

FILED UNDER SEAL

EXHIBIT 1

EXPERT REPORT OF DR. BRIAN N. LEVINE

State of New Mexico v. Meta Platforms, Inc., et al.

State of New Mexico, First Judicial District Court, Santa Fe County

No. D-101-CV-2023-02838

July 11, 2025

3 Methodology

3.1 Overview

This report is based on my training and experience as a computer scientist who works to thwart internet-based child sexual exploitation. The materials from Meta that I reviewed include topics that I have researched, studied, or have experience with from my interactions with law enforcement, practitioners in non-governmental organizations (NGOs), prosecutors, policymakers, and industry professionals focused on child rescue, particularly in online settings.

Internet-based child sexual exploitation takes place in many odious forms. **Child sexual abuse material (CSAM)**, often referred to as “child pornography”,⁵ is a still image or video that captures the sexual exploitation of minors.⁶ The **commercial sexual exploitation of children (CSEC)**, also called child sex trafficking, includes the recruitment and soliciting of a person under 18 years of age for the purpose of a commercial sex act.⁷ The exchange may be for items on an Amazon.com wish list posted to Instagram, cryptocurrency exchanged online, or other forms of consideration. It may involve an in-person meet up that was arranged online. Sexual abuses to children occur in myriad ways: they may be groomed to meet in person with an adult perpetrator and raped; or they may be extorted and self-produce abuse that is captured electronically by an adult perpetrator.⁸ CSAM may be traded for years, and the trauma and degraded health can last for decades.⁹ These crimes represent violations of children’s fundamental human rights.¹⁰

Nothing has changed the nature and prevalence of CSAM- and CSEC-related crimes more than the internet.¹¹ Meta’s platforms have played and currently play a key role in that change, as demonstrated by the many tens of millions of CyberTips reported to the National Center

⁵ “CSAM” is the preferred term because the abuse captured is not pornography — the acts are non-consensual and a violation of a child’s privacy and safety.

⁶ See METANMAG-005-00508087 “Child Safety 101 for MK [Messenger Kids] Integrity”, an internal presentation by Safety Policy Director Sarah Chang, which though focused on Messenger Kids, outlines the harms of child exploitation generally and on Meta’s platforms.

⁷ See more information at <https://www.justice.gov/humantrafficking>.

⁸ See page 65 of [Levine, Report to Congress, 2022] “Practitioners who speak to victims of coerced self-produced abuse find that they will tell a story of hurt and abuse that is as severe as the experiences of victims who are physically abused in person. ... One victim said to an investigator, ‘I don’t know what it is like to be raped, but I know what it’s like to rape myself.’”

⁹ “The impact on victims is devastating and lifelong 15 times more likely to develop a psychotic disorder such as schizophrenia; 4 times more likely to abuse drugs or contact mental health officials; 3 times more likely to attempt suicide” among other impacts write Sarah Chang and Ravi Sinha in an internal presentation (METANMAG-005-00508074).

¹⁰ See “Convention on the Rights of the Child” United Nations. Adopted November 20, 1989. <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child>.

¹¹ [Levine Report to Congress, 2022].

for Missing and Exploited Children (NCMEC) and law enforcement,¹² and an untold additional number that have gone unreported or undetected. There are many internet venues that allow people to share content, such as file-sharing networks, including the popular BitTorrent network and application.¹³

But Meta's platforms are in a distinctly different class of services, as their goal is to "help people connect".¹⁴ Meta's users are not limited to CSAM file sharing. First, users can also connect with others who are interested in CSAM/CSEC. Second, Meta has a strong history of proactively recommending content and accounts related to CSAM/CSEC to its users.¹⁵ When communities of adults interested in child sexual exploitation form, they have a unique internet-based opportunity to normalize their interests, train each other on how to groom children to sexual abuse and evade law enforcement, trade CSAM, connect on other venues, and create child sex trafficking markets.¹⁶

One of Meta's overarching goals is to recruit children between the ages of 13 and 18 to its platforms via the mobile devices they carry. It is also well-documented that Meta's products are used widely by children under 13 years of age.¹⁷ Meta has elected to deploy online

¹² In 2024, Facebook 8,590,357; Instagram, Inc. 3,320,008; and WhatsApp Inc. 1,851,086 (<https://www.missingkids.org/content/dam/missingkids/pdfs/cybertiplinedata2024/2024-reports-by-esp.pdf>) ; In 2023: Facebook 17,838,422; Instagram, Inc. 11,430,007; and WhatsApp Inc. 1,389,618 (<https://www.missingkids.org/content/dam/missingkids/pdfs/2023-reports-by-esp.pdf>).

¹³ "Forensic Investigation of Peer-to-Peer File Sharing Networks," Marc Liberatore, Robert Erdely, Thomas Kerle, Brian Neil Levine, and Clay Shields. DFRWS Annual Digital Forensics Research Conference, August 2010. https://dfrws.org/sites/default/files/session-files/2010_USA_paper-forensic_investigation_of_peer-to-peer_file_sharing_network.pdf.

¹⁴ See <https://www.meta.com/about/company-info/> (last retrieved June 2025).

¹⁵ For example, "Meta Is Struggling to Boot Pedophiles Off Facebook and Instagram" Jeff Horwitz and Katherine Blunt. December 1, 2023. <https://www.wsj.com/tech/meta-facebook-instagram-pedophiles-enforcement-struggles-dceb3548>. Full text: <https://archive.ph/kMH43>.

¹⁶ See "Child Abusers Run Rampant as Tech Companies Look the Other Way," Michael Keller and Gabriel Dance, November 9, 2019. "Images of the girl surfaced in a case reported to the authorities by a woman who had been conversing with a Michigan man on Facebook Messenger. The man had proposed that the woman and her children live as nudists, while also suggesting to her that incest was normal. He offered to move in with her along with his 13-year-old daughter, whom, he said, he had orally raped the night before. The man, Snehal Yogeshkumar Shah, had also been communicating on [Meta's] Messenger with other abusers, who recommended he download the Kik messaging app and create a Dropbox account to store his illicit material. The police found more than 400 illegal photos and videos in his Dropbox account and on his iPhone, including some of the West Coast girl. They also found chats on Kik between him and two young teenagers containing explicit imagery. He is now in prison."

¹⁷ See Brooke Auxier. "7 Facts About Americans and Instagram." Pew Research Center. <https://www.pewresearch.org/fact-tank/2021/10/07/7-facts-about-americans-and-instagram/> October 21, 2020; or "The Common Sense Census: Media Use by Tweens and Teens, 2021" Victoria Rideout, Alanna Peebles, Supreet Mann, Michael B. Robb. 2022. Common Sense.

platforms where adults can connect and communicate easily with children. Children on Meta products — in their teens, tweens, and in single digits — too often find themselves feeling worse mentally, as Meta’s studies show. Unfortunately, vulnerable children are sought after by predators because they more easily groomed into sexual exploitation.¹⁸ Meta’s own analysis demonstrates that it recommends minors to adults it has classified as “groomers” seeking child sexual exploitation.¹⁹

Meta’s platforms are among the most popular on earth,²⁰ bringing them enormous revenue but also creating a perfect storm of pain and abuse for children. The combination on Meta’s platforms of community-forming tools and recommendations, the size of its user base, the mixing of adults and children, weak validation of the age of users, the presence of a very motivated and adversarial set of perpetrators, and a demonstrable lack of ability to mitigate abuse on their platforms, has resulted in a scale of CSAM/CSEC incidents never seen before.

When Meta reacts to child sexual abuse on their platform, for example, by reporting incidents as a NCMEC CyberTip, there is a chance for a child to be rescued; when Meta fails to detect or report abuses hosted on their software and computers,²¹ the chances of rescue are lower for the child. Either way, the cost of the abuse is borne by the child, their families and communities, law enforcement, prosecutors, courts, social workers, and others — but not Meta. It’s a win-win for Meta: policies and practices that do not work to restrict or detect CSAM/CSEC-related users and content tend to allow a larger user base from which to earn ad revenue, and yet Meta does not pay the cost of any resulting incidents of child sexual abuse. The large number of CyberTips that Meta has reported to NCMEC is an indication that CSAM- and CSEC-related content and activity is attracting users to the platform. And any CSAM- and CSEC-related content that was not reported quickly, not reported due to failures, or has gone undetected increases the attraction. Any content that is published or exchanged, whether reported or not, serves to normalize sexual relationships between adults and minors, whether viewed by adults or minors.

https://www.common sense media.org/sites/default/files/research/report/8-18-census-integrate-d-report-final-web_0.pdf.

¹⁸ See “The 5 Stages of Predatory Sexual Grooming,” Vinita Mehta. Psychology Today. September 28, 2022. <https://www.psychologytoday.com/us/blog/head-games/202209/the-5-stages-predatory-sexual-grooming>.

¹⁹ See METANMAG-021-00151987: “22% of the teens recommended to Groomers, received a follow request from Groomers when the recommendation was shown to them. In other words, groomers clicked on the follow button 22% of the time after we recommended a teen to them.”

²⁰ “Teens, Social Media and Technology 2024”, Michelle Faverio and Olivia Sidoti. Pew Research Center. December 2024. <https://www.pewresearch.org/internet/2024/12/12/teens-social-media-and-technology-2024/>.

²¹ See Section 5.1.9 for a discussion of the problems NCMEC has had with Meta’s CyberTip reporting.

Meta is a quintessential technology company of our time. It offers services that exist only due to software they wrote, which is only available due to their deployment of computers and systems. They control the rules of social engagement, the terms of use, who is eligible to become a user, and who is connected in groups. There is no content exchanged between members other than through the software they control and deployed, and content is not transferred other than via computers they control. When new features are present, such as recommendations of “People You May Know”, it is because Meta elected to design and deploy them. When Meta offers end-to-end encrypted (E2EE) connections that prevent it from viewing its users’ contents and reporting abuse, it’s because they have decided to do so.

Detection and response to CSAM and CSEC is an active area of research and widely practiced by law enforcement, NGOs, and companies around the world.²² It often involves techniques from computer science. Meta has deployed a variety of computational methods for detecting and responding to these incidents, including tools such as PhotoDNA,²³ that recognize near duplicate images, and machine learning tools that can sometimes recognize new CSAM or CSEC activity.

But there is no silver-bullet computational method that Meta can apply to stopping CSAM and CSEC. Meta has a yearslong history of knowing about, qualitatively and quantitatively, the proliferation and prevalence of CSAM and CSEC on their online platforms. And Meta has for years known that stopping these problems, or even appreciably mitigating them, is a deep challenge that they cannot solve computationally, particularly given the perfect storm they’ve elected to create.

If Meta’s priority was the safety and privacy of children, they would change the setup of platforms that they control. Instead, as I show in this report, their practices are aligned to a for-profit perspective, and they prefer to assign themselves a low-bar for successfully protecting children from sexual exploitation and are satisfied with the resulting costs that are paid for by others in society.

There is a long history of child sexual exploitation on Meta’s platforms. These harms are often and repeatedly discovered by third parties — law enforcement, NGOs, advocacy groups, and journalists — and they are often “rediscovered” by Meta’s own employees year after year.

It is in this context that Meta introduced algorithmically based recommendation features that act as a force multiplier for perpetrators of CSAM and CSEC. These recommendation features include suggesting to adults, who are already known to Meta to attempt to groom children into sexual relationships (and yet are still users of the platform), that they follow

²² See, e.g., “Expert Report of Rickey Kitch.” I understand Mr. Kitch also played a central role in the State of New Mexico’s investigation into the easy availability of CSAM on Meta’s platforms even in the wake of many purported “fixes” by Meta, as documented in its complaint.

²³ See “An Overview of Perceptual Hashing,” Hany Farid. *Journal of Online Trust and Safety*, 1(1). 2021. <https://doi.org/10.54501/jots.v1i1.24>.

Like),⁵⁷ suggested posts,⁵⁸ and suggested Reels.⁵⁹ A typeahead/autocomplete feature suggests related search terms.

WhatsApp is a chat app that Meta owns. Users can send each other text, images, and video. They can communicate by live voice and video calls.⁶⁰

Meta's platforms, including Facebook Messenger, WhatsApp, and Instagram's Direct Messages, all feature a technology called "end-to-end encryption" (often called "e2ee" within Meta). In April 2016, WhatsApp became the first widely used Meta-owned platform (then under Facebook) to implement default end-to-end encryption for all chats, calls, and media.⁶¹ Around the same time, Facebook Messenger introduced an optional end-to-end encrypted chat feature called "Secret Conversations," though users had to manually enable it for each chat.⁶² In 2023, Meta deployed end-to-end encryption for Messenger by default and has

⁵⁷ "When someone visits your Page, they may see a Related Pages section on the left side of your Page, which may list Pages similar to yours." <https://www.facebook.com/help/514622715244231/>; See also "Facebook Feed Recommendations AI system" Meta.com <https://transparency.meta.com/features/explaining-ranking/fb-feed-recommendations/>,

⁵⁸ "How Instagram determines which posts appear as suggested posts" Instagram.com <https://help.instagram.com/381638392275939>.

⁵⁹ "What kind of reels you may see on Instagram," Instagram.com <https://help.instagram.com/1525585517644948>. "On Instagram, you can find and view public and suggested reels in places like the Reels tab, Explore, and in your feed. You'll see reels from accounts you follow and from suggested accounts. The suggested reels you see are recommended to you based on a variety of factors, including which accounts you follow and the reels you've watched, liked or engaged with recently."

⁶⁰ WhatsApp features are used in the sexual exploitation of children. For example, "South Carolina man pleads guilty to sexually exploiting two children in Virginia," U.S. Attorney's Office, Eastern District of Virginia. March 18, 2025 <https://www.justice.gov/usao-edva/pr/south-carolina-man-pleads-guilty-sexually-exploiting-two-children-virginia>: "Schoenmann and the Minnesota man used WhatsApp to discuss their sexual interest in young girls, and to trade images of their victims. Between July 2019 and July 2022, Schoenmann created at least eight images depicting MV1 [minor victim 1] engaged in sexually explicit conduct. MV1 was between six and eight years old when Schoenmann sexually exploited her. Schoenmann also created at least eight sexually explicit images of MV2 [minor victim 2], who was between less than a year old and two years old when Schoenmann sexually exploited her. Schoenmann distributed the sexually explicit images of his victims to the Minnesota man."

⁶¹ Metz, C., *Forget Apple vs. the FBI: WhatsApp Just Switched on Encryption for a Billion People*, Wired (Apr. 5, 2016), <https://www.wired.com/2016/04/forget-apple-vs-fbi-whatsapp-just-switched-encryption-billion-people/>.

⁶² ParentZone, *Facebook's Secret Conversations: a parent guide* (Sep. 13, 2024), <https://parentzone.org.uk/article/facebook-secret-conversations>; see also, e.g., METANMAG-002-00285620 (May 2020 "Monthly Product Update" contained a mention of a "Disappearing Mode Test" for Messenger which would include e2ee in ephemeral messaging despite "decrease abuse reporting,

stated that it is committed to deploying E2EE for Instagram Direct by default.⁶³ (E2EE communications is available on both apps, but in Instagram, the user has to enable it before it is active, thus it is not “by default”.)

Meta’s child safety controls are critically based on knowing the true age of its users.⁶⁴ Meta has had a variety of methods of validating user ages.⁶⁵ In 2013, this included manual review of accounts reported to be under 13 years of age. Meta reports that since 2022, it has partnered with Yoti, an external company that has a variety of methods including validating government ID and biometrics.⁶⁶

4.2 Overview of Meta’s CSAM/CSEC detection policies on its platforms

Although technology is rapidly adopted by perpetrators who sexually exploit children, it can also be an effective tool for supporting humans working to prevent, detect, investigate, and report CSAM and CSEC crimes. That said, CSAM/CSEC content that is encrypted cannot be accessed, inspected, or reported by Meta.

decrease value of those reports, and in cases of severe harm, make it close to impossible for victims to pursue legal action.”)

⁶³ Crisan, L., *Launching Default End-to-End Encryption on Messenger*, Meta (Dec. 6, 2023), <https://about.fb.com/news/2023/12/default-end-to-end-encryption-on-messenger/>; Jeff Horwitz and Katherine Blunt, *Meta Starts Fully Encrypting Messages on Facebook and Messenger App*, The Wall Street Journal (Dec. 6, 2023), <https://www.wsj.com/tech/meta-to-start-fully-encrypting-messages-on-facebook-and-instagram-a936c4f9>. “Direct messages on Instagram also will shift to such encryption by default somewhat later, likely in the new year, according to people familiar with the matter. A Meta spokesman said the company doesn’t have a date for that.”. Full text: <https://archive.ph/zOjB1>. See also “Meta refreshes promise to roll out default end-to-end encryption in Messenger this year. Protection for Instagram will come shortly after,” Makena Kelly. The Verge. August 22, 2023. <https://www.theverge.com/2023/8/22/23841490/meta-facebook-messenger-instagram-encryption-default>: “‘We remain committed to rolling our default end-to-end encryption for private conversations on Messenger in 2023, and shortly afterwards for Instagram,’ [Meta’s deputy privacy officer, Rob] Sherman wrote. ‘End-to-end encryption is the best technology we have today to protect people’s messages, and we also see it as an important reason why people might choose to use our products over competitors’.”

⁶⁴ See, e.g., Jayakumar Ex. 46 (“Youth Safety & Well-Being Deep Dive February 2021” presentation noting “Age verification is both an independent requirement and necessary foundation for all other requirements”).

⁶⁵ See page 8 of Meta’s Third Supplemental Responses and Objections to Second Set of Interrogatories.

⁶⁶ See <https://www.yoti.com/business/age-verification/>; I have no information on which of Yoti’s specific technologies are used by Meta.

restricting users to adults 18 years of age and older, would limit the Defendant’s potential to earn revenue from showing advertisements to minors.

From my review, Meta’s policies, statements, and operational practices do not follow best practices, but merely widely used practices of tech companies. Most glaringly, Meta does not value *impact over intent* — which means that in an evaluation, what matters is the impact of a company’s products and operation, and not merely good intentions.⁸⁷ Below, I highlight Meta claims to have an “industry-leading approach”,⁸⁸ have “zero tolerance for any form of child exploitive material”, and that it has “long invested in child safety efforts”.⁸⁹ In my training and in my field, the security, safety, and privacy of systems that are designed to protect either people or data are not evaluated by the intentions of the owners; what matters is an evaluation of the actual harms observed.

Another key aspect of best practices that Meta lacks is modern *threat models*. The view that a system is secure when it protects only its users is outdated; rather, all individuals who could be harmed by a system must be evaluated as part of a threat model. For example, imagine that two adults trade CSAM via Meta’s platforms and do so through secure logins and encrypted data transfer. Although the two parties have communicated securely and confidentially, the child whose rape is captured in the CSAM has been extensively harmed — despite being neither a party to the communication nor user of the platform. And by Meta’s own reporting, there have been tens of millions of incidents of children suffering from CSAM or CSEC on their platform, while it generated revenue from its users.

This modern threat model is *entirely missing* from CEO Mark Zuckerberg’s “vision and principles around building a privacy-focused messaging and social networking platform.”⁹⁰ Meta has driven toward a service where people have “confidence that no one else can access what they share”. And meanwhile, Mr. Zuckerberg has offered dichotomous statements acknowledging that the perpetrators of child sexual exploitation on Meta’s platforms are aggressive: “We work closely with law enforcement to find and stop bad actors. Still, no matter how much we invest or how effective our tools are, this is an adversarial space.”⁹¹ As

for Kids and Teens”, September 2024. Federal Trade Commission <https://www.ftc.gov/reports/look-behind-screens-examining-data-practices-social-media-video-streaming-services>.

⁸⁷ *Design for Safety*, Eva PenzeyMoog. A Book Apart Press. 2001. N.b., I understand that “intent” may have a legal meaning in this case; I am not trying to provide an opinion on corporate intent or that of any particular individual. I use the word “intent” in this document in the way that PenzeyMoog does: as a contrast to “impact”.

⁸⁸ See <https://about.meta.com/actions/safety/onlinechildprotection/> (last retrieved June 2025).

⁸⁹ “Instagram Suggested ‘Groomers’ Connect with Minors, FTC Says” Riley Griffin and Kurt Wagner. May 6, 2025. <https://news.bloomberglaw.com/privacy-and-data-security/instagram-urged-groomers-to-connect-with-minors-ftc-says>.

⁹⁰ “A Privacy-Focused Vision for Social Networking,” Mark Zuckerberg.

March 6, 2019. <https://about.fb.com/news/2019/03/vision-for-social-networking/>.

⁹¹ Hearing before The United States Senate committee on The Judiciary, January 31, 2024

Concurrently in March 2019, Zuckerberg posted a statement that his vision for the company's products included end-to-end encryption (E2EE).¹¹⁰ Once enacted, Meta would no longer be able to detect CSAM, CSEC, or any other type of violating content based on scanning content from users, whether text or images.

Internally, employees struggled to find a way to not lose the ability to detect and respond to harms to children due to the planned E2EE deployment. These struggles reveal some of the limitations that Meta's infrastructure had in protecting children, even ahead of the deployment of E2EE. (See Section 5.1.7, below, for a detailed discussion of Meta's E2EE efforts.)

For example, in April 2019, Meta began a "proactive risk investigation into inappropriate interactions with children [IIC] with a focus on preparing for end to end encryption."¹¹¹ Meta employees hypothesized that "we may be facilitating possible groomers finding young people through following, hashtags, explore, and search."¹¹² They continued, "Currently, in order to identify something as IIC we have to know the relative age of the two parties (adult and minor) and the two must be engaged in sexualized conversation. This has been historically challenging to enforce on IG both because we do not collect age of users and the detection of sexualized conversation usually occurs in messages which receive limited review and will no longer be accessible going forward (unless user reported)" because of the planned change to E2EE.

Several experiments at that time confirmed that hypothesis and several others.¹¹³ But still, Meta did not resolve to separate adults from minors on the platform or to raise the minimum age of the platform to 18 and over. A sample of the results is as follows.¹¹⁴

- "[We are] recommending groomers follow minors": Meta was recommending teen users to accounts they labeled as "groomers" and 22% of the time, the groomers followed the teen.
- "We are facilitating adults finding minors and minor-focused content through recommendations and explore. ... [REDACTED]"

2016 in the UK to garner more than three million posts last week on Instagram alone. ... Instagram, Facebook and Twitter had done little beyond attempting to restrict it to adults even though there are under 13s unofficially on their sites." <https://www.telegraph.co.uk/news/2019/08/03/lolita-phenomenon-allowing-men-groom-teenage-girls-without-fear/> And in September 2019, the Independent reported that "Sexual grooming crimes soar by a third as paedophiles turn to Instagram". <https://www.independent.co.uk/news/uk/home-news/sexual-grooming-instagram-social-media-children-paedophile-children-sex-a9099016.html>.

¹¹⁰ <https://www.theverge.com/2019/3/6/18253472/mark-zuckerberg-facebook-letter-privacy-encrypted-messaging>.

¹¹¹ METANMAG-002-02373893.

¹¹² METANMAG-021-00151987.

¹¹³ METANMAG-021-00151987.

¹¹⁴ All quotes from METANMAG-021-00151987.

Months later, another prominent journalist easily found even more child sexual exploitation on Meta's platforms. In February 2024, reporters from the New York Times investigating Instagram found:

[M]en who treat children's Instagram pages and subscription services as menus to satisfy their fantasies. They trade information about parents considered receptive to producing and selling 'private sets' of images. A [Telegram] group with more than 4,000 members was highly organized, with an F.A.Q. page and a Google sheet that tracked nearly 700 children, identifying them by hashtags to help members find them within the long chat history. The group's logo showed a child's hand in an adult hand.²⁶⁸

In other words, the problems that Mr. ██████'s 2019 email message delineated persisted five years later. The article continued:

The Times asked the Canadian Center for Child Protection, an organization that monitors online child exploitation, to review links and other potentially illegal material posted by the Telegram groups and elsewhere. The center identified child sexual abuse imagery involving multiple underage Instagram models from around the world, as well as sexualized videos of others, including a preteen girl wearing a thong and a young teenager raising her dress to show her bikini bottom.

Meta’s CSAM and CSEC Detection Efforts are Inadequate and Mismanaged. Meta has attempted to detect and mitigate CSAM and CSEC on its platforms, but their efforts have not been adequate, particularly when considering the number of children present on the platform, which I discuss in Section 5.1.5. There are numerous examples of their mismanagement of the problem.

A November 2019 task description states that “[redacted]
[redacted]
[redacted] We have seen that there is a percentage of these
users who continue to post such content numerous times despite our attempts to educate them
and remediate their behavior. In addition, NCMEC and Law Enforcement have questioned
why we allow this behavior to persist on our platform.”²⁶⁹ As discussed below, Meta also had
a policy [redacted]

were not against policy.²⁷¹

²⁶⁸ “A Marketplace of Girl Influencers Managed by Moms and Stalked by Men” Jennifer Valentino-DeVries and Michael H. Keller, Feb. 22, 2024 <https://www.nytimes.com/2024/02/22/us/instagram-child-influencers.html>.

²⁶⁹ METANMAG-075-00074341.

270 Cog05950.

²⁷¹ Cog05950.

In November 2017, Meta reconsidered its policy of retaining for merely five days deleted direct messages on Instagram. It offered that “When CO [community operations]/ LERT/ Online Safety teams find out through separate investigation/ classifiers/ police reports that child exploitation is taking place, they search for the minor and investigate media retroactively. IG’s storage currently is such that media that isn’t reported can be removed from our servers within 5 days - however in this case, violating and illegal media will almost never be reported.”²⁷² As a solution, content was retained to 30 days. The task was not fully completed until September 2018.²⁷³

In October 2020, an employee discovered that more than 4,500 Instagram accounts were involved in the grooming of children but were not removed from the system. In January, they also found “8000+ Pages that where [sic] an SRT IIC²⁷⁴ decision was taken, but no action taken on Pages/Accounts”.²⁷⁵

In March 2021, Product Manager [REDACTED] wrote a “Sync and Priorities” email message noting the launch of a new child safety measurement metric that will “help contextualize CEI data & measure minor sexualization prevalence.”²⁷⁶ He wrote that “The current hashtag enforcement (human review in SRT) has many gaps when it comes to more hidden abuse like child safety violations. ... This playbook once created can then be a documented best practices for determining CEI hashtags for SMEs and other ETEs working on child safety issues. The ultimate impact of this work will enable us to enforce on all CEI hashtags we previously couldn’t have with the current review protocol, and limit visibility of high-risk hashtag pages that help bad actors discover bad content/potential victims. ---> Abuse prevention).”²⁷⁷

And in an August 2021 internal chat,²⁷⁸ [REDACTED] (Meta software engineer) wrote “My understanding is that our IIC text classifiers, namely ‘Borat’, ‘Bunkerbuster’ (HPRG) and UTM, are only running on FB [Facebook] Messenger threads”, and [REDACTED] asks, “do we know if they *ever* ran on IG [Instagram] Direct? And if so why we stopped?”. [REDACTED] replies “I confirm borat and UTM models run only on FB. ... looking at the code HPRG also doesnt have it [i.e., never ran on IG].” Additionally, Bunkerbuster “was intermittently offline. Every so often Pouya would remember and trigger it. This was a huge backfill, as it had been offline for like a few months.”²⁷⁹ Elsewhere, employees denigrated the

²⁷² MT-IG-AG-NM-000●25487.

²⁷³ METANMAG-007-●0305682.

²⁷⁴ “SRT” is Meta’s “Single Review Tool”, a proprietary content moderation interface used by human moderators. “SRT IIC” would presumably be Meta’s review of instances of IIC.

²⁷⁵ METANMAG-051-●0021365.

²⁷⁶ METANMAG-051-●0125572.

²⁷⁷ METANMAG-051-●0125572.

²⁷⁸ METANMAG-010-●018●536.

²⁷⁹ METANMAG-079-●0018237.

system:²⁸⁰ “bunkerbuster is a very old classifier that is completely unusable post e2ee and it’s only for a specific market (English, North America I think?)”²⁸¹

Meta has collaborated with NCMEC to scan for known CSAM using MD5 cryptographic hashes.²⁸² In an urgent February 2022 email,²⁸³ [REDACTED] (Meta product manager) stated, “I’m writing to enlist some urgent support from IG [Instagram] to mitigate legal and legitimacy risks we’re facing in the Adult Sexual Exploitation, Child Safety and Terrorism space. ... Adult Sexual Exploitation (StopNCII.org), Child Safety (NCMEC) and Terrorism (GIFCT) space are all problem areas which participate in cross-industry hash-sharing collaborations. ... All three arrangements involve Meta receiving hashes from an external body (e.g. NCMEC), which we can use internally to match and remove violating content.” [REDACTED]

[REDACTED] continued, “[REDACTED]” [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED] Upon hearing the news, [REDACTED] responds, “if this is true, it’s a disaster - [REDACTED] Surely not?”²⁸⁵

And other internal Meta documents noted that “the previous client side development [of MD5 calculations] was done by integrity problem teams’ eng [engineers], in an ad-hoc manner. there is no foundational owner who is responsible for maintenance and quality” of this critical function, that the “original developers left the team or company,” and the “lack of granular owner and eng [engineering] resources” available exacerbated the issues.²⁸⁶

²⁸⁰ METANMAG-07500132827.

²⁸¹ A February 2019 document describes an effort to improve Meta’s computational methods for determining IIC (METANMAG-005-01844878): [REDACTED]

[REDACTED] In other words, Meta was missing a vast number of instances of IIC.

²⁸² I am unaware of the date when this began.

²⁸³ METANMAG-041-00000553. See also METANMAG-002-02313590 and METANMAG-003-00709444.

²⁸⁴ See also METANMAG-011-01059066 where the duration of the failure is clarified by [REDACTED]: “My understanding is this is the state that has been in place for at least a couple of years.”.

²⁸⁵ METANMAG-005-01835845.

²⁸⁶ METANMAG-002-01631826.

Many documents report that Meta is consistently under-resourced for addressing child safety. For example, an internal presentation entitled “Instagram Well-being H1 2019- planning” — a planning document for the first half of 2019 — acknowledged that Instagram’s integrity team was thin relative to the scope and importance of the work.²⁸⁷ Given resource limitations, “we will not be doing major proactive work” in areas like harassment, financial scams, credible threats of violence, impersonation, prostitution and sexual solicitation and forms of child exploitation, the presentation said.

In her presentation from September 2018 on Inappropriate Interactions with Children (that is discussed in Section 5.1.2), Ms. [REDACTED] cites resources as a blocker; e.g., “We not only need more review capacity, we need critical mass (i.e. minimum volumes) across markets in order to tackle this problem globally.”²⁸⁸

In a February 2019 email chain,²⁸⁹ Head of Data Analytics Brady Lauback states in a discussion about Integrity work at Instagram that “I don’t think IG Integrity was really ever funded.” VP of Integrity Guy Rosen responds, “You are correct, there is a growing realization this is underfunded. This was deliberate - I explicitly had the convo with Mark at HC planning, and he thought IG [Instagram] has another year or two to catch up. I think we are not sure that is the case any more.”

A 2019 presentation by [REDACTED] (a manager of Meta’s risk assessment programs) noted failures in Meta’s processes for responding to reports: “On February 9th, the Metro UK reached out to us about a profile on IG that they said included a long-shot of four naked children as its profile picture. ... A profile containing CEI was live on IG despite ~8k user reports. Escalations of individual content as apparent CEI did not result in the profile being disabled. The individual content wasn’t reviewed in the CEI queue. ~30 escalations of the profile for apparent CEI did not result in it being disabled. The profile was not reviewed in the CEI queue. ... At the time of this escalation, the CEI contractor queues had a total backlog of 508K jobs (including reactive and proactive). Now, the CEI contractor queues have a total backlog of 394.5K jobs.”²⁹⁰

During an internal chat in April 2020 on Meta’s handling of minor sex trafficking leads, [REDACTED] (a Meta Child Safety manager) writes about the lack of resources Meta has ascribed to addresses child exploitation on its platforms: “Jo and Amanda have been trying to get COINV [Community Operations Investigations²⁹¹] to work minor sex trafficking cases for 18 months and they couldn’t get any staff from them as COINV are so small. The CONIV got

²⁸⁷ See also reporting in “Instagram Suggested ‘Groomers’ Connect with Minors, FTC Says,” Riley Griffen and Kurt Wagner. May 6, 2025. Bloomberg News. <https://news.bloomberglaw.com/privacy-and-data-security/instagram-urged-groomers-to-connect-with-minors-ftc-says>.

²⁸⁸ METANMAG-064-00001083.

²⁸⁹ METANMAG-003-00678146.

²⁹⁰ METANMAG-002-02454647.

²⁹¹ See for example METANMAG-005-01214011.

staffing and there was no tooling or guidelines or anything at all prepared.”²⁹² [REDACTED] (a Meta data scientist) responds, “I’d like to force them to do what they should. Otherwise, everyone just leans on us as a crutch.”

In July 2020, [REDACTED] noted in an internal chat that “we’re starting to see multiple LERT escalations for the same user, which means different forces/branches of LE have notified us about the same person.”²⁹³ [REDACTED]: “So the concern here is that the underlying infra isn’t consolidating? Or that we’re really crap at finding them?” Reilly: “I wanted to flag it that we’re crap at identifying users if multiple branches of LE have to tell us about them” “There’s 5469 accounts being backfilled disabled for IIC...that means we hit disable on an SRT job in [Messages] reactive OS but the user didn’t [sic] disabled, and was happening from June 20th until last week...these users still actively IIC violating post when we should have disabled them.”

A presentation from October 2020 entitled, “Child Safety State of play” details many problems in Meta’s approach to preventing child sexual exploitation.²⁹⁴ A list of “vulnerabilities” includes “minimal protections”; “low accuracy” classifiers including for “minor sexualization” and “Trafficking”; “under resourced” allocations to analytics, content review, user education, platform growth; lack of “strong policies” for Groups; new “vectors for abuse” from new products; and poor reporting. The presentation lists “risks” as “Child safety issues - from exposure to ‘harmful’ content (e.g., SSI) to sexual exploitation - have given significant momentum to highly problematic legislation and regulatory and antitrust probes across countries that matter to us...” The presentation includes a discussion question: “Do we agree we are not properly resourced to address child safety issues?”

In August 2021, Meta completed an internal audit (IA) of Child Safety Operations, which was then discussed among leadership. A presentation of the results enumerated a detailed list of failures by the Child Safety team.²⁹⁵

- CEI [child exploitation imagery message ... threads may not be reviewed and enforced on timely [sic].
- The NCMEC reporting backlog is not consistently reviewed or timely resolved. At the time of testing, IA noted ~247K reports enqueued between May 2017 and July 2021 that were in ‘pending submission’ status.
- IA noted 77,829 [NCMEC reporting] failures aged between three to 10 months that had not been resolved. As a result, eligible reports may not be timely submitted to NCMEC.
- Currently, one FTE is dedicated to the NCMEC reporting process, which may not be sufficient for Facebook to consistently meet reporting requirements or maintain an

²⁹² METANMAG-051-00095496.

²⁹³ METANMAG-078-00015118.

²⁹⁴ MT-IG-AG-NM-000180935. This October 2020 date is based on discovery metadata.

²⁹⁵ METANMAG-005-01331915.

effective control environment [for] ... resolving reporting failures, clearing backlogs, monitoring cancelled reports, resolving reporting SEVs.

- [REDACTED]
- The Child Safety team does not have visibility on the surfaces covered by proactive detection classifiers that identify CEI, IIC, and MS related violating content on the platform. As a result, violating content may not be identified and actioned on.
- [REDACTED] impacting the Child Safety team's ability to detect violating content.
- [REDACTED]
- Reports submitted to NCMEC through the Single Review Tool (SRT) did not include a tag to indicate a human review as required by NCMEC. ... As a result, incomplete or unvalidated reports may be submitted to NCMEC.
- [REDACTED]
- [REDACTED]

In an email discussion of the results, leadership admits these are serious issues, but that because they are deploying E2EE perhaps some are not worth addressing. Specifically, a Meta Integrity and privacy director, [REDACTED], offered, "Reading the report: omg, this area is pretty on fire. Backlog of 1.00K's? No alerting set up, no model refreshes?" ... these are actually all big problems, but, we are choosing to do nothing about them because when E2EE comes online, our entire solution approach to Child Safety will change, as will our output to NCMEC, so fixing the 'old' solution for the next 3-12 months isn't worth it. Stay on course. (Again, I'd like to confirm that one with Chris, but may have to probe [REDACTED] instead)." [REDACTED] replied, "None of this surprises me. We have a lot of work to build these operational excellence things. ... Remember there are company critical prios like Youth and E2EE also relying on Child Safety so we have to decide which ones of these are PIs [priority ones] and which ones are not."²⁹⁶

More than a year later in September 2023, the Instagram Youth Well-being team was subject to an internal audit.²⁹⁷ The audit returned a rating of "Significant Improvements Needed" and included many failures:

²⁹⁶ METANMAG-005-01350162.

²⁹⁷ METANMAG-018-00005057.

[illegible]

- The IG Relevance Integrity team does not always flag non-recommendable media when the classifier scores exceed predetermined thresholds. As a result, users may be exposed to sensitive content via content recommendation surfaces.”

5.1.5 Meta tolerated minors on its platforms & knew many children under 13 years of age were users

Meta decided more than nearly two decades ago to allow children 13 years and older on its platforms and let them interact with adults. As a result of their decision, they can protect children only when they know their true age; and they can restrict adults from interacting with children only when they know the ages of adults.²⁹⁸ In other words, the features and controls that Meta has in place to protect children are not effective when they don't know the correct ages of users.

In my 2022 Report to Congress, I note that Android and Apple iPhone apps are used heavily by minors, and under-moderated apps are rich “hunting grounds” for individuals looking to exploit minors. This holds true here for Meta’s platforms.

As far back as 2007, a Reuters article stated, “Chief Privacy Officer Chris Kelly said Facebook has privacy features unlike other social network sites that segment users by age,

²⁹⁸ See, e.g., Jayakumar Ex. 46 (“Youth Safety & Well-Being Deep Dive February 2021” presentation noting “Age verification is both an independent requirement and necessary foundation for all other requirements.”)

extensive risks to children on Meta's platforms but in this "all hands" presentation, inspiration is driven by a need to deploy E2EE rather than bringing attention to child sexual exploitation. Ms. Davis states:

Multi-national and multi-sector bodies, including the Five Eyes, the EUIF, and the WePROTECT Global Alliance, are driving a coordinated campaign against our e2ee plans, which has caught the attention of our advertising clients as well as large foundations such as the Oak Foundation and Skoll, and driven shareholder inquiries. To push back against calls for back door access or to stop e2ee altogether, we must build child safety protections for minors and drive down child exploitation prevalence on our platform. [REDACTED]

She also states, "[REDACTED]". She adds that "Child Safety analytics is under-resourced" and that developing content independent CSAM detection tech" was "unfunded...and de-prioritized." "[REDACTED]" was also "de-prioritized."

By July 2021, Meta's age restrictions involved asking users their age, reminding parents, reporting by other users, and content reviewers look for accounts that appear to belong to people underage. A presentation states that they also use machine learning based on "multiple signals, including how old your followers are and how old the people you follow are. [Meta] also look[s] at the age you shared with us on other Facebook apps and apply it elsewhere, like if you share your birthday with us on Facebook but not on Instagram."³³⁷ In speakers notes, Safety Policy Director [REDACTED] writes that "Verifying someone's age and understanding age is one of our industry's most complex challenges. These are not new problems."

At this time, CyberTips to NCMEC from Meta were at an all-time high, indicating that the platforms are as popular as ever with persons who sexually exploit children. And even though child safety is critically dependent on knowing the age of the persons involved, they push forward by trying to improve their age classifiers rather than restrict their teen audience until the problem is addressed more successfully.

In a July 2021 email thread, regarding Mr. Zuckerberg's feedback on policy commitments related to safety on Meta's nascent VR platforms, leadership at Meta did not want to commit to identifying users under the age of 13 (which Meta calls "U13s").³³⁸ The proposed commitment was "Identify u13s and ensure youth are in age-appropriate experiences".

³³⁷ METANMAG-002-02462209.

³³⁸ METANMAG-059-00000318.

pages popular with children. Instagram, owned by Facebook, does not have the same restrictions. Until this past week, it had allowed users to send private messages to anyone, and a Times reporter was able to contact and video-chat with a 13-year-old girl who had a private account (the girl and her parents gave permission to conduct the test). After The Times asked about the policy, Instagram announced new features on Wednesday that allow users to block messages from people they do not follow.³⁵⁵ The company said it would also require users to enter their age when signing up.

The article is another example of Meta reacting to public pressure rather than conducting a safe-by-design analysis.

In a September 2020 email thread, Instagram Communities director [REDACTED] notes that “This new data shows that prevalence of ‘sex talk’ to minors is 38X on IG Direct vs Facebook Messenger, in US. This is largely due to a much larger potential victim pool on IG vs Messenger (32x) but is concerning. CI [the Content Integrity team] is reconsidering their investment in child safety mitigation and planning to invest more than previously planned into IG for 2021.”³⁵⁶ Similarly, in October 2020, Mr. Otaru notes that although prevalence of child sexual abuse is rampant on Instagram, they will merely “consider” some separation of minors and adults. He writes: “Recent prevalence metrics indicate that Tier 2 IIC is 3x more prevalent on IGD [Instagram Direct] than Messenger; on IGD, 500k non-adult accounts receive Tier 2 IIC every week. We will audit the measures in place on Facebook/Messenger and adapt high-impact measures for IGD. For example, we will consider restricting flagged IIC groomers from sending message requests to non-adult accounts.”³⁵⁷

5.1.7 Meta has deployed end-to-end encryption (E2EE) on its platform despite harms to children

Meta has regularly stated that it has “zero tolerance for child exploitation imagery”.³⁵⁸ Yet this stance is at odds with its deployment of end-to-end encryption (E2EE) of content exchanged with children and among its adult users. E2EE prevents the review, detection, prevention, and response of CSAM/CSEC as well as the provision of information to law

³⁵⁵ Meta has provided discovery responses stating that, as of early 2025, only about “0.08% of age 13-17 Users predicted to reside in the U.S. enabled” its opt-in “Limit Interactions” feature, which “limit[s] comments and direct message requests from people who do not follow them or only recently followed them on Instagram.” See Defendants’ Seventh Supplemental Objections and Responses to Plaintiff’s First Set of Interrogatories to Defendants (Interrogatories 1, 4, 5, 12).

³⁵⁶ MT-IG-AG-NM-000171201.

³⁵⁷ MT-IG-AG-NM-000004084.

³⁵⁸ See e.g., “Instagram baby-selling bust prompts call for ‘cyber patrols’,” Beh Lih Yi. Reuters. October 18, 2018. <https://www.reuters.com/article/world/instagram-baby-selling-bust-prompts-call-for-cyber-patrols-idUSKCNIMS0QD/>. Quoting “zero-tolerance policy prohibiting child exploitation” in <https://committees.parliament.uk/writtenevidence/106653/pdf/>.

enforcement upon request.³⁵⁹ Given the extraordinary number of incidents of child sexual exploitation that Meta itself has recorded on its platform, E2EE is a policy of tolerance of these abuses.

In 2016, end-to-end encryption was introduced on Facebook's WhatsApp platform (which at that time had over 1 billion users around the globe), intensifying debates over access to digital data for (among other reasons) law enforcement purposes.³⁶⁰ Facebook's Messenger platform also began optionally offering end-to-end encryption in 2016, but only made that setting its default in 2023, despite the urging of government officials and others to halt such implementation in order to facilitate law enforcement investigations.³⁶¹ Meta claimed that they "worked closely with outside experts, academics, advocates and governments to identify risks and build mitigations to ensure that privacy and safety go hand-in-hand."³⁶²

A variety of parties have informed Meta that deploying end-to-end encryption on its platforms would represent a danger to children. A demonstration of this danger is available from an examination of Meta's NCMEC CyberTip Report counts. For many years, Meta has submitted CyberTips to NCMEC based on various detection methods. The following plot shows the number of reports.³⁶³ The sudden drop in 2024 is due to Meta's deployment of EE2E, which decimated Meta's ability to detect CSAM and child sexual exploitation in the content of its users.

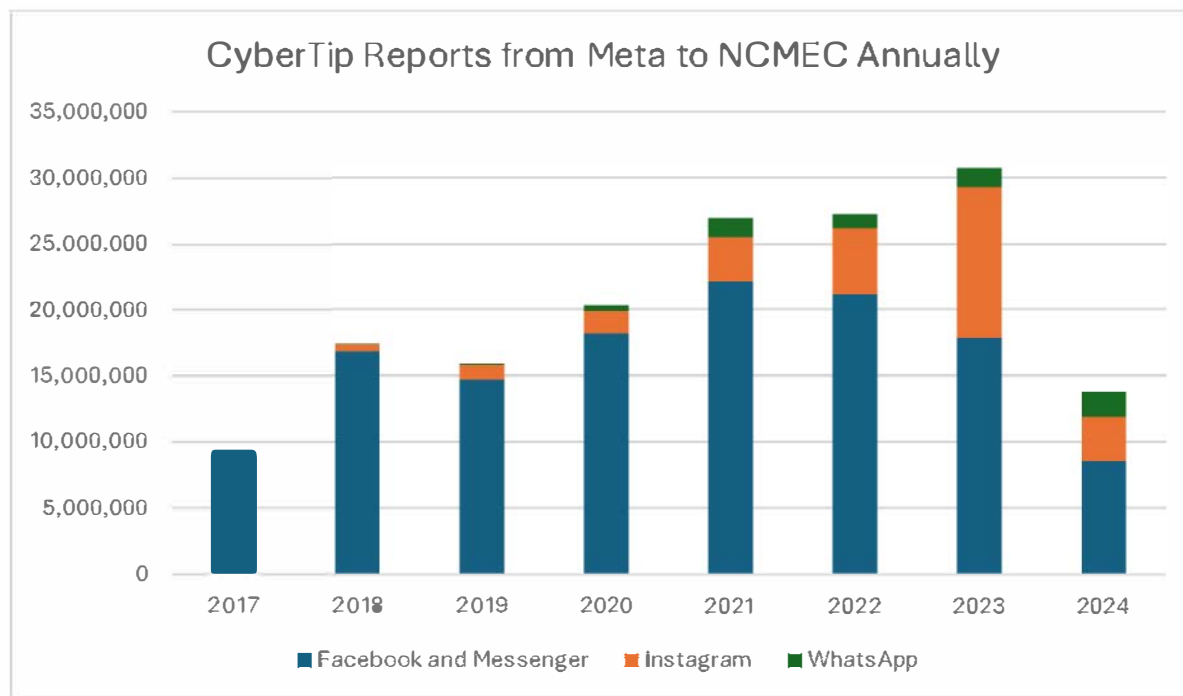
³⁵⁹ See "Child sexual abuse material and end-to-end encryption on social media platforms: An overview," Teunissen, C and Napier, S. 2022. *Trends & issues in crime and criminal justice* no. 653. Canberra: Australian Institute of Criminology. <https://doi.org/10.52922/ti78634>

³⁶⁰ See "WhatsApp Introduces End-to-End Encryption," M. Issac. New York Times. April 5, 2016. <https://www.nytimes.com/2016/04/06/technology/whatsapp-messaging-service-introduces-full-encryption.html>. See also "WhatsApp has an encrypted child abuse problem," Josh Constine. Tech Crunch. December 20, 2018. <https://techcrunch.com/2018/12/20/whatsapp-pornography/>

³⁶¹ Falconer, R., *Facebook Messenger to become fully encrypted*, Axios (Dec. 7, 2023), <https://www.axios.com/2023/12/07/facebook-messenger-encryption>; Fried, I., *U.S., allies fire up encryption fight with Facebook*, Axios (Oct. 4, 2019), <https://www.axios.com/2019/10/04/us-allies-encryption-fight-facebook>.

³⁶² <https://www.theguardian.com/technology/2023/dec/07/meta-facebook-messenger-end-to-end-encryption>.

³⁶³ The plot is based on METANMAG-002-02257779 and NCMEC public reports at <https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata>. At this time, I do not have data for 2016 or earlier.



NCMEC’s position on E2EE is and has been that “The use of end-to-end encryption would prevent the companies or any third-party from detecting illegal activity occurring on their platforms, including the activity of people who use the internet to perpetuate online demand for graphic sexual abuse material of children. We believe personal security is extremely important and support efforts to improve online privacy. But, if this solution is implemented with no exceptions for detecting child sexual exploitation, millions of incidents of abuse will remain hidden, leaving these young victims without any help or protection from these horrific crimes.”³⁶⁴

NCMEC has publicly commented on the significant decrease in CyberTip reports filled in 2024.

In 2024, the CyberTipline received 20.5 million reports of suspected child sexual exploitation [total from all companies]. While this is a significant number, the overall number of reports declined from 36.2 million reports received in 2023. This decline is especially concerning because the REPORT Act, which was enacted in 2024, mandates companies to report two additional forms of child sexual exploitation for the first time – child sex trafficking and online enticement. With additional mandatory reporting, we would have expected increased reporting, but our analysis identified multiple contributing factors to the decline, including a decrease in reports from certain platforms and a decline caused by further implementation of end-to-end

³⁶⁴ <https://ncmec.org/theissues/end-to-end-encryption>.

encryption (E2EE). NCMEC has long raised concerns about implementing E2EE without an exception for detecting child sexual exploitation.³⁶⁵

And NCMEC has commented on Meta's drop in CyberTip reports in a statement to Congress:

While the first two factors likely contributed to the decrease, the likeliest factor contributing to the overall drop in reported incidents in 2024, is a result of Facebook's implementation of end- to-end encryption (E2EE). Meta began to implement default E2EE in Facebook Messenger in December 2023 and completed that process in the summer of 2024.³⁶⁶ Even when NCMEC adjusts for Facebook's bundling and realignment of content it reports to the CyberTipline, the numbers demonstrate that Facebook reported approximately 6.9 million less reports in 2024 than it did in 2023. For years NCMEC has anticipated a negative impact on CyberTipline reporting as more online platforms adopted default E2EE on their platforms without adequate safeguards for child protection. When a platform voluntarily chooses to blind itself to child sexual exploitation by disabling its ability to detect and report abuse, it is not just losing a report – it is potentially losing a child. Every lost report can represent a child who may never be identified, rescued, or safeguarded. It means the child's ongoing abuse and repeated revictimization will continue unchecked, while offenders remain free to exploit more victims in the shadows.³⁶⁷

Ahead of this E2EE deployment, many others offered expert opinions. For example, in June 2019, Professor Hany Farid, a noted child safety expert and inventor of PhotoDNA, wrote in a public op-ed that "The adoption of end-to-end encryption would significantly hamper the efficacy of PhotoDNA. This is particularly troubling given that the majority of the millions of yearly Facebook reports to NCMEC's CyberTipline originate on Facebook's Messaging services. Blindly implementing end-to-end encryption will significantly increase the risk and harm to children around the world."³⁶⁸

In October 2020, The US Attorney General, the UK Secretary of State, Australian Minister for Home Affairs, New Zealand Minister of Justice, and the Canadian Minister of Public

³⁶⁵ <https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata>.

³⁶⁶ Internal citation to <https://about.fb.com/news/2024/03/end-to-end-encryption-on-messenger-explained>.

³⁶⁷ United States Senate Committee on the Judiciary Subcommittee on Crime and Counterterrorism "Ending the Scourge: The Need for the STOP CSAM Act" March 11, 2025 Testimony of Michelle DeLaune, President and CEO National Center for Missing & Exploited Children. https://www.judiciary.senate.gov/imo/media/doc/2025-03-11_testimony_delaune.pdf.

³⁶⁸ "Facebook's plan for end-to-end encryption sacrifices a lot of security for just a little bit of privacy" Hany Farid, June 16, 2019. Fox News. <https://www.foxnews.com/opinion/hany-farid-facebook-end-to-end-encryption-security-privacy>. See also Farid's statement to Congress: <https://www.congress.gov/116/meeting/house/110075/witnesses/HHRG-116-IF16-Wstate-FaridH-20191016.pdf>.

Safety and Emergency Preparedness co-signed a statement on EE2E.³⁶⁹ They wrote in part: “End-to-end encryption will simply close a curtain to what happens online, enabling encrypted platforms to become lawless environments where the lack of oversight and visibility into criminal activity emboldens offenders. We oppose privacy measures that fail to address how the internet is used to entice children into sexually exploitive situations and to traffic images and videos of children being raped and sexually abused.”

In an open response, CEO Mark Zuckerberg refused to distinguish the privacy and safety needs of minors from adults. Mr. Zuckerberg instead erroneously claimed that the needs of adults are the same as children’s, as if Meta lacked the ability to keep children’s accounts unencrypted. He wrote: “The ‘back-door’ access you are demanding for law enforcement would be a gift to criminals, hackers and repressive regimes, automatically creating a way for them to enter our systems, leaving every person on our platforms more vulnerable to real-life harm. It is simply impossible to create such a back-door for one purpose and not expect others to try and open it. People’s private messages would be less secure and the real winners would be anyone seeking to take advantage of that weakened security.”³⁷⁰

Meta has continually viewed the E2EE issue in extreme terms, planning all content to be encrypted. In reality, there are many options, including separating out communications of children and leaving them unencrypted. For example, the default could be that content is not end-to-end encrypted unless the person can prove they are adult. This choice would still allow, very dangerously, two adults to exchange CSAM on Meta’s platform, but it would be safer than the path Meta has taken.

A shareholder proposal in May 2020 demanded “the Board of Directors issue a report assessing the risk of increased sexual exploitation of children as the Company develops and offers additional privacy tools such as end-to-end encryption.”³⁷¹ Despite garnering a majority of non-management-controlled stock, the proposal, which Meta opposed, did not pass.³⁷²

³⁶⁹ <https://www.justice.gov/archives/opa/pr/international-statement-end-end-encryption-and-public-safety> or see METANMAG-002-00287053. See also a July 2019 letter from UK Home Office Secretary: MT-IG-AG-NM-000071466.

³⁷⁰ METANMAG-002-00283371; see also <https://www.cnbc.com/2019/12/10/facebook-rejects-ag-barrs-request-to-stop-messaging-encryption.html>.

³⁷¹ MT-IG-AG-NM-000255157; <https://www.telegraph.co.uk/news/2020/05/20/facebook-faces-major-shareholders-revolt-encryption-plan-could/>.

³⁷² Proxy Impact, “Online Child Safety Impacts and Harm Reduction”, <https://www.proxyimpact.com/facebook>.

interventions work before encryption. After this, we accept we will have no insight into ground truth. Disappointing, but we as a company have decided privacy is the priority.” The document’s H2 2020 learnings summary stated: “We are uncomfortable with the level of virality that our CEI [child exploitation imagery] enforcement systems sometimes allow: We learned that we detect & delete only ~50% of CEI at (or near) upload time, with the remainder retroactioned at a later date (sometimes weeks or months later).” [REDACTED]

Safety Policy Director [REDACTED]’s “Child Safety 101 for MK [Messenger Kids] Integrity” internal presentation from September 2021 outlines the risks and challenges related to child safety on Meta’s platforms focusing on abuse, exploitation, and inappropriate interactions within the Messenger Kids and broader Meta ecosystem.³⁹⁰ It highlights concerns about how Meta’s platforms are used for child exploitation, the organizational challenges in managing these risks, and the potential legal, regulatory, and reputational issues. For example, on Exploitation Risks: “In 2019, Messenger was responsible for over 80% of all reports of child sexual exploitation made by Facebook.”; on Systemic Issues and Policy Gaps: “Given the regulatory environment linked to our e2ee announcement... launching products without child safety mitigations plays into accusations that we are abandoning children to black boxes of abuse.”

In a September 2021 internal presentation entitled, “Strategy to address IIC; Child Safety CI & Safety Policy” by [REDACTED], [REDACTED], and Safety Policy Director [REDACTED], numerous risks and gaps are presented.³⁹¹ The viewpoint of the company is that “we are trying to launch IGY [Instagram youth] amidst a multi-government campaign using child safety concerns to try and thwart our plans to end-to-end encrypt messaging”. The presentation draws attention to the fact that detection, measurement, and enforcement of IIC problems will be thwarted severely by E2EE. The presentation identifies several “risks and gaps”:

- [REDACTED]
[REDACTED] and we do not know how well FB’s text detection will perform on IGY.”
- [REDACTED]
[REDACTED]
[REDACTED]

³⁹⁰ METANMAG-005-00508087.

³⁹¹ METANMAG-002-01715525.

- “Risk 3. Lack of human review capacity, quality and routing: Reactive user reporting from private surfaces are not always reviewed appropriately. Example: For example, on IG, reviewers have no tool available to report IIC to NCMEC from reactive report queues, and [REDACTED]”
- “Risk 4. Limited Surface coverage: We have limited understanding of the scope of Child Safety issues [REDACTED]”
- “Risk 5. Heuristics for safety: We cannot continue to rely on models validated against message content (e2ee is coming). Example: [REDACTED]”

92

Notably, for Risks 1 and 2 above, the answer to the question “What’s needed to unblock” is “capacity,” which implies that Meta has not allocated the capacity to address these risks.

In January 2021, Global Head of Safety Antigone Davis reminds her co-workers “We Should Care. When someone submits a report or tells us something is wrong, we should be responsive. We need to take action to prevent harm and provide support if harm has already occurred.”³⁹³ This is one of the only statements encouraging a survivor-centric approach I have seen from leadership in presentations. She also states that “Not All Threads Should be E2EE. If due to the nature of the product (i.e. size, semi-public nature, transactional) and/or vulnerability of the primary and intended consumers of a product we cannot substantially mitigate risks via prevention, detection or user-support in e2ee, we should consider exemption.”³⁹⁴ Which is an admission that E2EE is harmful to the safety and privacy of children on Meta’s platforms.

In a May 2022 presentation entitled, “Child Safety on Messenger”, Ms. [REDACTED] and Mr. Sinha provide an overview of the Safety Policy framework at Meta, focusing on risks such as bullying, harassment, child sexual exploitation, and grooming. It includes procedures for identifying risks during product reviews, particularly for new or updated app features, and highlights the safety vulnerabilities these products can introduce.³⁹⁵ “Children are being ‘surveilled and manipulated’ and ‘apps are pushing kids to spend unhealthy amounts of time online.’ - US Senate”. She writes, “Over 80% of CSAM reported by FB [Facebook] is from MSGR [Messenger]. At least 50% of CSAM deletes are a result of retroactive proactive detection (not caught at time of upload). Most of what’s shared in MSGR are images not yet banked. And with e2ee coming, an enforcement model which relies on finding bad content and bad actors after the harm is done is not enough, which is why we are pivoting to prevention.” And “Most of what’s shared in MSGR are new images for which we have no

³⁹² METANMAG-002-01715525.

³⁹³ METANMAG-002-00698539.

³⁹⁴ METANMAG-002-00698539.

³⁹⁵ METANMAG-005-00508074.



Figure 8. Screenshot sent by Mr. Horwitz to Meta from an account of a 15-year-old, pointing out the 9-week-old comment "I want to do her so bad.". Reproduced from MT-IG-AG-NM-000224142.

Meta's internal email chain discussing Mr. Horwitz's simple smoke test contains surprise from employees.⁴¹⁸ Liza Crenshaw makes clear that it's the recommender system that is surfacing this content: "I'm surprised we even recommend 'preteen' anything after the exercise we went through on autofills there. ... re: the comments like 'I wanna do her so bad', would our IIC classifiers not be intended to catch those?" She told the team she plan to contact Mr. Horwitz "Clarifying that [REDACTED] - emphasizing that this announcement is re: our intention to build one and making. ... the comments are exactly the type of behavior we're hoping to stop with these restrictions, but it may take time to get it right." While Meta has taken its time to "get it right", it has continued efforts to recruit and retain minors to its platforms.

5.1.9 Meta makes it difficult for users to report content and for law enforcement to respond

Meta has a history of failures related to reports from users of child exploitation and CyberTip reports to NCMEC.

First, Meta has made it difficult for users to report CSAM- and CSEC-related content. The Canadian Centre for Child Protection detailed the limitations of Facebook Messenger, Instagram, and WhatsApp in a 2020 report: "Because WhatsApp uses end-to-end encryption, reporting functions are the most minimal of any of the platforms reviewed here. ... There is no way to report images or videos sent within a DM at all."⁴¹⁹

⁴¹⁸ MT-IG-AG-NM-000224142.

⁴¹⁹ "Reviewing Child Sexual Abuse Material Reporting Functions on Popular Platforms," Canadian Centre for Child Protection. September 23, 2020. <https://protectchildren.ca/en/resources-research/csam-reporting-platforms/>.

There are documents that suggest backlogs in processing reports. For example, in a November 2020 internal discussion, data scientist [REDACTED] is “concerned we actually don’t know what is happening with dequeues.” Adding “I worry we’ll drop a legitimate IIC T1 (particularly when it’s been user reported) and do nothing.”⁴²⁰ Colleague [REDACTED] responds: “Oh sorry we were too slow to process your report so we deleted it after 28 days.” [REDACTED]: “right, that’s what worries me.” [REDACTED]: “A big portion of our time in Child Safety is always spent on what I like to call “Figure out what the Eff we’ve been doing for all these years and why it’s broken.” [REDACTED]: “YES.” [REDACTED]: “Based on very preliminary numbers it seems the escalation decisions in Messenger queues get 0 precision in our queues.” [REDACTED]: “wow. What do we think that means?” “That they are escalating not the right things?” [REDACTED]: “Bad review quality.” “Them missing stuff might also be a problem but so is the opposite.”

Similarly in a November 2020 discussion of issues with IIC escalations, [REDACTED] wrote “I would argue they are not even basic maturity in DMAT [Don’t Miss a Thing]... IIC is currently not even basic.” Adding “BTW this SEV is particularly painful. Thousands of minors have been reporting IIC T1 [Tier 1] abuse and even though we knew that there is IIC T1 going on (more than 50% of which is Sextortion which can lead to suicide) we haven’t done anything. We had a broken escalation path and no measurement. God knows what happened to those kids.”⁴²¹ [REDACTED] responds: “The perpetrators have been disabled, yes? Just checking, but I know we were meant to do so much more.” [REDACTED]: “In sextortion normally the abusers have a network of fake accounts so it’s probably not very helpful.” [REDACTED]: “Not I only have 2 FTCs staffing English escalations only, about 200+ hours a week requested but not on the way yet. 2 more English FTCs on the way by EOY but that’s it. Backlogs are 30000++ across IG, FB, etc. -> IG escalations were never worked, someone created a Queue before I started...” [REDACTED]: “I think we can all agree that this whole situation is Not Good and there are Several issues blocking us from correctly reporting all the T1 we know about to NCMEC/LE.”

And in a February 2021 internal chat regarding Inappropriate Interactions with Children (IIC) enforcement issues at Meta, software engineer [REDACTED] writes, “Out of 43 reactive reports that were escalated to the triage queues since 11/4, only 2 have been reviewed”⁴²² due to lack of capacity. Child safety manager [REDACTED]: “At the moment we are sending some blank reports to NCMEC w/ no messages attached for example – we can’t do that to MORE markets. The flow needs to work first.” [REDACTED]: “So because we don’t really have a working escalation flow for user reports, I would say our DMaT maturity is “at risk” here. The reasons being missing review capacity as well as insufficient eng support” “It’s that we don’t even have a working escalation flow right? Except for English markets, on FB only, and with significant backlogs.” [REDACTED]: “The issue is that multiple parts of our flow are broken.” [REDACTED]: “when we had this conversation in a big call before xmas, the end decision

⁴²⁰ METANMAG-051-00095317.

⁴²¹ METANMAG-078-00019520.

⁴²² METANMAG-051-00097719.

was we're not missing anything if we have a backlog & the escalation flow works.' "Only having 3 reactive escalations from English since November I would say is a problem..." [REDACTED]: "Is it a rate? Yes. The rate is 0." [REDACTED]: "But nowhere near as concerning as the fact that, even if we got more escalations, we would just add most of them to the backlog indefinitely." [REDACTED]: "Leadership wants us to fix it? They need to put up the resourcing." "So we'll need to figure out how to fix: - ranking in Messenger O/S for English, review quality in Messenger O/S for English, latency in IIC queues, capacity if we want to address other languages." [REDACTED]: "Well if something was reported to us in November and we start reviewing in July... I'm not sure it's any better than not reviewing at all." [REDACTED]: "This isn't a bomb threat, it's an adult having sex with a minor, they've still committed a crime even if time passes. These backlogs aren't new and are prioritized – there have been (bigger) backlogs here since before I joined the company..." [REDACTED]: "I don't disagree. Good thing we have the measure to let us know this is happen! Again, I think the measurement is working. The review/enforcement is lacking."

Relatedly, one of Wall Street Journal reporter Jeff Horwitz's concerns in his September 2023 email message to Meta was that his reports of CSAM and CSEC had gone unanswered, though months had passed: "While I know that some glitches have been fixed, activists I'm talking to as well as my own reports of terrible FB groups [REDACTED] suggest that the system is still a mess -- all those groups were ruled appropriate when I reported them. One activist I trust reported 12 separate pieces of clearly violating CSAM (his characterization of the content was flat illegal stuff) and was told all of them had been reviewed and were fine. But beyond missed calls, a lot of user reports seem to just disappear."⁴²⁴

And as I discussed in detail above, Meta's move to E2EE represents their choice to largely abandon their ability to find CSAM and CSEC in users' content.

Second, Meta's training on when to issue a report has been mismanaged: Meta likely underreported CSAM for years. Specifically, it's training materials to moderators stated, "When unsure whether a person is an adult or minor, err on the side of an adult."⁴²⁵ This was reported by the New York Times in 2022, who wrote: "The training document, obtained by The New York Times, was created for moderators working for Accenture, a consulting firm that has a contract to sort through Facebook's noxious content and remove it from the site. The age policy was first disclosed in California Law Review by a law student, Anirudh Krishna, who wrote last year that some moderators at Accenture disagreed with the practice, which they referred to as 'bumping up' adolescents to young adults. ... Representatives for

⁴²³ [REDACTED].

⁴²⁴ METANMAG-021-00494992.

⁴²⁵ METANMAG-005-01848502, "IIC Review Training, All Markets". These instructions also appear in METANMAG-038-00052496 "IIC Review Training, CO Market Teams", and METANMAG-011-01652638 "Sexual Harassment Review Training, Interactions Integrity.

Apple; Snap, the owner of Snapchat; and TikTok said their companies took the opposite approach of Meta, reporting any sexual image in which a person's age was in question.”⁴²⁶

These training instructions are particularly problematic as moderators are told to make expert anatomical judgements about the ages of individuals in potential CSAM and compare to drawings juxtapose descriptions such as “testicular volume between 12 and 20 ml”. The drawings illustrate the Tanner scale.

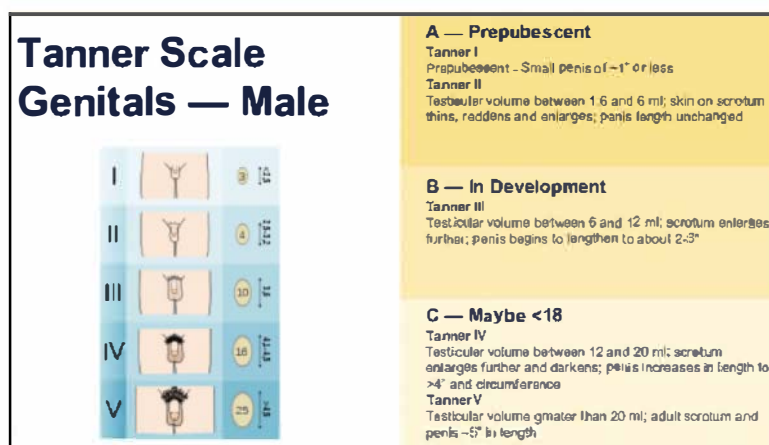


Figure 9. Reproduced from METANMAG-005-01848502.

The Times article criticizes Meta's use of the Tanner scale: “In a 1998 letter⁴²⁷ to the journal *Pediatrics*, Dr. Tanner said that using the stages to measure ‘chronologic age’ when analyzing child sexual abuse imagery was ‘wholly illegitimate.’ Dr. Tanner died in 2010. The co-author of the letter, Dr. Arlan L. Rosenbloom, now a retired pediatric endocrinologist, said in an interview that a child at 13 or 14 could be ‘fully developed’ under the Tanner stages. He also characterized Meta's approach as ‘a total misuse’ of the scale. Ms. Davis [Global Head of Safety] said the scale was widely used in the tech industry and was just one factor in estimating age. She acknowledged its limitations and said the company was exploring alternatives.”

Third, Meta's reporting is less useful to law enforcement than it could be. In a March 2020 presentation on Child Safety, it is stated that “We have some evidence that we may be

⁴²⁶ “Adults or Sexually Abused Minors? Getting It Right Vexes Facebook” Michael Keller. March 31, 2022. New York Times. <https://www.nytimes.com/2022/03/31/business/meta-child-sexual-abuse.html>; see also <https://doi.org/10.15779/Z38KW57J9B>.

⁴²⁷ The article links to <https://pubmed.ncbi.nlm.nih.gov/9882230/> “Misuse of Tanner puberty stages to estimate chronologic age.” Rosenbloom AL, Tanner JM. *Pediatrics*. 1998 Dec;102(6):1494.

over reporting even against our current liberal reporting guidelines,”⁴²⁸ thus wasting the time and limited resources of NCMEC and law enforcement.⁴²⁹

In March 2022, NCMEC’s General Counsel Yiota Souras wrote⁴³⁰ to Meta about “Widespread Flaws in Meta’s Reporting to the CyberTipline”. Her conclusion is that “Meta’s poor reporting practices have become a hindrance to child safety, rather than a solution.” She provides three examples of major problems.

- First, “a large volume of reports that appeared to involve compromised or hacked accounts distributing child sexual abuse material. ... it has caused confusion among the law enforcement community around the world and, as described below, has created strain on NCMEC’s operational resources.”
- Second, Meta submitted reports that “contained imagery of political unrest, cartoons, and inanimate objects. After NCMEC brought this to Meta’s attention, Meta requested that NCMEC flag almost 57,000 reports as erroneous. A majority of these reports had already been made available to law enforcement agencies globally.” Other documents clarify that the reports contained no CSAM.⁴³¹
- Third “over 600,000 otherwise valid reports submitted to the CyberTipline over the past 3 months were missing crucial information as a result of a Meta ‘bug’ and needed to be resubmitted. This error alone will result in a significant increased workload for law enforcement agencies who already are struggling to keep pace with the volume of reports.” Other documents clarify that the reports did not contain the “upload IP address” making them inactionable by law enforcement.⁴³²

The letter continues with blistering criticism of Meta’s practices.⁴³³ Meta’s failures cause legitimate reports to be “lost in the morass” of Meta’s erroneous reports. She writes that “Law enforcement is growing reluctant to trust CyberTipline reports submitted by Meta.” And that their errors create “tremendous operational strain on both NCMEC staff and on law

⁴²⁸ METANMAG-002-00285147.

⁴²⁹ See also, e.g., METANMAG-007-00020093 (2019 NCMEC CyberTipline Report on Facebook and Instagram that notes “submitted reports do not contain enough context or pertinent information to be actioned by law enforcement”; that inactionable reports caused NCMEC “CyberTipline analysts [to] spend a disproportionate amount of time identifying and manually processing these reports” which “takes time away from processing actionable, high valuable reports” and affects law enforcement’s “triaging and takes time away from processing actionable reports.”)

⁴³⁰ METANMAG-002-00461193.

⁴³¹ See METANMAG-002-02055355.

⁴³² See METANMAG-002-02055355.

⁴³³ The quality of Meta’s reporting to NCMEC led to backlash from state-based Internet Crimes Against Children (ICAC) Task Forces, a majority of which refused to accept certain types of reports from Meta for at least a period of time, and both prevented law enforcement from following up on real threats to children and diverted law enforcement resources to inactionable tips; see July 3, 2025 deposition of NCMEC representative Fallon McNulty.

enforcement ... Meta's flawed reporting is compromising the integrity of this process.” As a result, “NCMEC is gravely concerned about the impact on child safety and the potential scope and magnitude of these reporting errors, as well as Meta's apparent inability to put an end to the continual mistakes.”

And a January 2024 Guardian article reports that Meta and other companies are delaying the process of rescuing children by not examining the content of CSAM ahead of sending CyberTips to NCMEC: “‘If the company has not indicated when they report the file to NCMEC that they have viewed the file prior to making the report, we cannot open it,’ said NCMEC. ‘When we send it along to law enforcement, they cannot view it or open it without first serving legal process on the [social media company].’ ... In the last year, the major tech companies Meta, Twitter and Alphabet have all slashed jobs from their teams responsible for moderation. Such cuts ultimately result in fewer reports of child abuse material being reviewed by humans at the companies, placing an even more significant burden on law enforcement as the potential number of search warrants needed increases, said Korobov.”⁴³⁴

5.1.10 Meta allows workarounds to prohibited search terms, suggests sexualized searches, or may suggest alternative searches for banned content

In a July 2018 article, a former moderator stated that Meta had a policy of not shutting down accounts that shared CSAM: “Katz said there was a particularly sinister photo and video that popped up repeatedly in her queue. She said it featured two children between 9 and 12 years old standing facing each other, wearing nothing below the waist, and touching each other. It was clear, Katz said, that there was someone behind the camera telling them what to do. ‘It would go away and come back. It would appear at multiple times of the day,’ she said. ... At the time, Katz said, she was not asked to report the accounts sharing the material — something she said ‘disturbed’ her. ‘If the user’s account was less than 30 days old, we would deactivate the account as a fake account,’ she said. ‘If the account was older than 30 days, we would simply remove the content and leave the account active.’”⁴³⁵ Accounts left active are fodder for Meta’s algorithmic recommendation systems.

Meta’s failure to shut down Instagram accounts associated with the sexual exploitation of children was corroborated by a September 2019 article. The article notes the difficulty of reporting CSAM on Meta at that time.⁴³⁶ “Business Insider reported the [CSAM] videos

⁴³⁴ “Revealed: US police prevented from viewing many online child sexual abuse reports, lawyers say,” Katie McQue. January 17, 2024. <https://www.theguardian.com/technology/2024/jan/17/child-sexual-abuse-ai-moderator-police-meta-alphabet>.

⁴³⁵ “A former Facebook moderator says she had to review 8,000 posts a day and it made her numb to child porn,” Jake Kanter. Business Insider. July 9, 2018. <https://www.businessinsider.com/sarah-katz-what-like-to-work-as-a-facebook-moderator-2018-7>.

⁴³⁶ “Instagram’s new TV service recommended videos of potential child abuse,” Isobel Asher Hamilton. September 21, 2018. Business Insider. <https://www.businessinsider.com/instagram-tv-service-igtv-recommended-potential-child-abuse-2018-9>.