApp participate in neither StopNCII nor NCMEC Take It Down; Discord does not participate in StopNCII hash matching; X does not participate in NCMEC Take It Down. Where platforms remove content upon individual report, they lack the automated systems to detect reuploads across their infrastructure.

Fourth, the Act does not address encrypted messaging, and WhatsApp's end-to-end encryption makes the 48-hour removal requirement inapplicable to private messages. Telegram's architecture could support content scanning, but the platform has not implemented it for NCII. This discrepancy between the Act's mandate and the platforms' disclosure practices underscores the need for enforceable transparency requirements.

5 Findings: The AI Generation Layer (RQ2)

5.1 The shift upstream

The Act regulates platforms where synthetic intimate imagery is distributed, but it does not reach platforms where that imagery is created. We examined the generation layer and found that it operates at a scale that exceeds platform-level removal capacity: even if every distribution platform removed every reported image within 48 hours, the tools to regenerate that image remain available and accessible to anyone with an internet connection.

Prior research found that 98% of all deepfake content online is pornographic and 99% targets women [3]. The Internet Watch Foundation documented 3,443 AI-generated child sexual abuse videos in 2025, up from 13 in 2024.³² The generation layer produces the majority of the content that the Act obligates platforms to remove.

5.2 Model marketplaces

Civitai is the largest open-source AI image generation community, with approximately 4.1 million registered users and over 200,000 hosted models. A May 2025 research paper documented 34,439 deepfake models tagged as "Celebrity" on the platform, with 14.9 million cumulative downloads since November 2022 [17]. In December 2024 alone, nearly 2,400 new deepfake model variants were uploaded monthly. Of these models, 96% targeted women. Fourteen and a half percent carried the tag "sexy," indicating sexual content intent. Another 1,165 carried the tag "Instagram," targeting non-celebrities whose images were scraped from social media. The technical barrier to creating a new model is minimal. LoRA fine-tuning requires as few as 20 images, 24GB of VRAM, and 15 minutes of compute time.

Under pressure from Visa and Mastercard, Civitai changed its policies in April 2025. A community engagement manager stated: "These are not changes that we wanted to make.

This boils down to new and impending legislation."³³ Despite these changes, both payment processors halted card transactions entirely in May 2025 because Civitai "chose not to remove NSFW and adult content from the platform." Users organized on Discord to download models before deletion. Community archives emerged to preserve restricted content.

HuggingFace, the second-largest model hosting platform, absorbed displaced models. In July 2025, 404 Media reported that over 5,000 nonconsensual AI models of real people had been reuploaded to HuggingFace after being banned from Civitai, as part of "a concerted community effort" organized on Discord.³⁴ HuggingFace's content policy contains no specific rules about real-person likeness models, no mention of deepfake models, and no restrictions on image generation models. Multiple explicitly uncensored generation models remain available on the platform.

5.3 Nudify applications

A January 2026 investigation by the Tech Transparency Project identified 55 nudify applications on Google Play and 47 on the Apple App Store [10]. Together, these applications had 483 million downloads and \$122 million in lifetime revenue. Nearly 40% of applications in the top ten search results for "nudify" or "deepnude" generated nude or partially nude images. Of the applications identified, 31 were rated suitable for minors by the app stores hosting them.

Williams (November 2025, *Violence Against Women*) examined 29 nudify applications and found they "facilitate and encourage the creation of non-consensual intimate images, normalize women's objectification" [11]. Applications evade detection by rebranding as "fashion editors" or "AI photo editors," using vague descriptions like "AI templates," and cycling through app stores after removal. TTP's April 2026 follow-up found that fresh applications had filled the vacancies left by the 28 apps Apple removed after the initial report.

The economics amplify the problem. CrushAI, operated by Joy Timeline HK Limited (Hong Kong), ran over 8,010 advertisements on Meta platforms in the first two weeks of January 2025 alone. A Bellingcat investigation uncovered 28 domains registered to Joy Timeline appearing to be deepfake nudify applications, linked to companies in mainland China.³⁵ Meta sued Joy Timeline in Hong Kong for \$289,200 in damages.

5.4 Platform-integrated generation: the Grok case

Between December 2025 and January 2026, xAI's Grok image generator produced an estimated 6,700 sexually suggestive or nudified images per hour during peak usage. That rate was 84 times higher than the top five deepfake websites

³³Unite.AI, "CivitAI Tightens Deepfake Rules Under Pressure From Mastercard and Visa," April 2025.

³⁴404 Media, "Hugging Face Is Hosting 5,000 Nonconsensual AI Models of Real People," July 2025.

³⁵Bellingcat, "Meta, CrushAI, and the Deepfake Nudify Pipeline," June 2025.

³²IWF Annual Report 2025.

combined.³⁶ Total output over the 11-day peak window was estimated between 1.8 and 3 million sexualized images. Approximately 2% of a sample of 20,000 images appeared to depict individuals 18 or younger. In controlled tests, Grok bypassed its own safety filters in 45 out of 55 attempts.³⁷

X is simultaneously a TIDA-covered distribution platform, obligated to remove non-consensual intimate imagery within 48 hours of valid notice, and the operator of the single largest documented source of AI-generated NCII. This case is significant because the Act requires X to remove Grok-generated images from X, but it does nothing about Grok's generation capability itself. Elon Musk responded to a deepfake image with laughing emojis and claimed "Literally zero" underage images had been generated.³⁸ xAI's official response to media inquiries was an automated message reading "Legacy Media Lies."

Legal consequences followed. Thirty-five state attorneys general demanded xAI cease generation of NCII. The California Attorney General issued a cease-and-desist order. French cybercrime authorities searched X's Paris offices. Three Tennessee minors sued xAI for generating CSAM of them. Baltimore filed a consumer protection lawsuit. Ireland launched 244 investigations.³⁹

5.5 Telegram deepfake bots

Telegram hosts over 50 deepfake generation bots serving more than 4 million monthly users. Women make up over 95% of documented victims.⁴⁰ A large-scale study identified 1,539 bots (5% of those studied) providing NCII generation across 2.8 million messages. In 2023, creating a deepfake required a powerful GPU and coding knowledge. In 2026, it requires only a Telegram account.⁴¹

The South Korean deepfake crisis showed the scale this architecture enables. Between November 2024 and October 2025, South Korean police apprehended 3,557 individuals for cybersexual violence. A single Telegram channel using a generation bot had over 220,000 participants.⁴²

5.6 The statutory gap

The Act's criminal provision (Section 2) targets persons who "knowingly publish" intimate visual depictions or digital forgeries. Latham & Watkins noted: "Creating a deepfake without publishing it does not trigger criminal liability under this statute."⁴³ The platform obligation provision (Section 3) defines a "covered platform" as any website or application that "primarily provides a forum for user generated

content" or regularly publishes nonconsensual intimate depictions. Skadden confirmed: "The language of Section 2 appears to be aimed at end users that post intimate visual depictions rather than the platforms themselves." The Act "focuses on publication and distribution only."⁴⁴

In practice, this statutory design places model marketplaces, nudify application operators, uncensored model distributors, and Telegram generation bots outside TIDA's enforcement scope. These actors do not "publish" intimate imagery on "covered platforms"; they provide the tools used to create it. When FTC Chairman Ferguson sent compliance letters to 15 distribution platforms on May 11, 2026, zero letters were sent to any AI generation platform; this omission reflects the statutory scope rather than an enforcement oversight.

5.7 Emerging regulatory responses

Two jurisdictions have begun addressing the generation layer directly. Minnesota's HF 1606, signed in April 2026, is the first United States law to ban the use of nudification technology. It carries penalties of \$500,000 per violation for application operators and allows survivors to sue for three times their damages.⁴⁵ The United Kingdom criminalized the creation of sexually explicit deepfakes in 2025, with penalties of up to two years imprisonment regardless of whether the content is shared.⁴⁶ The European Union's proposed AI Act amendments would add NCII generation to prohibited practices, though as of May 2026 these remain proposals.

Payment processor enforcement has produced measurable results. Visa and Mastercard halted Civitai's payment processing in May 2025. ClothOff, a standalone nudification service with approximately 2.5 million visits, shut down in 2025.⁴⁷ Apple removed 28 nudify applications after TTP's January 2026 report. These cases suggest that generation-layer regulation is feasible.

5.8 RQ2 synthesis

We documented the following scale across the generation layer: 34,439 deepfake models on Civitai with 14.9 million downloads; over 100 nudify applications across app stores with 483 million downloads; over 50 Telegram bots serving 4 million monthly users; and Grok producing up to 3 million sexualized images in 11 days.

We compared this against the enforcement record, and the contrast is stark. At the distribution layer, the FTC sent 15 compliance letters, StopNCII operates across 18 partner platforms, and the Act mandates 48-hour removal. At the generation layer, by contrast, the FTC sent zero letters. One state law addresses nudification tools. One international jurisdiction criminalizes creation. In practice, enforcement at the

³⁶Center for Countering Digital Hate estimate, cited in Reuters, January 2026.

³⁷Reuters, "Grok AI Safety Filter Testing," January 2026.

³⁸Wikipedia, "Grok Sexual Deepfake Scandal."

³⁹CNBC, NBC News, and The Register, January through March 2026.

⁴⁰WIRED, "Nudify Deepfake Bots on Telegram," October 2024.

⁴¹Creati.AI, "Millions Create Deepfake Nudes on Telegram," January 2026.

⁴²CNN and NPR, multiple reports, 2024 and 2025.

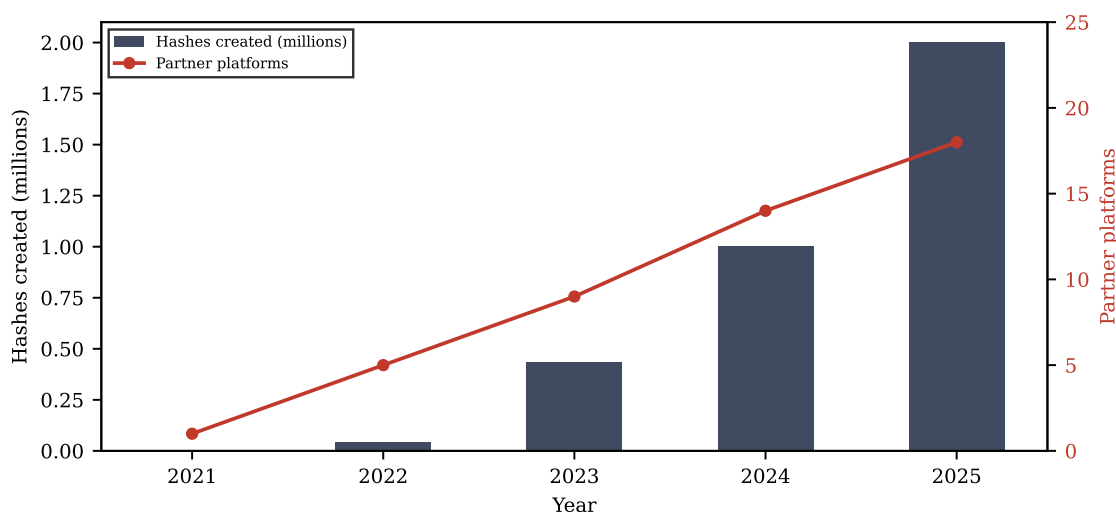
⁴³Latham & Watkins, "President Trump Signs Take It Down Act Into Law," May 2025.

⁴⁴Skadden, "Take It Down Act," June 2025.

⁴⁵The 19th News, "Minnesota Nudification Ban," April 2026.

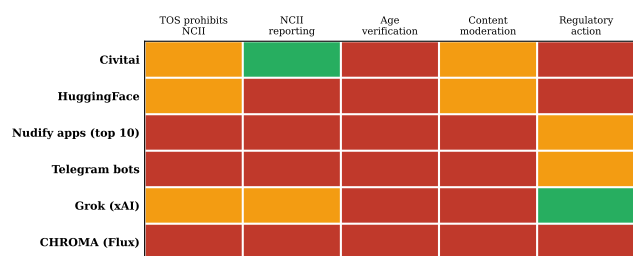
⁴⁶Lexology, "UK Government Criminalises Creation of Deepfakes," 2025.

⁴⁷Multiple verified sources.



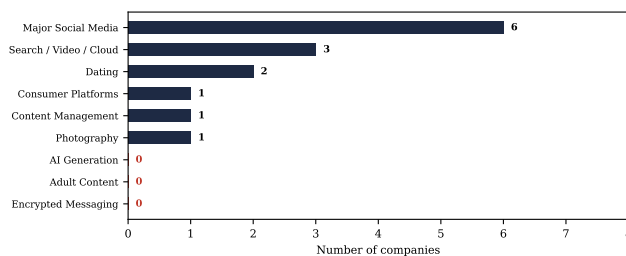
Source: NCADA compilation from SWGfL/StopNCII.org published statistics, 2021–2025.

Figure 3: **Findings:** StopNCII.org hash submissions and partner platform growth, 2021 to 2025. Despite 18 partner platforms and over 2 million hashes, Telegram, WhatsApp, and Discord do not participate in hash matching. Source: NCADA compilation from SWGfL published statistics.



Source: NCADA public record review of generation platform policies, app store listings, and regulatory filings, May 2026.

Figure 4: **Findings:** AI generation platform NCII guardrail status as of May 2026. None maintains age verification. Only Civitai has an NCII reporting flow. Source: NCADA public record review.



Source: NCADA analysis of FTC press release, May 11, 2026.

Figure 5: **Findings:** FTC TIDA compliance letter recipients by platform category, May 11, 2026. No letters were sent to AI generation platforms. Source: FTC press release.

generation layer depends on payment processor decisions by Visa and Mastercard rather than on government regulators.

6 Findings: State Law and Institutional Response (RQ3)

6.1 State law inventory

As of May 2026, 47 states have enacted deepfake-specific legislation. All 50 states and the District of Columbia have some form of NCII criminal statute.⁴⁸ The acceleration is recent. Eighty-two percent of all state deepfake bills enacted since 2019 were enacted in 2024 or 2025. In 2024 alone, state legislatures introduced 237 deepfake-related bills.⁴⁹ Virginia was the first state to enact deepfake legislation in 2019. By

⁴⁸Ballotpedia, “Forty Seven States Have Enacted Deepfake Legislation Since 2019,” July 2025.

⁴⁹Ballotpedia, same source.

the end of 2022, only 9 states had followed. The acceleration began in 2023 with the Westfield, Aledo, and Issaquah school incidents and reached its peak in 2024 with 22 new states enacting legislation.

The headline figure of 47 states overstates the actual protection available. Several states cover only minors (Mississippi, West Virginia, South Dakota, Kentucky). Others address only election-related deepfakes (New Mexico). A survivor of AI-generated intimate imagery in Kentucky has criminal recourse only if they are under 18.

6.1.1 Comprehensive state protections

Six states have enacted what this paper classifies as comprehensive protections. We define comprehensive as criminal penalties plus a civil cause of action plus explicit coverage of synthetic or AI-generated media.

California maintains multiple criminal statutes alongside Civil Code Section 1708.86, which establishes a private right

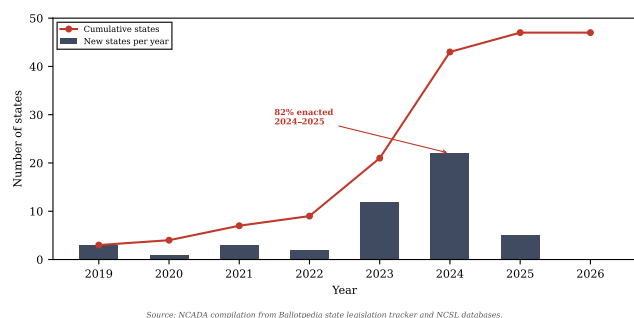


Figure 6: **Findings:** State deepfake law enactment, 2019 to 2026. The acceleration beginning in 2023 coincided with documented school incidents. Source: Ballotpedia state legislation tracker.

of action with platform liability provisions. Minnesota enacted separate criminal (Section 617.262) and civil (Section 604.31) statutes, and in April 2026 became the first state to ban nudification applications directly through HF 1606, carrying \$500,000 penalties per violation.⁵⁰ Colorado’s SB 25–288 provides criminal penalties alongside civil remedies with \$150,000 in liquidated damages. Washington enacted the broadest scope with no domicile limitation for victims. Florida provides criminal penalties of up to 15 years imprisonment alongside civil remedies with a \$10,000 minimum and platform liability requirements. New Jersey classifies deepfake NCII as a third-degree crime with fines up to \$30,000 plus a civil action pathway.

6.1.2 Dead zone states

Three states have no deepfake-specific NCII legislation: Alaska, Missouri, and Ohio. Alaska’s HB 47 passed the state House in February 2026 and remains pending. Missouri has multiple pending bills but none enacted. Ohio’s HB 185 sits in committee.

The Ohio dead zone matters in particular. In April 2026, an Ohio man became the first person convicted under the federal Take It Down Act for using AI to create and distribute NCII targeting adults and children.⁵¹ The conviction was federal because Ohio had no state law under which to prosecute. This case validates the Act’s role as a federal floor. Without the federal statute, the defendant would have faced no criminal liability.

6.1.3 The civil remedy gap

While 47 states have enacted criminal deepfake statutes, only an estimated 12 to 15 states provide a deepfake-specific civil cause of action. Approximately 20 to 25 provide a civil remedy for NCII more broadly.⁵² The gap matters. Criminal

prosecution requires state action, resource allocation, and the discretion of prosecutors who may deprioritize cases. A civil cause of action allows the survivor to pursue removal and damages independently.

The DEFIANCE Act (S. 1837), which passed the Senate unanimously on January 13, 2026 and awaits House action, would establish a federal civil cause of action with \$150,000 in liquidated damages (\$250,000 in cases involving stalking).⁵³ If enacted, the DEFIANCE Act would close the civil remedy gap for survivors in all 50 states.

6.1.4 Federal preemption

The Act does not include a preemption clause. The National Association of Attorneys General confirmed: “The Take It Down Act does not include a preemption clause, allowing state NCII laws to work concurrently.”⁵⁴ Survivors can pursue remedies under both federal and state law. States remain free to enact protections that exceed the federal floor.

6.2 School district investigation

6.2.1 Documented incidents

We identified 19 documented incidents of AI-generated or synthetic intimate imagery involving K–12 students in the United States between January 2023 and May 2026. These incidents span 12 states and involved victims ranging from middle school to high school age. Table 2 presents the incidents chronologically.

The earliest documented case occurred at Carmel High School in New York in February 2023. A student created a deepfake of a school principal. Administrators determined no law had been broken.⁵⁵ The Westfield High School incident in October 2023 led to the broadest policy response. Male students used AI to blend female classmates’ faces onto nude bodies and distributed images via Snapchat group chats. Over 30 girls were affected. No criminal charges were brought under then-existing New Jersey law. One victim, Francesca Mani, testified at the U.S. Senate Commerce Committee field hearing in Dallas in June 2024 and was invited as a guest to the 2024 State of the Union address. Governor Phil Murphy signed New Jersey deepfake legislation in April 2025.⁵⁶

At Aledo High School in Texas (October 2023), 9 girls were targeted. This incident directly catalyzed Representative Elliot’s introduction of the Take It Down Act. At Issaquah High School in Washington (October 2023), a 14-year-old student used a nudify application discovered on TikTok to generate images of at least 6 female classmates. The school kept the investigation internal and did not notify law

⁵⁰The 19th News, April 2026.

⁵¹NBC News, “First Person Convicted Under Law Criminalizing Intimate Deepfakes,” April 2026.

⁵²Figures from NCSL and Ballotpedia trackers; individual statute verification not conducted for all 50 states.

⁵³Congress.gov, S. 1837, 119th Congress.

⁵⁴NAAG, “Congress’s Attempt to Criminalize Nonconsensual Intimate Imagery,” 2025.

⁵⁵Multiple news sources, February 2023.

⁵⁶CBS New York, FOX 5, NJ State Bar Foundation, multiple sources.

enforcement. Police learned of the incident from three separate parents.⁵⁷

The Lancaster Country Day School case (November 2023 through March 2026) exposed a gap in mandatory reporting. Two male students created approximately 350 AI-generated nude images of nearly 50 female classmates using photos from Instagram, yearbooks, and TikTok. School administrators received an anonymous tip in November 2023 but did not notify law enforcement or file a ChildLine report until June 2024, after additional reports surfaced. The Lancaster County District Attorney determined that school officials “were not legally required to report because child-on-child harm is exempted” from mandatory reporting in Pennsylvania.⁵⁸ The head of school and the school board president resigned. Both students received probation, 60 hours of community service, and were ordered to pay \$12,000 in restitution.

At Sixth Ward Middle School in Louisiana (Fall 2025), 8 girls were targeted with AI-generated images. One victim was expelled from school for hitting the boy who was sharing the images.

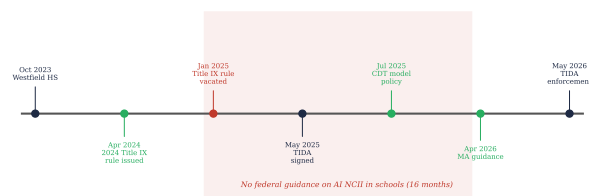
6.2.2 Institutional preparedness

The Center for Democracy and Technology published a national survey in 2024 [9]. The survey found that 40% of students know of a deepfake at their school. Sixty-two percent of teachers received no guidance from their school on how to handle NCII. Only 5% of teachers reported that their school provided resources to victims for having images removed from platforms. A 2025 RAND Corporation survey found that 13% of K–12 principals reported deepfake incidents at their schools and only 23% had updated school policies to address AI.⁵⁹ A Boston Globe investigation found that only 9 of 113 surveyed Massachusetts school districts mention AI in their harassment policies.⁶⁰

6.2.3 The Title IX vacuum

The 2024 Title IX rule issued by the Department of Education defined AI-generated NCII as a form of sexual harassment. It brought synthetic intimate imagery within the scope of schools’ federal civil rights obligations for the first time. On January 9, 2025, that rule was vacated nationwide.⁶¹ The reinstated 2020 regulations do not mention deepfakes, artificial intelligence, or synthetic media. As of May 2026, no replacement federal guidance addresses how schools should respond to AI-generated intimate imagery under Title IX.

The Hingham Middle School case (November 2025) illustrates the consequence: a student targeted by AI-generated images had their Title IX complaint dismissed. Governor Healey intervened publicly.



Source: NCADA regulatory timeline analysis from Federal Register, DOE publications, and state guidance documents.

Figure 7: **Findings:** Regulatory timeline for AI NCII in schools. The shaded region indicates 16 months without federal guidance. Source: Federal Register; DOE publications.

6.2.4 Available guidance

As of this writing, the total available guidance for school administrators on responding to AI-generated intimate imagery consists of the following: the CDT and ATIXA Model NCII Policy (July 2025), which is the only published K–12 model policy; the Massachusetts Department of Elementary and Secondary Education guidance (April 2026), which is the first state-level guidance; and nothing from the National School Boards Association, the American Association of School Administrators, or the federal Department of Education.⁶²

6.3 RQ3 synthesis

The state law landscape presents a paradox. Legislative activity is at an all-time high, with over 170 bills enacted since 2019 and 82% in the last two years. Yet the protections available to an individual survivor depend heavily on which state they live in, whether they are a minor, and whether they want criminal prosecution or a civil remedy. Three states provide no deepfake-specific protection at all. The federal floor that the Act provides is real (the Ohio conviction demonstrates this), but the Act offers no civil cause of action. The pending DEFIANCE Act would close that gap if enacted.

The school investigation reveals the widest institutional gap. Forty percent of students know of a deepfake at their school, yet 62% of teachers have received no guidance on how to respond, and only 5% can point to removal resources. The only K–12 model policy was published in July 2025 by CDT, a nonprofit, rather than by any government agency, and the 2024 Title IX rule that would have defined AI NCII as sexual harassment was vacated 16 months ago without any federal replacement. School administrators who want to respond correctly to an incident of AI-generated intimate imagery involving their students have almost nothing to consult.

7 Discussion

7.1 The remediation pathway in 2026

Our findings reveal a significant disparity between the regulatory architecture the Take It Down Act created and the reme-

⁵⁷404 Media, February 2024.

⁵⁸Philadelphia Inquirer, December 2024 and March 2026.

⁵⁹RAND Corporation, 2025.

⁶⁰Boston Globe, 2025.

⁶¹Holland & Knight, January 2025.

⁶²Searches for NSBA, AASA, and DOE model policies returned no results.

Table 2: **Findings:** Selected documented incidents of synthetic intimate imagery involving K–12 students, 2023–2026. This list is not exhaustive. Source: verified news coverage, court filings, and law enforcement records.

Date	School	State	Victims	Legal Outcome	Policy Change
Oct 2023	Westfield HS	NJ	30+	No charges; federal lawsuit	NJ deepfake law (Apr 2025)
Oct 2023	Aledo HS	TX	9	Investigation	Direct origin of TIDA
Oct 2023	Issaquah HS	WA	6+	No charges; school hid from police	WA state law (Jun 2024)
Nov 2023	Lancaster Country Day	PA	48+12	59 counts each; probation	Mandatory reporting debate
Dec 2023	Pinecrest Cove	FL	~24	First US arrest for AI NCII	N/A
Feb 2024	Beverly Vista MS	CA	16	5 expelled	School board resolution
2024–25	Cascade HS	IA	44	4 charged	Public victim statement
Fall 2025	Sixth Ward MS	LA	8+2	Victim expelled	N/A
Nov 2025	Hingham MS	MA	1	Title IX dismissed	Gov. Healey intervened
Dec 2025	Radnor HS	PA	Multiple	Investigation	AI ban; Gov. Shapiro roundtable
Feb 2026	Itasca ISD	TX	Multiple	Texas Rangers investigating	Pending

diation pathway it produces for victim-survivors of synthetic intimate imagery. We describe that pathway as it currently operates.

It begins at the generation layer. A person’s image is obtained from social media, a yearbook, or a screenshot, and is processed through one of the 34,439 deepfake models on Civitai, or one of the 102 nudify applications on app stores, or one of the 50+ Telegram bots serving 4 million monthly users. The technical barrier is minimal: LoRA fine-tuning requires as few as 20 source images and 15 minutes of compute time, and no federal law prohibits the creation itself. Only Minnesota and the United Kingdom have enacted generation-layer statutes, and the Act does not reach this stage.

The image is then distributed. If the survivor discovers the content (which depends on chance, monitoring services, or notification by others), they must navigate the platform’s reporting system. On Meta, the deepfake-specific reporting category is available only in Texas. On Telegram, no NCII reporting category exists. On WhatsApp, end-to-end encryption means the platform cannot access the content. On X, the Michigan study found that NCII reports filed through the platform’s own non-consensual nudity mechanism resulted in 0% removal over three weeks [1]. Only DMCA copyright claims produced action.

If the platform removes the content, removal does not prevent redistribution. Cuevas and Ribeiro (2026) demonstrated that removal from one platform leads to reappearance on others [7], and StopNCII’s hash-sharing network, while covering 18 partner platforms, excludes Telegram, WhatsApp, and Discord from hash matching. Because the generation tools remain available, the same content can be recreated at any time.

If the survivor seeks legal recourse, their options depend on geography. In 47 states, criminal prosecution is available. In 12 to 15 states, a civil cause of action exists. In three states (Alaska, Missouri, Ohio), no deepfake-specific law exists at all. The Act provides criminal penalties for publication and

a platform removal mandate, but no civil cause of action.

If the survivor is a student and the incident occurs at school, the response depends on whether administrators have guidance. One K–12 model policy exists (CDT, July 2025). One state (Massachusetts) has issued school-specific guidance. The 2024 Title IX rule that defined AI NCII as sexual harassment was vacated in January 2025. In Pennsylvania, the Lancaster County DA determined that school officials were not legally required to report student-generated AI CSAM because “child-on-child harm is exempted” from mandatory reporting.

7.2 TIDA’s design assumptions versus the 2026 reality

TIDA was designed with three assumptions, each of which the public record calls into question.

The first was that platforms would build compliant removal systems during the one-year compliance window. In practice, no platform has published a TIDA-specific compliance statement, only one (Discord) has published a TIDA-specific reporting page, and 9 of 10 platforms do not publish NCII-specific transparency data.

The second was that the FTC would have the capacity to enforce the law. FTC Chairman Ferguson testified in April 2026 that TIDA is “a high priority” and that the agency was building a complaint intake site and hiring “relevant personnel.” The FTC’s budget remains unchanged at \$425.7 million. An expert quoted in CyberScoop observed: “They are now in the business of regulating content moderation. That is hard work and not something they’re used to doing.”⁶³ Ferguson acknowledged that international offenders would be “not going to be super responsive to the civil law enforcement component of the FTC.”

The third was that the harm landscape would remain centered on distribution. The generation layer’s growth since the Act’s drafting has outpaced the law’s design: Civitai’s model

⁶³CyberScoop, “FTC Take It Down Act Enforcement,” May 2026.

library grew from a research curiosity to 14.9 million downloads, nudify applications accumulated 483 million installs, and Grok generated an estimated 3 million sexualized images in 11 days on the very platform the Act regulates.

7.3 Comparative analysis: TIDA and the EU DSA

The European Union’s Digital Services Act provides a structural comparison. The DSA imposes transparency reporting obligations on platforms as a compliance requirement rather than a voluntary practice, and the European Commission’s preliminary findings against Meta (October 2025) and its formal proceedings against X (January 2026) demonstrate active enforcement within the first year of the law’s application. By contrast, TIDA imposes a 48-hour removal mandate but does not require platforms to publish NCII-specific compliance data, which means the European Commission can identify noncompliance from published reports while the FTC can identify it only through individual investigations.

8 Recommendations

The following recommendations are anchored to specific findings presented in Sections 4 through 6.

8.1 For the Federal Trade Commission

The Commission should require platforms to publish NCII-specific transparency data as a condition of compliance, including removal volumes separated from broader content categories, average response times for NCII reports, and proactive versus reactive detection rates. Without mandated disclosure, the 48-hour standard is unverifiable. The Commission should also investigate Telegram’s compliance posture.

8.2 For Congress

Congress should amend the Act to address the generation layer. The statutory language targeting persons who “knowingly publish” does not reach model marketplaces, nudify application operators, or generation tools. Minnesota’s HF 1606 and the United Kingdom’s creation criminalization demonstrate that generation-layer regulation is feasible. Congress should also enact the DEFIANCE Act to provide a federal civil cause of action.

8.3 For platforms

Platforms should publish NCII-specific removal data in their transparency reports without waiting for regulatory mandate. Platforms not currently participating in StopNCII (Telegram, WhatsApp, Discord) should join the hash-sharing network. Platforms that received FTC compliance letters should publish TIDA-specific compliance statements.

8.4 For state legislatures

The three dead-zone states (Alaska, Missouri, Ohio) should enact deepfake-specific NCII statutes. States with criminal statutes but no civil cause of action should add a civil remedy pathway. States should consider Minnesota’s model of banning nudification tools directly.

8.5 For school administrators

Schools should adopt the CDT and ATIXA Model NCII Policy (July 2025) immediately as interim guidance. State education departments should issue school-specific guidance following Massachusetts DESE’s April 2026 model. The Department of Education should issue Title IX guidance addressing AI-generated intimate imagery. State legislatures should close the mandatory reporting gap exposed by the Lancaster Country Day case.

8.6 For civil rights attorneys

The per-platform findings in Section 4 provide documented evidence of the gap between published policies and actual enforcement. The Meta Oversight Board findings (auto-closed reports, failure to implement recommendations), the Michigan audit (0% removal under X’s NCII policy), and the Snapchat litigation (10,000 sextortion reports per month, 80% inactionable NCMEC submissions) are each citable in negligence and unfair trade practices claims.

9 Limitations and Future Work

This paper is a public record review, not an experimental audit. It cannot measure actual removal times, actual survivor experiences, or platform internal decision-making. The Michigan study by Qiwei et al. (2024) remains the only published controlled audit of NCII removal, conducted on a single platform (X) two years before enforcement [1]. Replication of that methodology across multiple platforms post-TIDA would advance the field.

The state law inventory relies on aggregate counts from the National Conference of State Legislatures and Ballotpedia trackers. Individual statute text was verified for the six most comprehensive states and the three dead-zone states, but not for all 50 states. The approximate figures for civil remedies (12 to 15 states with deepfake-specific civil causes of action) reflect tracker estimates, not exhaustive statutory analysis.

The school district investigation identified 19 documented incidents through verified news coverage and court filings. The CDT finding that 40% of students know of a deepfake at their school indicates that the documented record substantially undercounts the actual incidence.

The NCMEC CyberTipline data on AI-related reports must be interpreted with the Stanford correction. Pfeifferkorn demonstrated that 78% of H1 2025 “AI-related” reports consisted of Amazon scanning training data for known CSAM, not actual AI-generated content [16]. The headline figure of

1.5 million AI-related reports significantly overstates the volume of AI-generated CSAM.

NCADA's June 2026 follow-up audit will contribute original data on platform compliance by conducting a structured evaluation of NCII reporting systems across 10 platforms. That audit will include survivor testimony collected with consent and expert interviews with StopNCII, CCRI, Reality Defender, and academic researchers.

10 Conclusion

We found consistent and significant gaps across all three layers of the regulatory architecture that the Take It Down Act created. At the distribution layer, 9 of 10 platforms do not publish NCII-specific enforcement data, and the only controlled audit of any platform's NCII reporting mechanism found a 0% removal rate over three weeks [1]. At the generation layer, the tools producing synthetic intimate imagery operate at a scale the Act was not designed to address: 34,439 deepfake models on a single marketplace [17], over 100 nudify applications across app stores [10], and a platform-integrated AI tool that generated millions of sexualized images while the platform itself was subject to the Act's distribution requirements. At the institutional layer, 40% of students report awareness of deepfakes at their school [9], yet the only federal rule that would have classified AI-generated NCII as sexual harassment under Title IX was vacated in January 2025, and no replacement has been issued in the 16 months since.

The Act passed with bipartisan support of 409 to 2 and addresses a real gap in federal law. The evidence assembled in this review, however, highlights the need for amendments that extend the Act's reach to the upstream generation of synthetic intimate imagery, mandate the transparency reporting necessary for the FTC to verify compliance, and provide the institutional guidance that schools and survivors require. Minnesota, the United Kingdom, and individual civil society organizations have demonstrated that each of these gaps is addressable through existing legislative and policy mechanisms. Our study contributes to the broader understanding of platform accountability, content moderation processes, and how laws influence platform behavior in the context of non-consensual intimate media.

Acknowledgements

The authors thank Reality Defender for ongoing partnership in synthetic media detection research; the Center for AI and Digital Policy (ECOSOC-accredited) for institutional support; the United Nations Office on Drugs and Crime for partnership on transnational digital abuse; and Congresswoman Mikie Sherrill's office for collaborative work on the Digital Likeness Protection Act.

References

- [1] Li Qiwei, Shihui Zhang, Andrew Timothy Kasper, Joshua Ashkinaze, Asia A. Eaton, Sarita Schoenebeck, and Eric Gilbert. 2024. Reporting Non-Consensual Intimate Media: An Audit Study of Deepfakes. *arXiv:2409.12138*.
- [2] David Thiel, Renée DiResta, and Alex Stamos. 2023. Cross-Platform Dynamics of Self-Generated CSAM. Stanford Internet Observatory.
- [3] Security Hero. 2024. State of Deepfakes Report.
- [4] Danielle Keats Citron. 2018. Sexual Privacy. *Yale Law Journal* 128, 1870.
- [5] Mary Anne Franks. 2014. Drafting an Effective "Revenge Porn" Law: A Guide for Legislators. *SSRN Electronic Journal*.
- [6] Asia Eaton, Holly Jacobs, and Yanet Ruvalcaba. 2017. 2017 Nationwide Online Study of Nonconsensual Porn Victimization and Perpetration.
- [7] Cuevas and Ribeiro. 2026. Deepfake Pornography is Resilient to Regulatory and Platform Shocks. *arXiv:2602.02754*.
- [8] Lukas Struppek et al. 2022. Learning to Break Deep Perceptual Hashing: The Use Case NeuralHash. *ACM FAccT*.
- [9] Center for Democracy and Technology. 2024. In Deep Trouble: Surfacing Tech Powered Sexual Harassment in K–12 Schools.
- [10] Tech Transparency Project. 2026. Apple and Google Are Steering Users to Nudify Apps.
- [11] Kaylee Williams. 2025. Nudify Applications and NCII. *Violence Against Women*.
- [12] DarkGram. 2025. Characterizing Cybercriminal Activity on Telegram. *USENIX Security*. *arXiv:2409.14596*.
- [13] Semenzin and Bainotti. 2020. The Use of Telegram for Non-Consensual Dissemination of Intimate Images. *Journal of Gender Studies*.
- [14] Internet Watch Foundation. 2025. Annual Report. 3,443 AI-generated CSAM videos identified.
- [15] RAND Corporation. 2025. K–12 Principal Survey on AI-Generated Content in Schools.
- [16] Riana Pfefferkorn. 2025. Letter to NCMEC About AI CSAM Report Statistics. Stanford Center for Internet and Society.
- [17] Deepfakes on Demand: Characterizing Celebrity Deepfake Models on Civitai. 2025. *arXiv:2505.03859*.
- [18] Le and Doering. 2026. Perspectives on Non-consensual Deepfake Pornography: A Content Analysis of Reddit Discussions. *Sexuality & Culture*, Springer.

A State Deepfake Law Summary

Notable state provisions: Colorado SB 25–288 provides \$150,000 in liquidated damages. Minnesota HF 1606 (April 2026) is the first state to ban nudification applications directly, with \$500,000 penalties per violation. Florida provides criminal penalties up to 15 years and civil minimums of \$10,000. Washington has the broadest jurisdictional scope with no domicile limitation.

The DEFIANCE Act (S. 1837) passed the Senate unanimously on January 13, 2026 and is pending House action. It would establish a federal civil cause of action with \$150,000 in liquidated damages (\$250,000 for stalking-related cases). Enactment would close the civil remedy gap in all 50 states.

B FTC Compliance Letter Recipients

On approximately May 11, 2026, FTC Chairman Andrew N. Ferguson sent compliance letters to 15 companies advising them to comply with the Take It Down Act by the May 19, 2026 enforcement deadline. The FTC stated: “We stand ready to monitor compliance, investigate violations, and enforce the Take It Down Act.” Violations carry civil penalties of \$53,088 per violation under the FTC Act.

Notable absences: Telegram (largest documented vector for deepfake generation bots; approximately 50 moderators for 1 billion+ users); Pornhub/Aylo (operating under a separate \$5 million FTC consent order from September 2025); all AI generation platforms (Civitai, HuggingFace, nudify application operators); all uncensored model hosts.

C Per-Platform Research Protocol

Each of the 10 platforms was evaluated through the following 9-step protocol.

Step 1: NCII Policy. Locate and quote the platform’s published policy on non-consensual intimate imagery. Record the URL and date last updated. Note whether the policy addresses AI-generated or synthetic content.

Step 2: Deepfake/Synthetic Media Policy. If the platform maintains a separate policy for synthetic or AI-generated content, locate and quote it.

Step 3: Reporting Flow. Document the reporting mechanism for NCII from the platform’s help center. Record: form URL, required fields, whether an account is required, whether government ID is required, whether a deepfake-specific category exists, and any published response time commitments.

Step 4: Transparency Data. Locate the platform’s most recent transparency report. Extract all NCII-relevant metrics. Record the reporting period and publication date.

Step 5: StopNCII Participation. Verify the platform’s status on the StopNCII.org industry partner list.

Step 6: NCMEC Take It Down Participation. Verify the platform’s status on the NCMEC Take It Down participant list.

Step 7: FTC Enforcement. Determine whether the platform received a compliance letter on May 11, 2026.

Step 8: Documented Incidents. Search verified news sources for NCII or deepfake incidents involving the platform between January 2025 and May 2026. Verify each incident with at least two independent sources.

Step 9: Cross-Reference. Compare findings from Steps 1 through 8 to identify contradictions between published policies and documented enforcement, between transparency data and independent findings, and between partnership claims and verified participation. Gaps and contradictions are treated as findings.

Table 3: State deepfake and NCII law coverage summary as of May 2026. Source: NCADA compilation from NCSL and Ballotpedia trackers.

Category	States	Key Features
Comprehensive (criminal + civil + synthetic media)	California, Minnesota, Colorado, Washington, Florida, New Jersey	Criminal penalties, civil cause of action, explicit AI/synthetic media coverage, platform liability provisions
Criminal only (deepfake specific)	Approximately 30 additional states	Criminal penalties for distribution; no private right of action
NCII law without synthetic media coverage	Mississippi, West Virginia, South Dakota, Kentucky (minors only)	NCII criminalized but statutes do not cover AI-generated content, or apply only to minors
Election deepfakes only	New Mexico	Addresses election content only; does not cover intimate imagery
Dead zone (no deepfake law)	Alaska, Missouri, Ohio	No enacted deepfake legislation. First federal TIDA conviction occurred in Ohio (April 2026).

Table 4: FTC TIDA compliance letter recipients, May 11, 2026. Source: FTC Press Release.

Company	Platform(s)	Category	StopNCII
Amazon	Amazon, AWS, Twitch	E-commerce/Cloud/Streaming	No
Alphabet	Google, YouTube	Search/Video	Yes (Sept 2025)
Apple	App Store, iCloud	Consumer Electronics	No
Automattic	WordPress	Content Management	No
Bumble	Bumble, Badoo	Dating	Yes
Discord	Discord	Social/Gaming Chat	No (supporter)
Match Group	Tinder, Hinge	Dating	No
Meta	FB, IG, Threads, WA	Social Media/Messaging	Yes (founding)
Microsoft	Bing, Xbox	Search/Gaming/Cloud	Yes
Pinterest	Pinterest	Image Sharing	No
Reddit	Reddit	Social Media/Forum	Yes
SmugMug	SmugMug, Flickr	Photography	No
Snap Inc.	Snapchat	Social Media/Messaging	Yes
TikTok	TikTok	Social Media/Video	Yes
X Corp.	X (formerly Twitter)	Social Media	Yes (2025)