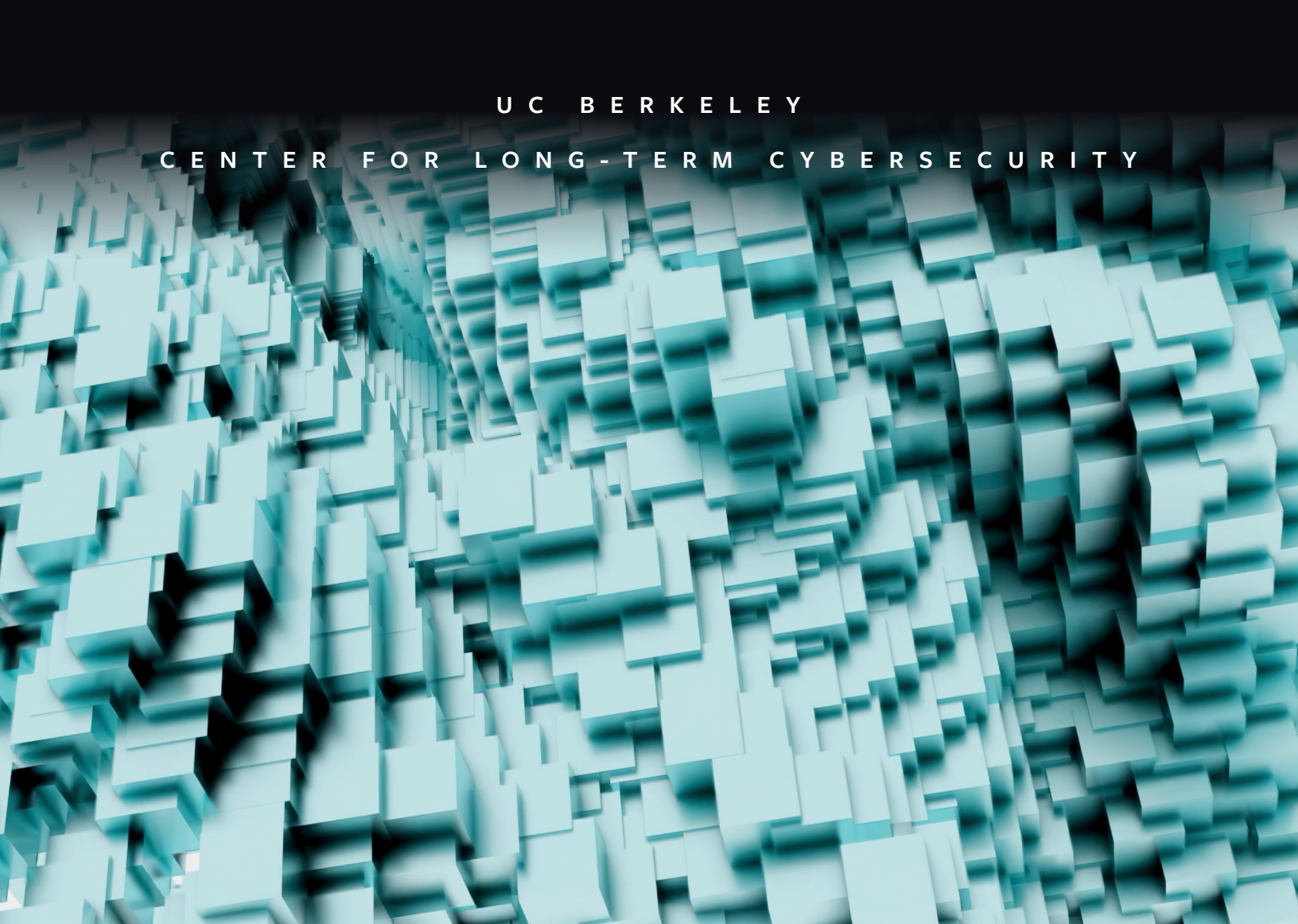




U C B E R K E L E Y
C E N T E R F O R L O N G - T E R M C Y B E R S E C U R I T Y

C L T C W H I T E P A P E R S E R I E S

Closing Gaps Across the Ecosystem

REGULATING THE TECHNOLOGIES THAT ENABLE AI-POWERED
NONCONSENSUAL INTIMATE IMAGES AND
CHILD SEXUAL ABUSE MATERIALS

C H A R L O T T E Y U A N , N A D A M A D K O U R

CLTC WHITE PAPER SERIES

Closing Gaps Across the Ecosystem

**REGULATING THE TECHNOLOGIES THAT ENABLE AI-POWERED
NONCONSENSUAL INTIMATE IMAGES AND
CHILD SEXUAL ABUSE MATERIALS**

CHARLOTTE YUAN, NADA MADKOUR

September 2026

The authors conducted the research as part of the program of professional education at the Goldman School of Public Policy, at the University of California, Berkeley. The judgments and conclusions are solely those of the author and are not necessarily endorsed by the Goldman School of Public Policy, the University of California, or any other agency.



Contents

EXECUTIVE SUMMARY v

INTRODUCTION 1

Key Terms 2

THREAT LANDSCAPE:

AI NCII AND CSAM PROLIFERATION AND REGULATORY GAPS 3

The AI NCII and CSAM Ecosystem 4

Disproportionate Harms on Women 5

Disproportionate Harms on Children 6

Current Policy Interventions and Limitations 7

METHODOLOGY 10

FINDINGS 11

Analysis of the AI NCII and CSAM Ecosystem 11

Policy Analysis 15

TAKE IT DOWN Act 16

State AI NCII and CSAM Laws: California, New York, Texas, and Utah 17

POLICY RECOMMENDATIONS 20

Overarching Recommendation—21

Distribution Layer 21

Federal-Level Recommendations—21

Deployment Layer 21

Distribution Layer 22

State-Level Recommendations—23

Deployment Layer 23

Distribution Layer 25

Model Layer 27

Other 28

CONCLUSION 29

APPENDIX 30

ACKNOWLEDGMENTS 37

Executive Summary

This paper proposes recommendations for policies that can be implemented in the United States to regulate technologies that enable the creation and dissemination of AI-powered nonconsensual intimate images (NCII) and child sexual abuse materials (CSAM). The recommendations are based on mapping existing laws against the AI NCII and CSAM value chain (i.e., the individuals and platforms supporting the production and distribution of these materials) and identifying regulatory gaps that lawmakers can address.

Deepfakes, or synthetic media developed with artificial intelligence (AI), have experienced unprecedented growth since 2018. The rapid proliferation of realistic AI-generated content has complicated the digital safety landscape and raised significant policy challenges, including elevated risks of sexual violence against women and children. AI has significantly lowered the barrier to entry for malicious actors seeking to create, modify, and distribute NCII and CSAM, promoting a network of abusive tools, malicious actors, and supporting infrastructure at scale that is extremely difficult to dismantle through existing regulatory approaches.

Robust policy guardrails are needed to mitigate the harms of AI-powered NCII and CSAM, which threaten individuals' privacy, financial stability, reputation, and psychological well-being. However, the regulatory landscape in the U.S. remains fragmented and is insufficient to address these risks. Despite a recent surge in legislative developments around deepfakes, lawmakers continue to struggle with determining and prioritizing the most critical intervention points. Regulatory gaps persist, as most laws focus on prosecuting individual creators and distributors, while the ecosystem of technologies, platforms, and economic incentives that enable the mass proliferation of AI-powered NCII and CSAM remains largely unregulated.

This paper examines the risks associated with AI NCII and CSAM through a review of relevant research, analysis of existing legislation, and engagement with experts on this topic. By tracking and evaluating 28 federal and state laws (in California, New York, Texas, and Utah), the paper identifies where current regulations fall short in mitigating harms, and which nodes in the value chain have the greatest potential for regulatory intervention.

While the TAKE IT DOWN Act was signed into law in May 2025, criminalizing the nonconsensual sharing of authentic or AI-manipulated intimate images, this law has proved to be insufficient in addressing the challenge. This paper offers recommendations for proactive regulations at both

the federal and state levels that would target the structural conditions that allow AI CSAM and NCII to proliferate. Following is a summary of key findings and recommendations:

- **Current legislative efforts overwhelmingly focus on punishing individual creators and distributors**, which limits accountability for key actors such as social media companies, open-source platforms, and model developers.
- An overarching recommendation is that lawmakers at both the federal and state levels should **remove the requirement that AI CSAM depict real identifiable children**. Creation or possession of AI-generated CSAM content should be sufficient grounds for prosecution.
- At the federal level, lawmakers should mandate that open-source model repositories implement a mechanism for the public to report models and datasets that contain CSAM and NCII, and establish standardized procedures for removing them.
- Future federal laws should **eliminate the requirement that perpetrators possess intent to distribute AI NCII and CSAM** in order to face civil or criminal liability.
- Lawmakers should **establish a civil right of action for victims to sue perpetrators** directly, complementing the TAKE IT DOWN Act and ensuring that legal action is available to those harmed.
- At the state level, lawmakers should pass laws that prohibit the advertisement of websites and applications that create or distribute AI NCII and CSAM; require that harmful content be taken down within a given time frame; and incentivize developers of AI models to meaningfully improve their safety practices.
- Lawmakers should also consider mandating AI literacy education at schools to equip children with an understanding of the harms of NCII and CSAM and the importance of consent. Improved education and awareness are essential for addressing cultural and behavioral root causes that regulation alone cannot reach.

Introduction

With the rise of artificial intelligence (AI), deepfake content has surged significantly since 2018. AI applications such as Grok, an AI chatbot developed by xAI, have facilitated the creation and large-scale dissemination of AI-manipulated non-consensual intimate images (NCII) of adults and child sexual abuse materials (CSAM).¹ Deepfakes have raised serious policy challenges, especially around the augmentation of risks of sexual violence against women and children. Sexually explicit materials remain a dominant category of deepfakes, accounting for 98% of content.² Research shows that nearly 96% of intimate deepfakes are nonconsensual, with 99% targeting women.³ Due to the disproportionate impact of these technologies on women and children, it is important to establish policy guardrails that protect individuals' financial stability, privacy, reputation, and psychological well-being.

The current U.S. regulatory landscape remains too fragmented to mitigate such harms at scale, leaving victims vulnerable to additional risks such as sextortion, a form of blackmail where perpetrators threaten to release intimate imagery unless victims pay money or provide additional sexual content. Regulations across states not only vary in scope, terminology, and enforcement, but also fail to situate the problem within a larger technological ecosystem, undermining interventions already in place. To combat the rapid proliferation of AI-powered NCII and CSAM, this paper aims to **(1) understand the types of risks of NCII and CSAM augmented by AI; (2) identify regulatory gaps at the federal and state levels**, focusing on New York, California, Texas, and Utah; and **(3) propose robust and actionable solutions for policymakers** based on an ecosystem approach that addresses all technologies and actors that facilitate the proliferation of AI NCII and CSAM.

To develop recommendations for policy measures, this research examined federal- and state-level regulations that have already been enacted in the U.S. as of July 2026. The paper investigated the following research questions:

¹ Chan, Kelvin. 2026. "EU Investigates Musk's AI Chatbot Grok over Sexual Deepfakes." *PBS News*. January 26. <https://www.pbs.org/newshour/world/eu-investigates-musks-ai-chatbot-grok-over-sexual-deepfakes>.

² Negreiro, Mar. 2025. "Children and Deepfakes." European Parliament. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775855/EPRS_BRI\(2025\)775855_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2025/775855/EPRS_BRI(2025)775855_EN.pdf).

³ Baekgaard, Kristine. 2024. "Technology-Facilitated Gender-Based Violence." *Georgetown Institute for Women, Peace and Security*. June 18. <https://giwps.georgetown.edu/resource/technology-facilitated-gender-based-violence/>.

1. How has AI amplified and augmented existing risks and harms of NCII and CSAM?
2. What are the existing regulatory gaps in the U.S.? What are the patterns of convergence and divergence across state-level deepfake regulations?
3. Within the AI NCII and CSAM ecosystem, which technologies or actors lack regulatory oversight?
4. What policy interventions can prevent and mitigate harms? What are potential challenges to implementing these solutions?

A major contribution of the paper is to propose policy solutions that **proactively prevent the generation of abusive materials from the beginning of the AI NCII and CSAM lifecycle**, rather than simply containing or removing the distribution of such content on the internet. By mapping existing regulations to potential intervention points in the deepfake ecosystem, we aim to help policymakers identify and address hidden technological nodes and actors that remain underregulated.

KEY TERMS

Below is an overview of key terms used in this paper:

- **Deepfakes:** AI-generated content depicting identifiable individuals that, for most viewers, are indistinguishable from real people.
- **AI-manipulated non-consensual intimate images (AI NCII):** Explicit visual material depicting adults, including AI-manipulated media, that is created, distributed, or threatened for release.⁴
- **AI-generated child sexual abuse material (AI CSAM):** Any sexually explicit visual depiction of a minor that is modified or entirely created using AI systems.⁵
- An **important distinction between AI NCII and AI CSAM** is that NCII is generated based on images of real humans, whereas CSAM can be created using both fictional content and images of real children.

4 Ding, Michelle L., Harini Suresh, and Suresh Venkatasubramanian. 2026. "How to Stop Playing Whack-a-Mole: Mapping the Ecosystem of Technologies Facilitating AI-Generated Non-Consensual Intimate Images." arXiv. doi:[10.48550/arXiv.2602.04759](https://doi.org/10.48550/arXiv.2602.04759).

5 Ciardha, Caoilte Ó, John Buckley, and Rebecca S. Portnoff. 2025. "AI Generated Child Sexual Abuse Material — What's the Harm?" arXiv. doi:[10.48550/arXiv.2510.02978](https://doi.org/10.48550/arXiv.2510.02978); "AI CSAM Report 2026: Harm Without Limits." 2026. *Internet Watch Foundation*. <https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/>

Threat Landscape: AI NCII and CSAM Proliferation and Regulatory Gaps

Prior research has demonstrated that the harms related to AI NCII and CSAM stem from a range of technologies, thereby requiring a shift toward policy solutions that span the full ecosystem of platforms and software that enable the production, modification, and distribution of harmful materials. However, current U.S. laws lack such a holistic design and fail to target key actors and technologies in the value chain, including open-source platforms, supporting infrastructure, social media platforms, model developers, and payment processors.

AI has increased risks associated with CSAM and NCII by lowering barriers to engagement and facilitating the normalization of extreme content.⁶ The creation of deepfakes previously required significant labor and niche technical expertise, such as coding skills, but AI has removed these barriers and allows non-technical users to function as both consumers and creators. Grok, for instance, has allowed its users to generate NCII at an unprecedented scale, with some reporting that the model generates about 6,700 NCII every hour.⁷ In 2024, the National Center for Missing and Exploited Children (NCMEC) reported an increase of 1,325% in AI-powered CSAM.⁸

Deepfakes have also augmented risks related to sextortion, which typically targets male victims between the ages of 14 and 17.⁹ Sextortion occurs when malicious actors use explicit images or videos to coerce victims into making payments. The process often begins with reaching out to individuals with explicit content of other people to manipulate the victim into sending their own materials.¹⁰ Victims are then threatened with sharing content publicly unless a ransom

6 Ibid.

7 Williams, Kaylee. 2026. "Grok Supercharges the Nonconsensual Pornography Epidemic." Tech Policy Press. January 14. <https://www.techpolicy.press/grok-supercharges-the-nonconsensual-pornography-epidemic/>.

8 Vaughan, Emma Henderson. 2025. "NCMEC Releases New Data: 2024 in Numbers." *National Center for Missing & Exploited Children*. May 8. <http://www.missingkids.org/content/ncmec/en/blog/2025/ncmec-releases-new-data-2024-in-numbers.html>.

9 Thorn. 2024. "Sextortion Survivors Share Their Stories to Help Other Victims." *Thorn*. November 4. <https://www.thorn.org/blog/sextortion-what-you-can-do/>.

10 "Malicious Actors Manipulating Photos and Videos to Create Explicit Content and Sextortion Schemes." 2023. Federal Bureau of Investigation. June 5. <https://www.ic3.gov/PSA/2023/psa230605>.

payment is received. With the rise of AI tools, malicious actors no longer need to persuade victims to send content but can easily modify existing images or videos.¹¹

AI tools like ChatGPT¹² and Gemini¹³ have increased access to the creation of deepfakes through an easy-to-use interface, further perpetuating AI NCII and CSAM at scale. Such developments amplify the urgency for policymakers to address these risks by regulating the technologies and actors involved.

THE AI NCII AND CSAM ECOSYSTEM

The rapid proliferation of AI tools has resulted in an entire deepfake ecosystem consisting of several unique actors (e.g., individual creators, developers, public platforms, and service providers), each of which plays a role in enabling large-scale production and distribution of AI NCII and CSAM. (The section of this paper entitled “Analysis of the AI NCII and CSAM Ecosystem,” provides an in-depth analysis of the ecosystem that enables the creation and distribution of AI NCII and CSAM.)

Effective policy interventions must situate the problem of AI-powered NCII and CSAM within the context of this broader ecosystem. Regulations targeting a single actor, or only a subset of actors, are unlikely to result in effective mitigation. For example, if one website or app is banned, users tend to simply migrate to another platform.¹⁴ This characteristic of NCII and CSAM poses significant challenges to suppressing activity, illustrating that simply shutting down public channels or targeting individual creators is not enough to prevent the creation and distribution of abusive content. In a table entitled “Taxonomy of Technologies and their Roles in the AIG-NCII Ecosystem,” Ding et al. (2026) map the specific technologies that facilitate AI-generated (AIG) NCII through creation, distribution, proliferation and discovery, supporting infrastructure, and monetization.¹⁵ For example, in the proliferation and discovery category, enabling technologies include deepfake creation communities, search engines, advertisement

11 “AI-Generated Child Sexual Abuse: The New Digital Threat We Must Confront Now.” 2025. *Thorn*. August 13. <https://www.thorn.org/blog/ai-generated-child-sexual-abuse-the-new-digital-threat-we-must-confront-now/>.

12 Vallance, Chris. 2026. “OpenAI Works to Stop ChatGPT Generating ‘sex Crime Scene’ Images.” *BBC*. June 17. <https://www.bbc.com/news/articles/c8o2ldjklzo>.

13 Rogers, Reece. 2025. “Google’s and OpenAI’s Chatbots Can Strip Women in Photos Down to Bikinis.” *WIRED*. December 23. <https://www.wired.com/story/google-and-openais-chatbots-can-strip-women-in-photos-down-to-bikinis/>.

14 Ding, Michelle L., Harini Suresh, and Suresh Venkatasubramanian. 2026. “How to Stop Playing Whack-a-Mole: Mapping the Ecosystem of Technologies Facilitating AI-Generated Non-Consensual Intimate Images.” arXiv. doi:10.48550/arXiv.2602.04759.

15 Ibid.

platforms, and app stores. Using such a framework can support the development of more effective policy interventions.

The deepfake ecosystem powers an inherently profitable market worth millions of dollars, which complicates traditional regulatory efforts and requires reform that limits market incentives. As of 2025, it is estimated that the current AI “nudifier” economy is worth over \$36 million,¹⁶ a profitable gateway that channels revenue into other digital sectors. Mainstream social media platforms such as Instagram generate significant revenue from hosting advertisements for nonconsensual AI nude apps,¹⁷ and historically, the porn industry has served as a main driver for the adoption of new technologies (e.g., VCRs, interactive gaming).¹⁸ As a result, the lucrative deepfake industry continues to attract investors, advertisers, and corporate partnerships, magnifying the prioritization of profit over safety.

DISPROPORTIONATE HARMS ON WOMEN

Deepfakes are being used as a tool for gender-based violence, which often leads to significant psychological, financial, and reputational harm to victims. Studies demonstrate that most AI-manipulated NCII target women with a 9:1 ratio compared with men.¹⁹ DeepNude, an app that “undresses” women with a single click (i.e., makes a clothed woman in an image appear naked) is unable to generate synthetic content of men because it was trained only on images of women.²⁰ The disproportionate harm to women also extends to political and public figures. Researchers reported that, across 11 websites containing NCII of U.S. Senators and Members of Congress, 25 out of 26 images were of women.²¹ While the majority of non-consensual deepfake content has historically targeted female celebrities, there is an accelerating trend toward the proliferation of images targeting ordinary individuals.²²

16 Mantzarlis, Alexios, and Santiago Lakatos. 2025. “AI Nudifiers Continue to Reach Millions and Make Millions.” *Indicator*. July 14. <https://indicator.media/p/ai-nudifiers-continue-to-reach-millions-and-make-millions>.

17 Maiberg, Emanuel. 2024. “Instagram Advertisements Nonconsensual AI Nude Apps.” 404 Media. April 22. <https://www.404media.co/instagram-advertisements-nonconsensual-ai-nude-apps/>.

18 Arlidge, John. 2002. “The Dirty Secret That Drives New Technology: It’s Porn.” *The Guardian*, March 3, sec. Technology. <https://www.theguardian.com/technology/2002/mar/03/internetnews.observerfocus>.

19 Ibid; Cuevas, Alejandro, and Manoel Horta Ribeiro. 2026. “Deepfake Pornography Is Resilient to Regulatory and Platform Shocks.” *arXiv*. doi:10.48550/arXiv.2601.09117iv.2602.02754; <https://arxiv.org/pdf/2602.04759>

20 Ajder, Henry, Giorgio Patrini, Francesco Cavalli, and Laurence Cullen. 2019. *The State of Deepfakes: Landscape, Threats, and Impact*. Deeptrace. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf.

21 Li, E. Rosalie, Benjamin Shultz, and Nina Jankowicz. 2024. “Deepfake Pornography Goes to Washington: Measuring the Prevalence of AI-Generated Non-Consensual Intimate Imagery Targeting Congress.” *American Sunlight Project*, December.

22 “Police Warn of Rising Threat from Sexual Deepfakes.” 2025. *National Police Chiefs’ Council (NPCC)*. November 24. <https://news.npcc.police.uk/releases/police-warn-of-rising-threat-from-sexual-deepfakes>.

AI NCII is also used to create “revenge porn,” non-consensual sharing of private sexual images driven by the motivation to retaliate against women.²³ In May 2026, two men were charged with using AI to process benign photos of women to create and publish NCII, marking one of the first major prosecutions targeting AI-powered revenge porn.²⁴ Many forums where these images are distributed are profit-driven, charging victims hundreds of dollars to remove intimate pictures.²⁵ Victims of these materials often suffer from mental health issues, including loss of trust in others, PTSD, anxiety, and depression. These impacts can also manifest as destructive coping mechanisms, such as avoidance, denial, self-harm, and substance abuse.²⁶

DISPROPORTIONATE HARMS ON CHILDREN

AI increases the accessibility and distribution of abusive content of children, especially in peer-to-peer contexts at schools. Research found that 79% of high school students who have been exposed to AI NCII reported that the content depicted a fellow student.²⁷ Currently, most schools do not have mechanisms in place to address these issues, despite their prevalence, and fail to handle incidents appropriately for traumatized victims.²⁸ The spread of harmful content at schools is partially because children are more prone to generating and distributing nudified images than adults, contributing to a sense of normalization among creators and distributors.²⁹

A common misconception is that AI CSAM poses less harm than non-synthetic CSAM. However, law enforcement agencies warn that the exponential rise of synthetic content is overwhelming current investigative capacities, making it harder to identify real victims.³⁰ Studies also show that AI not only enhances the ability of malicious actors to blackmail children using manipu-

23 Bates, Samantha. 2017. “Revenge Porn and Mental Health: A Qualitative Analysis of the Mental Health Effects of Revenge Porn on Female Survivors.” *Feminist Criminology* 12 (1). SAGE Publications: 22–42. doi:10.1177/1557085116654565.

24 Jeyaretnam, Miranda. 2026. “Two Men Charged Under A.I.-Revenge Porn Law: What to Know.” *Time*, May 22. <https://time.com/article/2026/05/22/ai-deepfake-revenge-porn-take-it-down-act-law/>.

25 “FTC, Nevada Obtain Order Permanently Shutting down Revenge Porn Site MyEx.” 2018. *Federal Trade Commission*. June 21. <https://www.ftc.gov/news-events/news/press-releases/2018/06/ftc-nevada-obtain-order-permanently-shutting-down-revenge-porn-site-myex>.

26 Ibid; Sciacca, Beatrice, Angela Mazzone, Magnus Loftsson, James O’Higgins Norman, and Mairéad Foody. 2023. “Nonconsensual Dissemination of Sexual Images Among Adolescents: Associations With Depression and Self-Esteem.” *Journal of Interpersonal Violence* 38 (15–16): 9438–64. doi:10.1177/08862605231165777.

27 Laird, Elizabeth, Maddy Dwyer, and Kristin Woelfel. 2024. “In Deep Trouble: Surfacing Tech-Powered Sexual Harassment in K-12 Schools.” Center for Democracy and Technology, September. <https://cdt.org/wp-content/uploads/2024/09/2024-09-26-final-Civic-Tech-Fall-Polling-research-1.pdf>.

28 Pfefferkorn, Riana, Shelby Grossman, Sunny Liu, and Jeffrey Hancock. 2025. “AI-Generated Child Sexual Abuse Material: Insights from Educators, Platforms, Law Enforcement, Legislators, and Victims.” <https://purl.stanford.edu/mn692xc5736>.

29 Ibid.

30 “Child Sexual Abuse Material Created by Generative AI and Similar Online Tools Is Illegal.” 2024. *Internet Crime Complaint Center*. March 29. <https://www.ic3.gov/PSA/2024/PSA240329>.

lated materials, but also enables more distressing content than traditional CSAM, exacerbating harms to children and psychological distress to content reviewers who regularly examine and investigate such material.³¹ For example, a high school student in Ohio committed suicide after a perpetrator sent a threatening text of an AI-generated nude image and demanded payment to withhold its publication.³² In 2024, NCMEC received over 500,000 reports of sextortion scams, with more than 100,000 reports involving AI-generated content in 2025.³³

CURRENT POLICY INTERVENTIONS AND LIMITATIONS

Despite recent legislative efforts and the bipartisan nature of the issue, regulation of AI NCII and CSAM remains fragmented and insufficient to provide robust guardrails. Federal legislation lacks a holistic ecosystem perspective that holds all actors accountable for content proliferation, resulting in several unaddressed gaps and limited effectiveness.

Federal Legislation: Signed into law in May 2025, the Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act, also known as the TAKE IT DOWN Act (TIDA), is the newest comprehensive standalone law against AI NCII and CSAM that has been enacted (as of July 2026). However, TIDA’s notice-and-takedown mandates only apply to intermediary platforms hosting user-generated content, instead of services that directly publish content. Another significant limitation of TIDA’s coverage is its exclusion of fully synthetic content.³⁴ Restricting statutory protections to only content that involves real identifiable individuals leaves a significant regulatory gap. The effects of this gap are reflected in an August 2026 case, where a federal appeals court was bound by a 2002 Supreme Court precedent to protect in-home possession of AI CSAM depicting fictional children on First Amendment grounds because no real children were involved.³⁵ Additionally, for deepfake depictions of adults, prosecutors have to either demonstrate that the creator intended to cause harm or that the publication of materials actually caused financial, psychological, or reputa-

31 Ciardha, Caoilte Ó, John Buckley, and Rebecca S. Portnoff. 2025. “AI Generated Child Sexual Abuse Material — What’s the Harm?” arXiv. doi:10.48550/arXiv.2510.02978; Struckman, Kara. 2025. “Combating AI-Generated CSAM.” *Wilson Center*. February 5. <https://chinafellowship.wilsoncenter.org/article/combating-ai-generated-csam>.

32 Hale, Rachel. 2025. “These Teenage Boys Were Blackmailed Online — and It Cost Them Their Lives.” *USA TODAY*. February 25. <https://www.usatoday.com/story/life/health-wellness/2025/02/25/teenage-boys-mental-health-suicide-sextortion-scams/78258882007/>.

33 Valdes, Nicole, and Kerry Breen. 2025. “A Teen Died after Being Blackmailed with A.I.-Generated Nudes. His Family Is Fighting for Change.” *CBS News*. May 31. <https://www.cbsnews.com/news/sextortion-generative-ai-scam-elijah-heacock-take-it-down-act/>.

34 Branum, Becca. 2024. “NDII Victims Deserve Help. Let’s Build an Effective Takedown System.” *Center for Democracy and Technology*. November 12. <https://cdt.org/insights/ndii-victims-deserve-help-lets-build-an-effective-takedown-system/>.

35 *United States v. Anderregg*. 2026. Seventh Circuit.

tional harm, which adds a burden to survivors and is often difficult to prove.³⁶ One study found that TIDA did not suppress NCII activity and even increased migration of content to other platforms.³⁷ Other federal laws have been introduced that could fill in some of the existing gaps, including the Disrupt Explicit Forged Images and Non-Consensual Edits Act (DEFIANCE Act), which would expand the existing federal civil cause of action to include synthetic NCII.³⁸ The Stop CSAM Act, introduced in 2025, would strengthen response to child sexual exploitation by enhancing NCMEC's CyberTipline reporting requirements, including whether the content was modified by AI.³⁹

State Legislation: Regulations at the state level have faced several legal challenges. Despite an acceleration of legislative activity, state policymakers have struggled to determine which measures to prioritize due to a lack of understanding of the technological ecosystem. In 2025, most of the bills introduced in state legislatures to regulate AI NCII and CSAM died in committee.⁴⁰ One reason many bills have failed to advance is that deepfake legislation is often subject to strict scrutiny under the First Amendment, especially when the government attempts to restrict speech based on the content.⁴¹ Legal protections on platforms present another obstacle: Section 230 protects online platforms from being held legally responsible for third-party content, largely insulating platforms from liability risks they would otherwise face as “publishers” or “speakers” of third-party content.⁴² Thus, bills that require takedowns or content moderation for platforms often face legal challenges.

As of July 2026, 46 states have enacted laws that criminalize AI CSAM.⁴³ The terminology and enforcement mechanisms in AI NCII and CSAM laws vary widely across states, presenting another major legislative challenge. For example, while New York chose to amend existing NCII

36 Sen. Cruz, Ted. 2025. *S.146 - 119th Congress (2025-2026): TAKE IT DOWN Act*. <https://www.congress.gov/bill/119th-congress/senate-bill/146>.

37 Cuevas, Alejandro, and Manoel Horta Ribeiro. 2026. “Deepfake Pornography Is Resilient to Regulatory and Platform Shocks.” arXiv. doi:10.48550/arXiv.2602.02754.

38 “H.R.3562 - 119th Congress (2025-2026): DEFIANCE Act of 2025.” 2025. Legislation. Congress. 2025-05-21. May 21. <https://www.congress.gov/bill/119th-congress/house-bill/3562>.

39 “Text - S.1829 - 119th Congress (2025-2026): STOP CSAM Act of 2025.” 2025. Congress. June 26. <https://www.congress.gov/bill/119th-congress/senate-bill/1829/text>.

40 S. Dawson, Gregory, Kevin C. Desouza, James S. Denford, and Marc Barda Picavet. 2025. “How Different States Are Approaching AI.” *Brookings*. August 18. <https://www.brookings.edu/articles/how-different-states-are-approaching-ai/>.

41 Osman, Hisham. 2025. “Required Deepfake Labeling: Contradicting or Supporting The First Amendment?” *Columbia Undergraduate Law Review*. December 6. <https://www.culawreview.org/journal/required-deepfake-labeling-contradicting-or-supporting-the-first-amendment>.

42 “Beyond Section 230: Principles for AI Governance.” 2025. *Harvard Law Review*. April 10. <https://harvardlawreview.org/print/vol-138/beyond-section-230-principles-for-ai-governance/>.

43 “State Laws Criminalizing AI-Generated or Computer-Edited CSAM.” 2026. *Enough Abuse*. <https://enoughabuse.org/get-vocal/laws-by-state/state-laws-criminalizing-ai-generated-or-computer-edited-child-sexual-abuse-material-csam/>.

and CSAM laws to cover AI-generated content, California and New Hampshire have established new laws that classify NCII as an invasion of privacy and a form of fraud.⁴⁴

Future legislation should prioritize mitigating potential legal barriers for victims. Law enforcement faces major legal barriers due to a lack of available tools to determine whether content is AI-generated, whether depictions of sexual violence are real, and who should be responsible for determining whether content is AI-generated.⁴⁵ For victims, tracking the perpetrators is a difficult task that requires expensive technologies and financial resources.⁴⁶ Additionally, many deepfake laws require proving intent to harm, which can be challenging to establish when perpetrators can claim self-gratification as the objective.

In short, AI has amplified the harms caused by NCII and CSAM by reducing the barriers to entry, facilitating large-scale production and circulation of abusive content, and fostering a profitable industry. This has led to disproportionate negative impacts on women and children. Current policy regulations remain fragmented and insufficient because they fail to tackle these issues from an ecosystem approach. To fill the gap, this paper proposes policy interventions addressing the entire AI NCII and CSAM value chain.

44 Williams, Kaylee. 2024. "US States Struggle to Define 'Deepfakes' and Related Terms as Technically Complex Legislation Proliferates." *Tech Policy Press*. September 12. <https://www.techpolicy.press/us-states-struggle-to-define-deepfakes-and-related-terms-as-technically-complex-legislation-proliferates/>.

45 Struckman, Kara. 2025. "Combating AI-Generated CSAM." *Wilson Center*. February 5. <https://chinafellowship.wilsoncenter.org/article/combating-ai-generated-csam>; Pfefferkorn, Riana, Shelby Grossman, Sunny Liu, and Jeffrey Hancock. 2025. "AI-Generated Child Sexual Abuse Material: Insights from Educators, Platforms, Law Enforcement, Legislators, and Victims." doi:10.25740/mn692xc5736.

46 Pirius, Rebecca. 2025. "Is Deepfake Pornography Illegal?" *Criminal Defense Lawyer*. December 2. <https://www.criminaldefenselawyer.com/resources/is-deepfake-pornography-illegal.html>.

Methodology

Understanding the current U.S. regulatory landscape is critical to designing effective policy interventions. This paper aggregated 28 standalone federal and state laws addressing AI NCII and CSAM and evaluated each against a set of criteria. Laws already enacted in California, New York, and Texas were examined because these states are leaders in regulating deepfakes, and each has adopted a distinct regulatory model.⁴⁷ To ensure the proposed policy recommendations can apply to smaller states, laws in Utah were included in the analysis to represent a middle ground of how most states can regulate AI.⁴⁸

After gathering a list of AI NCII and CSAM laws, we identified a set of parameters to assess the effectiveness of interventions and what might be lacking. Development of these criteria was based on a comprehensive analysis of legislative texts and desk research, then validated through semi-structured interviews with a diverse set of experts. Interviewees were selected for their expertise in deepfakes and online safety issues, and represented a range of perspectives from academia, industry, and advocacy organizations.

This paper focuses on enacted laws, and does not evaluate pending or failed legislation. Future research could include deeper analyses of the barriers that cause bills to die in committee. Finally, the policy recommendations in this paper were designed based on existing knowledge of threats and harms associated with AI NCII and CSAM. However, the rapidly changing nature of the policy landscape and technological development presents a major challenge to ensuring the proposed policy recommendations in this paper remain useful and effective in addressing future novel risks.

47 “Forty-Seven States Have Enacted Deepfake Legislation since 2019.” 2025. *Ballotpedia*. July 22. <https://news.ballotpedia.org/2025/07/22/forty-seven-states-have-enacted-deepfake-legislation-since-2019/>

48 “Utah’s Moderate Approach to AI Regulation.” 2025. *MultiState*. March 15. <https://www.multistate.ai/updates/vol-18>.

Findings

ANALYSIS OF THE AI NCII AND CSAM ECOSYSTEM

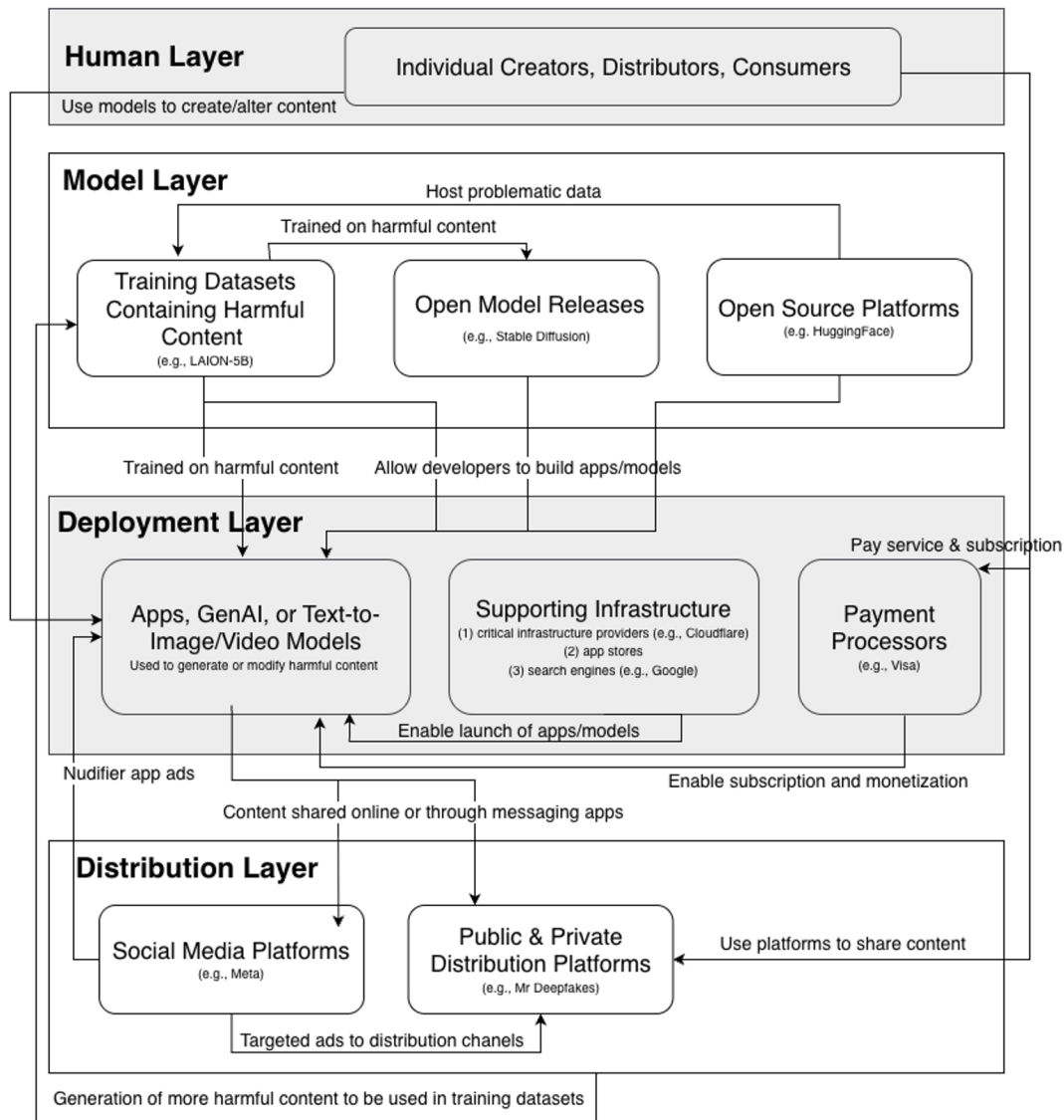
Effective policy interventions should situate the problem of AI NCII and CSAM within a larger ecosystem to address the range of technologies that facilitate the proliferation of abusive content. Building on existing literature and expert interviews, Figure 1 breaks down the ecosystem value chain. Technologies and actors in the ecosystem represented in Figure 1 are adapted from work by Ding et al. 2026, which provides a taxonomy of 11 categories of technologies that play a role in facilitating AIG-NCII.⁴⁹ Building on this taxonomy, we draw on the three-layer AI regulation framework introduced by Bailey and Barrington (2026),⁵⁰ comprising the Model, Deployment, and Distribution components of AI regulation. In particular, we adapt each layer to the AI CSAM and NCII regulatory context and add an additional “human” layer to incorporate individual creators and distributors. Finally, we map the proportion of enacted legislation onto each node, as shown in Figure 2.

The four layers in the deepfake ecosystem include: **(1) human, (2) model, (3) product deployment, and (4) distribution.** The *human* layer refers to individual creators and distributors of AI NCII and CSAM. The *model* layer enables the development of AI models and apps, including training datasets that contain harmful content (e.g., LAION-5B), open model releases (e.g., Stable Diffusion), and open-source platforms that host harmful training datasets (e.g., Hugging Face). The *product deployment* layer represents the software environment and user interface through which a model or app is accessed by end users (e.g., a web browser or app). Key technologies in this layer include generative AI or text-to-image/video models, payment processors, and supporting infrastructure, which includes technologies that enable the commercialization of apps/models for the public: infrastructure providers (e.g., Cloudflare, Google sign-in account), app stores, and search engines (e.g., Google). Finally, the *distribution layer* includes all technologies that enable the distribution or circulation of AI NCII and CSAM content, including social media platforms as well as public and private platforms. Figure 1 illustrates the AI NCII and CSAM value chain and the relationships among technologies.

49 Ding, Suresh, and Venkatasubramanian. “How to Stop Playing Whack-a-Mole.”

50 The figure presents an adapted version of the regulatory decomposition framework introduced by Bailey and Barrington 2026, under submission.

Figure 1. Breakdown of the AI NCII & CSAM Value Chain



The ecosystem starts with **training datasets** in the model layer that contain harmful content (e.g., LAION-5B, Common Crawl), which can be hosted on **open-source AI platforms** (e.g., Hugging Face, Civitai), making datasets available for malicious actors to develop models or apps that are primarily used for creating AI CSAM and NCII. One implication of this structure is that effective policy solutions should start by addressing the earliest stages of the model

lifecycle, which includes ensuring that the training data does not contain harmful content.⁵¹ **AI labs** (e.g., Stability AI, Black Forest Labs) that release open-weight models (e.g., Stable Diffusion, Flux) also play an important role in enabling nudifier developers. Making these weights publicly available creates safety vulnerabilities, allowing bad actors to download, modify, and fine-tune models.⁵²

In the deployment layer, **supporting launch infrastructure** encompasses technologies that enable the commercialization of apps/models for the public, including critical infrastructure providers (e.g., Cloudflare, Google account), app stores that host the nudifier apps, and search engines that help individuals to find and access AI CSAM and NCII websites. Additionally, **payment processors** (e.g., Visa, Apple Pay) are essential actors that allow individuals to make payments to for-profit nudifier apps that let them generate or modify content. **Apps, generative AI (GenAI), and text-to-image models** (e.g., Grok) are tools that enable users to create and modify AI NCII and CSAM content.

Once a product is launched, **social media companies** in the distribution layer host advertisements for nudifier apps, promoting access for the general public. For example, Indicator found that Meta promoted over 9,000 ads for 11 different nudify tools on Facebook and Instagram between September and December 2025.⁵³ After content is created using models, an individual may keep it for their own use, choose to distribute it on **public platforms** (e.g., Mr. Deepfakes), or share it privately through **messaging apps** (e.g., WhatsApp), facilitating the circulation of harmful content.

Figure 2 maps 28 enacted federal and state interventions onto each node in the value chain, highlighting regulatory gaps and identifying critical opportunities for policy action. As shown in the figure, legislative efforts are overwhelmingly concentrated on the human layer (78%), primarily targeting individual creators and distributors, followed by the deployment layer (33%) and distribution layer (26%). In contrast, none (0%) of the regulations analyzed in this paper addresses nodes in the model layer. This highlights an urgent need for increased regulatory

51 Researchers at Stanford found that multiple prominent machine learning training datasets contain CSAM, including Large-scale Artificial Intelligence Open Network's LAION-5B. It contained harmful content because it scraped the open web indiscriminately. See <https://doi.org/10.25740/kh752sm9123>.

52 For example, malicious actors can use the low-rank adaptation (LoRA) fine-tuning technique to create malicious content easily with additional layers that modify the behavior of a base model. See <https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/>

53 Mantzarlis, Alexios. 2025. "Nonconsensual Nude Generators Had Another Banner Year. What Will It Take to Defeat Them?" *Indicator*. December 4. <https://indicator.media/p/nonconsensual-nude-generators-had-another-banner-year-what-will-it-take-to-defeat-them>.

scrutiny over the deployment, distribution, and model layers that enable the sharing of AI CSAM and NCII content.

Figure 2. Mapping of Legislation to the AI NCII and AI CSAM Value Chain

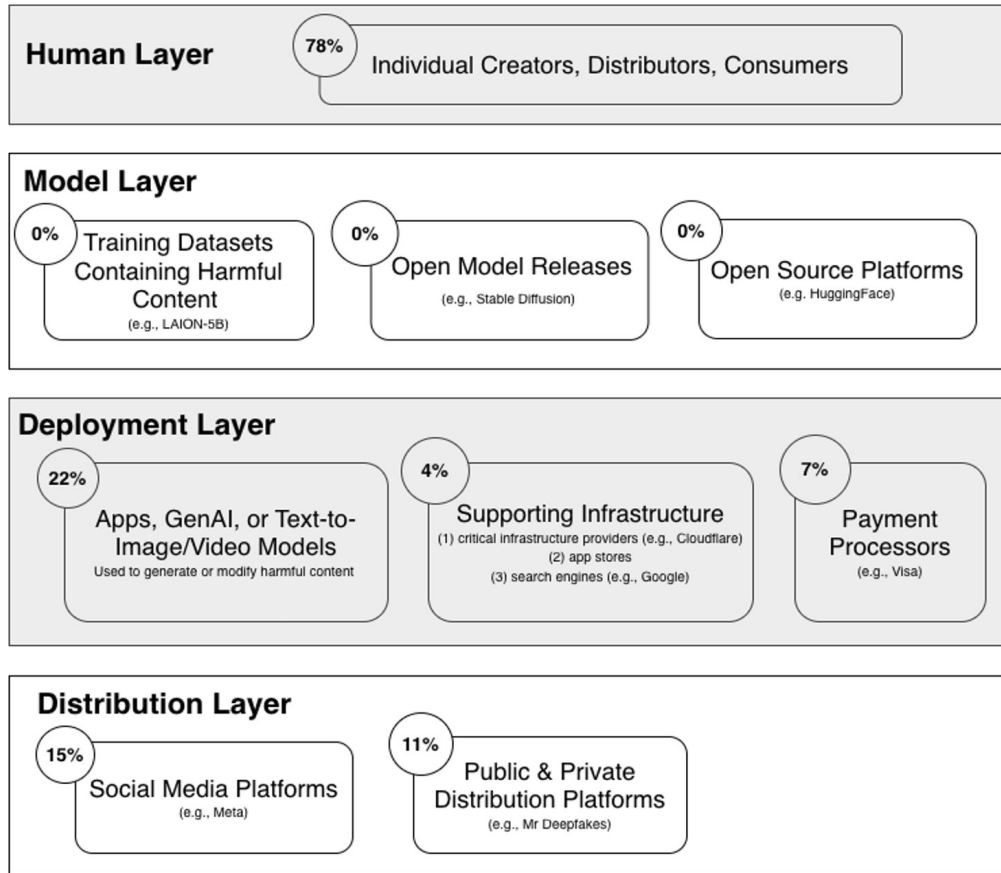
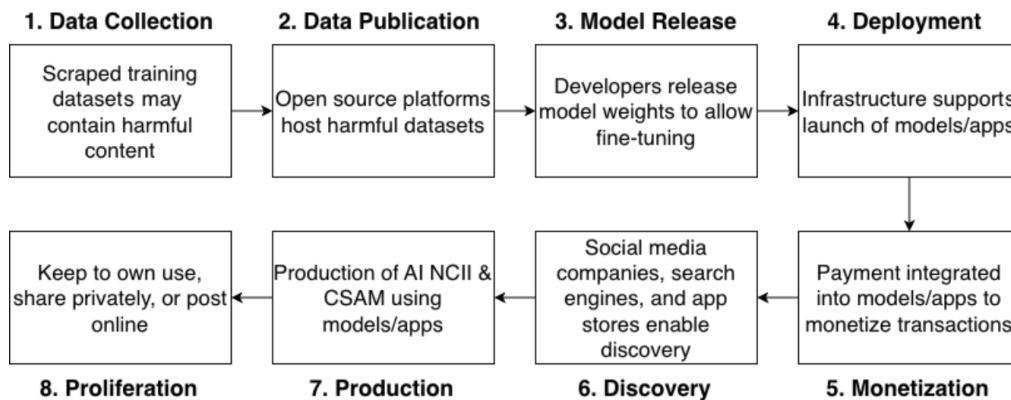


Figure 3 shows the typical lifecycle for AI NCII and CSAM, illustrating a pipeline of harm starting from data collection and curation that indiscriminately scrapes nonconsensual and toxic content from websites (Stages 1–2). AI labs and developers then train models on these datasets and publicly release model weights, allowing malicious actors to download and fine-tune architectures (Stage 3). Supported by infrastructure providers and payment processors that facilitate monetization, developers build and launch commercial nudify tools (Stages 4–5). Social media companies, search engines, and app stores drive user discovery (Stage 6). Finally, users can use these tools to generate and modify AI NCII and CSAM, which is either retained for personal use, shared in private messaging channels, or published on public forums (Stages 7–8).

Figure 3. AI NCII & CSAM Lifecycle



POLICY ANALYSIS

This section presents an overview of all laws enacted at the federal and state levels (in California, New York, Texas, and Utah) that address AI NCII and CSAM, as well as their implications for policy recommendations. A set of criteria was developed to evaluate the effectiveness of each law. Based on expert consultation and legislative text analysis, the following six parameters were used to evaluate the laws:

1. **Criminal or Civil: Is a case brought by the government to punish an offender (criminal), or brought by a private party to compensate the victim (civil)?** If a law is criminal, prosecutors bring the case on behalf of the state instead of representing the victim. Penalties can include fines and jail time. In contrast, some civil laws allow victims to sue their perpetrators directly. In these circumstances, civil penalties can be more effective than criminal penalties as they grant victims the ability to bring their own cases to court. However, civil penalties are limited to financial sanctions.
2. **Liability: Who can be sued or prosecuted?** Experts interviewed for this paper emphasized the importance of holding “hidden actors” (e.g., critical infrastructure providers) accountable for the proliferation of AI CSAM and NCII content. Identifying which actor is primarily prosecuted for violations of a law can provide insights into which technologies remain underregulated.
3. **New or Retrofit: Is new legislation created or amended based on existing law?** States tend to diverge on the models between new and statutory retrofit, making this parameter important to track and compare across states. Expanding existing NCII and CSAM laws to cover AI-generated or modified content is potentially faster, faces less

political risk, and builds on established precedent and enforcement infrastructure. At the same time, given that AI has significantly transformed the traditional safety landscape, new standalone laws are essential to move beyond current reactive protections and fill in loopholes, despite potential higher implementation costs, legal challenges, and political risks.

4. **Requiring Depiction of a Real Human: Does the law require proving that the content generated is indistinguishable from an identifiable real human, and/or does the law still apply even when the content is fully synthetic?** This criterion is important because proving that AI-manipulated content is indistinguishable from an actual person creates evidentiary barriers for victims and prosecutors.⁵⁴ For example, prosecutors may have to establish forensic evidence demonstrating how an image was generated and that it is sufficiently “indistinguishable” from an identifiable child to satisfy statutory standards, which can be technically difficult to determine.⁵⁵
5. **Requiring Intent to Harm: Does the law require the victim to prove that the offender intended to cause harm?** This parameter assesses whether a law imposes the burden of proof on victims, requiring that they prove the perpetrator had intent to harm, which is often difficult and easily contested.
6. **Requiring Intent to Distribute: Does the law require proving that the offender intended to distribute or upload content?** If a law requires proof of intent to distribute AI NCII or AI CSAM, it creates a loophole by not addressing content created for the perpetrator’s own usage. Experts highlighted that some AI CSAM and NCII content is created for personal use, not for distribution on public platforms.⁵⁶

TAKE IT DOWN Act

As of July 2026, the TAKE IT DOWN Act (TIDA) is the only enacted federal law that addresses both AI CSAM and NCII. This law represents the first major legislation to criminalize authentic and AI-generated NCII and CSAM content, treating them as legally equivalent. A key provision designates the Federal Trade Commission (FTC) to be in charge of requiring platforms to establish a reporting infrastructure and remove reported content within 48 hours, or

54 “Image-Based Sexual Abuse Laws: Combat Nonconsensual AI Deepfakes.” 2025. *RA/INN*. June 23. <https://rainn.org/rainns-recommendations-for-legislators/image-based-sexual-abuse-laws-combat-nonconsensual-ai-deepfakes/>.

55 Bechrakis, Efthimia Maria. 2026. “When the Image Is Fake but the Harm Is Real: Children’s Rights in the Age of AI-Generated Sexual Exploitation.” *American Bar Association*. June 4. <https://www.americanbar.org/groups/litigation/resources/newsletters/childrens-rights/when-image-is-fake-harm-real/>.

56 Authors’ stakeholder interviews 2026.

face penalties.⁵⁷ The FTC began enforcing the law in May 2026 and established the website TakeItDown.ftc.gov to allow victims to report platforms that fail to address NCII and CSAM removal requests within 48 hours.⁵⁸

A major limitation of TIDA is its reactive approach, as the law only addresses content that has already been created. Interviewees indicated that the law primarily targets public distribution channels through which users upload and share harmful materials, and largely fails to prevent content generation.⁵⁹ Mandating platforms to take down harmful content alone is insufficient to address the issue because AI NCII and CSAM can easily be migrated to other websites when one shuts down. Furthermore, TIDA requires that materials subject to prosecution depict “identifiable individuals,” which can be reliably interpreted as real adults and children (and not AI-generated images of people who do not exist). TIDA also requires proof that perpetrators intended to distribute content, which imposes a legal burden on victims.

State AI NCII and CSAM Laws: California, New York, Texas, and Utah

In California, New York, Texas, and Utah, a total of 27 laws have been enacted to address AI NCII or AI CSAM (as of July 2026). Table 2 in the Appendix provides a summary of these statutes.

Across all 27 laws, there is an even split between AI CSAM (nine laws) and AI NCII (nine laws). The remaining nine laws target both categories of content, showing that states are in some cases responding to the problem as a threat of sexual violence to both minors and adults. After analyzing each law based on a set of parameters, a pattern emerges: there is an **even split between criminal and civil laws**, and most enacted laws are statutory retrofits. There is also an overwhelming legislative focus on individual creators and distributors, and many laws place a high evidentiary burden on victims. As a result, the laws systematically underprotect against harm from the technologies that facilitate the creation and distribution of AI CSAM and NCII.⁶⁰

In California and New York, legislation is distributed almost equally between criminal and civil frameworks. This hybrid approach is effective in criminalizing AI NCII and CSAM while also

57 In April 2026, the U.S. Attorney’s Office for the Southern District of Ohio sentenced a man who had installed over 24 AI platforms and 100 AI web-based models to harass victims, representing the first national conviction under this Act: <https://www.justice.gov/usao-sdoh/pr/columbus-man-pleads-guilty-cyberstalking-exes-creating-ai-generated-obscene-material>

58 “FTC Begins Enforcing the TAKE IT DOWN Act.” 2026. *Federal Trade Commission*. May 18. <https://www.ftc.gov/news-events/news/press-releases/2026/05/ftc-begins-enforcing-take-it-down-act>.

59 Authors’ stakeholder interviews 2026.

60 For a complete evaluation of legislation based on different key parameters, see Table 3 in the Appendix.

providing victims a private right of action to seek financial compensation and demand the removal of content.

	California	New York	Texas	Utah	Total
Criminal Laws	5/9 laws (56%)	1/2 laws (50%)	5/8 laws (63%)	6/8 laws (75%)	17/27 laws (63%)

Most enacted AI NCII and CSAM laws are statutory retrofits, especially in New York (two out of two laws) and Utah (seven out of eight laws). New and retrofit laws both have their own advantages. Expanding current laws to include AI-generated or modified content tends to face less political friction and administrative burden for law enforcement to extend established legal requirements. On the other hand, new statutes can address underregulated nodes and mitigate emerging challenges as tools evolve. Effective regulation requires a state to enact new standalone legislation while also extending existing laws.

	California	New York	Texas	Utah	Total
Retrofit Laws	6/9 laws (67%)	2/2 laws (100%)	4/8 laws (50%)	7/8 laws (88%)	19/27 laws (70%)

Federal laws and most state-level interventions fail to address the different technologies and actors involved in the AI NCII and CSAM ecosystem. The following table shows that the **majority of laws** (21 out of 27 laws) **target individual creators and distributors** and fail to hold other actors (e.g., social media companies) accountable. Targeting individual actors may have higher political feasibility and lower administrative burden, but the impact on mitigating AI NCII and CSAM is limited. On the other hand, some states (e.g., California’s AB 621) have taken action to fill this gap by creating liability for third parties, payment processors, and social media platforms.

	California	New York	Texas	Utah	Total
Laws Targeting Individual Creators and Distributors	7/9 laws (78%)	2/2 laws (100%)	5/8 laws (63%)	7/8 laws (88%)	21/27 laws (78%)

Since **most enacted laws** (19 out of 27) **require AI NCII and CSAM to be indistinguishable from an actual, identifiable human**, victims and prosecutors bear the evidentiary burden of authenticating digital evidence. For example, when content is heavily edited or entirely AI-generated, traditional detection methods often fail to identify hyperrealistic AI CSAM,

making it technically difficult to establish that a real child is depicted.⁶¹ Compared to AI NCII laws, most AI CSAM laws do not have this requirement, suggesting that legal protections for children are stronger than for adults.

	California	New York	Texas	Utah	Total
Laws Requiring Depiction of Real Human	6/9 laws (67%)	2/2 laws (100%)	5/8 laws (63%)	6/8 laws (67%)	19/27 laws (70%)

Nine out of 27 laws require establishing intent to cause harm, which inevitably provides a defense for perpetrators who may claim personal gratification as their primary motivation for generating sexually explicit content. Notably, **only a few statutes require proving intent to harm**, thereby removing such a barrier.

	California	New York	Texas	Utah	Total
Laws Requiring Intent to Harm	3/9 laws (33%)	1/2 laws (50%)	2/8 laws (25%)	3/8 laws (38%)	9/27 laws (33%)

Proving that creators intended to distribute materials is often difficult. Many stakeholders emphasized the important **distinction between creators and distributors**: not all creators upload content online, but rather use materials themselves or share them privately. Eliminating the requirement to prove intent to distribute allows the law to penalize the upstream creation phase rather than merely targeting downstream distribution. There is a significant variation among states in the number of laws that require demonstrating intent to distribute. Texas leads in this parameter, as none of the state's enacted laws requires proving intent to distribute.

	California	New York	Texas	Utah	Total
Laws Requiring Intent to Distribute	6/9 laws (67%)	2/2 laws (100%)	0/8 laws (0%)	5/8 laws (63%)	13/27 laws (48%)

61 “AI CSAM: The Growing Digital Threat to Children.” 2025. *Our Rescue*. December 5. <https://ourrescue.org/resources/child-exploitation/csam/ai-csam>.

Policy Recommendations

To effectively combat AI NCII and CSAM, both ex-ante and ex-post approaches are essential. While ex-ante solutions focus on preventing the initial production and dissemination of harmful content on the internet, ex-post interventions aim to contain, mitigate, and remove materials that have already been produced or circulated to avoid further proliferation.

The recommendations proposed in this paper prioritize the deployment layer, followed by the distribution layer, and lastly the model layer.⁶² Product deployment has the most opportunity for intervention, as it has potential to regulate a range of powerful actors in the industry (e.g., app stores, payment processors).⁶³ Intervening at this layer can be more effective by targeting the ecosystem where harmful content is generated and commercialized, preventing abuse before it spreads. Nevertheless, regulating social media companies and other distribution platforms remains a crucial tool to curb further proliferation of AI NCII and CSAM after they have been produced and disseminated. Despite the current lack of regulation at the model layer, it is not recommended to prioritize this because of technological complexity and opacity, meaning even developers often cannot fully explain why a given input produces a given output. This interpretability problem — the difficulty of tracing which internal parameters generate specific outputs — poses significant challenges for policymakers and developers to intervene and implement safety guardrails.

Because states can function as incubators for new regulation by running pilot programs to shape legislation before it is scaled nationwide, it is recommended that policymakers and advocacy groups primarily focus on state policies.⁶⁴ States also play an important role in drafting and enacting bills tailored to their local needs. At the same time, federal oversight remains critical in streamlining regulations and reducing fragmentation across states, ensuring baseline protections across jurisdictions.

62 The report draws on the three-layer framework introduced by Bailey and Barrington (2026), comprising the Model, Deployment, and Distribution components of AI regulation.

63 Authors' stakeholder interviews 2026.

64 Tyler, Charles W., and Heather K. Gerken. 2022. "THE MYTH OF THE LABORATORIES OF DEMOCRACY." *Columbia Law Review*. <https://www.columbialawreview.org/content/the-myth-of-the-laboratories-of-democracy/>.

OVERARCHING RECOMMENDATION

Distribution Layer

Remove the requirement that AI CSAM content depict real children for it to be illegal.

The challenge of identifying AI-altered content presents a systemic risk to the traditional child safety landscape, especially for CSAM investigators who may struggle to distinguish between AI and authentic content.⁶⁵ Therefore, the requirement of proving that AI-manipulated depictions are indistinguishable from real children creates a significant loophole for CSAM creation and distribution, as it allows content to be proliferated as long as it is “not realistic enough.” Similarly, coverage being limited to CSAM depicting identifiable individuals fails to account for the ways AI advancements have complicated traditional legal boundaries, which is reflected in the August 2026 Seventh Circuit ruling that exempted the in-home possession of AI CSAM because no real children were involved. To strengthen existing policy measures, legislation should remove such requirements.

At the federal level, TIDA requires that materials depict real, identifiable humans to be considered illegal, posing an evidentiary burden of authenticating digital evidence. However, at the state level, several laws in California and Texas have eliminated this prerequisite. California’s AB 1831 extends the existing law that prohibits the creation and distribution of depictions of children under 18 to cover content altered or generated by AI, including synthetic materials.⁶⁶ Similarly, Texas SB 1621 closed the “no real child” loophole by expanding the offenses of transmission and possession of AI CSAM materials.⁶⁷

FEDERAL-LEVEL RECOMMENDATIONS

Deployment Layer

1. Impose liability on platforms that host harmful models or training datasets.

Many open-source platforms host models and training datasets that contain CSAM and NCII materials, allowing malicious actors to build models designed for nudification.⁶⁸

65 Levine, Alexandra S., Reily Griffin, and Kurt Wagner. 2026. “AI Has Emboldened Child Predators, and Investigators Can’t Keep Up.” *Bloomberg*. April 22. <https://www.bloomberg.com/features/2026-ai-child-predators-law-enforcement/>; Authors’ stakeholder interviews 2026.

66 “AB-1831 Crimes: Child Pornography.” 2024. *California Legislative Information*. September 30. https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB1831.

67 “Texas Senate Bill 1621.” 2026. *LegiScan*. <https://legiscan.com/TX/text/SB1621/id/3140545>.

68 Authors’ stakeholder interviews 2026.

Between 2022 and 2025, over 34,000 deepfake model variants were available on popular repositories, and these were downloaded almost 15 million times.⁶⁹ Since these platforms play a crucial role in facilitating the production of AI CSAM and NCII, it is essential to establish liability and penalties for failing to adopt practices designed to limit such materials. For example, platforms should be required to implement a mechanism for the public to submit reports and establish response procedures, including processes for reviewing submissions and promptly taking down models or training datasets that contain or facilitate the creation of harmful content.

Australia's eSafety Commissioner successfully forced Hugging Face, a platform hosting open-source models and datasets, to comply with the Australia Online Safety Act by updating its terms of service to require all uploaded models to mitigate risks related to CSAM.⁷⁰ If a platform fails to take appropriate enforcement action when users violate the terms, it faces potential penalties, including fines up to 49.5 million Australian dollars.⁷¹

Distribution Layer

2. **Eliminate the requirement that perpetrators must have intent to distribute.**

A major gap of TIDA is that it does not address perpetrators who create content for personal use, but rather focuses on those who distribute materials on public channels. TIDA also requires that intent be established at the time when the content was produced, which limits prosecution when intent develops or changes after initial production.⁷² To address this, future federal statutes should eliminate intent to distribute as a prerequisite for liability.

The Enhancing Necessary Federal Offenses Regarding Child Exploitation (ENFORCE) Act, introduced in August 2025 and passed in the Senate in December 2025, would close this loophole by criminalizing the production of AI CSAM regardless of intent to distribute, while ensuring that offenders face consistent consequences.⁷³ Passing this legislation would successfully close the loophole and strengthen protections for victims.

69 Stephan, Adrianna. 2025. "A Weapon Against Women in Politics: Reining in Nonconsensual Synthetic Intimate Imagery." *New America*. September 18. <https://www.newamerica.org/insights/a-weapon-against-women-in-politics/>.

70 "eSafety Action Prevents Services 'nudging' Australian School Children." 2025. eSafety Commissioner. November 27. <https://www.esafety.gov.au/newsroom/media-releases/esafety-action-prevents-services-nudging-australian-school-children>.

71 Werner, Jeremy. 2025. "eSafety Forces Major 'Nudify' Provider Out of Australia as Crackdown on AI-Generated Child Exploitation Expands." *BABL AI*. December 10. <https://babl.ai/esafety-forces-major-nudify-provider-out-of-australia-as-crackdown-on-ai-generated-child-exploitation-expands/>.

72 Authors' stakeholder interviews 2026.

73 "Texas Senate Bill 1621." 2026. LegiScan. <https://legiscan.com/TX/text/SB1621/id/3140545>.

3. **Establish a private right of action to allow victims to sue perpetrators.**

TIDA consists of both criminal and civil components, as it criminalizes AI NCII and CSAM while requiring covered platforms to establish notice-and-removal mechanisms under the enforcement of the FTC. However, the law does not include a private right of action allowing a victim to file a lawsuit to seek compensation for harm. Criminal cases often present high barriers for victims navigating the court system and typically do not provide compensation for damages (e.g., job loss, mental health issues). Additionally, since the judicial process requires presentation of evidence, harmful content that victims do not want exposed is inevitably viewed by more people.

Therefore, a victim-centered civil law that guarantees the right of action is essential to fill these gaps and lower the barrier to entry for smaller cases.⁷⁴ If enacted, the DEFIANCE Act would serve as a victim-centered civil counterpart of TIDA by providing individuals the right of action to sue offenders who produce or distribute materials. The DEFIANCE Act was introduced in the House in May 2025 and unanimously passed the U.S. Senate in January 2026.

STATE-LEVEL RECOMMENDATIONS

Deployment Layer

1. **Ban advertisements of AI nudify apps on social media platforms.**

Social media platforms make millions of dollars each year promoting advertisements for nudify apps that provide tools to create sexually explicit materials without consent. Between November 2025 and January 2026, Meta removed 344,000 ads for nudifier apps across Facebook and Instagram.⁷⁵ However, just eight days following Meta's report, Indicator found the company ran over 900 new ads for 10 different nudifiers.⁷⁶ Because Meta and other social media platforms act as gateways for consumers to potentially use nudify apps, an outright ban on hosting advertisements for such services would effectively reduce users and thus the proliferation of abusive content.

With faster legislative cycles, states can pass legislation to limit nudification apps. For example, in May 2026, Minnesota passed HF1606, which bans AI nudification tech-

74 Authors' stakeholder interviews 2026.

75 "Integrity Reports, H1 2026." 2026. Meta. March 19. <https://transparency.meta.com/reports/integrity-reports-h1-2026/>.

76 Silverman, Craig. 2026. "Briefing: UK Gov Mulls an AI Label Mandate." *Indicator*. March 20. <https://indicator.media/p/briefing-uk-gov-mulls-an-ai-label-mandate>.

nology, including advertisements or promotions of deepfake websites, applications, software, or programs that can nudify images or videos.⁷⁷ Through this law, victims can file civil lawsuits to seek compensatory and punitive damages, injunctive relief, and attorney’s fees.

2. **Extend liability for AI NCII and CSAM to launch infrastructure providers.**

Based on the AI NCII and CSAM ecosystem value chain (Figure 1), launch infrastructure that supports the commercialization of apps or models includes critical infrastructure providers, such as cloud platforms and app stores. While their role is often under-recognized, these actors facilitate the monetization and proliferation of deepfake models and apps. Tech Transparency Project, a watchdog that investigates corporate accountability, found that the Apple and Google app stores each has hosted dozens of nudify apps and enabled 483 million downloads, thereby making more than \$122 million in revenue as of April 2026.⁷⁸ Following legal pressure and investigations, both companies said the search term “nudify” was disabled in their app stores.⁷⁹ According to NCMEC, Cloudflare is one of the main contributors to sexual abuse by providing internet infrastructure for prominent deepfake pornography sites.⁸⁰ These supporting infrastructure actors not only increase visibility to such apps but also support their profit generation.

Most existing AI CSAM and NCII legislation does not impose liability on these actors; only 4% of enacted laws in the four states analyzed provide pathways for prosecuting infrastructure providers that host harmful content. To fill this gap, states should design interventions that require each actor in the launch infrastructure to take appropriate mitigation actions. App stores should critically evaluate and prohibit nudifier apps from appearing on their platforms, search engines should block harmful websites that share or generate AI NCII or CSAM from showing up in search query results, and infrastructure providers should terminate services with platforms contributing to the creation and distribution of AI NCII and CSAM. Finally, financial institutions should block transactions for identified AI nudify services.

77 “MN HF1606.” 2026. *BillTrack50*. <https://www.billtrack50.com/billdetail/1846425>.

78 “Apple and Google Are Steering Users to Nudify Apps.” 2026. *Tech Transparency Project*. April 15. <https://www.techtransparencyproject.org/articles/apple-and-google-are-steering-users-to-nudify-apps>.

79 Wells, Georgia, and Rachel Wolfe. 2026. “AI Supercharges Deepfake Nudes—Unleashing a New Form of Bullying Among Kids.” *Wall Street Journal*, June 14, sec. Tech. <https://www.wsj.com/tech/ai-deepfake-nudes-bullying-school-d242b8d4>.

80 “Cloudflare.” 2026. *NCOSE*. <https://endsexualexploitation.org/cloudflare/>.

An example of a state-level law that imposes liability on these hidden actors is California AB 621, which holds liable platform operators, service providers, and intermediaries that knowingly or recklessly facilitate sexually explicit deepfake services.⁸¹ This effectively forces third-party providers to terminate their relationships with nudify platforms.

Mitigations at the financial layer represent another potential intervention. In August 2025, a bipartisan coalition of 47 state and territory attorneys general issued demands to leading technology and financial companies (e.g., Google, Visa, PayPal) to implement more robust measures against the proliferation of AI NCII, including requesting that payment platforms explain how they identify and remove payment authorization for AI NCII content.⁸² Following a *New York Times* report that accused PornHub, a pornography website, of hosting NCII and CSAM, both Visa and Mastercard blocked transactions to the platform.⁸³ Policymakers can expand this enforcement model to restrict payment-processing access for any websites that create, host, or distribute AI NCII and CSAM.

In implementation, laws aimed at regulating internet platforms might face free speech challenges due to Section 230 of the Communications Decency Act, a federal safe-harbor law that protects online intermediaries from liability to ensure free speech.⁸⁴ For example, Texas HB 581 includes a specific exemption that protects internet service providers, search engines, and cloud service providers from liability solely for providing access or connection to a website.⁸⁵

Distribution Layer

3. **Require companies to deplatform harmful materials within a specific time frame.**

AI has amplified existing risks and harms of NCII and CSAM by increasing access to tools used to generate materials, lowering technical barriers, and making synthetic

81 “AB 621: Deepfake Pornography.” 2025. *CalMatters*. October 13. https://calmatters.digitaldemocracy.org/bills/ca_202520260ab621.

82 Jeter, Lisa. 2025. “State and Territory Attorneys General Urge Tech and Payment Platforms to Address Deepfake Exploitation.” *National Association of Attorneys General*. August 26. <https://www.naag.org/press-releases/state-and-territory-attorneys-general-urge-tech-and-payment-platforms-to-address-deepfake-exploitation/>.

83 Nelson, Eshe. 2020. “Visa and Mastercard to Investigate Financial Ties to Pornhub.” *The New York Times*, December 7, sec. Business. <https://www.nytimes.com/2020/12/07/business/visa-mastercard-pornhub.html>.

84 Waheed, Noor. 2024. “Section 230 and Its Applicability to Generative AI: A Legal Analysis.” *Center for Democracy and Technology*. September 4. <https://cdt.org/insights/section-230-and-its-applicability-to-generative-ai-a-legal-analysis/>.

85 “HB 581 - 89th Legislature.” 2026. *Texas Policy Research*. <https://www.texaspolicyresearch.com/bills/89th-legislature-hb-581/>.

content more realistic. Research by the Internet Watch Foundation found that the number of AI-generated child sexual abuse videos increased by 26,385% from 2024 to 2025.⁸⁶ Even though the majority of deepfakes appear on public websites,⁸⁷ our analysis suggests that only 11% of state laws require the removal of content from these platforms, and 15% specifically require removal from social media platforms.

To control the proliferation of these materials, states that do not currently have take-down laws should establish mandatory removal of AI NCII and CSAM from platforms that host the content, especially social media platforms and deepfake sites. Once a platform receives a verifiable report that these materials are hosted there, the company should be required to respond and remove the material promptly (e.g., within 48 hours).

States should also mandate that platforms take down content that facilitates malicious actors from practicing sextortion. Major platforms, such as Telegram and YouTube, have circulated instructional manuals, often titled “sextortion guides,” that train people how to exploit children, with advertisements offering specialized coaching for additional payment.⁸⁸ To combat these increasingly professionalized crimes, states should require platforms to block and remove such content promptly. For example, California AB 981 and Utah HB 276 both require platforms to establish notice and takedown procedures for victims, including by providing reporting mechanisms and temporarily block or remove reported AI-generated explicit content within statutory windows.

As part of these mandates, regulators should ensure that platforms implement trauma-informed reporting design. For example, platforms should improve the accessibility of reporting pathways by ensuring that submission forms or other information can be found within two clicks from the home page. In order to file a report, some platforms currently require a victim to view and re-engage with the abusive content, which can retrigger the traumatic experience. Such features should be updated to allow a trusted representative to report on a victim’s behalf.⁸⁹

86 “AI CSAM Report 2026: Harm Without Limits.” 2026. *Internet Watch Foundation*. <https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/>.

87 “2023 State Of Deepfakes: Realities, Threats, And Impact.” 2023. *Security Hero*. <https://www.securityhero.io/state-of-deepfakes/>.

88 Brooks, Libby. 2024. “Sextortion Guides and Manuals Found on Telegram and YouTube.” *The Guardian*, August 22, sec. UK news. <https://www.theguardian.com/uk-news/article/2024/aug/22/video-sextortion-guides-and-manuals-found-on-tiktok-and-youtube>.

89 Authors’ stakeholder interviews 2026.

Model Layer

4. Enhance data integrity and content provenance across the model lifecycle.

Implementing restrictions that limit users' capacity to generate or modify harmful content represents a proactive intervention, embedding safeguards directly into the product's architecture to prevent abuse before harm occurs. Experts interviewed emphasized the importance of making technical improvements to ensure models cannot generate harmful materials in the first place, which requires AI developers to build more effective safety designs.⁹⁰ This poses significant challenges, as traditional risk evaluation methods, such as red teaming and adversarial testing, cannot be applied to CSAM due to the legal prohibitions that prevent evaluators from generating or manipulating illicit material.⁹¹

Policymakers should enforce data integrity and content provenance obligations across both input and output sides of model development, requiring developers to verify that training datasets are free of illegal material and to document the origins of generated content. On the input side, states should mandate that all AI training data undergo standardized, pre-training cryptographic hashing, a process that turns images and videos into unique digital fingerprints that can be automatically matched against databases of known CSAM and NCII, allowing removal of harmful material before training occurs. This may include leveraging existing infrastructure such as Thorn's Safer platform, which provides API-based hashing-and-matching tools to scan content against databases of verified CSAM hash values.⁹² Public disclosures of dataset composition more broadly can also enhance transparency and allow regulators to assess compliance; for example, California's AB 2013 requires companies to publicize high-level summaries of how they source and filter training data.⁹³

On the output side, policy should mandate that AI developers comply with content provenance standards that allow tracing AI-generated content back to its source. AI has complicated the traditional CSAM investigation landscape, which relies on background landscapes and geographic clues to locate victims. With AI tools available for offenders to manipulate these forensic indicators, investigators face increasing challenges in iden-

⁹⁰ Ibid.

⁹¹ Lyu, Mengyao, Yuhong Yang, Haiwen Hong, Hui Chen, Xuan Jin, Yuan He, Hui Xue, Jungong Han, and Guiguang Ding. 2024. "One-Dimensional Adapter to Rule Them All: Concepts, Diffusion Models and Erasing Applications." In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7559–68. Seattle, WA, USA: IEEE. doi:[10.1109/CVPR52733.2024.00722](https://doi.org/10.1109/CVPR52733.2024.00722).

⁹² "CSAM Detection for Child Safety." 2026. Safer. <https://get.safer.io/csam-detection-tool-for-child-safety>.

⁹³ "AB-2013 Generative Artificial Intelligence: Training Data Transparency." 2024. September 30. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240AB2013.

tifying victims depicted in AI-manipulated content.⁹⁴ To address this, developers should be required to embed metadata and latent watermarks into model outputs to preserve traceability. For example, California’s SB 942 requires AI platforms with over one million monthly users to provide free AI-content detection tools that verify if content was created or modified by the provider’s systems and embed latent watermarks (i.e., hidden metadata that identify whether the content is artificial, even if the visible watermark is stripped away). Providers must also give users the option to add a visible manifest watermark.⁹⁵

Other

5. Promote AI literacy to combat harms in schools.

Experts interviewed noted that there has been a troubling demographic shift in the creation and distribution of AI NCII, as content is increasingly produced by younger creators.⁹⁶ Traditionally, CSAM and NCII have been created by adults. With the increased accessibility of digital tools and reduced barriers to entry, the younger generation can now use AI tools to create content depicting their peers and adults. For example, in 2023, students in a middle school in Indianapolis created deepfake nude content of their teacher, eventually leading the teacher to quit her job.⁹⁷

The increasing generation of harmful content by minors complicates the regulatory landscape and calls for improved education on deepfake technologies. This may include teaching and training students about the risks and harms associated with deepfake technologies, the importance of consent, and the disproportionate impact on women and children.

Policy measures could emulate California’s AB 873, an integrated curriculum model that requires the state to add media literacy to curriculum frameworks across existing academic subjects.⁹⁸ While the statute primarily aims to address AI-driven misinformation by adding media literacy to the curriculum, similar legislation can require K-12 schools to educate children about the risks and harms of AI NCII and CSAM.

94 Authors’ stakeholder interviews 2026.

95 “Senate Bill No. 942.” 2024. *California Legislative Information*. September 20. https://leginfo.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942.

96 Authors’ stakeholder interviews 2026.

97 Nickel, Dana. 2024. “AI Is Shockingly Good at Making Fake Nudes — and Causing Havoc in Schools.” May 28. <https://subscriber.politicopro.com/article/2024/05/artificial-intelligence-deepfake-nudes-schools-00160183>.

98 Jones, Carolyn. 2023. “Recognizing Fake News Now a Required Subject in California Schools.” *CalMatters*, November 10, sec. K-12 Education. <https://calmatters.org/education/k-12-education/2023/11/fake-news-california-school/>.

Conclusion

AI has transformed the creation and distribution of NCII and CSAM from a problem of individual bad actors into an industrialized ecosystem that spans multiple technologies and platforms. Most current laws have not been updated to reflect this new reality, however. This analysis mapped 28 enacted laws onto the NCII and CSAM value chain, revealing that enforcement is concentrated on creators and distributors, with a limited number of regulatory interventions targeting the platforms and infrastructure that enable the mass production, distribution, and monetization of content. The recommendations in this paper aim to help shift policy toward a more effective framework that spans the entire AI NCII and CSAM value chain.

Given the rapid development of technology and the current lack of understanding of the AI NCII and CSAM ecosystem, future research is needed to study how different technologies across the value chain interact, and how these compounding effects enable the proliferation of harmful content. Interdisciplinary collaboration among researchers, advocates, and industry is required to understand the evolving risk and harm landscape and design enduring policy solutions as the ecosystem transforms.

Appendix

Table 1: General Description and Key Parameters of the TAKE IT DOWN Act

Legislation Number	Date Introduced	Type	Title	Sponsor Party	Statute Summary	Criminal or Civil?	Liability	New or Retrofit?	Require Depiction of Real Human?	Require Intent to Harm?	Require Intent to Distribute?	Penalty
S146	2025-01	NCII and CSAM	TAKE IT DOWN Act	R	The first major federal response to AI NCII and CSAM. Criminalizes the publication of AI non-consensual intimate imagery. Upon receiving a take-down request, platforms must remove the content within 48 hours.	Criminal	Platforms, individual creator(s), and distributor(s)	New	Yes	Yes	Yes	Up to two years imprisonment for cases with adult victims and up to three years for cases involving minors, plus fines and forfeiture of property used to commit the crime

Table 2. General Description of AI CSAM & NCII Laws in California, New York, Texas, and Utah

Legislation Number	State	Date Introduced	Type	Title	Sponsor Party	Summary
AB 1831	California	2026-01	CSAM	Crimes: child pornography	D	Extends existing law that prohibits creation and distribution of depictions of individuals under 18 to include content that “is digitally altered or generated by the use of artificial intelligence.”
SB 926	California	2026-01	NCII and CSAM	Crimes: distribution of intimate images	D	Makes it a crime to “intentionally create and distribute...any photo realistic image, digital image, electronic image, computer image, computer-generated image, or other pictorial representation” of sexual intimate content that is indistinguishable from “an authentic image of the person depicted.” Covers both developer and user liability. Expands the revenge porn law to include AI-generated content.
AB 621	California	2025-02	NCII	Deepfake pornography	D	Expands monetary civil liability, and recasts and expands the civil code to include AI CSAM. Provides that a minor cannot consent to creation and distribution of deepfake porn.

CLOSING GAPS ACROSS THE ECOSYSTEM

Legislation Number	State	Date Introduced	Type	Title	Sponsor Party	Summary
SB 981	California	2024-01	NCII	Sexually explicit digital images	D	Requires social media platforms to establish a mechanism for users to report NCII. Once reported, platforms must temporarily block content during investigation and permanently remove it if confirmed.
AB 602	California	2019-04	NCII	Depiction of individual using digital or electronic technology: sexually explicit material: cause of action	D	Creates a right of action against an individual who intentionally distributes NCII, and authorizes the plaintiff who suffers harm to seek injunctive relief and recover attorney's fees, as well as specified monetary damages.
SB 942	California	2024-01	NCII and CSAM	California AI Transparency Act	D	Requires large GenAI systems within California to make available a free AI detection tool and offer users the option to include a disclosure in content created or altered by the provider's system.
SB 1381	California	2024-02	CSAM	Crimes: child pornography	D	Expands the scope of existing provisions that prohibit the production, development, or possession of CSAM to include content that is digitally altered or generated by the use of AI.
AB 1874	California	2024-01	NCII and CSAM	Crimes: disorderly conduct	R	Increases the punishment for a second or subsequent offense of conviction of disorderly conduct, including AI-generated content.
AB 1962	California	2024-01	NCII and CSAM	Crimes: disorderly conduct	D	Expands the definition of disorderly conduct to include "the distribution of those images recorded, captured, or otherwise obtained by the person distributing the image without the authorization of the person depicted."
SB S1042A	New York	2023-01	NCII	Prohibits unlawful dissemination or publication of intimate images created by digitization and of sexually explicit depictions of an individual.	D	Criminalizes the creation and dissemination of intimate video images and "images created by digitization and of sexually explicit depictions of an individual" intentionally.
S05959	New York	2019-05	NCII and CSAM	Establishes the right of publicity and provides for a private right of action for unlawful dissemination or publication of a sexually explicit depiction of an individual	D	Provides a private right of action for unlawful dissemination or publication of certain sexually explicit depictions. Establishes a cause of action against perpetrators.
SB 20	Texas	2025-02	CSAM	Relating to the creation of the criminal offense of possession, promotion, or production of certain obscene visual material appearing to depict a child.	R	Criminalizes the creation and dissemination of intimate video images and "images created by digitization and of sexually explicit depictions of an individual" intentionally.
HB 2700	Texas	2023-02	CSAM	Relating to prosecution of certain criminal offenses prohibiting sexually explicit visual material involving children.	R	Expands the scope of existing child pornography statutes to include content that is digitally altered or generated by AI.

C L O S I N G G A P S A C R O S S T H E E C O S Y S T E M

Legislation Number	State	Date Introduced	Type	Title	Sponsor Party	Summary
SB 1621	Texas	2025-02	CSAM	Relating to prosecution and punishment of certain criminal offenses prohibiting sexually explicit visual material involving depictions of children, computer-generated children, or other persons; creating criminal offenses; increasing criminal penalties.	R	Expands the offenses of possession or promotion of CSAM to include AI-generated images.
HB 581	Texas	2024-11	CSAM	Relating to the creation of artificial sexual material harmful to minors	R	Creates grounds for lawsuits against a commercial entity that generates CSAM and NCII. Businesses providing explicit AI-generation tools need to comply with age verification laws and obtain consent from people used in the produced image.
HB 149	Texas	2025-03	NCII	Texas Responsible Artificial Intelligence Governance Act (TRAIGA)	R	Prohibits AI systems designed for “restricted purposes,” including encouragement of self-harm, unlawful discrimination, constitutional rights infringement, and CSAM generation.
SB 441	Texas	2024-11	CSAM	Relating to criminal and civil liability related to sexually explicit media and artificial intimate visual material; creating a criminal offense; increasing a criminal penalty	D	Amends civil liability laws to address the production, solicitation, disclosure, or promotion of AI NCII without consent. Introduces specific liability for website owners and payment processors facilitating such content and provides legal protections for claimants.
SB 1361	Texas	2023-03	NCII	Relating to the unlawful production or distribution of sexually explicit videos using deep fake technology, creating a criminal offense	D	Creates a Class A misdemeanor offense for a person to knowingly produce or distribute a deepfake video that appears to depict a person.
HB 3133	Texas	2025-02	NCII	Relating to user reports of explicit deep fake material on social media platforms	D	Requires social media platforms to provide a complaint system that allows users to submit and track reports about explicit deepfake material. Mandates that platforms respond to such reports by immediately confirming receipt, removing the reported content and its duplicates, and providing a written update to the user within seven days.
HB 148	Utah	2024-03	NCII and CSAM	Artificial Pornographic Images Amendments	R	Amends the Sexual Exploitation Act to include computer-generated videos.
HB 238	Utah	2024-03	NCII and CSAM	Sexual Exploitation of a Minor Amendments	R	Amends the definition of “child sexual abuse material” to include AI-generated content.
SB 66	Utah	2024-03	NCII and CSAM	Criminal Offense Amendments	D	Amends the definition of “counterfeit intimate image” in the criminal code to include AI-generated images.

C L O S I N G G A P S A C R O S S T H E E C O S Y S T E M

Legislation Number	State	Date Introduced	Type	Title	Sponsor Party	Summary
SB 256	Utah	2026-02	NCII	Identity Protection Modifications	R	Amends provisions related to libel and slander to address AI content by clarifying that defamation law applies to AI-generated content.
HB 276	Utah	2026-01	NCII	Artificial Intelligence Modifications	R	Enacts new provisions to address the creation and distribution of NCII and establishes standards for digital content provenance. Prohibits platforms that use AI to create or modify intimate images from distributing such images. Imposes requirements on “covered platforms,” such as social media sites, to establish notice and takedown procedures for AI NCII.
HB 289	Utah	2026-01	CSAM	Child Sexual Abuse Material Amendments	R	Updates the existing statutes and definitions of CSAM to include visual depiction that are generated or created using AI.
SB 144	Utah	2025-03	NCII and CSAM	Sexual Crimes Amendments	R	Expands the crime of Sexual Exploitation of a Minor to include accessing with the intent to view CSAM.
HB 13	Utah	2024-12	NCII and CSAM	Sexual Extortion Amendments	R	Updates sexual extortion laws to include threats of distributing fake intimate images generated by AI or other digital tools.

Table 3. Key Parameters of AI CSAM and NCII Laws: California, New York, Texas, and Utah

Legislation Number	State	Type	Criminal or Civil?	Liability	New or Retrofit?	Requires Depiction of Real Human?	Requires Intent to Harm?	Requires Intent to Distribute?	Penalty
AB 1831	California	CSAM	Criminal and Civil	Individual creator and distributor	Statutory retrofit	No	No	No	Imprisonment up to one year, or by a fine not exceeding \$2,000.
SB 926	California	NCII and CSAM	Criminal and Civil	Individual distributor	Statutory retrofit	Yes	Yes	Yes	Imprisonment not exceeding one year, or by a fine not exceeding \$2,000.
AB 621	California	NCII	Civil	Public platforms, apps/models, individual creator and distributor, infrastructure provider	Statutory retrofit	Yes	No	Yes	Civil penalty of \$30,000 to \$50,000, act committed with malice of \$150,000 to \$250,000.
SB 981	California	NCII	Civil	Social media platforms	New	Yes	No	Yes	Up to \$2,500 per violation.
AB 602	California	NCII	Civil	Individual creator and distributor	New	Yes	No	Yes	Can recover statutory damages between \$1,500 and \$30,000, aggravated damages up to \$150,000 if the act was committed with “malice” (intent to harm).
SB 942	California	NCII and CSAM	Civil	apps/models	new	Not applicable	No	No	\$5,000 per violation, per day.

CLOSING GAPS ACROSS THE ECOSYSTEM

Legislation Number	State	Type	Criminal or Civil?	Liability	New or Retrofit?	Requires Depiction of Real Human?	Requires Intent to Harm?	Requires Intent to Distribute?	Penalty
SB 1381	California	CSAM	Criminal	Individual creator and distributor	Statutory retrofit	No	No	Yes	A fine not to exceed \$1,000, county jail time for a first offense to a felony fine of up to \$10,000 and state prison time for repeat offenses.
AB 1874	California	NCII and CSAM	Criminal	Individual creator and distributor	Statutory retrofit	Yes	Yes	No	Imprisonment in county jail for up to one year or a fine not exceeding \$2,000 dollars, or both.
AB 1962	California	NCII and CSAM	Criminal and Civil	Individual creator and distributor	Statutory retrofit	Yes	Yes	Yes	Imprisonment in county jail not exceeding one year, or a fine not exceeding \$2,000, or both. If the victim was a minor, the violation is punishable by imprisonment in county jail not exceeding one year, or by a fine not exceeding \$2,000, or both.
SB S1042A	New York	NCII	Criminal and Civil	Individual creator and distributor	Statutory retrofit	Yes	Yes	Yes	Imprisonment up to one year, or a fine not exceeding \$1,000.
S05959	New York	NCII and CSAM	Civil	Individual creator and distributor	Statutory retrofit	Yes	No	Yes	Civil damages, including money for the harm caused and an order to remove the content.
SB 20	Texas	CSAM	Criminal	Individual creator and distributor, combination/group	New	No	No	No	Tiered system: first offense of 2 to 10 years in prison or \$10,000, second offense of 2 to 20 years in prison or \$10,000.
HB 2700	Texas	CSAM	Criminal	Individual creator and distributor	Statutory retrofit	Yes	No	No	Imprisonment ranging between 2-20 years, fine of \$500.
SB 1621	Texas	CSAM	Criminal	Individual creator and distributor	Statutory retrofit	No	No	No	Minimum of 180 days of imprisonment up to 10 years, max fine of \$10,000.
HB 581	Texas	CSAM	Civil	Apps/models	Statutory retrofit	Yes	No	NA	\$10,000 per day that the entity operates a website in violation of the age verification requirements, \$10,000 per instance, max \$250,000.
HB 149	Texas	NCII	Civil	Apps/models	New	No	No	No	No less than \$10,000 and no more than \$12,000 per violation to \$80,000–\$200,000 per incurable violation, plus \$2,000–\$40,000 per day for continuing violations.

CLOSING GAPS ACROSS THE ECOSYSTEM

Legislation Number	State	Type	Criminal or Civil?	Liability	New or Retrofit?	Requires Depiction of Real Human?	Requires Intent to Harm?	Requires Intent to Distribute?	Penalty
SB 441	Texas	CSAM	Criminal and Civil	Individual creator and distributor, payment processor, public platforms, social media platforms, apps/models	Statutory retrofit	Yes	Yes (individual creators)	No	Between 2–10 years in prison; victims can recover damages, punitive damages, and attorney’s fees.
SB 1361	Texas	NCII	Criminal	Individual creator and distributor	New	Yes	Yes (intent to deceive)	No	Imprisonment up to 1 year, fine up to \$4,000.
HB 3133	Texas	NCII	Civil	Social media platforms	New	Yes	No	NA	Up to \$10,000 per violation.
HB 148	Utah	NCII and CSAM	Criminal and Civil	Individual creator and distributor	Statutory retrofit	Yes	Yes	Yes	Adults: first offense is Class A Misdemeanor, second or subsequent offense is third-degree felony. Minors: possession is a second-degree felony, production/distribution is a first-degree felony.
HB 238	Utah	CSAM	Criminal	Individual creator and distributor	Statutory retrofit	No	No	Yes	Possession/viewing: second-degree felony. Distribution/production: second-degree felony if the depicted minor is 14+. First-degree felony if the minor is under 14.
SB 66	Utah	NCII and CSAM	Criminal	Individual creator and distributor	Statutory retrofit	Yes	Yes	Yes	Distribution is Class A Misdemeanor for a first offense, third-degree felony for subsequent offenses.
SB 256	Utah	NCII	Civil	Individual creator and distributor	Statutory retrofit	Yes	Yes	Yes	Injunctive relief, damages alleged and proved, exemplary damages, and reasonable attorney’s fees and costs.
HB 276	Utah	NCII	Civil	Apps/models, public platforms, social media platforms	New	Yes	No	Yes	Actual damages (for emotional distress), punitive damages (for egregious cases), attorney’s fees.

CLOSING GAPS ACROSS THE ECOSYSTEM

Legislation Number	State	Type	Criminal or Civil?	Liability	New or Retrofit?	Requires Depiction of Real Human?	Requires Intent to Harm?	Requires Intent to Distribute?	Penalty
HB 289	Utah	CSAM	Criminal	Individual creator and distributor	Statutory retrofit	No	No	No	Possession: third-degree felony. Production/distribution: second-degree felony. Aggravated (for victims under 14): first-degree felony.
SB 144	Utah	NCII and CSAM	Criminal	Individual creator and distributor	Statutory retrofit	Yes	No	No	Raised the penalty from a second-degree felony to a first-degree felony.
HB 13	Utah	NCII and CSAM	Criminal	Individual creator and distributor	Statutory retrofit	Yes	No (requires intent to coerce)	N/A	Standard extortion: third-degree felony. Aggravated extortion: second-degree felony (adult). Aggravated (child victim): first-degree felony.

Acknowledgments

We would like to express our sincere gratitude to Dr. Meredith Sadin, Assistant Research Professor at the UC Berkeley Goldman School of Public Policy, for her invaluable guidance throughout the development of this paper.

We are also deeply grateful to the experts and practitioners who generously shared their time and perspectives during the interview process, including Sarah Elizabeth Barrington, Michelle L. Ding, Amanda Goharian, Megan Griffin, David Evan Harris, Rachel Keels, Alexios Mantzarlis, and Omny Miranda Martone. Participants do not necessarily endorse the recommendations or considerations included in this report.

We thank Chuck Kapelke for editing, web, and media support, and Nicole Hayward for design and formatting of this document. Special thanks to Ann Cleaveland for providing a home and intellectual support for this work at the UC Berkeley Center for Long-Term Cybersecurity.



CLTC

Center for Long-Term
Cybersecurity

UC Berkeley