TRANSPARENCY HUB
AN APPLIED SOCIAL MEDIA LAB PROJECT

Mapping the Burden:

Toward Survivor-Centered Reporting Systems and Policies for Non-Consensual Intimate Images

Grace Harlan, Alice Jo, Cailin Crockett [1]

Design by Malvika Dwivedi

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

Table of Contents

Abstract

Advances in generative AI have dramatically lowered the barriers to creating and distributing sexual "deepfakes" or non-consensual intimate images (NCII)—a form of technology-facilitated gender-based violence (TFGBV) that disproportionately targets women and marginalized populations. Yet as the technology enabling NCII scales, the reporting systems meant to address it remain opaque, inconsistent, and burdensome for victim-survivors. Using the walkthrough method developed by Light, Burgess, and Duguay (2018), we conducted a comparative analysis of NCII reporting interfaces and policies of seven major online platforms: Facebook, Instagram, Snapchat, TikTok, Reddit, X, and Google Search. We find that current NCII reporting systems are structurally onerous, unclear, and re-traumatizing for survivors. Platform design choices—interfaces, categorization, architecture—are overwhelmingly reactive: most platforms privilege keeping content reported as NCII visible rather than presumptively removing it to protect survivors, and few provide transparency on removal processes or perpetrator accountability. Collectively, these design choices replicate the loss of control and agency survivors experience when their images are created and shared without their consent. Absent clear structural changes to platform design—changes we anticipate would align with the legislative intent of the United States' TAKE IT DOWN Act and complementary regulations in the United Kingdom and other global contexts—the current approach maintains the burden of harm on survivors. We provide recommendations to improve platform reporting processes, as well as to hold the wider technological ecosystem accountable for their role in the creation and promotion of synthetic NCII. Ultimately, the technical, legal, and societal approach to NCII must shift from treating it as an unfortunate but inevitable byproduct of online life to establishing standards and coordinated infrastructure for rapid cross-platform detection and removal.

[1] Grace Harlan and Alice Jo are co-authors listed alphabetically by last name. Cailin Crockett, Harvard Berkman Klein Center affiliate, is an independent expert and consultant to a range of industry, non-profit, and governmental entities, including StopNCII.org and the Cyber Civil Rights Initiative. Her contributions to this white paper as an advisor and third co-author reflect her personal opinion, and not those of external organizations with which she is affiliated.

1. Introduction

In late 2025, xAI's chatbot Grok generated an estimated 3 million sexualized images of adults and children within days of X integrating its image editing feature onto the main interface of the platform.[2] X's decision to embed generative artificial intelligence (AI) with deepfake capabilities into its interface enabled the creation and spread of non-consensual intimate imagery (NCII) at scale. This issue is not unique to X, as "nudify" apps have proliferated across the internet. A 2025 study found that searches for "deepnude," "nudify," and "undress app" on Google, Yahoo, and Bing all yielded at least one search result that led to tools for NCII generation within the first 20 results.[3] The emergence of these apps has made it possible to transform a single image of someone's face into a hyperrealistic digital sexual forgery within minutes.[4]

NCII abuse occurs when perpetrators create, threaten to share, or distribute non-consensual sexually explicit images and videos. This content is not only limited to authentic media, but also includes manipulated private photographs and synthetic, AI-generated sexual imagery. NCII is a form of image-based sexual abuse (IBSA) that objectifies, humiliates, and often silences victim-survivors; NCII and IBSA occur within a broader continuum of technology-facilitated gender-based violence (TFGBV), with links to offline forms of abuse.[5]

While NCII creation is possible at little to no cost to the perpetrator, it has significant and lasting harm to the person depicted. Both real and synthetic NCII violate what MacArthur Genius Scholar Professor Danielle Citron terms "intimate privacy"—the norms that regulate access to our bodies, desires, and intimate lives—and can cause persistent psychosocial harms, impacting survivors' personal, social, and professional lives.[6]

NCII abuse is deeply gendered. While significant research has focused on IBSA involving minors, this paper examines the experiences of adult victim-survivors in navigating online reporting processes and image removal requests.[7] According to the latest CDC data, one in ten American women experienced some form of technology-facilitated sexual violence in the past year, and more than 2.3 million adult women had their intimate images shared without their consent.[8] While the CDC data does not distinguish between authentic and synthetic NCII, it is clear that the magnitude of this form of sexual violence is only growing with the influx of generative AI tools. For example, a 2025 study identified nearly 35,000 publicly downloadable deepfake model variants, with 96% modified specifically to target and undress women.[9]

[2] Center for Countering Digital Hate, Grok Floods X with Sexualized Images of Women and Children, Center for Countering Digital Hate, January 22, 2026, https://counterhate.com/research/grok-floods-x-with-sexualized-images/

[3] Chiara Puglielli and Anne Craanen, The Ecosystem of Nonconsensual Intimate Deepfake Tools Online (Institute for Strategic Dialogue, 2025), https://www.isdglobal.org/digital-dispatch/the-ecosystem-of-nonconsensual-intimate-deepfake-tools-online/.

[4] Britt Paris and Joan Donovan, "Deepfakes and Cheap Fakes." Data & Society, September 18, 2019. https://datasociety.net/library/deepfakes-and-cheapfakes/.; Cassidy Gibson et al., "Analyzing the AI Nudification Application Ecosystem," 2025, https://www.usenix.org/system/files/usenixsecurity25-gibson.pdf.

[5] Carmela Mento et al., "Psychological Violence in Image-Based Sexual Abuse (IBSA): The Role of Psychological Traits and Social Communications–A Narrative Review," Healthcare (Basel, August 22, 2025) 2025 13 no. 17, 2083, https://doi.org/10.3390/healthcare13172083.

[6] Danielle K. Citron, The Fight for Privacy: Protecting Dignity, Identity and Love in the Digital Age, (W. W. Norton & Company, 2022), xii.

[7] Gabriella De Guzman, "Deepfake Nudes Are a Harmful Reality for Youth: New Research from Thorn," Thorn, March 3, 2025, https://www.thorn.org/blog/deepfake-nudes-are-a-harmful-reality-for-youth-new-research-from-thorn/; Sandra Cortesi et al., Frontiers in Digital Child Safety: Designing Child-Centered Digital Ecosystems That Support Rights, Agency, and Well-Being (TUM Think Tank, 2025), https://tumthinktank.de/en/project/frontiers-in-digital-child-safety/.

[8] Ruth W. Leemis et al., The National Intimate Partner and Sexual Violence Survey (NISVS): 2023/2024 Sexual Violence Data Brief (Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, 2025), 6, https://www.cdc.gov/nisvs/media/pdfs/sexualviolence-brief.pdf.

[9] Will Hawkins, Brent Mittelstadt, and Chris Russell, "Deepfakes on Demand: The Rise of Accessible Non-Consensual Deepfake Image Generators," FAccT '25: Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency, June 23, 2025, 1602–14, https://doi.org/10.1145/3715275.3732107.

Like other forms of gender-based violence, survivors often face significant barriers to seeking resources and recourse. Stigma, self-blame, and unhelpful, or even retraumatizing, interactions with law enforcement discourage many survivors from coming forward or pursuing legal remedies for NCII, and available options remain highly contingent on location and legal context.[10] When survivors do attempt to seek redress from online platforms, they frequently encounter reporting systems that are difficult to navigate and ineffective. Refuge, the UK's largest domestic abuse support organization, conducted a study and found that 95% of interview participants were unsatisfied with platform responses.[11] Such experiences reflect broader patterns of platforms being slow to act or unresponsive to urgent requests for content removal.

Our analysis is situated within the policy landscape shaped by the US federal TAKE IT DOWN Act (TIDA) and draws on existing survivor-centered research to evaluate whether current systems deliver on content removal, the most fundamental form of relief. Our contributions for making the current online ecosystem less adversarial towards NCII survivors are as follows:

- Using the walkthrough method developed by Light, Burgess, and Duguay, we analyze NCII reporting interfaces across Facebook, Instagram, Snapchat, TikTok, Reddit, X, and Google Search, surfacing the design choices that shape survivors' most urgent priority: the swift removal of their images.[12] Our analysis finds that current reporting processes are cumbersome, opaque, and reactive, placing the burden of redress on survivors. While TIDA is a significant legislative improvement, its promise depends on effective implementation alongside structural changes to platform design.
- Our findings guide concrete recommendations that include centering the survivor in the reporting process, standardizing reporting categories across platforms, implementing trauma-informed communication practices, and increasing transparency. These recommendations largely complement recent FTC guidance on how covered platforms can comply with TIDA.[13]
- We propose additional strategies that future online platforms can use to shift from response-based systems that treat NCII as an unfortunate but inevitable byproduct of online life toward a "safe by design" approach that treats rapid, trauma-informed NCII removal as a baseline obligation.

The ease with which perpetrators create, post, and spread AI-generated NCII stands in sharp contrast to the maze-like reporting systems survivors must navigate to remove it—a disparity that reflects a reactive approach by platforms and policymakers alike. We find redesigning the ecosystem to reduce the demand and distribution of NCII in the first place is possible. We urge industry and regulators alike to treat NCII as a serious, foreseeable harm that platforms have the responsibility and power to prevent.

[10] V. Karasavva, "The Frequency, Nature, Impact, and Coping Strategies of Nonconsensual Intimate Image Dissemination Victimization: A Scoping Review," Trauma, Violence, & Abuse (2025), https://doi.org/10.1177/15248380251383940.
[11] Refuge, Marked as Unsafe: How Online Platforms Are Failing Domestic Abuse Survivors (Refuge, 2022), 5–8, https://refuge.org.uk/wp-content/uploads/2022/11/Marked-as-Unsafe-report-FINAL.pdf; 29% of survivors they interviewed reported experiencing intimate image abuse on social media.
[12] Ben Light, Jean Burgess, and Stefanie Duguay, "The Walkthrough Method: An Approach to the Study of Apps," New Media & Society 20 no. 3 (2018): 883, https://journals.sagepub.com/doi/10.1177/1461444816675438; Erika Rackley et al., "Seeking Justice and Redress for Victim-Survivors of Image-Based Sexual Abuse," Feminist Legal Studies 29, no. 3 (2021): 317, https://doi.org/10.1007/s10691-021-09460-8.
[13] Andrew Ferguson, "TIDA Stakeholder Letter," Federal Trade Commission, 2026, https://www.ftc.gov/system/files/ftc_gov/pdf/TIDA-Stakeholder-Letter.pdf.

2. Background

Role of the Technical Ecosystem

From a technical standpoint, NCII is most often shared via large, public online platforms that use algorithmic recommender systems that aim to personalize and amplify content. These platforms play a major part of the "malicious technical ecosystem" that facilitates the spread and consumption of NCII, but they are not alone.[14]

Commercial Impact

The commercialization of both real and synthetic non-consensual intimate sexual content by nudify tool developers, online forums, and dedicated websites, has rapidly expanded through the use of mainstream service providers and platforms. From cloud infrastructure providers, to payment processors, and messenger platforms that host and deliver services, as well as accept payments, the NCII ecosystem generates $36 million a year.[15] Search engines enable the widespread discovery of NCII websites—such as MrDeepFakes—and nudify apps, both in the form of advertisements for genAI tools, and directly surfacing content.[16] AI nudifier apps rely on advertising on social media platforms and deploy customer referral schemes such as referral links on platforms like Reddit and X, allowing them to accumulate over 24 million unique visitors in September 2023 alone.[17] Behind this commercial ecosystem are open-source AI models, apps, and genAI chatbots that facilitate the creation of NCII.[18]

A 2026 study by the Tech Transparency Project (TTP) showed that even though Google and Apple explicitly forbid nudify apps in their terms of service, they have not in practice taken the same action for their advertising and recommendation systems. In total, the nudify apps surfaced in TTP's app store searches have been downloaded 483 million times and made more than $122 million in lifetime revenue.[19] Technical solutions to proactively prevent NCII—including blocking nudify app searches—exist across this ecosystem. However, without legal and regulatory incentives, industry efforts remain unenforceable and therefore voluntary, suggesting a choice by platforms to prioritize profit over protection. Promisingly, the bipartisan Senate Intelligence Authorization Act for FY2027 passed committee in May 2026 and included Sec. 711, which prohibits acquisition and use of any AI models procured by the intelligence community from generating image-based sexual abuse of minors (CSAM) and adults (NCII). Citing the NIST AI Risk Management Framework: GenAI Profile as a standard for how AI vendors can implement safeguards against their models generating IBSA, the bill requires the intelligence community to vet AI models using this or a comparable framework, and to remove from national security systems any applications that fail to demonstrate safeguards for IBSA. This bill goes further than TIDA in its aim to prevent the spread of intimate image abuse, and may incentivize AI companies to better safeguard their systems against this particular misuse.

[14] Michelle L. Ding and Harini Suresh, "The Malicious Technical Ecosystem: Exposing Limitations in Technical Governance of AI-Generated Non-Consensual Intimate Images of Adults," arXiv, 2025, arXiv.2504.17663.

[15] Alexios Mantzarlis and Santiago Lakatos, "AI Nudifiers Continue to Reach Millions and Make Millions," Indicator, July 14, 2025, https://indicator.media/p/ai-nudifiers-continue-to-reach-millions-and-make-millions.

[16] Michelle L. Ding, Harini Suresh, and Suresh Venkatasubramanian, "How to Stop Playing Whack-a-Mole: Mapping the Ecosystem of Technologies Facilitating AI-Generated Non-Consensual Intimate Images." arXiv, 2026, arXiv:2602.04759; Catherine Han et al., "Characterizing the MrDeepFakes Sexual Deepfake Marketplace," arXiv, 2024, arxiv:2410.11100v3. Note: MrDeepFakes, a popular forum for deepfake creators, ceased operations in May 2025.

[17] Santiago Lakatos, "A Revealing Picture." Graphika, December 8, 2023, https://graphika.com/reports/a-revealing-picture.

[18] Ding, Suresh, and Venkatasubramanian, "How to Stop Playing Whack-a-Mole," 6.

[19] "Apple and Google Are Steering Users to Nudify Apps," Tech Transparency Project, April 15, 2026, https://www.techtransparencyproject.org/articles/apple-and-google-are-steering-users-to-nudify-apps.

Survivor Experience

While platforms that host and disseminate NCII content are only a part of the ecosystem, they often interact most directly with survivors. As Li et al. (2025) argue, these platforms wield unprecedented power as "crime scene, evidence locker, judge, jury," hosting the abuse, controlling access to the evidence, and determining if, when, and how the content is removed.[20] As it stands, the reporting process is cumbersome, re-traumatizing, and often ineffective, resulting in many survivors not reporting the content at all.[21]

As this paper explores, each platform varies significantly in its approach to defining NCII, structuring the reporting process, and handling takedown requests. This discrepancy renders responses to survivors obscure, mixed, and challenging for survivors to understand, in many cases delaying the timeline to remedy. While we focus on these platforms to emphasize the importance of survivor-centered approaches, our goal is to identify common gaps across the industry, as well as promising developments aligned with the timing of recent legislation.

The TAKE IT DOWN Act

Following years of advocacy from victim-survivors, legal experts and activists, the US passed the bipartisan federal TAKE IT DOWN Act (TIDA) in May 2025 with an enforcement deadline for platform responsibilities under the law of May 19, 2026. The law prohibits the online publication of intimate visual depictions of a) an adult subject where publication is intended to cause or does cause harm to the subject, and where the depiction was published without the subject's consent or, in the case of an authentic depiction, was created or obtained under circumstances where the adult had a reasonable expectation of privacy; or b) a minor subject where publication is intended to abuse or harass the minor or to arouse or gratify the sexual desire of any person. The law requires covered platforms to "make reasonable efforts" to remove non-consensual intimate visual depictions "no later than 48 hours after receiving the notice," including "known identical copies"; it also requires platforms to "provide a 'plain language' explanation of its notice-and-removal process on its site."[22] These are all important wins that survivors and their allies have long advocated for.

Despite the law's plain language requirement and recent FTC guidance instructing platforms to "make it easy for people to submit a removal request," both lack specific guidance for implementation.[23] Further, TIDA does not require online platforms to disclose NCII in their transparency reports, meaning it will remain a challenge to assess the scale of the issue on each platform.

The Cyber Civil Rights Initiative, which operates the national 24/7 helpline for NCII, identified three key concerns with TIDA: first, the law contains no safeguards against false complaints, which risks making the statute "highly susceptible to abuse," including by false or malicious reports. Second, the law prohibits claims against covered platforms for "good faith" removal, meaning that those with lawful content unlawfully removed could have no recourse, posing potential free speech concerns.

[20] Qiwei Li et al., "Platforms as Crime Scene, Judge, and Jury: How Victim-Survivors of Non-Consensual Intimate Imagery Report Abuse Online," CHI '26: Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems, April 13, 2026, 2, https://doi.org/10.1145/3772318.3791115.
[21] Sherry Hakimi, "Tools for Reporting Online Violence Are Broken. Here's How to Fix Them," Tech Policy Press, August 28, 2025, https://www.techpolicy.press/tools-for-reporting-online-violence-are-broken-heres-how-to-fix-them/.
[22] Congressional Research Service, "The TAKE IT DOWN Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images," Congress.gov, May 20, 2025, https://www.congress.gov/crs-product/LSB11314; A covered platform is any public-facing website, app, or online service that primarily hosts user-generated content or that publishes, curates, or makes available NCII in the regular course of business.
[23] Ferguson, "TIDA Stakeholder Letter," 2.

Third, the law's definition of covered platforms does not apply to platforms that "consist primarily of content that is not user generated but is preselected by the provider."[24]

Further, by narrowly focusing the criminal penalty on the distribution of NCII, without also explicitly criminalizing its creation or monetization, the US continues to have a legislative loophole that sustains the generation of NCII.[25] TIDA's definition of a "covered platform" lacks clarity regarding the obligations of AI companies who build tools that enable the generation of NCII, but do not identify as hosts for user-generated content. This shortcoming is concerning since TikTok, Meta, X, Apple, Google, etc. continue to surface advertisements for nudify apps on their platform feeds as well as host websites dedicated to the creation and commercial consumption of NCII (despite these sites often violating platforms' stated content moderation policies). In this key respect, the US lags behind its peers. For example, the United Kingdom, South Korea, and Australia, have gone a step further in criminalizing the creation of NCII.[26] The UK Data Act of 2025 criminalizes creation, and their recent amendment to the Crime and Policing Bill will match TIDA's 48-hour timeline for removing images flagged as NCII.[27]

While TIDA and similar laws globally are a step in the right direction, more attention must be paid to the larger ecosystem perpetuating these harms. Ultimately, survivor-centered reform requires coordinated action across platform design, industry standards, and legislation to enforce rapid cross-platform detection and removal. Holistically addressing NCII will also require preventing it in the first place by reducing capabilities for its creation via genAI and reducing demand via commercialization.

3. Methodology

The Walkthrough Method

While public awareness of NCII as a form of online harm has increased, far less attention in academic literature has been paid to the reporting process itself—the interface language, categorization options, and feedback loops that platforms present to someone in crisis. Recent studies about the reporting processes for online harms and NCII have begun to address this gap: Li et al. (2025) conducted trauma-informed interviews with survivors to document the harms of platform reporting; Flynn et al. (2025) used focus groups alongside platform walkthroughs to examine the reporting process for harmful and offensive content on Facebook, Instagram, and X; the Center for Democracy & Technology's (CDT) July 2025 Rapid Response report assessed NCII policies and reporting tools across eight platforms.[28] We add to this body of work by foregrounding the user experience of navigating reporting interfaces, particularly in light of the May 2026 enforcement of TIDA.

[24] Cyber Civil Rights Initiative, "CCRI Statement on the Passage of the TAKE IT down Act (S. 146)," April 28, 2025, https://cybercivilrights.org/ccri-statement-on-the-passage-of-the-take-it-down-act-s-146/; Trudi K. Sundberg, "Federalizing NCII Regulation: The Take It Down Act's Approach to Criminalization, Platform Liability, and Threats to Disseminate," Georgia Law Review 59: no. 3 (2025), https://digitalcommons.law.uga.edu/glr/vol59/iss3/9.

[25] Julia Hörnle, "Fighting the beast of image-based sexual abuse. Part 1—the criminal law and platform regulation," International Journal of Law and Information Technology, Volume 34, 2026, eaag006, https://doi.org/10.1093/ijlit/eaag006.

[26] Clare McGlynn and Rüya Tuna Toparlak, "The 'New Voyeurism': Criminalizing the Creation of 'Deepfake Porn'," Journal of Law and Society 52, no. 2 (2025): 204–228, https://doi.org/10.1111/jols.12527.

[27] UK Department for Science, Innovation and Technology, "Tech firms will have to take down abusive images within 48 Hours under new law to protect women and girls," GOV.UK, February 19, 2026, https://www.gov.uk/government/news/tech-firms-will-have-to-take-down-abusive-images-within-48-hours-under-new-law-to-protect-women-and-girls.

To do this, we use the walkthrough method developed by Light, Burgess, and Duguay (2018), a qualitative framework for analyzing digital platforms through systematic, step-by-step engagement with their interfaces from a user perspective. We adapt this approach to study the interfaces that survivors face on platforms while recognizing that as researchers simulating these processes, our experience cannot fully reflect the emotional stakes, urgency, or trauma that shape how these systems are actually experienced by individuals in crisis.

The walkthrough method directs researchers to be attentive to three dimensions: 1) the environment of expected use, or who the platform imagines its users to be; 2) the technical walkthrough, or the sequences of screens, prompts, and options encountered when performing a task; and 3) the platform's governance structures that shape users' interactions with the platform and each other, as demonstrated in their public-facing documents.[29] By walking though the NCII image removal process on seven platforms between February and May 2026—documenting the language, required steps, reporting categories, and platform responses—we surface key insights about design choices that shape the reporting experience.

This paper focuses on Facebook, Instagram, Snapchat, TikTok, Reddit, X, and Google Search.[30] Each platform ranks highly on the monthly user count, operates public communication channels, and qualifies as a covered platform subject to the jurisdiction of TIDA.[31] At the time of publication, most of these platforms are partnered with StopNCII.org—a voluntary hash registry that enables survivors to securely upload their intimate images and prevent them from being posted on participating platforms.[32] Our assessment of these seven platforms also explores whether their reporting mechanisms handle the particular challenges of AI-generated NCII, where content can be produced at scale, duplicated endlessly, and may not fit neatly into the categories platforms currently offer.

4. Environment of Expected Use

The walkthrough method's first analytical layer examines the environment of expected use: how platforms represent themselves, who they imagine their users to be, and what they envision them doing. This framing reveals the distance between the digital experiences platforms promise and the reality for survivors of NCII abuse.

[28] Li et al., "Platforms as Crime Scene, Judge, and Jury"; Asher Flynn et al., "Content Moderation and Community Standards: The Disconnect between Policy and User Experiences Reporting Harmful and Offensive Content on Social Media," Policy & Internet 17, no. 3 (2025): 1343, https://doi.org/10.1002/poi3.70006; Becca Branum and Mi Yeon Kim, Rapid Response: Building Victim-Centered Reporting Processes for Non-Consensual Intimate Imagery (Center for Democracy and Technology, July 24, 2025), https://cdt.org/insights/rapid-response-building-victim-centered-reporting-processes-for-non-consensual-intimate-imagery/.

[29] Light, Burgess, and Duguay, "The Walkthrough Method," 883.

[30] While not a traditional social media platform, Google Search handles user-generated content and NCII reports directly.

[31] "Biggest Social Media Platforms by Users," Statista, 2025, https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

[32] Meta, Snapchat, TikTok, Reddit, X, and Snapchat are official partners with StopNCII. For a complete list of StopNCII industry partners, see here.

Implications for NCII

Platforms consistently present themselves as spaces of connection, creativity, entertainment, and safety. For example:

- **Meta** states that it has "responsibility to promote the best of what people can do together by keeping people safe and preventing harm."[33]
- **X** describes itself as "your trusted digital town square where conversations unfold in real time" and promises content delivered "raw and unfiltered."[34]
- **Google** states they are committed to "significantly improving the lives of as many people as possible" and deliver reliable and relevant information in the "most useful" way.[35]

The mismatch between platforms' mission statements and the inadequacy of their responses to harms that directly impact their core users is not merely an unfortunate coincidence, but a direct manifestation of social media platform design.

The algorithms underpinning the major platforms are designed to maximize engagement at all costs, leveraging insights from behavioral studies that show people tend to react more to negative, extreme, and polarizing content. Beknazar-Yuzbashev et. al (2025) found respondents who encountered more toxic posts were 18% more likely to click to view the comment sections of the posts, building on prior research showing how online social networks provoke moral outrage.[36] By creating a hostile environment for users who are more likely to be the target of online harms and nudging susceptible perpetrators towards harmful content, algorithmic feeds make the virality of NCII an endemic feature of online platforms.

Conflicting Incentives

Women are both the demographic most affected by NCII online and many platforms' largest user base. Women stand out in their use of several platforms, including Facebook, Instagram and TikTok.[37] X and Reddit are found to have predominantly male users, who report using the platform primarily for entertainment (whereas Meta users tend to cite keeping in touch with loved ones).[38] NCII perpetrators are overwhelmingly male, and research consistently finds that they frame NCII creation as "entertainment" or a "technical challenge"—motivations rooted in misogyny and status-seeking that thrive when platforms fail to treat NCII as the abuse it is.[39] NCII creates a hostile online environment that undermines equitable participation in online services, including by chilling speech, compromising privacy, and in some cases threatening survivors' physical safety.[40] Such content poses both immediate and long-term harm to its targets, especially if widely distributed, and creates a systemic

[33] Meta, "Company Info," Meta.com, accessed April 2026, https://www.meta.com/about/company-info/.
[34] X Corp., "X," App Store, accessed April 2026, https://apps.apple.com/us/app/x/id333903271.
[35] Google, "Our Approach – How Google Search Works," Google.com, accessed April 2026, https://www.google.com/intl/en_uk/search/howsearchworks/our-approach/.
[36] George Beknazar-Yuzbashev et al., "Toxic Content and User Engagement on Social Media: Evidence from a Field Experiment," CESifo Working Paper No. 11644, 2022, https://ssrn.com/abstract=5130929; M.J. Crockett, "Moral outrage in the digital age." Nature Human Behaviour 1 (2017): 769–77, https://doi.org/10.1038/s41562-017-0213-3.
[37] Jeffrey Gottfried and Eugenie Park, Americans' Social Media Use 2025 (Pew Research Center, November 20, 2025), https://www.pewresearch.org/internet/2025/11/20/americans-social-media-use-2025; For more information on platform user base statistics, see Figure A1 in the Appendix.
[38] Ibid.; Colleen McClain, Monica Anderson, and Risa Gelles-Watnick, How Americans Navigate Politics on TikTok, X, Facebook and Instagram (Pew Research Center, June 12 2024), https://www.pewresearch.org/internet/2024/06/12/how-americans-navigate-politics-on-tiktok-x-facebook-and-instagram.
[39] Jaron Mink, Lucy Qin, Elissa M. Redmiles, "'Unlimited Realm of Exploration and Experimentation': Methods and Motivations of AI-Generated Sexual Content Creators," arXiv, 2026, arXiv: 2601.21028v1.
[40] Security Hero, "2023 State of Deepfakes: Realities, Threats, and Impact," 2023, https://www.securityhero.io/state-of-deepfakes/.

risk across digital platforms for gender-based harassment and intimidation.[41]

Despite the evidence that enragement drives engagement, it would be rational to assume that platforms also have a business incentive to make online environments as positive as possible for their target user-base.[42] The National Organization for Women found in a 2026 study that more than 1 in 3 Millennial women and almost half of Gen Z respondents reported experiencing online abuse.[43] The study found a high rate of self-limiting, whether through self-censorship (39%) or deleting or restricting social media accounts (44%). Similarly, 32% of women who reported experiencing harassment tried to delete personal information online as a reaction to abuse.[44] A handful of major online platforms dominate our social, political, economic and community life, making opting out of participation not a truly viable alternative.[45] Other jurisdictions—such as the European Union's pairing of the Digital Services Act with the Digital Markets Act—offer a guide to addressing these parallel obstacles to safer online environments. While the DSA focuses on user safety, content moderation, and platform accountability, the DMA ensures fair competition by imposing strict rules on large "gatekeeper" platforms.[46] Specific focus on NCII system response improvements via TIDA is necessary but insufficient for a complete ecosystem effect that will reduce risk for harm in the first place.

As outlined in the following section, our walkthrough findings reveal significant gaps between companies' stated commitments and the experience they deliver to survivors.

[41] Immediate harms include physical security risks posed by online threats (co-occurring with other forms of abuse like doxxing, or real life sexual violence), and long-term harms include the mental health impacts of experiencing any form of trauma, online or offline. The EU's Digital Services Act recognizes gender-based violence as a systemic risk that providers of very large online platforms (VLOPs) and very large online search engines (VLOSEs) must assess and mitigate. The European Commission is currently investigating whether X properly assessed and mitigated risks prior to deploying Grok AI.

5. Walkthrough Findings and Recommendations

Our findings and recommendations center on platform experience and policies, with an eye toward structural, trauma-informed reforms that shift the burden of redress away from survivors and make NCII reporting clearer, faster, and more accountable.

A. Language - Inconsistent NCII Terminology Across Platforms

Analyzing the NCII reporting process on platforms via the walkthrough method entails close reading and analysis of community guidelines, as well as platform rules and guidelines around permitted and forbidden activity types.

Where NCII Appears in Guidelines

Without using a search engine, these guidelines can be hard to find, particularly in-app. Moreover, most platforms do not directly possess a standalone NCII policy (see Figure 1). Thus, the issue is not only the lack of consistent terms for NCII used across the industry, but also a patchwork for where this information is housed on platforms.[47]

[42] Steve Rathje, Jay J. Van Bavel, and Sander van der Linden, "Out-group animosity drives engagement on social media," Proc Natl Acad Sci U S A, 118 no. 26 (2021), https://pubmed.ncbi.nlm.nih.gov/34162706/.
[43] National Organization for Women Incogni Research Team Online Abuse Against American Women is Escalating, March 6, 2026, https://now.org/wp-content/uploads/2026/03/Incogni-Online-Abuse-Survey-2026.pdf
[44] Ibid
[45] Burton Ong and Ding Jun Toh, "Digital Dominance and Social Media Platforms: Are Competition Authorities up to the Task?" IIC - International Review of Intellectual Property and Competition Law 54, no. 54 (2023), https://doi.org/10.1007/s40319-023-01302-1.
[46] European Parliament, "EU Digital Markets Act and Digital Services Act Explained," European Parliament, December 14, 2021, https://www.europarl.europa.eu/topics/en/article/20211209STO19124/eu-digital-markets-act-and-digital-services-act-explained; In the landmark 2026 K.G.M. v. Meta et al. case, a jury found Meta and YouTube negligent in designing intentionally addictive algorithms—a consumer product liability framework that may serve as a blueprint for holding platforms accountable for harm caused by their algorithms if not the content itself while working around Section 230 immunities.
[47] For a one-stop location to check platform community guidelines, privacy policies, and terms of service, see Transparency Hub, a platform designed to help people explore, compare, and better understand the data practices of consumer-facing social and technology applications.

Inclusion of AI-Generated Content in Definitions of NCII

As NCII becomes easier to create synthetically, platforms must explicitly include AI-generated content as part of their NCII definition. As it stands, this gap has yet to be addressed by several platforms (see Figure 1).

- **Reddit, TikTok, and Meta** explicitly include in their definition of NCII that it applies to AI-generated images.
- **Meta, Snap, and Reddit** explicitly state that offering nudifying services is not permitted.
- **X** still uses the language that NCII refers to pictures that were "taken or appear to be taken" without consent, as opposed to including the term "created."
- **Google**'s community guidelines policy has not been updated since 2023 and does not directly reference AI-generated images, though it does clarify that you can report NCII within the reporting channels.

How Non-Consensual Content is Defined

Examining platforms' community guidelines, we found that while some platforms explain what they mean by non-consensual, none go as far as to define consent.

- **Meta** defines NCII in its Stop Sextortion page, but not explicitly in its community guidelines. They delineate markers for non-consent, saying posts must meet one of the following criteria: 1) Vengeful context (such as caption, comments or page title); 2) Independent sources such as law enforcement records, media reports (such as leak of images confirmed by media) or representatives of a survivor of NCII; 3) Report from a person depicted in the image or who shares the same name as the person depicted in the image. If it meets one of those 3 criteria, the image must also be "non-commercial" and "containing nudity or sexual acts."
- **Google** follows similar criteria, but says all should be met, not just one of the three: 1) The imagery shows you (or the individual you're representing) nude, in a sexual act, or in an intimate state; 2) You (or the individual you're representing) didn't consent to the imagery or the act and it was made publicly available or the imagery was made available online without your consent; 3) You aren't currently being paid to commercialize this content online or elsewhere.
- **Reddit, along with Google and Meta**, also specifies that the content must be non-commercial.[48]

There should be a low barrier threshold for initial content removal. Meta's criteria, including the simplest marker—the person flagging NCII for takedown is the person depicted—supports a culture of believing the reporter and hedging bets to remove the image immediately. This approach enables platforms to perform back-end checks to verify this is the case and reupload it if they determine it is not NCII. Initial removal should be predicated on an attestation by the reporter that they have been depicted without their consent, or that an intimate image of them intended to be private was published without their consent. Through a strong appeal process—which would enable individuals to understand why the platform took the action they did, and allow the individuals to seek a re-review—initial removal would not signify permanent removal of lawful content.

[48] The emphasis on "non-commercial" content does not clearly state that individuals featured in commercial imagery can withdraw their consent to continued distribution. This creates an industry gap, as certain adult-content platforms, like OnlyFans, have specific policies that note commercial sexually explicit imagery can become NCII if the person depicted withdraws their consent.

From a survivor-centered, privacy-preserving lens, presumptive removal is the preferred approach: the harm of keeping actual NCII online far outweighs the temporary removal of content later found not to qualify. Unfortunately, TIDA does not require platforms to provide any standardized appeals process.

How NCII is Defined

While all the platforms we reviewed have some description of NCII, only TikTok, Meta, and Reddit have definitions listed (see Figure 1). Moreover, they provide varying framings around their descriptions. On Meta, X, and Snap, the framing suggests that a revenge or harassment angle is required for nude content to be treated as NCII. Reddit and Meta include the term "deepfake," whereas Snapchat uses "AI-generated sexual content." Thus, even across platforms that forbid synthetic content, the language varies. These incongruent policies have a two-fold harm: survivors may struggle to find relevant information during moments of crisis, and other users may fail to fully understand their obligations to refrain from the non-consensual distribution of intimate imagery.[49]

It is worth noting that while the industry clearly struggles to define NCII and consent, so do governmental policies. As demonstrated by our walkthrough (Figure 1), language in community guidelines is entirely unique to each platform, with different taxonomies of harm outlined on their sites. The inconsistency of language and definitions thus presents an issue not only for survivors navigating platforms but also for survivors navigating the law, as definitions within the legal system across state and federal codes vary. Industry and civil society alike find the absence of a commonly workable definition to be a barrier.[50] Article 16 of the UN Cybercrime Convention offers a starting point, defining "intimate image" as a visual recording of a person over 18 made by any means that is sexual in nature, was private at the time of recording, and in respect of which the person depicted maintained a reasonable expectation of privacy.[51]

[49] Branum and Kim, Rapid Response,14–23; The eight platforms include X, Facebook, Instagram, Reddit, Discord, Pornhub, Xvideos, and OnlyFans. 22. The CDT found that across these platforms, 31 distinct terms were used to encompass NCII.
[50] Ding, Suresh, and Venkatasubramanian, "How to Stop Playing Whack-a-Mole," 12.
[51] United Nations: Office on Drugs and Crime, "UN Cybercrime Convention - Full Text," //www.unodc.org/unodc/en/cybercrime/convention/text/convention-full-text.html.
[52] Ferguson, "TIDA Stakeholder Letter."

Anticipated Improvements Under TIDA

The requirement in TIDA for platforms to use plain language in their NCII policies should improve the user's ability to navigate complex guidelines.[52] This would also help ensure that all users, not just survivors, understand that NCII is unacceptable. Also, complying with TIDA will require policies to align with the law's definition of NCII, even though platforms are not required to use the same definition, which should also enable more comparable data across platforms' transparency reports (if they choose to include NCII in those reports).

RECOMMENDATIONS

Platforms should align across industry and, in dialogue with experts, survivors, researchers, civil society, etc., develop a common interoperable definition of NCII. The common definition should also explicitly recognize AI-generated NCII and align with prevailing legal standards and definitions. The definition should be consistent throughout community guidelines and reporting pages, not just in one or the other. Moreover, a clear definition of consent, and the lack of it, should be provided.

Figure 1. Platforms' Definitions of NCII

Platform	In What Category NCII Policy Appears in Community Guidelines	Does it Explicitly Define NCII?	How Harm is Framed in Guidelines	Includes Synthetic/ AI-generated Content?	"Nudify" / Offering to Create Allowed?	Does the Definition of NCII (or Term Used By Platform) Include Consent?
Meta	**Adult Sexual Exploitation**	Yes (more clear on Stop Extortion page than community guidelines)	Sexual exploitation; consent signals (revenge, reports, verification)	Yes	Explicitly not allowed to offer	Partial - explains how they identify non-consensual
X	**Adult Content → NCII policy**	Partial - through examples, not written explanation	Non-consensual sharing framed via harassment/revenge context	No (not updated since 2021)	Not mentioned in guidelines	No
TikTok	**Safety and Civility** → Adult Sexual Abuse	Yes	Image-based sexual abuse, sextortion, non-consensual acts	Says "real or edited" but does not explicitly mention AI	Not mentioned in guidelines	Partial - explains how they identify non-consensual
Snap	**Sexual Content**	Partial - through examples, not written explanation	Sexual exploitation, harassment, and threats	Yes	Explicitly not allowed to offer	No
Reddit	**Rule 3** → **linked policy**	Yes	Non-consensual sharing, exploitation, solicitation (incl. deepfakes)	Yes	Explicitly not allowed to offer	No
Google	**Abuse Program Policies and Enforcement** → NCII	Partial - through examples, not written explanation	Private nude, sexually explicit, or intimate images or videos	No (not updated since 2023)	Not mentioned in guidelines	No (explains how they identify non-consensual in reporting page, not guidelines)

B. Survivor Re-Engagement with Abusive Content

For any online harm, reporting requires some degree of re-engagement with the abusive content, which is not unique to NCII or to platform reporting systems. What compounds the toll in NCII cases, however, is that survivors are expected to search for, track, and report to each platform where content depicting them has spread, making the process extremely retraumatizing. Beyond its psychological toll, repeated re-engagement carries practical risks: searching for one's own NCII can generate engagement signals that boost its visibility in search results and recommendation systems—amplifying the very harm the survivor is trying to contain. In cases involving intimate partner violence (IPV) or coercive control scenarios, such searches may be surfaced to abusers through stalkerware or spyware, escalating safety risks.[53] This risk is heightened with AI-generated NCII, as perpetrators can rapidly produce and repost new variants from the same source material once they become aware of survivors' efforts to remove the content.

At the same time, survivors who refrain from taking action face further harms, and for survivors never made aware that NCII of them exists, there is no choice to make at all. Survivors who disengage are left in the dark about where the abusive content exists, how widely it has spread, and whether it continues to circulate. This persistent ambiguity produces what McGlynn and Toparlak (2025) call "social rupture," an ongoing sense of vulnerability and loss of control, with chilling effects on behavior long after the initial abuse.[54] Survivors thus face a compounded dilemma: endure the trauma of re-exposure to report every known instance, refrain from reporting and live with the

anxiety of not knowing, or—when NCII is suspected but cannot be located—both at once.

Barriers to Accessing the Reporting System

For survivors who report to platforms, the first barrier is access itself. Most platforms require reporters to hold an active account to report content in-app, meaning survivors must accept terms of service and provide personal data to the platform hosting their NCII content in order to request its removal. All platforms we examined offer at least one account-free pathway, but with varying degrees of accessibility and anonymity.

- **Facebook, Instagram, and TikTok** provide dedicated standalone reporting forms accessible without an account.[55]
- **Reddit, Google, and Snapchat** also offer account-free options but subtly discourage them—Reddit prefers logged-in reports and does not accept screenshots as evidence, Google prompts users to sign in "for faster, more accurate help," and Snapchat notes that omitting a username may limit its ability to investigate.[56]

Beyond account requirements, platform architecture shapes what can be reported and removed at all. Encryption directly affects the surface area available for NCII reporting and removal: Instagram's recent decision to remove end-to-end encryption from its messenger capabilities will expand that surface area, though some advocates have cautioned this comes at a cost to user privacy.[57] As such, infrastructure decisions, alongside platform design, ultimately determine whether survivors can access support.[58]

Navigating Multiple Reporting Pathways

Survivors often encounter multiple reporting pathways—a standard in-app or in-browser flow for general community standards violations, and a specialized form for NCII—without clear guidance about which to use.

- **Facebook, Instagram, and Snapchat** offer multiple reporting pathways but provide no clarity about whether one receives faster review, more specialized attention, or different moderation criteria. Snapchat goes further by asking users to report both in-app and through the separate web form, noting that doing both "will help us provide a better response"—placing additional burden on survivors.
- **TikTok** directs adult NCII reports through the in-app reporting flow while reserving a dedicated form to report "child sexual abuse, non-consensual intimate imagery, or explicit deepfakes," combining the report form for CSAM and adult NCII.
- **X** routes NCII reports through a form for "Safety and Sensitive Content," providing the option, "I'd like to submit a US Take It Down Act Report."

---

[53] Other coercive control cases may involve the creation of NCII for commercial purposes in which those depicted are not being compensated.
[54] McGlynn and Toparlak, "The 'New Voyeurism'," 12.
[55] Meta, "Report Non-Consensual Intimate Images (NCII) on Meta Platforms," Meta Help Center, accessed April 2026, https://www.meta.com/help/policies/1437976901029950/; TikTok, "Report a problem," TikTok Support, accessed May 2026, https://www.tiktok.com/support/faq_detail?id=7581820702655978040&category=web_account.
[56] Reddit, "How do I report something if I don't have a Reddit account?" Reddit Help, accessed April 2026, https://support.reddithelp.com/hc/en-us/articles/360058758291-How-do-I-report-something-if-I-don-t-have-a-Reddit-account; Google, "Report a problem: Request personal content removal from Google Search," Google Search Help, accessed April 2026, https://support.google.com/websearch/contact/content_removal_form?sjid=6616573576786095416-NA; Snap, "Report a Safety Concern," Snapchat Safety Center," accessed April 2026, https://values.snap.com/safety/safety-reporting.
[57] Belle Torek, "The Rollback of Instagram Encryption, and What It Means for Survivor Safety," Safety Net Blog, Safety Net Project, May 11, 2026, https://www.techsafety.org/blog/2026/5/11/the-rollback-of-instagram-encryption-and-what-it-means-for-survivor-safety.
[58] Reducing encryption expands the surface area available for NCII detection and content moderation, but encrypted communications can be a critical tool for survivors' safety and autonomy—particularly those fleeing IPV. See NNEDV's Understanding Encryption: A Guide for Victim Service Providers (2021).

In contrast to the current system, the "no wrong door" principle holds that multiple entry points should direct survivors to the support and resolution they need regardless of which pathway they take. The principle is adapted from the crime victim rights' field and signifies "a system [that] can be created that offers a 'seamless web' of services" where "there are no wrong doors" for survivors to enter into a responsive network of help.[59] While multiple pathways can increase accessibility, they can also introduce confusion when their relationship to each other is unclear. When platforms signal (explicitly or implicitly) that one reporting method will carry more weight than another, this principle breaks down.

Admittedly, this practice often reflects a practical reality: reports filed by logged-in users, for instance, may be easier to action because the platform has a means of follow-up contact if additional information is needed to locate the NCII content. Platforms should nonetheless be transparent about these differences while ensuring that no survivor is disadvantaged simply by how they chose to come forward.

Inconsistent Categorization of NCII During the Reporting Process

Once inside the in-app reporting flow, users on most platforms must select a single category from a dropdown menu of pre-defined violations.[60] NCII cases, however, frequently overlap with other harms—doxxing, impersonation, harassment, hate speech, and threats to physical safety—particularly when identifying information accompanies the imagery.[61] The categorization systems we examined reveal striking inconsistencies in how platforms name and route NCII as a distinct harm.

- **Reddit, Google, and Snapchat** offer the most direct entry points: Reddit is the only platform using "Non-consensual intimate media" and "US Take It Down Act" as top-level categories; Google's in-browser flow routes users directly to specialized NCII reporting via "It shows a sexual image of me"; and Snapchat lists "They leaked/are threatening to leak my nudes" as a main reporting category.[62]
- **Facebook, Instagram, TikTok, and X** embed NCII within broader violation categories: Facebook and Instagram route reports through "Bullying, harassment or abuse" or "Adult content" with NCII-specific sub-options, while X bundles NCII under "Private or Non-Consensual Content" alongside other privacy violations; TikTok routes NCII through "Violence or abuse" and "Sexual content."

These categorical inconsistencies create a critical usability failure. Trauma-informed interface design calls for minimizing friction and cognitive demand when users are vulnerable and seeking help—yet survivors must decode each platform's idiosyncratic reporting schema at the moment of crisis when clarity matters most.[63] Through this lens, explicitly naming NCII and relevant legal frameworks like TIDA as main reporting categories—as Reddit demonstrates—is a promising way of affirming to survivors that their experience is recognized and legally actionable.

[59] Dr. Marlene Young and John Stein, "History of the Crime Victims' Movement in the United States | Office of Justice Programs," 2004, https://www.ojp.gov/ncjrs/virtual-library/abstracts/history-crime-victims-movement-united-states.
[60] See Figure A3 in the Appendix for a list of all possible reporting categories a survivor is presented with in each platform we examined.
[61] Flynn et al., "Content Moderation and Community Standards," 1343.
[62] Note: this language implies the nude was in possession of the survivor and leaked, not that a perpetrator used AI to create it.
[63] eSafety Commissioner Australia, "Technology, gendered violence and Safety by Design," 2024, https://www.esafety.gov.au/sites/default/files/2026-02/Designing-for-Safety_Preventing-child-sexual-exploitation-and-abuse-online_CSEA-toolkit.pdf?v=1779455842782.

Descriptive Input

Some platforms offer free-text fields where survivors can provide testimony or contextual details.

- **X**'s "US Take It Down Act Report" form prompts users to "please provide more details about what's happening."
- **Snapchat**'s in-app reporting flow and web form instructs reporters to provide a "description of the violation," and to "try to be as detailed as possible."
- **Google** Search includes an optional space for "helpful context" in both the specialized reporting flow and in-browser reporting process.

From a trauma-informed perspective, free-text fields honor survivor agency by creating space to convey the layered nature of abuse in ways that dropdown menus cannot capture. However, platforms should be transparent about whether such fields are optional, how information will be used, and whether it will be reviewed by a human or automated system. Without that clarity, survivors cannot know whether the effort and potential retraumatization of providing context will affect their case; if it does not, asking for it is not only a waste of their time but risks doing further harm.

Post-Report Communication and the Absence of Human Support

After submitting a report, survivors enter a distressing phase of waiting with little information about the decision-making process. Some platforms acknowledge the emotional difficulty in their automated responses: Reddit thanks users for "looking out for yourself and your fellow redditors" making "Reddit a better, safer, and more welcoming place for everyone," and X notes "we know it wasn't easy." However, such language can come across as disingenuous when survivors receive no substantive follow-up. Refuge's 2022 study found that over half of those who reported NCII abuse received no response from the platform at all, and 41% reported content more than three times in an attempt to elicit a reply.[64]

No platform we examined provided a clear timeline for resolution or a channel for follow-up questions. Meta, TikTok, and X each offer vague commitments: Meta states a "specialized team will review your report as quickly as possible," X promises notification after "a few days," and TikTok says it will "review and take action, if there is a violation" with no clear description of such actions. The lack of a clear timeline or communication channel for follow-up questions leaves survivors unable to know whether their report has been seen, is under review, has been deprioritized, or closed without action.

More fundamentally, the entire reporting process lacks meaningful human connection. Survivors navigating profound violations of their privacy, autonomy, and safety encounter interfaces that are automated, impersonal, and nontransparent. While some automation at the first level of review is necessary and appropriate given the volume of content platforms must moderate, Li et al. found that survivors wanted more human involvement in sensitive decision-making, not less.[65] Platform employees, however, may not be best positioned to provide this deeper support. Beyond the specialized training required to provide trauma-informed assistance, platforms face an inherent conflict of interest: their interpretations of community standards violations may not align with how survivors define or experience NCII.

[64] Refuge, Marked as Unsafe, 7.
[65] Li et al., "Platforms as Crime Scene, Judge, and Jury," 24.

This underscores the critical role of third-party victim-serving organizations—such as the Cyber Civil Rights Initiative and the UK Revenge Porn Helpline—as independent bridges between survivors and platforms that offer psychosocial support, safety planning, and direct platform outreach. Yet their impact depends on survivors being able to find them.

Linking to external resources for counseling, legal guidance, or regional image takedown support is a best practice called for by survivors and advocates, but coverage across the platforms we examined remains inconsistent (see Figure 2).

Google Search provides the most comprehensive list of external resources while X notably offers none at all. Only four of the seven platforms cite the Cyber Civil Rights Initiative, and only Meta platforms refer users to the US National Domestic Violence Hotline—a significant gap given the well-documented overlap between NCII and intimate partner violence, where survivors may need support extending well beyond image removal.[66] Until such referrals to specialized support become a consistent standard rather than a platform-by-platform choice, the reporting process will remain incomplete.

Figure 2. External NCII Support Resources Referenced by Platforms

Platform	Where Resources Appear	External NCII Support Resources Listed
Facebook and Instagram	Post-report screen, **Meta Help Center**, **Crisis support resources**	Image removal & prevention: StopNCII.org, TakeItDown.ncmec.org, Thorn Reporting navigation & legal guidance: Cyber Civil Rights Initiative, Lila.help Crisis & emotional support: Crisis Text Line, National Domestic Violence Hotline, Love Is Respect
X	No support resources	No support resources
TikTok	**TikTok Community Support Resources**	Image removal & prevention: StopNCII.org, TakeItDown.ncmec.org Crisis & emotional support: US National Sexual Assault Hotline (and country-specific equivalents) Advocacy & education: Futures Without Violence
Snapchat	Post-report screen, **Safety Resources and Support on Snapchat**, Snapchat Safety Center: **What you need to know about financial sextortion**	Image removal & prevention: StopNCII.org, TakeItDown.ncmec.org Reporting navigation & legal guidance: Revenge Porn Helpline Regional resources: "Sexual Risks & Harms Resources"–country-based guides covering both CSAM and adult NCII
Reddit	**Reddit Help: Responding to non-consensual sharing of intimate media**	Image removal & prevention: StopNCII.org Reporting navigation & legal guidance: Cyber Civil Rights Initiative Crisis Helpline and International Resources Crisis & emotional support: Crisis Text Line
Google	Post-report screen, **Google Search Help Center**	Image removal & prevention: StopNCII.org Regional resources: North America (Cyber Civil Rights Initiative), Asia (Digital Rights Foundation, SG Her Empowerment, Women's Human Rights Institute of Korea), Europe (Revenge Porn Helpline)

Anticipated Improvements Under TIDA

Though TIDA does not address most of the issues outlined above for post-report communication, the 48-hour required window to remove content reported as NCII will undoubtedly improve the current, open-ended response survivors receive after submitting a report.

RECOMMENDATIONS

- **Allow non-account holders to report within platforms:** Survivors who have left a platform for safety reasons, or who never used it, should not be locked out of seeking support. At minimum, all platforms should provide a dedicated, clearly labeled NCII reporting form accessible without an account.
- **Signpost dedicated NCII reporting channels:** Where multiple reporting methods exist, platforms must make clear through repeated signposting that there is no wrong door to support and eliminate pressure to report through multiple channels. Specialized NCII forms should be integrated directly into in-app reporting when a user selects NCII as their issue, or the option should be clearly surfaced at the point of reporting.[67]
- **Allow multi-category reporting to capture overlapping harms:** NCII rarely occurs in isolation. Reporting systems should allow survivors to select multiple harm categories simultaneously, alongside an optional free-text field, ensuring the full scope of abuse is captured rather than forcing survivors to choose a single category that may not reflect their experience.
- **Invest in trust and safety infrastructure:** Platforms must design the reporting flow, communication during review, and post-decision follow-up with an assumption that behind every NCII report is a person in distress. This requires sustained investment in trained human reviewers at a moment when major platforms are pivoting to automation, and trust and safety teams have faced mass layoffs.[68] Resourcing these teams also means addressing who builds them: the underrepresentation of women and marginalized groups in leadership, engineering, and product design perpetuates the gender bias that NCII disproportionately exploits.[69]
- **Embed support resources and establish trusted relationships with victim-serving organizations:** Platforms should include links to helplines and specialist services directly within reporting flows and post-report screens, and recognize these organizations as trusted flaggers whose reports are escalated for review. Such partnerships can strengthen both the quality of platform response and support available to survivors.
- **Integrate trauma-informed, human-in-the loop communication practices:** All reporting-related communications should use clear, plain, non-judgmental language. Platforms should provide explanations when content is not removed, offer accessible appeal mechanisms, and ensure survivors can escalate from automated systems to human review. Timely case updates—drawing from victim-notification best practices in the criminal legal system—would represent a step toward rights-based communication frameworks and accountability that current systems lack.

[66] Nicole Henry and Gemma Beard, "Image-Based Sexual Abuse Perpetration: A Scoping Review," Trauma Violence & Abuse 25, no. 5 (2024): 3994. https://doi.org/10.1177/15248380241266137.

[67] The suggestion here is not to eradicate anonymous reporting options, but to either integrate them in-app when survivors click NCII as the issue they are reporting, or inform the survivor of the option.

[68] Rachel Elizabeth Moran et al., "The End of Trust and Safety?: Examining the Future of Content Moderation and Upheavals in Professional Online Safety Efforts" in Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (ACM, 2025), 1–14, https://doi.org/10.1145/3706598.3713662.

[69] SWGfL, Model National Framework for Addressing Non-Consensual Intimate Images, 8. https://swgfl.org.uk/resources/model-national-framework-for-addressing-non-consensual-intimate-images/.

C. Filtering - Hiding Harm Only From the Survivor

After a report is submitted, the interim measures platforms offer are limited. TikTok and X state that reported content will no longer appear on the reporter's feed or timeline but will remain visible to others while Meta, Snapchat, Google, and Reddit do not address what will happen with the content in the interim. Similarly, Reddit and X allow reporters to immediately block the posting user, and TikTok offers the option to mute them, while Meta, Snapchat, and Google offer no such options.

Consequences of Only Removing Content for the Reporter

Leaving content viewable on other users' feeds until a final moderation decision is reached is a harmful and avoidable default setting. X's 2023 "Freedom of Speech, Not Reach" policy frames this as a middle ground between the "binary 'leave up versus take down' approach," reducing visibility of potentially violative content rather than removing it right away.[70] This approach likely reflects First Amendment concerns about restricting potentially protected speech—a framing that privileges the expressive rights of the potential perpetrator over those of the survivor. As Citron and Penney (2019) have argued, NCII abuse operates as a form of silencing, driving survivors off platforms.[71] Hiding reported content from public view during review would better balance these competing interests, significantly limiting the reach and harm of the flagged post while a decision is made.

Some platforms offer the option to block the user once the survivor has reported their content, but blocking is an imperfect short-term measure that also puts the onus on the survivor. Li et al. found that after a survivor used repeated blocking as a tactic against her perpetrator, Reddit prevented her from filing a report against them—a design failure that punished the survivor for seeking self-protection[72] Both filtering and blocking imply that only the survivor experiences harm from the content, obscuring its wider social harm. Broader gender-based violence (GBV) research has found evidence that social and gender norms that promote misogyny and violence against women and girls create risk for victimization and perpetration. Recent studies have demonstrated a link between exposure to online misogyny and acceptance of traditional gender roles that tend to normalize gender-based violence. A 2025 survey by King's College London and Ipsos reveals a marked regression in attitudes on gender roles and the status of women and girls in society among Gen Z men and boys, attributed to a lifetime of exposure to social media.[73] Kiesler et al. (2012) suggest that people learn the norms of an online community by observing behaviour and its consequences.[74] Filtering violative content from individual feeds rather than removing it from the platform overall may therefore reduce opportunities for social learning and positive bystander prevention, inadvertently limiting the organic enforcement of norms that deters future abuse.

[70] X Safety, "Freedom of Speech, Not Reach: An Update on Our Enforcement Philosophy," X Blog, April 17, 2023. https://blog.x.com/en_us/topics/product/2023/freedom-of-speech-not-reach-an-update-on-our-enforcement-philosophy.
[71] Danielle K. Citron and Jonathon W. Penney, "When Law Frees Us to Speak," Fordham Law Review 87 (2019), https://scholarship.law.bu.edu/faculty_scholarship/632/.
[72] Li et al., "Platforms as Crime Scene, Judge, and Jury," 12.
[73] King's Global Institute for Women's Leadership and Ipsos UK, Gen Z Men and Women Most Divided on Gender Equality, Global Study Shows (King's College London, March 5, 2025), https://www.kcl.ac.uk/news/gen-z-men-and-women-most-divided-on-gender-equality-global-study-shows.
[74] Li et al., "Platforms as Crime Scene, Judge, and Jury," 10.

Anticipated Improvements Under TIDA

Hash-matching technology (through initiatives like StopNCII—see Figure 3 for more information) enables industry to share verified NCII across platforms, creating an even stronger case for immediate removal rather than leaving content visible during deliberation, as content will be identified as NCII by the survivor. Meta announced they will begin sharing the hashes of the verified non-consensual images they remove by feeding those hashes into StopNCII, in order to remove images faster and "prevent the re-sharing of these images across different online platforms, even if someone hasn't pre-emptively uploaded the hashed image themselves to StopNCII.org."[75]

Given Meta's role in contributing technical engineering, funding, and policy support to facilitate StopNCII's hash-matching infrastructure in 2021, this new development to proactively share hashes suggests the catalytic impact of increased government scrutiny and regulation on industry to address NCII.[76] Recent guidance from the FTC on TIDA compliance explicitly directs covered platforms to "consider sharing your hashes" with StopNCII.org for images and videos of adults, in order to "prevent the reappearance of intimate content you already removed from your platform."[77] Additionally, in May 2026, Ofcom in the UK officially recommended that tech firms use automated detection technology and StopNCII to reduce the spread of NCII.[78] While Meta's announcement about sharing hashes with StopNCII is a promising step, hash-sharing remains voluntary and inconsistent across platforms.

RECOMMENDATIONS

- **Adopt a "remove-first, verify-later approach" during the review process:** Allowing reported content to remain visible while under review can prolong harm for both survivors and potential viewers. Rather than relying on downstream filtering or suppression of flagged content, platforms should prioritize immediate removal of NCII reports to prevent further dissemination and compounding harm.
- **Scale hash-sharing across platforms following existing models:** Platforms should adopt (and regulators should require) tested technical interventions for tracking and verifying NCII through common hash-matching databases, such as StopNCII. Universal platform adoption of hash-matching infrastructure would enable survivors to convert NCII content into a hash just once, notifying all participating platforms, without having to repeatedly re-engage with it. Since most platforms have not yet opted into such systems, survivors currently report within each platform individually, multiplying both the psychological burden and the practical risks of re-exposure.

[75] Meta, "Intimate Image Abuse and Sextortion," Meta Safety Center, accessed April 2026, https://www.meta.com/safety/topics/bullying-harassment/nci-i/?srsltid=AfmBOoqVjoQE4qvrdlR3u9o-3g75Ibhihrl7_mrbmBlhUj43aP3R1BNw.
[76] Becca Branum, Aliya Bhatia, and Belle Torek, Digital Fingerprints, Human Stakes: Governing NCII Hash-Matching (Center for Democracy and Technology, April 9, 2026), 22, https://cdt.org/insights/digital-fingerprints-human-stakes-governing-ncii-hash-matching/.
[77] Ferguson, "TIDA Stakeholder Letter."
[78] Ofcom, "Platforms Should Use Detection Technology to Stop Spread of Illegal Intimate Images Online, under Strengthened Ofcom Codes," May 18, 2026, https://www.ofcom.org.uk/online-safety/illegal-and-harmful-content/platforms-should-use-detection-technology-to-stop-spread-of-illegal-intimate-images-online-under-strengthened-ofcom-codes.

Figure 3. Overview of StopNCII.org

StopNCII.org: A Survivor-Centered Tool for Cross-Platform NCII Removal [79]

What: StopNCII.org, operated by the UK charity South West Grid for Learning (SWGfL), is a cross-platform hash-matching service that enables survivors to proactively prevent the non-consensual sharing of their intimate images and videos. The system generates a unique "digital fingerprint" or "hash" of a known image directly on the survivor's device, transmitting only that non-identifying code—not the image itself—to participating platforms to detect and remove matching content and prevent future uploads.[80] This provides a pathway for survivors to reclaim control over their images.

Strengths: Privacy-preserving, non-invasive model limits who must view the content during the takedown process, reducing retraumatization; Creates a pathway for survivors to reclaim control over the process; Accepts verified hashes of NCII from partnered platforms and survivor service providers (NGOs) participating in the NCII Global Clearing Center to expand its repository and enable cross-platform coordination to contain the spread of NCII at scale.[81]

Limitations: Reach is limited to platforms that voluntarily participate; StopNCII cannot circumvent the role of platforms in helping survivors remove images after they've already been posted and circulated without consent; Screenshots and AI-manipulated variants may require survivors to submit multiple hashes, and video hashing in particular has to be precise (edited videos will not be captured through a hash of the original content); Can only hash media that survivors can access and submit; Cannot help survivors locate content they suspect exists but cannot find, nor detect content in private or encrypted channels.[82]

StopNCII is an important mechanism for containing the spread of NCII across platforms. However, it does not eliminate the need for more comprehensive, survivor-centered NCII reporting, removal, and prevention infrastructure within platforms themselves, nor does it undercut the need for initiatives like SWGfL's NCII Global Clearing Center, which enables trusted NGO partners to directly access and use the StopNCII system.

[79] This figure was shared with the StopNCII team for review prior to publication to ensure factual accuracy. Their feedback is reflected here.
[80] Branum, Bhatia, and Torek, Digital Fingerprints, Human Stakes.
[81] SWGfL, "Introducing the Global Clearing Centre: Strengthening the Global Response Towards Intimate Image Abuse," 2025, https://swgfl.org.uk/magazine/introducing-the-global-clearing-centre-strengthening-the-global-response-towards-intimate-image-abuse/.
[82] Branum and Kim, Rapid Response; While StopNCII specifies that AI-generated images can be uploaded, its messaging—including phrases in their explainer video like "are you worried someone might share your intimate images?"—centers on leaked private images and should be updated to explicitly encompass AI-generated NCII.
[83] X, Global Transparency Report: H2 2024 (X Corp., 2024), https://transparency.x.com/en/reports/global-reports/2025-transparency-report; Reddit, Transparency Report: January to June 2025 (Reddit, Inc., 2025), https://redditinc.com/policies/transparency-report-january-to-june-2025-reddit.
[84] Center for Countering Digital Hate, Grok Floods X.
[85] Meta, Community Standards Enforcement Report: Adult Nudity and Sexual Activity (Meta Platforms, Inc., 2025), https://transparency.meta.com/reports/community-standards-enforcement/adult-nudity-and-sexual-activity/.
[86] X, "Non-Consensual Nudity Policy," X Help Center, accessed April 2026. https://help.x.com/en/rules-and-policies/intimate-media.

D. Insufficient Transparency and Accountability

Transparency Reports

Transparency reports offer one of the few windows into platform content moderation activity, but as currently structured, they obscure more than they reveal about NCII. The overwhelming majority of platforms reviewed do not disaggregate NCII as a distinct category in their public reporting, instead burying it within broad categories like "Adult Nudity and Sexual Activity" (Meta), "Adult Sexual and Physical Abuse" (TikTok), or "Sexual Content" (Snapchat)—none of which distinguish between consensual and non-consensual material, let alone AI-generated NCII. Only X and Reddit clearly separate NCII: X reported 76,392 non-consensual nudity reports in the second half of 2024, and Reddit documented 137,536 reports of non-consensual intimate media in the first half of 2025, with action taken on 49.7% of cases.[83]

Even where disaggregated data on NCII exists, significant gaps undermine its value. For example:

- **X**'s most recent report notably does not cover the GrokAI deepfake incident between December 2025 and January 2026.[84]
- **Meta**'s report between October-December 2025 shows that users appealed 894,000 Meta enforcement actions for adult nudity and sexual activity content, of which 750,000 pieces of content were later restored.[85] The striking rate of content restoration raises questions about the accuracy of initial enforcement.

Without standardized and disaggregated reporting, researchers, policymakers, and advocates have no reliable basis for tracking whether platforms are improving, stagnating, or regressing in their response to NCII.

Transparency with Survivors During the Reporting Process

Beyond what platforms disclose publicly, there is a parallel accountability failure in what they communicate to survivors who file reports. The majority of platforms offer no visibility into whether a perpetrator's account has been suspended, whether re-uploads have been detected, or what measures exist to prevent recidivism. X promises to "immediately and permanently suspend any account that we identify as the original poster of intimate media that was created or shared without consent," with temporary lock-outs after a first warning and permanent suspension for repeated violations.[86] On paper, this is the most specific perpetrator accountability policy of any platform reviewed. However, there is no way of knowing the extent to which X reliably enforces this. None of the platforms reviewed outline a mechanism for holding perpetrators accountable should they create new accounts to continue posting NCII, leaving a significant gap in accountability. Platforms should standardize and clearly communicate the penalties for distributing and creating NCII during the reporting process itself, and demonstrate in their transparency reports what was done to prevent long-term harm. Timely updates to survivors on the status of their cases would improve the accountability that current systems lack.

Anticipated Improvements Under TIDA

While TIDA does not explicitly require platforms to publish transparency reports, it empowers the FTC to file civil suits against non-compliant platforms—creating indirect incentives for platforms to document and disclose how they identify, verify, and remove NCII in order to demonstrate compliance.

RECOMMENDATIONS

- **Publish disaggregated transparency data specific to NCII reporting:** Platforms should report NCII incidents as a discrete category; disaggregate AI-generated from authentic NCII; include removal rates for verified NCII, appeal outcomes, and number of accounts suspended. As TIDA enforcement begins, platforms should also disclose the percentage of reported content that was reviewed and found not to be NCII, given the law's lack of safeguards against bad faith reporting.
- **Clearly communicate and enforce accountability measures for users who violate platform terms of service by posting NCII:** Platforms should outline when and how perpetrators' accounts are removed or permanently disabled in order to shape long-term practices of discouraging NCII creation and distribution among its users.
- **Ensure hash-matching transparency and accuracy:** Platforms who use hash-matching systems should publish transparent data on verified NCII removed compared to total reports received, and build review mechanisms to catch mistaken matches—ensuring the system's benefits for survivors are not undermined by erosion of trust in its accuracy.

6. Recommendations Beyond Platforms: Extending Responsibility to the Full Ecosystem

In our analysis, we found that the ongoing issue is that platform-level measures remain fundamentally reactive, focused primarily on hiding content as the central remedy for survivors. Stronger safety-and-privacy-by-design practices will be more effective if implemented in tandem with reducing the demand and audience for NCII in the first place. Eliminating AI-generated NCII will require interventions with a "multistakeholder, whole-of-society lens."[87] This begins with legislation and enforcement by regulatory agencies, the criminal legal system, and compliance and alignment among social media and AI companies' values, definitions, processes, and design.

That work demands a more holistic, ecosystem-wide approach to tackling image generation, distribution, and monetization—a key gap in TIDA, the UK's Online Safety Act, and other recent country-level laws aimed at curbing image-based sexual abuse. As others have rightly critiqued, while TIDA importantly criminalizes the sharing of AI-generated image-based sexual abuse of identifiable persons alongside authentic NCII, it fails to criminalize its creation. It also does not explicitly extend image removal requirements to AI developers and deployers, which are not clearly included under the law's definition of covered platforms (i.e., they do not host user-generated content in the same way as social media platforms or search engines).

The platforms we examined are just a part of the wider ecosystem that allows NCII to persist. Regulators can and should address the multilayered ecosystem of harm.

Important next steps include:

- **Regulate and enforce accountability for generative AI developers:** Standardize and require guardrails to reduce the risk of NCII creation, which is technically feasible through feedback loops, iterative stress-testing strategies, semantic guardrails, responsibly sourced training data, etc.[88]

- **Regulate and enforce accountability for infrastructure providers** responsible for hosting deepfake generation AI models, payments for NCII, distribution via easily discoverable forums; it is necessary for regulators to shut down the monetization of NCII and ban nudification apps (as EU, UK, and some US states have proposed to do).[89]

- **Policymakers should support international regulatory harmonization:** Congress should update TIDA and related legislation to address the creation of NCII, in addition to distribution, as the UK and South Korea have done.[90] This will help hold platforms to a consistent global standard.

- **Policymakers should emphasize cultural and societal prevention strategies** including funding and formal educational programs that educate the public about NCII, building core values around privacy, bodily autonomy, consent, and gender equality.[91]

[87] SWGfL, Model National Framework.
[88] Ibid.; Humane Intelligence, Foreign, Commonwealth & Development Office, Department for Science, Innovation and Technology, and the Global Partnership for Action on Gender-Based Online Harassment and Abuse, Digital Violence, Real World Harm: Evaluating Survivor-Centric Tools for Intimate Image Abuse in the Age of Generative AI (GOV.UK, 2025), https://www.gov.uk/government/publications/digital-violence-real-world-harm-evaluating-survivor-centric-tools-for-intimate-image-abuse-in-the-age-of-generative-ai.
[89] Minnesota state law HF 1606 passed in 2026 includes a provision to ensure companies would not be liable for general products by including an exemption from liability for companies where the "technical skill of a user" is required to edit or manipulate an image, like Adobe's photoshop.
[90] The United Kingdom's Data Use and Access Act and South Korea's Sexual Violence Prevention and Victims Protection Act of 2024.
[91] SWGfL, Model National Framework.

7. Conclusion

We conducted our research at a turning point in the landscape of combatting non-consensual intimate image abuse. TIDA and similar laws globally represent progress in establishing platform accountability for the spread of NCII. Our walkthrough analysis identifies common strengths and critical gaps in platform management of NCII reporting and takedown—validating survivors' and advocates' longstanding concerns that, without legal requirements or strong improvement incentives, the status quo has been wholly insufficient.

The findings are consistent across platforms. Most do not offer a dedicated NCII reporting category in in-app reporting flows, forcing survivors to route reports through ambiguous catch-all categories like "adult content" or "nudity and sexual activity." Crucially, not all platforms we examined include AI-generated content in their definition of NCII, or forbid offering to create or advertise synthetic AI. Where multiple reporting pathways exist, platforms fail to clarify whether certain channels receive faster review, more specialized attention, or different moderation criteria. Survivors must decode platform-specific language and navigate branching decision trees while managing the spread of their images across the internet.

This design violates a basic principle of trauma-informed interface: when users are vulnerable and seeking help, systems should minimize friction and cognitive demands, not require survivors to become fluent in each platform's idiosyncratic schema at the moment of greatest need.

Cross-platform coordination through StopNCII.org offers a meaningful but incomplete remedy: voluntary participation limits its reach, and it cannot substitute for platforms' own responsibility to remove content already circulating without consent. TIDA and related legal developments serve as incentives for companies to adopt hash-matching systems. Most platforms still fail to disaggregate NCII in their transparency reports, making it impossible for researchers, policymakers, or survivors themselves to assess whether their system is improving.

TIDA's focus on platform reporting and takedown is a necessary first step, but it is insufficient to produce the ecosystem-level change needed to reduce the risk for harm in the first place. The broader wave of social media litigation represents an encouraging development that could further shift platform incentives toward prevention rather than reaction, predicated on employing a consumer product safety lens that promotes safer online experiences.

The Grok xAI incident illustrated why such a shift is necessary: in March 2026, three Jane Does filed a class-action lawsuit against xAI, arguing that NCII content depicting them was created and processed through the company's own systems rather than simply uploaded by users.[92] This case highlights that platform accountability must extend beyond content moderation to the design decisions that allow for NCII to spread.

Recognizing NCII as a preventable and addressable harm is more urgent than ever as AI capabilities continue to lower the barriers for perpetrating abuse. We urge platforms to honor their own stated commitments to user safety: designing reporting processes that center survivor experience, investing in trust and safety infrastructure, standardizing NCII definitions and transparency reporting, and building proactive detection systems that are not dependent on survivors identifying and reporting every instance of their own abuse. Curbing NCII will require coordinated action not only across platforms but across the larger sociotechnical ecosystem—from AI model repositories to search engines to payment processors. We hope this paper contributes to that effort by making the case for more proactive, survivor-centered reporting processes and policies that meet the scale and urgency of the harm.

[92] Annika K. Martin et al., "Class Action Complaint," United States District Court Northern District of California San Jose Division, March 16, 2026, https://cdn.arstechnica.net/wp-content/uploads/2026/03/Doe-v-xAI-Complaint-3-16-26.pdf.

Appendix

A1. Platform User Demographics and How Companies Describe Their Product Experience

Meta (Instagram/Facebook) - In the United States, 62.3% of Instagram's users are reported to be between 18-34 and 55.4% users are female.[93] Facebook's user base is likewise largely composed of users aged 24-35[94] Meta states that it has "responsibility to promote the best of what people can do together by keeping people safe and preventing harm."[95] In a 2024 study by Pew, a majority of Facebook and Instagram users said they use the apps to connect with family and friends.[96]

Reddit - 44% of US Reddit users are aged 18–29, with 59.8% of Reddit users identifying as male.[97] Reddit describes itself as for everyone with a mission to "empower communities and make their knowledge accessible to everyone."[98] 72% of US Reddit users use the platform primarily for entertainment.

Snapchat - Snapchat has an even younger target demographic, with 90% of users aged 13-24.[99] Snapchat describes itself as an app that "empowers people to express themselves, live in the moment, learn about the world, and have fun together" and "the easiest and fastest way to communicate the full range of human emotions with your friends without pressure to be popular, pretty, or perfect."[100]

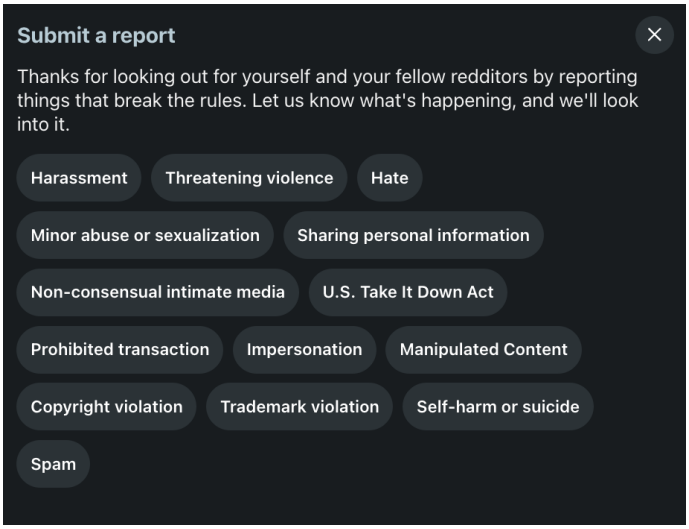
TikTok - 66% of TikTok users are aged 18-34, with 55.7% of the platform's users being male.[101] TikTok's stated mission is to "inspire creativity and bring joy," marketing itself as a space for "endless discovery" that is "for you, by you."[102] In Pew's 2024 study, 95% of respondents cited entertainment as their main reason for use and 71% of TikTok users said that harassment is a problem on the platform.[103]

[93] "Instagram Demographics in 2026: Key Audience and Creators Stats," Phyllo, February 2022, 2026, https://www.getphyllo.com/post/instagram-demographics-audience-creators-stats.
[94] Ibid
[95] Meta, "Company Info," Meta.com, accessed April 2026, https://www.meta.com/about/company-info/.
[96] McClain, Anderson, and Gelles-Watnick, "How Americans Navigate Politics."
[97] Cole Furrh, "Reddit Statistics 2026," InterTeam, January 28, 2026, https://www.interteammarketing.com/blog/reddit-statistics-2026.
[98] Reddit, "Reddit Audience Insights," Reddit for Business, accessed April 2026, https://www.business.reddit.com/audience-insights.
[99] Megan Morreale, "Snapchat Statistics for 2025: Usage & Trends," Sprout Social, June 24, 2025, https://sproutsocial.com/insights/snapchat-statistics/; Note: Though this paper examines adult NCII, we note that Snapchat is the most frequently cited platform where minors experience image-based sexual abuse.
[100] Snap, Inc., "Snapchat: Chat with Friends," App Store, accessed April 2026, https://apps.apple.com/rs/app/snapchat-chat-with-friends/id447188370.
[101] Jacqueline Zote, "46 TikTok Stats to Inform Your 2026 Strategy," Sprout Social, March 9, 2026, https://sproutsocial.com/insights/tiktok-stats/.
[102] TikTok, "About," TikTok.com, accessed April 2026, https://www.tiktok.com/about?lang=en; TikTok, "TikTok - Videos, Music & LIVE," App Store, accessed April 2026, https://apps.apple.com/sr/app/tiktok-videos-music-live/id835599320.

X - On X, 69.6% of users are aged 18-34, with more than half being 25-34, and male users make up 64.4% of the platform's audience.[104] X describes itself as "your trusted digital town square where conversations unfold in real time" and promises content delivered "raw and unfiltered."[105] 81% of users in the Pew study said they use X for entertainment, but notably in comparison to Facebook, Instagram, and TikTok, more users said they go to X to keep up with politics and 73% noted harassment as an issue on the app.[106]

Google Search - Users aged 25-34 dominate Google traffic, accounting for 26.85%, followed by the age group of 18-24 contributing 21.12% of visits.[107] Google states they are committed to "significantly improving the lives of as many people as possible" and deliver reliable and relevant information in the "most useful" way.[108]

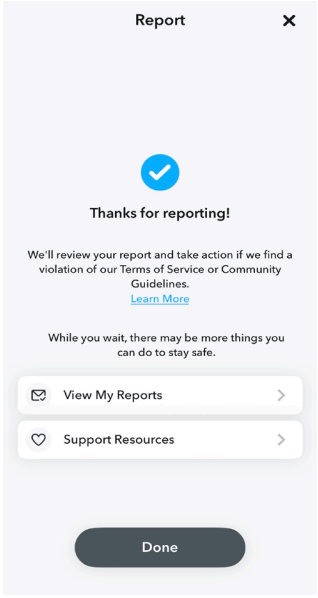
A2. Post-Report Screens



A2.1 X Post-Report Screen

Note: Screenshot from walkthrough of X's in-app reporting process captured on May 18, 2026. Post-report screen does not link external resources, states to the reporter that the post is removed from their timeline but not others'.

Date of Capture: May 18, 2026



A2.2 Snapchat Post-Report Screen

Note: Screenshot from walkthrough of Snapchat's in-app reporting process captured on May 18, 2026. Post-report screen links external resources under "Support Resources," and states that no action will be taken until Snapchat can review the content.

Date of Capture: May 18, 2026

[103] McClain, Anderson, and Gelles-Watnick, "How Americans Navigate Politics."
[104] Jacqueline Zote, "27 Twitter (X) Stats to Know in Marketing in 2026," Sprout Social, March 10, 2026, https://sproutsocial.com/insights/twitter-statistics/.
[105] X Corp., "X," App Store, accessed April 2026, https://apps.apple.com/us/app/x/id333903271.
[106] McClain, Anderson, and Gelles-Watnick, "How Americans Navigate Politics."
[107] Robert A. Lee, "Google Usage Statistics 2025: Key Trends and Data Insights," SQ Magazine, January 19, 2026, https://sqmagazine.co.uk/google-usage-statistics/.
[108] Google, "Our Approach – How Google Search Works," Google.com, accessed April 2026, https://www.google.com/intl/en_uk/search/howsearchworks/our-approach/.

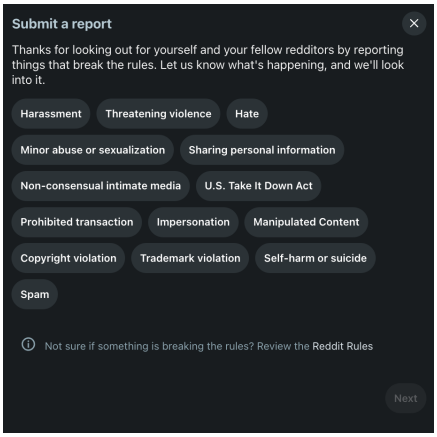
Appendix

A3. Platforms' In-App Reporting Categories

We found that the examined platforms include NCII across a number of varied harm types, with little consistency within and among platforms, likely creating confusion for survivors who report NCII across multiple sites. As of the publication date, platforms are continually changing reporting categories, making navigating such systems more complex.

Facebook	Problem involving someone under 18 Bullying, harassment or abuse Suicide or self-harm Violent, hateful or disturbing content Selling or promoting restricted items Adult content Scam, fraud or false information Intellectual property I don't want to see this Reporting specific harms: In the US, you can create a detailed report for something that contains intimate imagery.	Bullying, harassment, or abuse > Threatening to share my nude images / Seems like sexual exploitation / Seems like human trafficking / Bullying or Harassment Adult content > Threatening to share my nude images / Seems like prostitution / My nude images have been shared / Seems like sexual exploitation / Nudity or sexual activity
Instagram	I just don't like it Bullying or unwanted contact Suicide, self-injury or eating disorders Violence, hate or exploitation Selling or promoting restricted items Nudity or sexual activity Scam, fraud or spam Intellectual Property In the US, you can create a detailed report for something that contains intimate imagery.	Bullying or unwanted contact > Who is being harassed? Me / Someone Else / For both Are you under or over 18 / How is it bullying or unwanted contact? Threatening to share or sharing nude images / Bullying or harassment / Use of my image without consent / Spam Nudity or sexual activity > Threatening to share or sharing nude images /Seems like prostitution / Seems like Sexual Exploitation / Nudity or sexual activity
X	Adult Sexual Content Spam Hate, Abuse, or Harassment Child Safety Violent Speech Illegal and Regulated Behaviors Impersonation Private or Non-Consensual Content Suicide or Self-Harm Terrorism or Violent Extremism Civic Integrity	Private or Non-Consensual Content > Report content under the US Take It Down Act / Threatening to share or sharing private personal information without permission / Threatening to share or sharing a sexual, nude, or intimate photo/video of me or someone without permission / Sharing a photo/video of me that I do not want on the platform
TikTok	Inappropriate and irrelevant search I don't like it Violence or abuse Hate and harassment Sexual content Misinformation or AI-generated content Suicide and self-harm Regulated goods and activities Frauds and scams Counterfeits and intellectual property Undisclosed branded content Other	Violence or abuse (post may involve sexual or physical abuse, violence, criminal behavior, or human exploitation, including content that harms or endagers minors, such as child sexual abuse material (CSAM) or predatory behavior) > Just goes to submit, no further drop down Sexual Content (to report child sexual abuse content or sexually explicit digital identify theft affecting you or a child, submit a separate report form) Misinformation or AI-generated content (post may involve harmful misleading information, election interference, or misleading AI-generated content that could endanger people or disrupt civic processes)
Snapchat	I just don't want to see it Bullying, harassment and defamation Nudity & sexual content They leaked/are threatening to leak my nudes Threats, violence & dangerous behavior Hate speech, terrorism & violent extremism Drugs & weapons Suicide & self-harm False information or deceptive practices My IP is being infringed They are under the minimum age for using Snapchat Other	Bullying, harassment and defamation > I'm being bullied, harassed or defamed Someone else is being bullied, harassed or defamed / I'm being sexually harassed > Add a comment to your report (optional) Nudity & sexual content > It's an inappropriate Snapchat of me / It's an inappropriate Snapchat of someone else / It involves a child > Add a comment to your report (optional) They leaked/are threatening to leak my nudes > Add a comment to your report (optional)

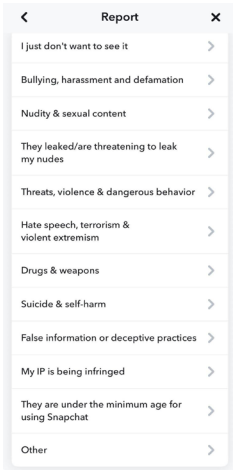
Reddit	Harassment Threatening violence Hate Minor abuse or sexualization Sharing personal information Non-consensual intimate media US Take It Down Act Prohibited transaction Impersonation Manipulated content Copyright violation Trademark violation Self-harm or suicide Spam	Non-consensual intimate media US Take It Down Act > Links to US Illegal Sharing of Non-Consensual Intimate Media Reporting Form
Google Search	It shows a sexual image of me It shows a person under 18 It shows my personal information and I don't want it there It's another kind of legal or policy violation	It shows a sexual image of me > We're here to help… directly to specialized NCII reporting flow



Reddit's In-App Reporting Categories

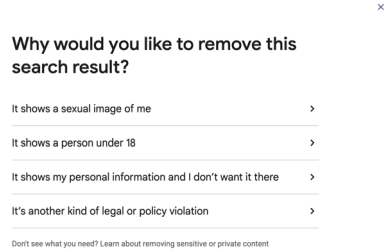
Note: While there are still several categories NCII could fall under (most explicitly "non-consensual intimate media" and "US Take It Down Act" among others), it is still best practice to have at least one clear reporting category related to NCII.

Date of Capture: May 18, 2026



Snapchat's In-App Reporting Categories

Date of Capture: May 18, 2026



Google Search's In-App Reporting Categories

Date of Capture: May 18, 2026