

Filed Under Seal

EXHIBIT A

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

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>.

of trauma can also affect law enforcement tasked with reviewing CSAM in connection with investigating and prosecuting offenders.) Meta reports hiring “around 40,000 people devoted to safety and security efforts”⁷⁴ and reports “15,000 moderators”.⁷⁵ More recently, some machine learning models have been trained to recognize new CSAM with varied success. However, machine learning models can be expensive to operate. In general, companies use all these techniques concurrently.⁷⁶

Detecting CSEC is much more difficult. It can vary in form quite widely. Humans can engage in evasion, though it is often not sophisticated in practice. In general, any feature that allows users to communicate, form connections, and form communities, can and is leveraged to engage in CSEC and CSAM trafficking. Perpetrators that are motivated will take advantage of opportunities to connect with anyone across the globe to trade content, coerce content, or sell content.

According to Meta’s responses to interrogatories, Meta’s classifiers include those that attempt to identify adult nudity and various child safety categories.⁷⁷ Based on classifiers, some content is removed, other content is not removed but no longer recommended to users.⁷⁸ Note that classifiers have an accuracy rate and they are not infallible. Meta first launched a classifier to action “adult nudity & sexual activity” (“ANSA”) content at least as early as January 2016.⁷⁹ Meta first launched a classifier to action “child safety” (“CS”) content at

Journal of Psychosocial Research on Cyberspace, 17(4), Article 8. <https://doi.org/10.5817/CP2023-4-8>; Bleakley, P., Martellozzo, E., Spence, R., & DeMarco, J. (2023). Moderating online child sexual abuse material (CSAM): Does self-regulation work, or is greater state regulation needed? *European Journal of Criminology*, 21(2), 231-250. <https://doi.org/10.1177/14773708231181361> (Original work published 2024); Spence, R., Martellozzo, E., & DeMarco, J. (2024). Content moderator coping strategies: Associations with psychological distress, secondary trauma, and well-being. *Journal of Media Psychology: Theories, Methods, and Applications*. Advance online publication. <https://doi.org/10.1027/1864-1105/a000454>; and “Content Moderator Mental Health and Associations with Coping Styles: Replication and Extension of Previous Studies,” Ruth Spence and Jeffrey DeMarco. *Behav. Sci.* **2025**, 15(4), 487; <https://doi.org/10.3390/bs15040487>.

⁷⁴ Letter from Meta to US Senate Committee on the Judiciary. April 19, 2024

https://www.judiciary.senate.gov/imo/media/doc/2024-01-31_-_qfr_responses_-_zuckerberg1.pdf

⁷⁵ Reportedly, Meta had 7,500 moderators in August 2018 (<https://www.vice.com/en/article/how-facebook-content-moderation-works/>); 15,000 moderators in June 2020

(<https://www.forbes.com/sites/johnkoetsier/2020/06/09/300000-facebook-content-moderation-mistakes-daily-report-says/>); and 15,00 in June 2022

(<https://transparency.meta.com/enforcement/detecting-violations/people-behind-our-review-teams/>).

⁷⁶ See, e.g., <https://www.technologycoalition.org/knowledge-hub/update-on-voluntary-detection-of-csam>.

⁷⁷ The interrogatories are focused on the latest classifiers and do not provide a history of Meta’s technology.

⁷⁸ Fourth supplement to first set of interrogatories. 7:23

⁷⁹ Fourth supplement to first set of interrogatories. 25:2–3.

organized offender enterprises.¹⁸⁴ I write “it is worth studying these cases in detail because they exemplify some of the worst-case scenarios for children and investigators — scenarios that have resulted in part from the lax moderation and regulation environment that is in place today.”¹⁸⁵

Unfortunately, large groups of perpetrators continually form and re-form on Meta’s apps. Facebook Groups are a feature where users can interact with other users with the same interests. For almost a decade, journalists, shareholders, and researchers have noted that Facebook and Instagram Groups are focused on child exploitation. In February 2016, a BBC reporter found “a whole group too - called ‘we love schoolgirlz’ - with obscene content - and that did not get taken down [by Meta].”¹⁸⁶ Because of the article, Will Gardner (CEO of Childnet International) privately wrote to Global Head of Safety Antigone Davis and asked “is there or can there be a mechanism to proactively look at these titles?”¹⁸⁷ He also asked whether more could be done than PhotoDNA, and he asked if the image could be taken down in the comments related to “sexual content involving minors”.

One year later in 2017, the BBC reported that it “reported dozens of photos to Facebook, but more than 80% were not removed. They included images from groups where users were discussing swapping what appeared to be child abuse material. ... the BBC used the report button to alert the company to 100 images which appeared to break its guidelines. They included: pages explicitly for men with a sexual interest in children; images of under-16s in highly sexualised poses, with obscene comments posted beside them; groups with names such as ‘hot xxxx schoolgirls’ containing stolen images of real children; an image that appeared to be a still from a video of child abuse, with a request below it to share ‘child pornography’. Of the 100 images only 18 were removed.”¹⁸⁸ (Meta reported the BBC reporter to law enforcement after some of the images found were forwarded to the company.)

In several presentations, Meta recognized the playbook followed by groups related to child sexual exploitation. For example, included in presentations entitled, “Follow-up discussion on messaging, encryption, and child safety September 2020”¹⁸⁹, “Facebook Child Safety Update, February 2021”¹⁹⁰ and “Meta and CyberTips, December 2021”¹⁹¹ is the following.

- “Examples of Patterns for CEI Group / Admin (WhatsApp):

¹⁸⁴ This term is used instead of “organized crime” as these groups were not organized to make money.

¹⁸⁵ [Levine, Report to Congress, 2022].

¹⁸⁶ “Paedophiles use secret Facebook groups to swap images,” Angus Crawford. BBC News (February 12, 2016). <https://www.bbc.com/news/uk-35521068>.

¹⁸⁷ METANMAG-002-00331837.

¹⁸⁸ “Facebook failed to remove sexualised images of children,” Angus Crawford. BBC News. March 7, 2017 <https://www.bbc.com/news/technology-39187929>.

¹⁸⁹ METANMAG-002-00697829.

¹⁹⁰ METANMAG-002-00698586.

¹⁹¹ METANMAG-002-02118883.

materials as the groups I had reported. ... A week later the largest groups began disappearing. But within months, new ones just as large had replaced them.

Four years from Mr. ██████'s report, in January 2024, Prof. Putnam published a follow-on piece, "Facebook (Still) Has a Child Predation Problem", writing:²⁰²

In the fan groups I found, it was clear how this material could be used to normalize children performing sexually for an audience via smartphone camera. ... Other posts were explicitly sexual in their approach. For instance, in the following screenshot, the background cartoon image shows a girl lying on her stomach with a man's hand reaching from behind to cover her mouth, in a pose suggesting a sex act is underway. The words say: 'If you see your initial [all 26 letters are shown in the image] I'll show it to you in private [eggplant emoji].] The post that followed in the feed called for replies with this invitation: 'Where are the brave and daring girls to play truth or dare?' Attempting to get a sense of the scale of the problem, over the course of a few days I documented twenty different active groups, ranging in size from 12,000 to 60,000 members. Most of these groups were presented to me as 'related groups' from previous groups I visited. A few of them are groups that appeared when I searched for the phrase 'si ves tu edad'—'if you see your age...' In just a few hours, I generated a 114 page compilation of screenshots of child-endangering content from these twenty groups, which as of this writing have over 558,000 members.... At a minimum, it is shocking that no algorithms have been put in place to be triggered by this explicit and constant evidence of the presence of underage children on the platform, and of adult users' age-explicit solicitation of them.

5.1.4 Meta has failed to adequately address persistent, prevalent, and severe CSAM/CSEC on its platforms, and its CSAM/CSEC Detection Efforts are Inadequate

As discussed in the foregoing sections, Meta knew of the scale of the problems it has with detecting and addressing child sexual exploitation on its platforms but did not react with sufficient urgency. Materials provided in discovery show a pattern: the severe dangers to the privacy and safety of minors on Meta's platforms were noted by employees, and the same problems are again found, unaddressed, years later by other employees, and often by journalists and advocates. This pattern shows that Meta knew of serious problems related to

²⁰² "Child Endangerment in 'Los Picus' Fan Groups: Facebook (Still) Has a Child Predation Problem" Lara Putnam, Tech Policy Press. January 23, 2024. <https://www.techpolicy.press/child-endangerment-in-los-picus-fan-groups-facebook-still-has-a-child-predation-problem/>.

CSAM and CSEC, but it lacked the interest or did not have the ability (due to a lack of skills or due inadequately allocating resources) to detect and sufficiently address these problems.

The combination of the enormous popularity of Meta’s platforms and its potent recommender systems has created opportunities for motivated individuals to commit sexual crimes against children and form communities that train each other and normalize their abhorrent behavior. As I wrote in my 2022 Report to Congress: “*Routine activity theory*”²⁰³ predicts that, to put it simply, when motivated individuals have greater, unmoderated access to CSAM online, they will likely download it; and when motivated individuals have greater, unmoderated access to victims online, they will likely sexually exploit them. Several studies have shown that people planning to commit a child sex crime select victims based on ease of access, perceived vulnerabilities, and attraction.²⁰⁴ In short, when greater access to CSAM or victims is coupled with a lack of supervision that would thwart the crime, the result is more crime.”²⁰⁵

In September 2017, a decade since child sexual exploitation concerns had been raised for Meta’s products,²⁰⁶ Meta conducted a “Red Team Exercise” focused on “Sexualize Minors”.²⁰⁷ The experiment found 135 “bad groups” of “minor sexual fetish types” involving real minors. The red team reported the groups to Meta to scrutinize the outcome: “70% of the groups were left up”. They documented many other failures of Meta’s operations.

- “28% of jobs where reps ignored bad groups that did violate our policies are attributable to mistakes in enforcement of our sexual solicitation and CEI policies.
- 97% of non violating ignores contained images of minors posted by adults who had no apparent relationship to the minor
- 9% of groups were escalated from OS to the Safety team, for a specialized CEI review. One group was escalated 91 times. None of the groups were deleted until reviewed as part of the investigation.
- 47% of ignored violating groups did not have past deletions taken into account.
- 35% of the admins of groups we reported had created at least one other bad group.

²⁰³ Internal citation to “Social Change and Crime Rate Trends: A Routine Activity Approach,” Lawrence E. Cohen and Marcus Felson. *American Sociological Review* 44(4): 588-608, 1979. <https://www.jstor.org/stable/2094589>.

²⁰⁴ Internal citation to “A content analysis of pedophile manuals,” Jessica R. Blalock and Michael L. Bourke, *Aggression and Violent Behavior*, Volume 64, 2022. <https://doi.org/10.1016/j.avb.2020.101482>.

²⁰⁵ [Levine, Report to Congress, 2022].

²⁰⁶ See e.g., “New Scrutiny for Facebook Over Predators,” Brad Stone. *New York Times*. July 30, 2007. <https://www.nytimes.com/2007/07/30/business/media/30facebook.html>; “Conn. AG: Facebook Fails to Protect Young Users,” Reuters. July 31, 2007. <https://www.foxnews.com/story/conn-ag-facebook-fails-to-protect-young-users>; “Facebook, N.Y. attorney general reach accord” Caroline McCarthy. Oct 16, 2007. CNET. <https://www.cnet.com/culture/facebook-n-y-attorney-general-reach-accord/>.

²⁰⁷ METANMAG-002-02424826.

- 97% of bad groups were created by adult users, and all of the groups were sexual in nature. 9% of the members were minors, and they are exposed to mature or sexual content.
- Consistency of enforcement against sexualized minor groups is below 30%.”²⁰⁸

An important conclusion of the analysis was that “We likely have many more sexualized minors groups on the site now, and we need to remove them.” Yet, the groups persisted as did Meta’s operational failures, as I show below.

As discussed in Section 5.1.2, in September 2018, Product Manager [REDACTED] presented her findings that the PMYK feature recommended minors to adults known to be “IIC accounts”. As part of her presentation, she revealed Meta’s very loose rules that allowed adults to contact children:

[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

[REDACTED] reported this age rule in an email message to a group of colleagues on June 25, 2018 and no one complained that she misrepresented this formulation.

Progress: ...

Refreshing the set of adult-minor age rules that guide decision making for IIC manual reviews (Antigone Davis, [REDACTED]

[REDACTED]) [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

²⁰⁸ METANMAG-002-02424826.

²⁰⁹ METANMAG-002-00251258.

²¹⁰ METANMAG-011-00029753. The list of names are listed in square brackets in the original. See also METANMAG-007-00026606 for Guy Rosen’s comments (“I think the point about the old rule is really secondary and not that important. I would just do a table with columns for adult age; and max violating minor age”, indicating he agrees implicitly that this is the rule used), as well as METANMAG-002-02123309 which also reports the rule: “Age rules: 3 versions circulating (2x - 7, chart, new chart with broader changes).”

²¹¹ “Who Is Too Young or Too Old for You to Date?” Theresa E. DiDonato. Psychology Today. May 2, 2014.

years old. But Meta's [REDACTED]

Documents show that Meta was aware it did a poor job with recidivist perpetrators on its platform (i.e., users who were banned for child sexual exploitation related activity or content but who then created a new account). In an August 2018 memo, Meta employees reported an analysis of 4,500 accounts that were disabled for Inappropriate Interactions with Children (IIC) between June 4–10. Using in-house technology to identify “single user multiple account” (SUMA) instances, Meta determined that at least 30% of the IIC accounts had another IIC action by July 9, one month later.²¹² (Notably, Meta's internal documents show that it may have vastly underreported the extent of SUMA accounts on its platforms, including those such created, as described here, to facilitate IIC.²¹³)

And Meta knew that its recommendation systems helped recidivists remain in communities. In an August 2018 email message, data science director [REDACTED] draws the following conclusions while enumerating the poor “soft” policies that Meta had in place at the time.²¹⁴ She notes that about 8,000,000 accounts monthly had been labelled as having Inappropriate Interactions with Children (IIC) “Tier 1 and Tier 2” accounts (though surely more had gone undetected). Meta internally defined these terms as follows: “Tier 1 (Real-world harm): Imminent or ongoing harm to a minor, such as intent to meet a child for sex or sextortion. This is illegal and we report to NCMEC as per our legal obligations. IIC Tier 2 (Online harm): Sexual conversation with a minor (colloquially known as ‘grooming’).”²¹⁵

²¹² METANMAG-002-00251258.

²¹³ METANMAG-005-01172946 (November 2021 “SUMA Newsletter” stating that “SUMA and SUMA MAU [Monthly Active Users] are common, especially on IG,” noting that “on IG, 46% of monthly active accounts belong to users with other IG account(s),” that “on FB [Facebook], 28% of monthly active accounts belong to users with other FB account(s),” and that “20% of SUMA MAU belong to users who have used 2+ accounts in the last 30 days.” In contrast, Facebook’s public documents over the same time period stated that for “the fourth quarter of 2021, we estimated that duplicate accounts may have represented approximately 11% of our worldwide MAUs” for Facebook, a far lower total. *See* Meta Earnings Presentation Q4 2021, page 20. https://s21.q4cdn.com/399680738/files/doc_financials/2021/q4/Q4-2021_Earnings-Presentation-Final.pdf).

²¹⁴ METANMAG-002-01453351; note that this email chain is related to [REDACTED]’s efforts from METANMAG-002-00251258 discussed in Section 5.1.2.

²¹⁵ METANMAG-002-01715525; *See also* METANMAG-002-01715525: “IIC Tier 1 (Real-world harm): Imminent or ongoing harm to a minor, such as intent to meet a child for sex or sextortion. This is illegal and we report to NCMEC as per our legal obligations. IIC Tier 2 (Online harm): Sexual conversation with a minor (colloquially known as ‘grooming’).”

In the email, [REDACTED] writes about the lack of resources Meta has to address the problem and the many possible mitigations they do not engage in:²¹⁶

The problem is larger than we thought: there are ~8M MAP [Monthly Active Persons] accounts who engage in IIC (Tier 1 + Tier 2); There is work in progress to ensure we can identify these accounts with high enough precision to action; But these efforts are slow in part because we need more labelers; To get more labelers, we need to get space for them in Mountain View....

For soft actions, should we be more aggressive? Already testing: We will remove minors from PYMK [People You May Know] for flagged accounts; We will block adults from exchanging N&P [nudity and pornography] with minors (should this be a more severe action?). Should we also experiment with blocking other top victim discover channels: Preventing flagged accounts from seeing profiles; Preventing flagged accounts from seeing minors in search results; [REDACTED]

[REDACTED] Putting some limits on how many contacts associated with minors somebody can import. For recidivism, should we be more aggressive: Should we block mobile devices (less likely to be an internet café) affiliated with IIC actioned accounts from creating a new FB account? The piece in red is crazy.²¹⁷

The “piece in red” appears²¹⁸ to be: “we find that the top PMYK suggestion source is ‘friends of users who shared devices (~30%), which indicates that these are likely recidivist IIC accounts and PMYK is helping them rebuild their social graph’. That is, Meta’s recommendation systems are ensuring that persons using Meta for child sexual exploitation are in persistent communities.

A notable April 2019 email message from Messenger RTC team member [REDACTED] describes many findings related to the sexual exploitation of children on Meta platforms.²¹⁹ [REDACTED] provides a good amount of detail — it’s striking how many repeated times over the years each of these inadequate qualities of Meta’s platforms are found, as I point out below. After a litany of issues, [REDACTED] concludes by saying that Meta’s recommendation system is integral to the problem.

- “Instagram is being used by bad actors to find and distribute child exploitative images (CEI), to send sexualizing comments to minors, and to encourage minors into sexual activity (grooming).”
- “... we do not always delete these accounts”
- “there are many cases where we don’t catch CEI media thus it continues to be distributed.”

²¹⁶ METANMAG-002-01453351.

²¹⁷ METANMAG-002-01453351.

²¹⁸ The document is greyscale, but there is only one text section that is highlighted.

²¹⁹ METANMAG-007-00120921.

A plethora of other examples demonstrate these themes, from before and after ██████████ internal email. Meta’s operations may intend to seek out the removal of CSAM but in the end they allow for communities to form consisting of people who sexualize children. Internal documents and investigative news articles, listed below, show how Meta’s recommendation systems bring together communities of persons, and that recommendations by Meta help

²²⁰ Wickr is an app for one-on-one messaging app that offers end-to-end encryption. See <https://en.wikipedia.org/wiki/Wickr> (last retrieved June 2025).

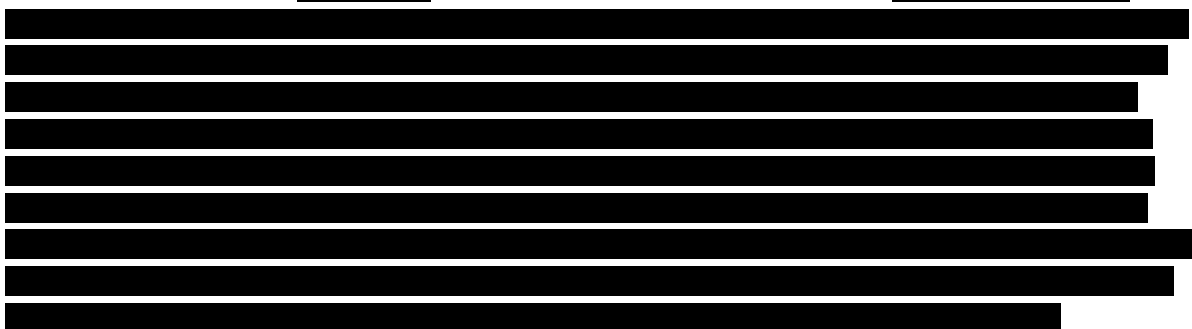
²²¹ This is not the first time that Meta was concerned with “borderline CEI”. See e.g., METANMAG-002-00348534; in March 2017, Sarah Jacobstein of UNICEF USA summarized what her organization knew of Meta’s operational history of letting communities of “borderline CEI” form (in relation to a BBC article) writing: “The ‘sexualization of minors’ policy is relatively new for Facebook and it has been tightened up over the past year. If the images were reported now, they would be removed. Facebook is continuing to work on the issue of comments that may sexualize an innocent photo of a minor. They are continuing to work on using word algorithms and have a firm violations policy — letting individuals know that that if they make similar comments again, their account will be deactivated.” The article is likely: “Facebook failed to remove sexualised images of children”, BBC News. Angus Crawford. March 7, 2017. <https://www.bbc.com/news/technology-39187929>.

²²² See also METANMAG-051-00034456.

recidivists re-build their connections. These communities, once formed, will then use other platforms to exchange content such as Mega.nz.²²³ Meta's poor management of the problem continues year to year, as detailed below.

A full year later, the problems in [REDACTED] email message had not been addressed. In April 2020 [REDACTED] uses [REDACTED] observations and more as the basis of a proposed experiment.²²⁴ [REDACTED] writes that Meta's infrastructure is recommending CSAM/CSEC content to its users, and also that Meta is failing to detect CSAM. [REDACTED] writes that "CEI-related accounts are being recommended to users on our Account Recommendations surfaces ... ~85,000 potential Borderline-CEI accounts were discovered and targeted for Non-Recommendable tier. Many of these accounts are not obviously violating, but they consume violating content and enable a network of distribution amongst CEI consumers and producers. ... there are many cases where we don't catch CEI media thus it continues to be distributed."

In July 2020, the eSafety Commissioner's Office of Australia wrote to Meta with urgent information.²²⁵ Just as [REDACTED] observed more than a year earlier, [REDACTED]



[REDACTED] Figure 3 is one example among many included in the letter.

²²³ [REDACTED] is referring to a well-used practice [REDACTED]

[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

²²⁴ METANMAG-005-01181299.

²²⁵ METANMAG-002-00436879.

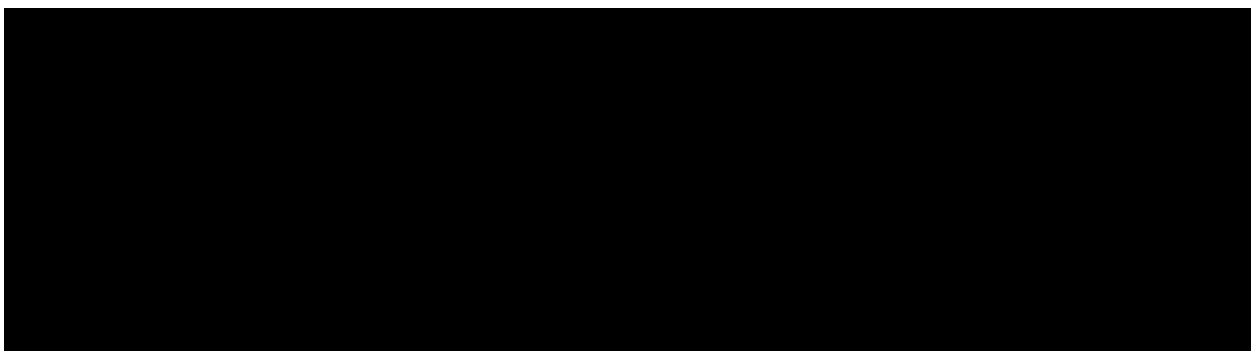


Figure 3. An example sent to Meta by Australia's eSafety Commissioner's Office. Reproduced from METANMAG-002-00436879. I have redacted the face of the minor.

In a March 2020 internal email message, [REDACTED] (a Senior Director of Communications, EMEA, for Instagram) again brings up the persistent problem that Meta's own systems are recommending accounts related to child sexual exploitation as soon as accounts of that types are found; in response Meta reacts in an ad hoc fashion. [REDACTED] states that "famous comedian Oliver Kocher (1.7m [followers]) and his wife Amira (650k [followers]) were sent CEI accounts by some community members and have since discovered dozens more. ... They describe ... how CEI accounts stay up for a long time and grow accounts with up to 7000 followers. Since posting the video, they have sent us 25+ more CEI accounts that they mostly find through our 'Suggestions' in the app and going through followers and comments."²²⁶

In response, employees suggest tactics that, presumably, would provide some immediate mitigation of the problem but that were not on the table until this public attention. And yet Meta for years knew of these problems. Instagram's Head of Safety and Well-being Policy, Vaishnavi Jayakumar, writes she is "in strong agreement that we should over-enforce at the moment and lower our thresholds." Shortly thereafter, Michael Rothschild notes that Meta is unable to fix the problem easily, writing, "[REDACTED]"

[REDACTED] For context, with COVID our CO [Community Operations] queues for CEI/IIC are backed up significantly and with [REDACTED] We're even considering to turn off recommendations as a break glass." Mr. Rothschild's statement is squarely naming recommendations as the cause of these child sexual exploitation harms.

In an October 2020 internal memo, Data Scientist [REDACTED] notes that Meta's strategies for mitigating the sexual exploitation of children result in either "reactive or ad-hoc solutions".²²⁷ Further, [REDACTED] notes that Meta's recommendation systems are connecting predators with each other and with minors. He states, [REDACTED]

²²⁶ METANMAG-002-01826128.

²²⁷ METANMAG-020-00000043.

the harmful effects of not doing enough to limit discoverability range across many different kinds of violations, are exceptionally visible, and often force us into reactive or ad-hoc solutions.”

Meta platforms are not limited to CSAM distribution; they have also persistently hosted child sex trafficking. In a March 2021 email, Market Specialist ██████ becomes another Meta employee to discover the same persistent themes: Meta’s own recommendation systems are encouraging perpetrators to find each other; that Meta has “difficulties” correcting their systems; that accounts from minors and adults are involved in selling CSAM/CSEC via Mega.nz and are easily discoverable; and that recidivism is a “large issue”.²²⁸ ██████ makes it clear that Meta is not used only for CSAM distribution: users are “selling”. A small sample of the problems she specifically enumerates are as follows.

- “We are working on several abuse prevention initiatives targeting the discoverability surfaces identified to be at higher risk [REDACTED]”

[illegible][illegible]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] 229

²²⁸ METANMAG-078-00020508.

²²⁹ METANMAG-078-00020508.

Even though almost two years had passed since [REDACTED] email message (discussed above), [REDACTED] lists those and many other Instagram product gaps, such as: [REDACTED]

[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]²³⁰

In March 2019, [REDACTED] of the National Center on Sexual Exploitation wrote to Meta about child sex trafficking on Instagram:

Last December, I was asked to meet with three survivors of domestic minor sex trafficking here in DC, the oldest being just 15 years old ... Their abuse and exploitation were greatly facilitated using your platform. They showed us message after message they were still getting from strangers asking them for sex. ... We have since reached out to a number of other survivor-serving organizations and are hearing the same things about the prevalence of Instagram used to recruit, groom, and then sell young people.²³¹

In response, copying only Meta employees, Facebook Data Science Manager [REDACTED] wrote, "I believe everyone knows that I'm an advisor at an anti-human trafficking non-profit. The moment Backpage was taken down,²³² all of their data showed that IG had become the leading two-sided marketplace for human trafficking. This is something I have brought up, and it's a challenging situation, which will be made untenable (more) with e2e [end-to-end encryption; discussed below]. This is a problem for which centralization will assist, but the dynamics of IG will demand product interventions beyond the standard detect -> remediate -> measure integrity framework."²³³ Head of Public Policy for Instagram Karina Newton responded, "I think you are exactly right".²³⁴

Despite this warning from [REDACTED], Meta employees found the sexual exploitation of children on Meta from [REDACTED]

[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

²³⁰ METANMAG-078-00020508.

²³¹ METANMAG-002-00559617.

²³² "Backpage.com, a now-shuttered internet forum for prostitution ads that included ads depicting sex work of children" See <https://www.justice.gov/usao-cdca/pr/justice-department-agrees-215-million-settlement-agreement-related-assets-internet>.

²³³ METANMAG-002-00559617.

²³⁴ METANMAG-002-00559617.

²³⁵ METANMAG-051-00034456.

██████²³⁶ Meta employees struggled to address the problem initially: “██████
██████
██████
██████²³⁷

██████ offers “To respond to ██████’s concerns and Antigone’s ‘What is the possible argument for not taking these down?’ Our gap was that the accounts often have not *yet* done anything to breach Child Safety policy and we also would not fan out on requests from brands to take down accounts.” Later, in November 2020, ██████ writes “It’s quite disturbing how often this happens, and every time our teams feel they have fight to get these cases prioritised. ... Setting aside the obvious risks for our users, if anyone from the client decided to take this to the press, there’s no doubt it would cause a major PR issue.”²³⁸

“In 2020, according to a report by US-based not-for-profit the Human Trafficking Institute, Facebook was the platform most used to groom and recruit children by sex traffickers (65%), based on an analysis of 105 federal child sex trafficking cases that year. The HTI analysis ranked Instagram second most prevalent...”, wrote McQue and McNamara in the Guardian almost three years later, in an April 2023 article on how “How Facebook and Instagram became marketplaces for child sex trafficking”.²³⁹ The Guardian article is one of a series that report on Meta’s failures to protect children. The April 2023 article states:

Tina Frundt [founder of a center for survivors of child sex trafficking] who has helped hundreds of children ...since ... 2008, says that the first thing she now does when a young person is referred to her is to ask for their Instagram handle. Other social media platforms are also used to exploit the young people in her care, but she says Instagram is the one that comes up most often.

The reporters directly interviewed five teenage survivors, asking whether

any of them been sexually exploited on social media or had explicit videos or pictures posted of them? They glanced at each other and burst out laughing. Yes, they said, of course. All the time. One girl said she felt that ‘nobody at Instagram cares, they don’t care what’s posted. They don’t care shit about us.’ Frundt claims that she is constantly asking Instagram to close accounts and take down exploitative content of kids in her care.

²³⁶ METANMAG-079-00174396.

²³⁷ METANMAG-075-00684440.

²³⁸ METANMAG-075-00684440.

²³⁹ “How Facebook and Instagram became marketplaces for child sex trafficking”, Katie McQue and Mei-Ling McNamara The Guardian, April 27, 2023.

<https://www.theguardian.com/news/2023/apr/27/how-facebook-and-instagram-became-marketplaces-for-child-sex-trafficking>.

Citing government prosecutions, the April 2023 Guardian article notes that Meta platforms are the most popular for child sex trafficking:

According to Goldworm [former assistant district attorney in Boston], while his investigations involved every social media platform, Meta platforms were the ones he encountered most often. Six other prosecutors in several different states told us that, in their experience, Facebook and Instagram are being widely used to groom children and traffick children. Five of these prosecutors spoke of their anger over what they felt were Meta's unnecessary delays in complying with judge-signed warrants and subpoenas needed to gather evidence on sex trafficking cases. "We get a higher rate of rejected warrants from Facebook than any other electronic service provider," claimed Gary Ernsdorff, senior deputy prosecuting attorney for King County, Washington state. "What I find frustrating is that the exchange can delay rescuing a victim by a month."

About one year later, in a March 2024 article in the Guardian, journalist McQue writes:

For the past year, the Guardian has been tracking Instagram accounts hosting content that advocates such activity as well as those that encourage violence against the women under the control of a pimp – someone who makes money from selling others for sex. The accounts identified often use hashtags as well as code phrases commonly associated with sex work to make it easier for buyers to locate them. ... According to child safety experts, the dangers to children and teenagers of accounts promoting sexual exploitation and 'pimp culture' must be taken more seriously if vulnerable young people are to be better protected online. ... Tina Frundt ... said that she has tried to alert Meta and other social media companies to the dangers of pimp accounts and the potential links to sexual exploitation and trafficking. 'In my experience, when you see these hashtags, it really is pimps who are running these accounts,' said Frundt. 'They recruit online on a daily basis. There is an impact to the propaganda that pimps put out. They make it seem funny when it's actually a reality. They desensitise people so that the youth don't understand what they're getting into'.²⁴⁰

In the article, we see another example of Meta determining in an ad hoc fashion what fraction of content is acceptably dangerous: "A Meta spokesperson said: '... the majority of content under the #pimp hashtag is about comedy, clothes or music.'" In other words, Meta found it acceptable to maintain the #pimp hashtag as long as only a minority of such content violates safety policies. According to the article, "As a result of the Guardian's investigation, Meta has closed down some Instagram accounts and prevented users from searching for

²⁴⁰ "'Pimps' use Instagram to glorify sexual violence and abuse, investigation finds", Katie McQue, Mar 5, 2024, The Guardian. <https://www.theguardian.com/global-development/2024/mar/05/pimps-use-instagram-to-glorify-sexual-violence-and-abuse-investigation-finds>.

hashtags and emojis associated with pimp accounts that were used on more than 350,000 posts, the apparent minority of posts that were not actioned prior to the Guardian article.”²⁴¹

Years later, in 2023, Meta is persistent in recommending CSAM- and CSEC-related content. A series of articles from the Wall Street Journal reiterate and reveal more details of these persistent themes: Meta’s recommender systems are connecting users to CSAM- and CSEC-related accounts, groups, chats and more; Meta, one of the most resourceful and resource-rich companies on the planet, is either unable to fix the system or unwilling to fix it, even with pressure from US senators and representatives and the attention of the world’s media.

In June 2023, an article in Wall Street Journal detailed how

Instagram connects pedophiles and guides them to content sellers via recommendation systems that excel at linking those who share niche interests, the Journal and the academic researchers found. ... In its reporting, the Journal consulted with academic experts on online child safety. Stanford’s Internet Observatory ... produced an independent quantitative analysis of the Instagram features that help users connect and find content. The Journal also approached UMass’s Rescue Lab,²⁴² which evaluated how pedophiles on Instagram fit into the larger ecosystem of online child exploitation. Using different methods, both entities were able to quickly identify large-scale communities promoting criminal sex abuse.²⁴³

The article details how hashtags are used to link a community of predators to each other: “Search terms such as #pedobait and variations on #mnsfw (‘minor not safe for work’) had been used to tag thousands of posts dedicated to advertising sex content featuring children,

²⁴¹ The McQue article continues, “Court papers seen by the Guardian from the prosecution of the 25-year-old child trafficker Aryion Jackson state that Instagram gave Jackson easy access to teenage girls who he groomed online and then trafficked into sexual exploitation. ... Jackson and other criminal gang members uploaded videos and images of girls under their control to Instagram Stories, a feature on the site where content disappears after 24 hours. Prospective buyers were then able to respond through the platform’s direct message function.” And that “another man [Kevondric Fezia] convicted of sex trafficking children in 2022 ... advertised the sexual services of girls as young as 14 over Instagram and Instagram Stories. ... Yet the child-trafficking activities taking place on Jackson and Fezia’s Instagram accounts were not picked up by Meta’s content-moderation software or staff, and their business activities continued for months. Law enforcement were alerted to their crimes only when the survivors and families contacted the police.”. See also “Two from Texas sentenced for bringing 14-year-old girl to La. to be sex trafficked,” KPLC Digital Team. KPLC News. June 24, 2022 <https://www.kplctv.com/2022/06/24/two-texas-sentenced-bringing-14-year-old-girl-la-be-sex-trafficked/>.

²⁴² This is my research lab.

²⁴³ “Instagram Connects Vast Pedophile Network” Jeff Horwitz and Katherine Blunt. June 7, 2023. Wall Street Journal. <https://www.wsj.com/tech/instagram-vast-pedophile-network-4ab7189>; full text <https://archive.ph/UKGXJ>.

rendering them easily findable by buyers, the academic researchers found.” In many cases, Instagram has permitted users to search for terms that its own algorithms know may be associated with illegal material. In such cases, a pop-up screen for users warned that “These results may contain images of child sexual abuse,” and noted that production and consumption of such material causes “extreme harm” to children. The screen offered two options for users: “Get resources” and “See results anyway.”

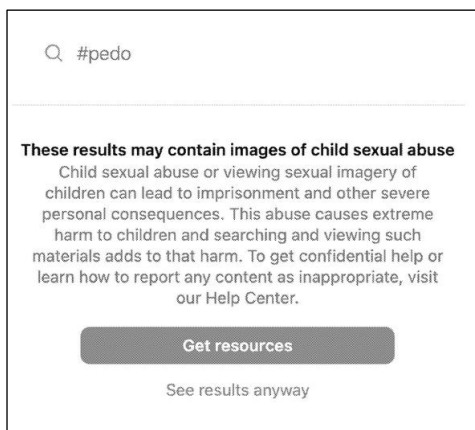


Figure 3. Screenshot of the interstitial reporter Jeff Horwitz gave as an example to Nick Clegg. Note that Mr. Horwitz’s example is for the hashtag “#pedo”. From METANMAG-055-00036192.

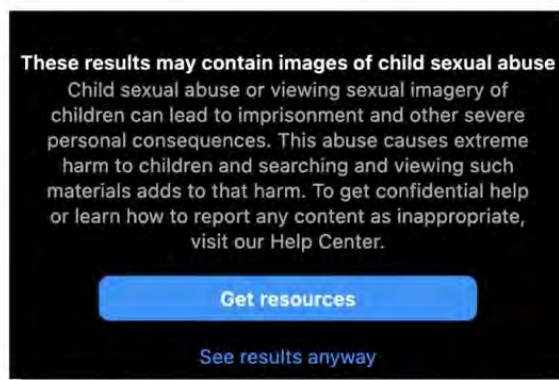


Figure 4. A higher-quality, color version of the interstitial that appeared in the Wall Street Journal article. See Footnote 243.

Notably, this warning is consistent with the policy that Meta had in place in 2017. In a presentation entitled “Instagram Product Policy & CO [Community Operations] Escalation Paths Overview”, [REDACTED] (a Meta Public Policy Manager) et al. write: “Instagram has developed multiple hashtag tools to prevent them from being used to spread violating material and to warn the community when they view hashtags associated with sensitive content. Interstitials²⁴⁴ when content is sensitive (self-injury, gun sale related); Restricted to ‘top only’ when the hashtag contains large amounts of abusive content; Blocked in rare cases when the hashtag itself violates or is consistently and overwhelmingly used for abuse.”²⁴⁵ Screenshots from the presentation are reproduced below. Note that the warning offers to “See Posts Anyway” and that abusive content still shows the “top” results.²⁴⁶

²⁴⁴ An interstitial is a pop-up text (or video) that appears ahead of the content a user requested that must be clicked to remove (or watched before it disappears).

²⁴⁵ MT-IG-AG-NM-000018658.

²⁴⁶ See all METANMAG-005-00010318 where Mr. Horwitz writes to Nick Clegg: “The above [Figure 4 in this document] brought to mind Antigone Davis’s comment to the Guardian a few weeks back that ‘we’ve been able to identify the kinds of searches that people do when they’re searching for trafficking content, so that when people search for that, we will pop up with information to divert them or to let them know that what they’re doing is illegal activity’. I very much understand why offering resources and reminding a user that their search queries are noticeable is a good policy. I

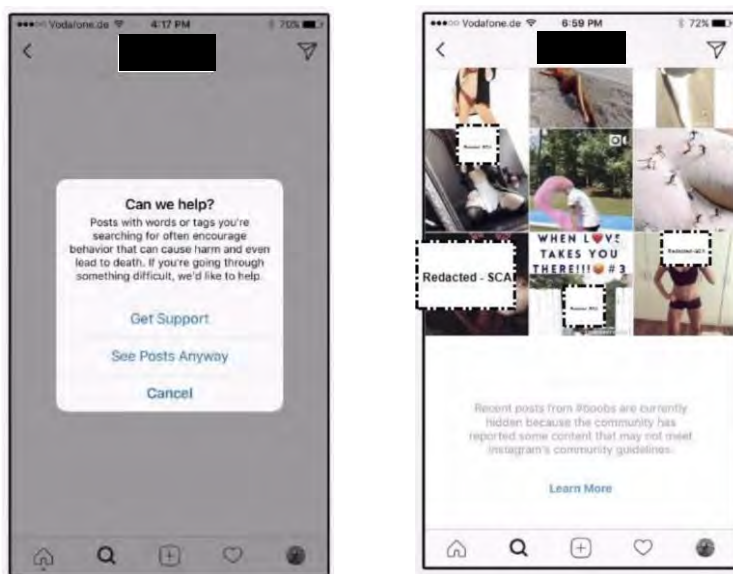


Figure 5. Reproduced from MT-IG-AG-NM-000018658.

The June 2023 article detailed, and as I observed directly while sitting with Jeff Horwitz, that thousands of accounts on Instagram related to child sexual exploitation were easily found through hashtags. Once an account was found, Meta would recommend more accounts to the user. Mr. Horwitz wrote, “Test accounts set up by researchers that viewed a single account in the network were immediately hit with ‘suggested for you’ recommendations of purported child-sex-content sellers and buyers, as well as accounts linking to off-platform content trading sites. Following just a handful of these recommendations was enough to flood a test account with content that sexualizes children.”

This June 2023 article captures many problems with child sexual exploitation on Instagram that were noted in [REDACTED] email message from four years earlier (discussed above). The article notes the large communities of predators and CSAM traders that formed on Meta platforms: “The pedophilic accounts on Instagram mix brazenness with superficial efforts to veil their activity, researchers found. Certain emojis function as a kind of code, such as an image of [REDACTED] n”—or one of ‘cheese pizza,’ which shares its initials with ‘child pornography,’ according to Levine of UMass.”

I observed directly as part of the WSJ investigation the same things as [REDACTED]: accounts focused on CSAM/CSEC were easily locatable and numerous. [REDACTED]

have had a hard time understanding, however, why after informing people they’re about to open a box of child porn IG would want to let them go ahead and open it.” He’s referencing a quote from “How Facebook and Instagram became marketplaces for child sex trafficking”, Katie McQue and Mei-Ling McNamara. The Guardian. April 27, 2023. <https://www.theguardian.com/news/2023/apr/27/how-facebook-and-instagram-became-marketplaces-for-child-sex-trafficking>.

[REDACTED]

The Wall Street Journal article references a Stanford technical report that concludes among other details that “Instagram’s user suggestion recommendation system also readily promotes other SG-CSAM accounts to users viewing an account in the network, allowing for account discovery without keyword searches. Due to the widespread use of hashtags, relatively long life of seller accounts and, especially, the effective recommendation algorithm, Instagram serves as the key discovery mechanism for this specific community of buyers and sellers.”²⁴⁸ The Stanford 2023 whitepaper also matches [REDACTED] email message. They write: “Actual content delivery [of CSAM] appears to happen on file sharing services such as Dropbox or Mega—links to these services are often seen in ‘proofs’ of transactions organized over DMs.”

Mr. Horwitz contacted Meta ahead of publication, and in an internal email, employees agreed with the WSJ piece.²⁴⁹ Indeed, the article reports that “A Meta spokesman acknowledged that Meta had received the reports and failed to act on them.” In a May 2023 internal email discussion in response to Mr. Horwitz’s email message, [REDACTED] wrote about how Meta’s operational “gaps”, deficient “policy”, and poor human “reviewer guidance” are responsible for these problems though they have been documented for years as shown above.²⁵⁰ Ms. [REDACTED] wrote:

Jeff has accurately identified previously unknown/unintended gaps in our reactive enforcement when violating content is reported by users. ... We have so far attributed the reactive enforcement gaps that Jeff has surfaced to technical errors that we’ve opened a SEV to address - reports not reaching reviewers - and to human error, where reviewers incorrectly actioned reports. ... [REDACTED]

[REDACTED] ... Investigating Jeff’s reports surfaced enforcement gaps when content is reported by users - gaps that we need to close given imperfect proactive detection and the severity of the violation type. Of the 428 reports Jeff appears to have made from various test accounts, we found that we correctly actioned only 25%, or 105 reports. Of the 75% of reports (323 reports) that we failed to review, remove and/or apply appropriate account-level consequences, we are conducting a comprehensive root cause analysis and have so far attributed the errors to the following issues:

²⁴⁷ METANMAG-002-00436879.

²⁴⁸ “Cross-Platform Dynamics of Self-Generated CSAM,” David Thiel, Renee DiResta, and Alex Stamos. June 6, 2023. <https://cyber.fsi.stanford.edu/io/publication/cross-platform-dynamics-self-generated-csam>.

²⁴⁹ METANMAG-082-00010159.

²⁵⁰ METANMAG-082-00000699.

1. Errors with auto-closing. We do not intend to auto-close (without human review) any user report that is tagged as ‘involving a child’ - every report should be reviewed. We have discovered 69 reports where this happened and have opened a SEV to mitigate and resolve those issues. (There were a number of reports that were properly eligible for auto-closure as they were reported for different violation types or not tagged as involving a child)

2. Reviewer guidance. We found that 104 reports were reviewed by and incorrectly ignored by human reviewers, [REDACTED]

[REDACTED] We are updating the reviewer guidance and our systems to [REDACTED]

[REDACTED]

[REDACTED]

policy will also implement a regular cadence of updating guidance to reviewers.²⁵¹

The article correctly reports my experience: “Following the company’s initial sweep of accounts brought to its attention by Stanford and the Journal, UMass’s Levine checked in on some of the remaining underage seller accounts on Instagram. As before, viewing even one of them led Instagram to recommend new ones. Instagram’s suggestions were helping to rebuild the network that the platform’s own safety staff was in the middle of trying to dismantle.”

My experience is entirely consistent with the state’s investigation of Meta that appears in New Mexico’s complaint from December 2023, such as paragraphs 68 through 94 and the case studies presented afterwards.²⁵²

In the leadup to the WSJ publication, due to discussions with Mr. Horwitz, Meta engaged in a “sprint” that resulted in many changes. It is unclear why Meta could attempt this effort and make these changes only due to public pressure rather than from its own numerous prior analyses of the same issues. A June 2023 email from Andrea Saul enumerates the actions.²⁵³

- “We’ve included the list of shorter term actions we have taken as a result of the XFN [cross functionality] sprint that are complete or will be a focus leading up to the publication of the story....
“1. [TRRM] We are going to temporarily show empty search results and/or block hashtags for a number of non-violating terms (i.e underage, teen, teen model). This list is based on terms we believe may be the focus of targeted scrutiny in the days following publication of the article. Given the popularity of some of these terms, we believe this will be noticed and plan to confirm reactively. We will plan to lift these measures at the conclusion of the product-led child safety taskforce.

²⁵¹ METANMAG-082-00000699.

²⁵² NMAG-SM-0371076; *see also* NMAG-PLTF-META-0003420, NMAG-PLTF-META-0003424, and NMAG-PLTF-META-0003427, as well as the Expert Report of Rickey Kitch.

²⁵³ METANMAG-055-00035193.

- 2. [TRRM] We will continue the XFN war room through next week to respond to real-time content and product issues surfaced by the reporter, Stanford Internet Observatory, and others after the story publishes. This is in addition to the daily sweeps we've been conducting, which will continue through next week. Information that is surfaced will be shared with the Child Safety Task Force to inform the work and exit criteria.
- 3. [REACTIVE ENFORCEMENT] We fixed the specific and most glaring issue of auto-closing reports that are tagged as "involves a child." The SEV is still open so we can conduct more analysis and a wider audit of auto-close rules to proactively uncover additional gaps.
- 4. [REACTIVE ENFORCEMENT] By early this week we will update our systems so that reviewers can see terms we have identified as coded language used by pedophiles and to take that into context when reviewing content.
- 5. [REACTIVE ENFORCEMENT] A reminder was sent to all OS reviewers about using account-level context in review and known child safety related terms.
- 6. [REACTIVE ENFORCEMENT] We have submitted for approval on a stat saying we disabled 397 out of the 448 accounts identified by the Stanford Internet Observatory and submitted 90 reports to NCMEC, as well as 5 enhanced manual reports, some of which, we are told, resulted in law enforcement action. We are conducting a Strategic Network Disruption (SND) using the accounts as our seed set.
- 7. [POLICIES] We blocked an additional —2700 terms we've identified as violating under the new adjustments to our hashtag and search queries policies for terms that are descriptive of sexual activity and associated with children, or Child Sexual Exploitation, Abuse or Nudity, even if those terms are not violating on their own. We'll let Jeff know we have made thousands of additions to the list this week.
- 8. [CONTENT] We are re-launching Instagram's child exploitation interstitial with updated language that makes it clear in the header that child sexual abuse is illegal, and we will not allow people to click through to see results from these searches.
- 9. [PROACTIVE ENFORCEMENT] We have approval on a stat saying we removed over 90,000 accounts due to CSAM-specific Strategic Network Disruptions from 2020-2022 from 2020-2022.
- 10. [PROACTIVE ENFORCEMENT] We are creating a monthly exercise between ops, product and policy to consistently update hashtag, search terms, and reviewer protocols with new coded language from pedophilic networks to keep ahead of those attempting to evade our systems."²⁵⁴

The article and its results received a great deal of attention worldwide.²⁵⁵ Following the article, letters expressing grave concerns were sent to Meta from the United States Senate's

²⁵⁴ METANMAG-055-00035193.

²⁵⁵ See Meta's collection of some of the attention at METANMAG-021-00025009.

Committee on Commerce, Science, and Transportation,²⁵⁶ United States Senate Committee of the Judiciary,²⁵⁷ and U.S. Senators Richard Blumenthal (D-CT) and Marsha Blackburn (R-TN).²⁵⁸ I have not seen Meta's responses to these letters. But it is notable that even after this level of attention, Meta still could not adequately fix these problems.

Specifically, in September 2023, Mr. Horwitz emailed Meta again and provides many examples of the same issues happening, particularly where Meta is recommending child sexual exploitation.²⁵⁹ He notes that Stanford was able to find hashtags that replaced the ones that were taken down, revealing Meta's ad hoc approaches rather than fundamental changes. He writes,

Stanford's done a bit of followup and found that the company just still isn't tracking this stuff. [REDACTED] is alive and well. And from there it couldn't be easier to find [REDACTED]

He also reports how, in July, he reported to Meta:

a picture of a girl hog tied and gagged inside a washing machine. The second one is an account entirely devoted entirely to extreme closeups of a girl's crotch" but they had not been reviewed. ...

Broadly, I know the company is working on trying to prevent recommending pedophilic content accounts but some spot checks don't seem to suggest it's an especially successful system. [REDACTED]

²⁵⁶ <https://www.commerce.senate.gov/services/files/1CEAE1E3-2AA0-44F7-B54F-FB9F9B79FC1D%3E>.

²⁵⁷ Re: Report of Instagram's Algorithm Facilitating Pedophile Networks. June 28, 2023 https://www.lgraham.senate.gov/public/_cache/files/3f6802b7-254d-477c-8df0-a0c5d303a6b0/2023.06.28---meta-letter-sjc.pdf.

²⁵⁸ https://www.blumenthal.senate.gov/newsroom/press/release/blumenthal-and-blackburn-demand-meta-answer-for-instagrams-failures-to-protect-children-from-predators_exploitation

²⁵⁹ METANMAG-021-00494992; see also an earlier email from Mr. Horwitz as METANMAG-010-00364985 and METANMAG-010-00147591. Later in January 2024 METANMAG-021-00548772

²⁶⁰ [REDACTED]

Below I've re-printed the screenshots that Mr. Horwitz sent Meta of his searches for "child links" (left), and the results when he searched accounts for same term. (right)



Figure 6. Screenshots sent by Mr. Horwitz to Meta in September 2023, reproduced from METANMAG-021-00494992.

Mr. Horwitz continued,

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

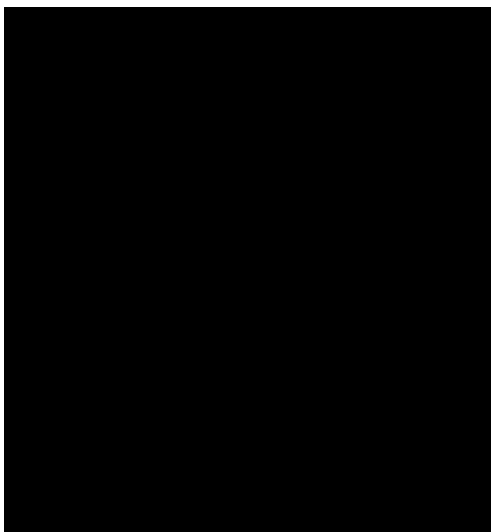


Figure 7. Cropped screenshot sent by Mr. Horwitz to Meta in September 2023, reproduced from METANMAG-021-00494992.

In an October 2023 internal chat, Comms [Communications] and i3 [Integrity, Investigations, & Intelligence] Child Safety members discussed various concerns, including “enforcement figures from the GRO investigation on the AYSF [accounts you should follow] case” originally brought up by Mr. Horwitz, to whom Meta recommended 172 CP seller accounts.²⁶¹ Several employees added clarity.

- The i3 Child Safety team confirmed that none of Mr. Horwitz’s accounts had what Meta calls a “malicious child safety actor” (MaSCA) score, even though the accounts he was following and interacting with were strictly focused on child sexual exploitation.²⁶²
- [REDACTED] questioned, “can we take a look here while ops is pulling the accounts from the video for why his accounts don’t have a MaSCA score [REDACTED] [REDACTED]”
- [REDACTED] responds by explaining the limitations of Meta’s approach: [REDACTED]
[REDACTED]
[REDACTED]

²⁶¹ METANMAG-003-00521329; (WIT) Nick Clegg, (Meta)(Vol. 2) - 03-21-2025.LEF_exhibit 63.pdf.

²⁶² Meta moved to this term from “IIC”; see METANMAG-002-01429719. I have not seen documents that detail how this heuristic is computed. See also, e.g., METANMAG-011-01239859 (2023 internal document regarding “High Risk Actors” and their relationship with MaSCA); METANMAG-028-00004296 (discussion regarding how [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

We have not been able to include the wider set from the taskforce in this list as the precision is varying.”

- [REDACTED] notes that she easily recreates Mr. Horwitz’s results (apparently this type of simple “smoke testing” is not performed at Meta): “Also, just flagging that I’m using my test user account to surface the accounts, and also started being recommended some myself.” And adds two new accounts to the take down list.
- [REDACTED] summarizes the status: “In addition to the 12 accounts featured in Jeff’s video, and the 204 IG accounts we disabled from preliminary fan-outs, GRO reviewed 1.6K users that were device linked to the original disabled accounts. We conducted subject element reviews (profile photo, username, and bio), and disabled 987 additional IG accounts for CSAM trade. In total, the investigation resulted in 1203 IG account takedowns. The majority of the users were based in India and engaging in trading/selling of telegram links.”
- She also writes: [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

It’s worth repeating that in 2019, [REDACTED] wrote internally that “[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]”²⁶³

Yet, years later, in late December 2023, employee [REDACTED] wrote in an internal chat that [REDACTED]
[REDACTED]²⁶⁴

Despite Meta’s efforts, five months later in December 2023, the WSJ checked again and found the same and similar problems that they identified six months earlier (and that [REDACTED] had raised in 2019) had persisted; writing:²⁶⁵

- “Five months later, tests conducted by the Journal as well as by the Canadian Centre for Child Protection show that Meta’s recommendation systems still promote such content. ... Company documents reviewed by the Journal show that senior Meta executives earlier this year instructed the company’s integrity team, which is responsible for addressing user safety issues, to give priority to objectives including reducing ‘advertiser friction’ and avoiding mistakes that might ‘inadvertently limit well intended usage of our products.’ Those objectives were listed as above traditional

²⁶³ METANMAG-007-00120921.

²⁶⁴ METANMAG-051-00047830.

²⁶⁵ “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>. Meta’s collection of stories: METANMAG-010-00443262.

safety work focused on harmful content, such as child exploitation, in planning documents viewed by the Journal”

- “On a Journal Instagram test account, Meta wouldn’t allow search results for the phrase ‘Child Links,’ but the system suggested an alternative: ‘Child Pornography Links.’ After Meta blocked that term following a query by the Journal, the system began recommending new phrases such as ‘child lingerie’ and ‘childp links.’”
- “Recently, Facebook’s ‘Groups You Should Join’ feature has suggested topics such as kidnapping, ‘dating’ children as young as 11 and even chloroforming women. After the Journal reported the chloroform groups to Meta, the company took them down. Within two weeks, however, an 1,800 member group specifically devoted to chloroforming children was back in Facebook’s ‘Groups You Should Join’ recommendations. In it, users posted pictures of girls with rags held over their faces. One member posted a picture of a smiling young girl with the caption of ‘is it suitable for kidnapping?’ Fellow group members agreed that the girl was. After the Journal flagged that group, Meta took it down as well.”

Meanwhile, a concurrent November 2023 Wall Street Journal investigation again shows problems with Meta recommending child sexual abuse:

Journal reporters set up the Instagram test accounts as adults on newly purchased devices and followed the gymnasts, cheerleaders and other young influencers. The tests showed that following only the young girls triggered Instagram to begin serving videos from accounts promoting adult sex content alongside ads for major consumer brands, such as one for Walmart that ran after a video of a woman exposing her crotch. ... Current and former Meta employees said in interviews that the tendency of Instagram algorithms to aggregate child sexualization content from across its platform was known internally to be a problem. Once Instagram pigeonholes a user as interested in any particular subject matter, they said, its recommendation systems are trained to push more related content to them. ... ‘Time and time again, we’ve seen recommendation algorithms drive users to discover and then spiral inside of these online child exploitation communities,’ McDonald said, calling it disturbing that ads from major companies were subsidizing that process.²⁶⁶

(Lianna McDonald is the executive director of the Canadian Centre for Child Protection.) This reporting is also consistent with New Mexico’s investigation of Meta’s platform, which appears in the complaint.²⁶⁷

²⁶⁶ “Instagram’s Algorithm Delivers Toxic Video Mix to Adults Who Follow Children,” Jeff Horwitz and Katherine Blunt Nov. 27, 2023. <https://www.wsj.com/tech/meta-instagram-video-algorithm-children-adult-sexual-content-72874155>; full text at <https://archive.ph/gwtwO>.

²⁶⁷ See New Mexico’s Amended Complaint for Abatement and Civil Penalties and Demand for Jury Trial; see also NMAG-PLTF-META-0003420, NMAG-PLTF-META-0003424, and NMAG-PLTF-META-0003427, as well as the Expert Report of Rickey Kitch.

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 [REDACTED] 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 of [REDACTED]

[REDACTED]²⁷⁰ Additionally, [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-000025487.

²⁷³ METANMAG-007-00305682.

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-051-00021365.

²⁷⁶ METANMAG-051-00125572.

²⁷⁷ METANMAG-051-00125572.

²⁷⁸ METANMAG-010-00180536.

²⁷⁹ METANMAG-079-00018237.

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]

[REDACTED] “ [REDACTED]
[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-075-00132827.

²⁸¹ A February 2019 document describes an effort to improve Meta’s computational methods for determining IIC (METANMAG-005-01844878); [REDACTED]
[REDACTED]
[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]
[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 HI 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), [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: “[REDACTED] 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?” [REDACTED]: “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]
[REDACTED]
[REDACTED]
[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] over time impacting the Child Safety team's ability to detect violating content.
- [REDACTED]
[REDACTED]
[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]
- [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 [REDACTED], 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 P1s [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.

- [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
- [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
- [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
- [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED] potential violating verified accounts may not be classified as IIC potential violators

- 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.”)

organization and region and make it hard for adults to contact users under the age of 18.”²⁹⁹ Of course, that is only possible if the age data is correct.

Meta’s policies assert that its family of apps require users to at least be 13 years old in the US.³⁰⁰ Confusingly, Meta’s advertisements on the Apple App Store states that Facebook,³⁰¹ Instagram,³⁰² Messenger,³⁰³ and WhatsApp³⁰⁴ are for ages “12+”. The Messenger Kids mobile app is rated for ages 4 and older, but its advertisement, unlike the other apps, is careful to clarify “4+, Made for Ages 9–11”.³⁰⁵ Meta’s Messenger application for MacOS is incorrectly advertised as being appropriate for ages “4+”.³⁰⁶

Quantitative studies regularly conclude that Meta’s platforms have millions of users under the age of 13.³⁰⁷ A 2011 Consumer Reports study found that “Of the 20 million minors who actively used Facebook in the past year, 7.5 million of them were younger than 13, according to projections from Consumer Reports’ latest State of the Net survey. Facebook’s terms of service require users to be at least 13 years old. Also among this group of minors using Facebook, more than 5 million were 10 and under.”³⁰⁸ A 2020 Pew research study found that “some 5% of parents of a child age 11 or younger said that, as far as they know, their child uses Instagram. When considering only parents with a child ages 9 to 11, however, this share was 11%.”³⁰⁹ A 2021 study from Common Sense media reported on the “percent of all 8- to 12-year-olds who have ever used: Instagram 10%; Facebook 8%.”³¹⁰ Similarly, a 2021 study

²⁹⁹ “Conn. AG: Facebook Fails to Protect Young Users,” Fox News, July 31, 2007.

<https://www.foxnews.com/story/conn-ag-facebook-fails-to-protect-young-users>.

³⁰⁰ For example see Facebook: <https://www.facebook.com/help/958848942357089>; Instagram <https://www.facebook.com/help/instagram/966909308115586/confirming-your-age-on-instagram/>; and WhatsApp <https://faq.whatsapp.com/240694875565320>. All last retrieved July 2025.

³⁰¹ <https://apps.apple.com/us/app/facebook/id284882215> last retrieved July 2025.

³⁰² <https://apps.apple.com/us/app/instagram/id389801252> last retrieved July 2025.

³⁰³ <https://apps.apple.com/us/app/messenger/id454638411> last retrieved July 2025.

³⁰⁴ <https://apps.apple.com/us/app/whatsapp-messenger/id310633997> last retrieved July 2025.

³⁰⁵ <https://apps.apple.com/us/app/messenger-kids/id1285713171> last retrieved July 2025.

³⁰⁶ <https://apps.apple.com/us/app/messenger/id1480068668> last retrieved July 2025.

³⁰⁷ See also “At Meta, Millions of Underage Users Were an ‘Open Secret,’ States Say,” Natasha Singer. November 25, 2023. New York Times.

<https://www.nytimes.com/2023/11/25/technology/instagram-meta-children-privacy.html>

³⁰⁸ <https://www.consumerreports.org/media-room/press-releases/2011/05/cr-survey-75-million-facebook-users-are-under-the-age-of-13-violating-the-sites-terms/> “CR Survey: 7.5 Million Facebook Users are Under the Age of 13, Violating the Site’s Terms” May 10, 2011 Consumer Reports magazine.

³⁰⁹ “8 Facts About Americans and Instagram.” Pew Research Center. October 21, 2020. Brooke Auxier. <https://www.pewresearch.org/fact-tank/2021/10/07/7-facts-about-americans-and-instagram/>

³¹⁰ “The Common Sense Census: Media Use by Tweens and Teens, 2021” Victoria Rideout, Alanna Peebles, Supreet Mann, Michael B. Robb. 2022. Common Sense. https://www.commonsensemedia.org/sites/default/files/research/report/8-18-census-integrated-report-final-web_0.pdf.

from Thorn found from a survey of 2,002 minors that 40% of those under the age of 13 used Instagram at least once a day, and 59% of those ages 13–17 used it daily.³¹¹ A 2023 peer-reviewed study by Raffoul et al. estimated that over 25 million U.S. users of Facebook and Instagram are under the age of 18, including over 6 million children under the age of 13 (with over \$5 billion in estimated advertising revenues derived from them).³¹²

Prior to 2021, Facebook relied almost exclusively on the ages that users self-reported at sign-up.³¹³ Facebook’s system for verifying minors in 2008, which put the onus of verification on users, was as follows. “Facebook employs a system of peer verification for users who identify themselves as under 18. This system relies on answers to questions accompanying friend requests to help determine if the user sending those requests attends a particular high school or knows the people he or she is contacting. Accounts are either verified or disabled based on these answers.”³¹⁴

In 2008, Facebook’s danger to children on its network was so significant that 48 Attorneys General came together to change Facebook’s policies via an agreement: “the first time a Facebook user wants to change his or her age, website staff will review their profile to determine whether the change is appropriate. ... It will remove groups for incest, pedophilia, cyberbullying and other violations of the site’s terms of services, as well as expel from the site individual violators of those terms.”³¹⁵

In April 2012, Meta was planning deployment of “Project Kid,” which had the goal of registering “25M MAU [Monthly Active Users] within a year,” including users under 13.³¹⁶ In their justification, the Meta presentation acknowledges that kids are already on the platform by lying about their age, and sought to normalize children to social networking from before they could speak (no minimum age is apparent in the presentation). They write: “allowing <13s [under 13] to join Facebook without lying allows us to give them the

³¹¹ “Responding to Online Threats: Minors’ Perspectives on Disclosing, Reporting, and Blocking” Thorn and Benson Strategy Group. May 2021.

https://info.thorn.org/hubfs/Research/Responding%20to%20Online%20Threats_2021-Full-Report.pdf; see also a related European survey: <https://www.netfamilynews.org/kids-teens-social-networking-in-europe>.

³¹² “Social media platforms generate billions of dollars in revenue from U.S. youth: Findings from a simulated revenue model,” Raffoul A, Ward ZJ, Santoso M, Kavanaugh JR, Austin SB (2023) . PLOS ONE 18(12): e0295337. <https://doi.org/10.1371/journal.pone.0295337>.

³¹³ Facebook, How Do We Know Someone Is Old Enough to Use Our Apps?, July 27, 2021, <https://about.fb.com/news/2021/07/age-verification/>.

³¹⁴ https://cyber.harvard.edu/sites/cyber.law.harvard.edu/files/ISTTF_Final_Report.pdf

³¹⁵ “Attorney General Larry Long Announces Nationwide Agreement with Facebook” News release, South Dakota Attorney General’s Office. May 8, 2008.

https://news.sd.gov/news?id=news_kb_article_view&sys_id=6b1f0d001b5869506e4aa97ae54bcbbd.

³¹⁶ METANMAG-002-02352081.

appropriate protections involving parents with their children's use of the site allows them to teach their child the correct norms of use ('training wheels' for social networking)".

In 2013, CNN reported that "if someone under 18 wanted to bypass the setting on Facebook before today, they could easily lie about their age when signing up for an account. Children under 13 are not officially allowed to sign up for a Facebook account, though they can skirt the rules in the same way. When someone underage does sign up for an account, Facebook assumes they have the permission of at least one guardian but does not verify it in any way."³¹⁷

The reason for CNN's story was Facebook's announcement that "Starting today, people aged 13 through 17 will also have the choice to post publicly on Facebook. While only a small fraction of teens using Facebook might choose to post publicly, this update now gives them the choice to share more broadly, just like on other social media services. In addition, teens will be able to turn on Follow so that their public posts can be seen in people's News Feeds. As always, followers can only see posts they are in the audience for."³¹⁸ Accordingly, this opened to teens (and children under 13 on the platform) the chance to be followed by adult strangers who exploit children, as I have covered elsewhere in this report.

For years, Meta's method of detecting children under 13 years of age was based on user-provided data only and ad hoc heuristics. Even in August 2018, as a presentation entitled "Underage Enforcement Update" states, Meta detected users as under 13 years old using heuristics.³¹⁹ Specifically, Meta looks to see if either: [REDACTED]

[REDACTED]

[REDACTED]

Meta was also internally aware that children under 13 were on Facebook and Instagram. A January 2018 memo, entitled "H2/H1 2018 Youth Team Review" notes that

The majority of US kids with phones use social media by age 9. Parents know about and connect to U13 [children under 13 years of age] on social media. Facebook is likely U13's first social media app (driven by parents). U13s use multiple social media apps, but struggle to articulate FB's value. ... approximately 56% of the global teen population is active on FB each month. ... Looking at people we predict to be 13 and 14 today, we can estimate that there were 4M people under 13 in 2015 on IG.

³¹⁷ "Facebook changes privacy settings for teens," Heather Kelly. CNN. October 31, 2013. <https://www.cnn.com/2013/10/16/tech/social-media/facebook-teens-privacy/index.html>.

³¹⁸ "Teens Now Start With 'Friends' Privacy for New Accounts; Adding the Option to Share Publicly," Facebook Newsroom. October 16, 2013. <https://about.fb.com/news/2013/10/teens-now-start-with-friends-privacy-for-new-accounts-adding-the-option-to-share-publicly/>.

³¹⁹ METANMAG-002-01410104.

This represents around 30% of all 10-12 years old in the US. Finally, assuming similar numbers today, 75% of US teens are MAP on IG.³²⁰

The document makes clear Meta's mission: "Three years from now we want to be the most relevant social products for Youth worldwide. To get there, we are focusing on each youth life stage, 'Kid' (6-10), 'Tween' (10-13) and 'Teen' (13+) and are building social products for them."

A September 2018 report confirms Meta's ad hoc heuristics were not effective.³²¹ The report, entitled, "Content Governance and Enforcement: Transparency", contains a section on company risks from child safety issues: "... people (especially policymakers and law enforcement) do think we create the opportunity for children to be groomed. Until we have really strong mechanisms for preventing inappropriate adult/minor contact, I think we risk generating more negative press than positive – especially when we do very little to keep U13s [children under 13 years of age] off our platform."

It was not until December 2019 that Instagram introduced controls on users' ages, and even then, they relied on users to self-report their date of birth. Karina Newton, then Instagram's Head of Public Policy, announced to a Safety Advisory Board (SAB) in an email message that "Historically, we didn't require people to tell us their age because we wanted Instagram to be a place where everyone can express themselves fully - irrespective of their identity."³²²

In the email, Newton stated that Instagram was announcing two updates as part of their ongoing work to keep young people safe: "Starting today, we will require new users to let us know their age before they can sign up for Instagram." Despite the sentence above about people expressing themselves fully, Ms. Newton indicates that she understood knowing users' real ages is critical to keeping the platform safe: "Understanding how old people are on Instagram is a central part of the work we're doing to keep young people safe. It will allow us to create age-appropriate experiences for our youngest community members, while living up to our longstanding rule to ensure people who are underage do not have access to Instagram. In the coming weeks, they would be rolling out a new setting in direct messages that would allow people to choose to no longer receive messages or be added to chat groups by people they didn't follow."³²³

Shortly after Instagram's belated implementation of a stated-age check, Will Gardner from ChildNet wrote to Meta in February 2020. He zeros-in on the critical problem that Meta has

³²⁰ MT-IG-AG-NM-000020872. Month based on document metadata.

³²¹ MT-IG-AG-NM-000038465. Month was determined via document metadata.

³²² MT-IG-AG-NM-000124720.

³²³ 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).

from not enforcing age controls: child safety controls across Meta platforms are not effective when they don't know the ages of users.³²⁴ He writes "it is such a good step that there is a need to declare age for Insta. ... Now this opens up the possibility of enabling additional protections for young users." He lists issues that young people had raised during his work on schools, including with Digital Leasers. Specifically: As of February 2019, they found that "42% of 8–17s in the UK said that 'Someone put me in a group chat that I didn't want to be in without asking me (like a Whatsapp group)'." He also said that "Users should be asked consent before being added to a group, and be able to see who else is in that group [...] including those in the group who the user has blocked."; "Children want to be able to permanently decline a group chat"; "You can report a group but not an individual in a group"; and "The admin of a Group cannot delete the group" As for blocking, he stated "there are still some things that the person you have blocked can do" including: "still see if you are online", "see your updated profile image," and "able to see each others messages in a group, and even chat you in a group". Additionally, "Children we speak to are still struggling to be able to report on WhatsApp." For Messenger, "vulnerable young people [...] are struggling with the amount of access Messenger gives."

In a presentation entitled, "Safety Policy: 2019 Safety & Well-Being Goals" Global Head of Safety Antigone Davis states in the notes: "What are our current weaknesses? Safety. We have a weak system for keeping U13s [under 13 years of age] off the platform and verifying age for protection of minors. We have a weak system for thwarting IIC [Inappropriate Interactions with Children] (early stage of classifier, little in app education for U18 [under 18] /new users re: identifying problematic interaction, few programs to build skills and resiliencies needed). ... We offer almost no in-product victim support. We are behind on integrity across growth vectors (PYMK [People You May Know], GYSJ [Groups You Should Join], PYML [Pages You May Like]). We lack a robust/showcase safety by design process."³²⁵ And similarly, in an August 2020 update, Meta AI engineer [REDACTED] writes "Age determination is a pain. This is not new, we knew this but we still have tremendous problems with age liars and age confessions. We are experimenting with a more advanced version of detecting age confessions."³²⁶

In documents, Meta states that that prior to this date, users who were not logged into the system — and therefore Meta could not know their age — could view a great deal of content: "In 2019, Meta instated new limits on the amount of content a logged-out user could see when accessing Instagram via web browser. Prior to this period there were some restrictions on the types of content that were visible to logged-out users (e.g., logged-out users could not access private pages or direct messages)."³²⁷

³²⁴ MT-IG-AG-NM-000124720.

³²⁵ METANMAG-002-02396133.

³²⁶ METANMAG-005-00458148.

³²⁷ METANMAG-002-00230805.

A February 2020 presentation entitled, “Youth Privacy Update” offers the status of Meta’s attempts to keep children under 13 (“U13s”) off their platform.³²⁸ Speaker notes confirm that “Proactive removal of u13s is the lynchpin to addressing public criticisms around our perceived failure to prevent the sexual abuse and grooming of minors, bullying and harassment of minors, and as a platform where minors are exposed to suicide and self-injury content and suicide-promotional communities.”

The presentation states that Meta’s “Current Approach to Age Management” is to simply request the users age at registration for Instagram and Facebook without verification. Meta is “Not proactively enforcing, except for when underage users are community reported.” The presentation notes that a “prediction model” is “under escalation” that could be used to “checkpoint likely underage users”. A related document dated March 2020 notes that [REDACTED]

[REDACTED]
[REDACTED]
[REDACTED].”³²⁹

A February 2020 email thread among VP of Facebook’s News Feed John Hegeman, the Head of the Facebook App Fidji Somo, and others with the subject “Age Management Decisions” outlines upcoming decisions regarding age management on Facebook and Instagram, emphasizing increased external pressure regarding under-13 users and teens.³³⁰ Deputy Chief Privacy Officer Rob Sherman notes that there is a “general understanding that an unknown number of u13s are on FB and especially IG, and lying about their age.” He notes that while steps have been taken to update age flows, the decision on how to handle age falsification was paused due to limited resources.

Mr. Sherman proposed discussing the following at a March 2020 meeting: whether to require users to state their age at registration instead of defaulting to over 13, whether to develop classifiers to proactively identify under-13 users for human review, and whether to consider other proactive enforcement measures.³³¹ He highlighted regulatory pressure to adopt stricter age verification, including requiring ID, treating all users as minors without effective age enforcement, and potentially raising the age of consent to 16 or 18. Sherman emphasized that taking action is necessary to “gain credibility needed to influence upcoming regulation in this area.”

An October 2020 presentation discussed above (“Child Safety State of play”) notes many limitations of Meta’s ability to manage underage users.³³²

³²⁸ METANMAG-010-00250808.

³²⁹ METANMAG-082-00000404.

³³⁰ METANMAG-002-00157461.

³³¹ METANMAG-002-00157461.

³³² MT-IG-AG-NM-000180935. This October 2020 date is based on discovery metadata.

- [REDACTED]
[REDACTED]
[REDACTED]
- U18 [under 18] enforcement: age gating relies on [REDACTED]
[REDACTED]; lack checkpoint
- Age verification: pushed to device, slow advancement, dependent on x-industry cooperation
- Age models: [REDACTED]
[REDACTED]
[REDACTED]
- Defaults, gating, rules: inconsistent, need wholesale review
- Minors in-product education: minimal
- Content: inappropriate/harmful content and experiences for minors”

During an October 2020 internal chat, [REDACTED] (Facebook data scientist) writes: “[REDACTED]

[REDACTED]³³³ And Instagram Director of Data Science Darius Kilstein replies, “[REDACTED]

[REDACTED] The employees make clear they know their technology is poor; [REDACTED] writes, “... our investigators have given feedback that

In a February 2021 discussion with colleagues about “younger kids (10-12 yos)”, [REDACTED] (UX researcher) noted that “there are definitely kids this age on IG [Instagram] ... they aren’t supposed to be on IG at all.”³³⁴

And two years after Ms. Newton’s email message discussed above,³³⁵ in a March 2021 presentation entitled “Messenger Wellbeing, All Hands,” Global Head of Safety Antigone Davis states the same critical problem, calling out Meta’s “[REDACTED]

[REDACTED]³³⁶ She makes clear

³³³ METANMAG-002-01314875; *see also, e.g.*, METANMAG-002-00013903 (“The issue here is that people lie, and we know that’s quite prevalent... Especially groomers - they lie to bypass our detection and even help them find minors.”)

³³⁴ MT-IG-AG-NM-000193948.

³³⁵ MT-IG-AG-NM-000124720.

³³⁶ METANMAG-002-00698654.

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]
[REDACTED]
[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.

Mr. Zuckerberg's feedback was "Could we revise this to: 'We will provide tools to parents to ensure u13s are in safe experiences,'" while "Boz [Andrew Bosworth]" noted "We're weak here. In theory parental controls help us on this path but they are delayed." The message states that the result of the discussion was to change the policy language to Mr. Zuckerberg's suggested wording. Notably, in the same thread, Mr. Zuckerberg's feedback on child safety goals of "Provide all developers w/ core tech to detect and prevent harmful behavior," and "Build cross-industry integrity tools to prevent worst actors in AR/VR [augmented reality/virtual reality]" was to walk it back: "Can signal, but aspirational" for both.

The effects of Meta's inability to validate ages is striking. For example, in a September 2021 write up, Meta data scientists Alison Lee and [REDACTED] discover that users ages 15 and younger are creating videos with partial nudity via Meta's TikTok-like video product Reels.³³⁹ They found that "Reels featuring kids/tweens, and teens (i.e., 0-15) dancing and posing are pervasive, and often show up on people's Reels experience, even in cold starts. While much of this content is benign (they may be uninteresting or uncomfortable for users, but not necessarily harmful), some may include risky elements (i.e., partial nudity, suggestive dance moves that are borderline CEI) that make both viewers uncomfortable and the subjects at risk...."

In a December 2021 report of the project's outcomes, Meta employees found that 20% of the accounts were "personal accounts of a teen girl, with no apparent parental supervision."³⁴⁰ And 39% were "personal accounts of an estimated under-13 minor with no apparent parental supervision. ... It's important to note that all of these accounts were marked as either 'adults' or 'unknown' by our internal stated age, which would mean that account-level age ranking and privacy treatments would leave these users vulnerable."

A September 2021 Wall Street Journal article quotes Instagram head Adam Mosseri as agreeing publicly that "you can lie about your age now [on Instagram]."³⁴¹ The same article states that Meta "had removed more than 600,000 accounts in the past three months for violating its rules on age limits," citing Meta spokesman Andy Stone as the source of the statistic.

5.1.6 Meta's platform features leave minors vulnerable and therefore at risk for exploitation

"Kyle Robinson is one year into serving a 10-year sentence at a federal prison in Massachusetts for sex trafficking two teenagers, one only 14 years old. ... Instagram, he said, was his platform of choice. 'I find the girls that have pride in themselves, but maybe don't

³³⁹ METANMAG-011-01747923.

³⁴⁰ METANMAG-018-00001967.

³⁴¹ "Facebook's Effort to Attract Preteens Goes Beyond Instagram Kids, Documents Show," Georgia Wells and Jeff Horwitz. Wall Street Journal. Sept. 28, 2021. <https://www.wsj.com/articles/facebook-instagram-kids-tweens-attract-11632849667>; full text: <https://archive.ph/8JNx6>.

hash set and it's only after our CEI proactive detection catches it or it's posted somewhere on our public platform and someone reports it that it's possible for us to bank the image, run detection across threads and find that hundreds or thousands of copies have already been distributed via MSGR." And "High Level Takeaway: Our current strategy focused on detection and enforcement is not working well today, and will slowly stop working as we move towards E2EE."

Overall, Meta enjoys an important economic benefit from enabling E2EE: if more of Meta's traffic is encrypted, then it has far less need for human reviewers and moderators. Moderators are expensive,³⁹⁶ and Meta has employed about 15,000 moderators annually.³⁹⁷ Moreover, those moderators have suffered secondary trauma from exposure to the material on Meta's platforms, and some have sued Meta over these harms.^{398,399,400} Indeed, earlier this year, Meta announced sweeping reductions in its reliance on human content moderation teams,⁴⁰¹ which

³⁹⁶ "The Silent Partner Cleaning Up Facebook for \$500 Million a Year," Adam Satariano and Mike Issac. August 31, 2021. New York Times.

<https://www.nytimes.com/2021/08/31/technology/facebook-accenture-content-moderation.html>

³⁹⁷ August 2018: 7,500 moderators (<https://www.vice.com/en/article/how-facebook-content-moderation-works/>); June 2020: 15,000 moderators

(<https://www.forbes.com/sites/johnkoetsier/2020/06/09/300000-facebook-content-moderation-mistakes-daily-report-says/>); June 2022: 15,000 moderators:

<https://transparency.meta.com/enforcement/detecting-violations/people-behind-our-review-teams/>.

³⁹⁸ "I was a content moderator for Facebook. I saw the real cost of outsourcing digital labour," Sonia Kgomo. The Guardian. February 12, 2025.

<https://www.theguardian.com/commentisfree/2025/feb/12/moderator-facebook-real-cost-outsourcing-digital-labour> "For two years, I spent up to 10 hours a day staring at child abuse, human mutilation, racist attacks and the darkest parts of the internet so you did not have to. It was not just the type of content I had to watch that gave me insomnia, anxiety and migraines, it was the quantity too."

³⁹⁹ "'I didn't eat or sleep': a Meta moderator on his breakdown after seeing beheadings and child abuse," Rachel Hall and Claire Wilmot. The Guardian. April 27, 2025.

<https://www.theguardian.com/technology/2025/apr/27/meta-moderator-on-the-cost-of-viewing-beheadings-child-abuse-and-suicide>.

⁴⁰⁰ "Meta faces Ghana lawsuits over impact of extreme content on moderators," Rachel Hall and Claire Wilmot. The Guardian. April 27,

2025. <https://www.theguardian.com/technology/2025/apr/27/meta-faces-ghana-lawsuits-over-impact-of-extreme-content-on-moderators> "Meta is facing a second set of lawsuits in Africa over the psychological distress experienced by content moderators employed to take down disturbing social media content including depictions of murders, extreme violence and child sexual abuse."

⁴⁰¹ "Meta shelves fact-checking in policy reversal ahead of Trump administration", Katie Paul, Supantha Mukherjee and Deborah Mary Sophia. Reuters. January 7, 2025.

<https://www.reuters.com/technology/meta-ends-third-party-fact-checking-program-adopts-x-like-community-notes-model-2025-01-07/>; 5Rights Foundation, Meta's rollback on safety measures puts children at risk, Jan. 8, 2025, <https://5rightsfoundation.com/meta-rollback-on-safety-measures-puts-children-at-risk/>.

CEO Mark Zuckerberg has openly acknowledged means “we’re going to catch less bad stuff” despite the proliferation of “child exploitation” on its platforms.⁴⁰²

5.1.8 Meta delayed implementing safety features until years after their need was first identified

Meta has deployed many features on its platforms that are relevant to my report. Meta’s response to Interrogatory No. 11 provides summary details. Below, I chronologically list features that are most relevant to child safety from that response on Instagram as the response did not provide the same information for its other platforms. Note that Instagram allows for public accounts where anyone can see the posted content, and private accounts where the viewers are restricted to specific other Instagram accounts.

- 2012 (no month provided): users are first able to delete specific comments on their posts.
- Dec. 2016: The option to turn off commenting was extended to all Instagram users.
- Dec. 2016: Instagram allows private accounts to remove followers they had previously approved
- Jun. 2017: Instagram employs a filter that will block certain offensive comments on posts and in live video, for English only.
- Sep. 2017: Users with public accounts can choose who can comment on their posts.
- May 2018: A filter hides comments containing attacks on a user’s appearance or character, as well as threats to a user’s well-being or health.
- July 2019: Instagram warns users that a comment they have drafted may be offensive.
- October 2019: Instagram allows users to “restrict” others from posting comments on their account that third parties can see, and from seeing if the post was read.
- May 2020: users can manage who can tag or mention them on Instagram.
- May 2020: In Messenger, “young people” are warned to be cautious when interact via direct messages with accounts who Instagram believes to be an adult who has exhibited “suspicious behavior”, and offers to “end the conversation or block, report, or restrict the adult.”
- Mar. 2021: The tool warning about direct messages with “suspicious adults” is deployed for Instagram as well.
- Apr. 2021: Instagram begins automatically filtering Direct Message requests that contain offensive words, phrases, and emojis.
- July 2021: Instagram allows users to determine how much content that is “sensitive” but not against policy appears in their Explore field.

⁴⁰² “Transcript: Mark Zuckerberg Announces Major Changes to Meta’s Content Moderation Policies and Operations,” Justin Hendrix. Tech Policy Press. January 7, 2025.
<https://www.techpolicy.press/transcript-mark-zuckerberg-announces-major-changes-to-metas-content-moderation-policies-and-operations/>.

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>.

through Instagram’s official reporting function. Because there were no obvious criteria for alerting the company to potential child exploitation, they were logged as ‘nudity or pornography.’ The videos remained online for five days. It was only after Business Insider contacted Instagram’s press office that the content was removed. By this time, the two videos — and other versions uploaded by the second user — had more than 1 million views. Instagram left the accounts that posted the videos active, however. Business Insider asked why the accounts were left up, as Instagram has a ‘zero-tolerance policy’ on child abuse. Instagram said the policy applied to the content and not to the account uploading it.” In an internal discussion of the article, it was noted that the account owner of one of the videos shared a device with an account that was previously disabled for “CEI/ CEI solicitation and sexual exploitation for minor sexualization”.⁴³⁷ This again documents that Meta is not carefully restricting predators on its platforms.

An internal email from June 2020 from Director of Product Management for Integrity Sayed Otaru provides another example of an employee noting that by letting users stay on the platform even after they “engage in IIC [Inappropriate Interactions with Children]”, Meta is allowing communities of perpetrators to form. This quote also exposes the extent of Meta’s problem, noting that each day 50,000 “underage accounts” are subject to IIC in English-speaking markets alone. Yet leadership does not prioritize these issues. Specifically, Mr. Otaru writes:

[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]
[REDACTED] The CI [Content Integrity] team has concluded a study (the numbers have not yet been formalized) and currently show that 500,000 IG underage accounts receive IIC on a weekly basis. The IG prevalence number is 3x Messenger’s.⁴³⁸

Data Scientist [REDACTED] corrects the statistic as “The metric is 500k victims per DAY in English markets only” [emphasis in original]. Product Manager [REDACTED] responds, “If we have confidence in this metric, we should publish it and hold ourselves accountable to it.” Well-Being Director of Engineering Yoav Shapira responds, “Child Safety is explicitly called out as a non-goal in our H2 plans. So if we can do something here, cool. But if we can do nothing at all, that’s fine too.”⁴³⁹

⁴³⁷ METANMAG-051-00015509: “@ [REDACTED] is the responsible account for the video in question ... It looks like [REDACTED] was disabled for CP_DNR on 4-28-18. @ [REDACTED] and @ [REDACTED] share associated devices. Also [REDACTED] (IGU 561384950) changed the username from @ [REDACTED] to [REDACTED] on 2016-06-04.” Definition: “cp_dnr for CEI/ CEI solicitation and sexual exploitation for minor sexualization”. More details of CP_DNR at METANMAG-002-01584619.

⁴³⁸ MT-IG-AG-NM-000160510 and MT-IG-AG-NM-000160523.

⁴³⁹ MT-IG-AG-NM-000160520.

Rather than reconsidering whether they can provide safety to minors, another team is given a chance at mitigation in the latter half of 2020.⁴⁴⁰ The conversation includes a discussion of how IIC is not the highest priority for Meta (election interference is labelled as the highest priority). Well Being Director of engineering [REDACTED] states that “total prevalence reduction was 23%”, calling “a *major* win”, yet this result leaves majority of IIC instances unmitigated. He adds “we will hold the line on IIC T2 [Meta’s term for “grooming of a minor”] prevalence on IG & Messenger, and respond to any spikes, but will not be goaling on prevalence reduction” even though he states that Instagram has “approximately 3x the size of the Messenger problem”.⁴⁴¹

In a November 2018 status update, ██████████ reports that until that date, Meta’s recommender algorithms were helping disabled accounts rebuild their network.⁴⁴² He writes: “if an IIC (Inappropriate Interaction with Children) [user] has an account disabled and creates/uses another one using the same device, such feature could inadvertently help them rebuild their friend graphs via signals from the disabled account on the same device. In this launch, we stopped using signals from disabled IIC accounts to help make friend suggestions for other accounts on the same device”.

Years later, in November 2023, New Mexico’s investigation determined that

[REDACTED]
[REDACTED]
[REDACTED] [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED] [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED]

In the report, the author provides an example of an account where the majority of the hashtags on the account were already “restricted”, including [REDACTED].⁴⁴⁵ Another example was of an account that were suspended bragging that they were back in hours, writing “i’m backkkk, got sused [suspended] this mornin ,, hmu 2 [hit me

⁴⁴⁰ METANMAG-002-00577650.

⁴⁴¹ MT-IG-AG-NM-000160510; and see METANMAG-002-01771928 from October 2020, which states that, “A recent exploration into IIC on IGDM, revealed that there are ~38x as many victims on IG as on FB. This is mostly due to there being many more minors on IG messaging with adults. Despite this, CEI deletes on IG are ~21x lower and IIC auto-enforcing on IG is 90x lower compared to on FB.”

⁴⁴² METANMAG-002-00861611.

443 NMAG-PLTF-META-0003420.

⁴⁴⁴ “Cheese Pizza” has the same first letters as “child pornography” (i.e., CSAM).

445 NMAG-PLTF-META-0003420.

up to] get some stuffs, [REDACTED]
[REDACTED] and they included: [REDACTED]
[REDACTED]
[REDACTED]
[REDACTED] A third example was an account with [REDACTED]. Notably this is months after the Wall Street Journal's "Vast Array" article showing the same type of problem.

In January 2024, the same investigation notes that Meta is ineffective in closing accounts and provides specific examples: “ [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED] 448

⁴⁴⁶ NMAG-PLTF-META-0003424.

⁴⁴⁷ See NMAG-PLTF-META-0003433 and NMAG-PLTF-META-0003434.

⁴⁴⁸ See New Mexico’s Amended Complaint for Abatement and Civil Penalties and Demand for Jury Trial, at page 33.

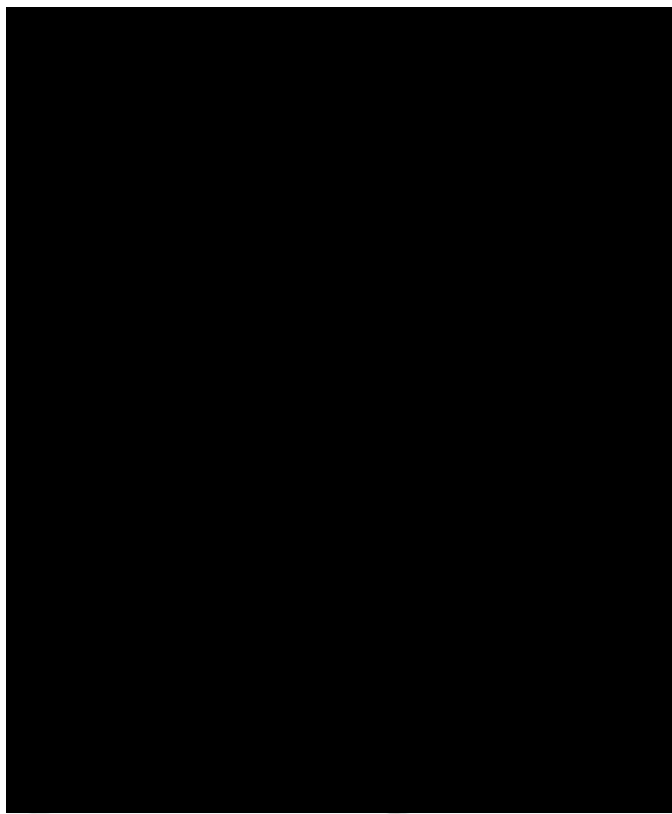


Figure 10 taken from New Mexico's Amended Complaint for Abatement and Civil Penalties and Demand for Jury Trial, at page 33.

All in all, these problems and the size of the platform have created a problem that cannot be addressed adequately by law enforcement. State law enforcement, such as the ICAC Task Forces are understaffed when compared to the level of incidents that occur on Meta's platforms. In my 2022 Report to Congress, in which I interviewed ICAC commanders and others in law enforcement, I concluded, "almost no agency can keep up with the number of tips they receive and leads they could pursue."⁴⁴⁹ In addition, the secondary trauma experienced by law enforcement can result in high turnover and a continual need for training of new investigators.⁴⁵⁰ The Expert Report of Rickey Kitch confirms that for the New Mexico ICAC Task Force and affiliated investigators, cases of child exploitation require intensive effort and a large amount of resources and training.⁴⁵¹

⁴⁴⁹ [Levine, Report to Congress, 2022].

⁴⁵⁰ [Levine, Report to Congress, 2022].

⁴⁵¹ See Expert Report of Rickey Kitch.

5.1.11 Meta's policies on reporting harms to children are exaggerated

Meta's public statements about child sexual exploitation are misleading. They state very prominently in their "Policy Details" for "Child Sexual Exploitation, Abuse, and Nudity" that "We do not allow content or activity that sexually exploits or endangers children. When we become aware of apparent child exploitation, we report it to the National Center for Missing and Exploited Children (NCMEC), in compliance with applicable law."⁴⁵² This wording has been posted since at least November 19, 2020 according to Meta's website.⁴⁵³

In an August 2022 email message, Content Policy team member [REDACTED] writes "Updating the policy ... was deliberate. We updated the policy line that gives us the most coverage and conveys that we take 'activity' into account when assessing accounts participating in child sexual exploitation, including seeking out CSAM." That is, the term "child sexual exploitation" is larger than CSAM images/videos and includes seeking out CSAM.⁴⁵⁴

Meta has repeated the vague statement "When we become aware of apparent child exploitation, we report it to NCMEC" many times to reporters without clarification; e.g., in June 2017 to Motherboard,⁴⁵⁵ in February 2018 to WNBC,⁴⁵⁶ in July 2019 to the Telegraph,⁴⁵⁷ and in April 2020 to NBC Universal.⁴⁵⁸ An internal "Child Safety Research" presentation from October 2018 is broader still, stating that "When we become aware of a credible threat or sexual exploitation of a minor, we notify NCMEC."⁴⁵⁹

However, internally, employees state that not all "child exploitation" is reported to NCMEC. Employees show they understand that [REDACTED] (see below).

⁴⁵² Meta, "Child Sexual Exploitation, Abuse, and Nudity"

<https://transparency.meta.com/policies/community-standards/child-sexual-exploitation-abuse-nudity> (last retrieved June 2025).

⁴⁵³ One must click on "Nov 19, 2020" to see the wording at this URL:

<https://transparency.meta.com/policies/community-standards/child-sexual-exploitation-abuse-nudity>

⁴⁵⁴ For example, Thorn defines "child sexual exploitation" (which is more specific than "apparent child exploitation") as "a broad term that encompasses all forms of child sexual exploitation. Sometimes CSEA (Child Sexual Exploitation and Abuse) is also used. This may include many of the other terms on this list, such as child sexual abuse material, child sex trafficking, and others like sextortion. This is one of the broadest ways to talk about the issue as a whole, rather than the individual and nuanced pieces." From "Decoding the Acronyms Used in Fighting Child Sexual Abuse. Thorn. March 17, 2023. <https://www.thorn.org/blog/decoding-the-language-of-child-sexual-exploitation-acronyms-to-know/>.

⁴⁵⁵ METANMAG-002-01400547.

⁴⁵⁶ MT-IG-AG-NM-000022350.

⁴⁵⁷ MT-IG-AG-NM-000070712.

⁴⁵⁸ MT-IG-AG-NM-000138960.

⁴⁵⁹ METANMAG-002-02259384.

As discussed elsewhere in this document, a 2016 presentation⁴⁶⁰ notes that [REDACTED]
[REDACTED]
[REDACTED] And in an August 2018 email message, a Meta employee
writes that “[REDACTED]
[REDACTED]

⁴⁶¹

In a March 2018 presentation entitled, “Sexual Exploitation: CEI Operational Guidelines for Outsourcing FB & IG” presented by the “CO [Community Operations] Safety” team, instructions are [emphasis in original] to [REDACTED]
[REDACTED]

- [REDACTED]
 - [REDACTED]
 - [REDACTED]
 - [REDACTED]
 - [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]
- [REDACTED]

Even years later, in a January 2024 internal chat, [REDACTED] asked whether “[REDACTED]
[REDACTED]
[REDACTED] responds, “[REDACTED]
[REDACTED]

⁴⁶³

Similarly, an April 2023 article from the Guardian reports that “Between 2009 and 2019, Meta reported just three cases as suspected child sex trafficking in the US to NCMEC, according to records disclosed in a subpoena request seen by the Guardian.”⁴⁶⁴ Meta disputes this number, and the article notes that “A Meta spokesperson claimed that over the past

⁴⁶⁰ MT-IG-AG-NM-000013685.

⁴⁶¹ METANMAG-011-01772307.

⁴⁶² Cog05950.

⁴⁶³ METANMAG-005-01727319.

⁴⁶⁴ “How Facebook and Instagram became marketplaces for child sex trafficking”, Katie McQue and Mei-Ling McNamara. The Guardian. April 27, 2023.
<https://www.theguardian.com/news/2023/apr/27/how-facebook-and-instagram-became-marketplaces-for-child-sex-trafficking>.

decade, the company had reported ‘tens of thousands of accounts which violated our policies against child sex trafficking and commercial child sexual abuse material to NCMEC.’ When we put these claims to NCMEC, it said that it had not received “tens of thousands” of reports of child trafficking from Meta, but had received that number related to child abuse imagery.”

5.1.12 Meta set up features that are exploited by predators to garner revenue from sexually exploiting children

At least two reports have illuminated the problems with Meta creating a revenue stream for its users.

In February 2024, a Wall Street Journal noted that “Meta last year began a broad rollout of tipping and paid-subscription services, part of an effort to give influencers a financial incentive for producing content Meta staffers found that its algorithms promoted child-modeling subscriptions to likely pedophiles, and in some cases parents discussed offering additional content on other platforms, according to some people familiar with the investigations. ... [Meta] chose to build an automated system to prevent suspected pedophiles from being given the option to subscribe to parent-run accounts. The technology didn’t always work, and the subscription ban could be evaded by setting up a new account.”⁴⁶⁵ The article continued “In two instances, the Journal found that parent-run accounts were cross-promoting pinup-style photos of children to a 200,000-follower Facebook page devoted to adult-sex content creators and pregnancy fetishization.”

And a March 2024 Guardian article investigated Meta Pay’s use in child exploitation crimes. “Meta Pay, known as Facebook Pay before rebranding in 2022, is a peer-to-peer payment service enabling users to transfer money over the company’s social networks. ... Through reviewing documents and interviewing former Meta content moderators, a Guardian investigation has found that payments for child sexual abuse content taking place on Meta Pay are probably going undetected, and unreported, by the company... ‘When Meta Pay is linked to Messenger or Instagram, the messages associated with payments could uncover illicit behaviors,’ said Krupena. ‘Now that this context is removed, the implications are significant. It almost feels like encryption is inadvertently facilitating illicit activity. This opens many opportunities for criminals to hide in plain sight.’”⁴⁶⁶

⁴⁶⁵ “Meta Staff Found Instagram Tool Enabled Child Exploitation. The Company Pressed Ahead Anyway”, Jeff Horwitz and Katherine Blunt Feb. 22, 2024 <https://www.wsj.com/tech/meta-staff-found-instagram-subscription-tool-enabled-child-exploitation-the-company-pressed-ahead-anyway-a18e81e6>; full text <https://archive.ph/iKJzP>.

⁴⁶⁶ “How Facebook Messenger and Meta Pay are used to buy child sexual abuse material. Court documents and interviews detail the products’ role in alleged exploitation – and how some payments go undetected.” Katie McQue Guardian March 22, 2024. <https://www.theguardian.com/global-development/2024/mar/14/facebook-messenger-meta-pay-child-sexual-abuse-exploitation>.