Meta

April 11, 2022

John F. Clark & Yiota Souras
NCMEC
333 John Carlyle Street Suite #125
Alexandria, Virginia 22314-5950
USA
(Sent via electronic mail)

Dear John and Yiota,

Thank you for your letter setting forth your concerns regarding recent technical anomalies that increased the overhead on our shared goal of protecting children and preventing and removing Child Sexual Abuse Material (CSAM) from our platform. With this reply we lay out the steps already taken to remediate these issues, most of which you are familiar with as we worked through them with your team. We also outline next steps to develop more long-term solutions and put in place a model designed to prevent and/or more quickly detect issues of this kind in the future.

Before doing so, however, we think it necessary to respond to the suggestion in your letter that we are treating our reporting obligations as a "compliance exercise." Nothing could be further from the truth. The suggestion flies in the face of Meta's industry-leading record when it comes to protecting children from exploitation online. We do more to meet our reporting obligations than any other member of industry, something NCMEC's leadership has noted on numerous occasions both publicly and privately.

We begin by setting forth a reminder of our deep investment and engagement in service of our mission to protect children, as well as our history of collaborating with and supporting NCMEC in furtherance of our shared goals.

Meta's Commitment To Protecting Children

Our body of work spans more than a decade and is responsible for many child rescues, identification of apparent offenders, takedown of CSAM at scale, and systemic advancement for NCMEC, law enforcement, and industry in support of our shared goal. Over time this investment has only grown, and is highlighted by NCMEC's most recent report showing that

Defendant's Exhibit
DX2916
Case Number:
D-101-CV-2023-02838

Meta's reporting in 2021 comprised 26 million of the total 29 million reports received by NCMEC.

The reports at the heart of your expressed concerns are an outlier within our record of extraordinary efforts, which reflect our company's and employees' life's work, commitment, and passion. More than 40,000 people work on security and safety at Meta, including specially trained teams with backgrounds in law enforcement, online safety, analytics, and forensic investigation, who review and report findings to the National Center for Missing and Exploited Children. In addition to this human power, we use an array of industry-leading tools and technologies to fight child sexual exploitation online across all three prongs of our approach to child protection -- prevention, controls and response.

Meta's Industry-Leading Reporting Efforts

- **Proactive Detection of CSAM.** While there is no obligation under U.S. law to search for CSAM, Meta has invested deeply to widely scan for this material across surface areas to identify, remove, and report it. This commitment alone far surpasses the efforts of all other companies – many of whom use detection technology across just a single service, or not at all.
    - **Photo DNA.** We have been using Microsoft's PhotoDNA technology since 2011 to thwart the sharing of known child sexual imagery on our platform, and remove and report it. We were the first large company to broadly adopt this technology and we evangelized its use, mentoring other companies on how to implement it on their services. Our adoption and championing of this technology revolutionized the global fight against CSAM.
    - **AI.** In addition to using PhotoDNA, we have built and deployed numerous other technologies to help us identify this content, images as well as video. Technology that not only identifies known content but classifiers that help identify previously unknown and/or new content.
    - **Google Content Safety API.** In 2020, to further advance this effort and put in place a best in class system, we partnered with Google to deploy their machine learning technology on our platforms to prioritize and detect previously unknown CSAM, building on our own tech to detect previously unknown CSAM.
- **Proactive Detection of Inappropriate Interactions with Children.** Our work has not stopped with image and video detection. We've invested in tools to detect and report inappropriate interactions with children and worked with stakeholders, including NCMEC, to advance needed legislative advancement to allow us to report these incidents. As a result of this work, we have helped thwart numerous cases of imminent or ongoing abuse through cybertips reflecting in-depth analysis and investigation.
- **Strategic Investigations and Network Disruptions.** We have invested in teams of highly specialized investigators who submit detailed Priority 1 and 2 manual reports, which have resulted in many child rescues and identification and prosecution of those responsible. Indeed, this month's feedback from NCMEC included approximately 150 arrests originating

from Meta CyberTips. We've also submitted numerous reports detailing takedowns of networks of bad actors.

- **Robust, Detailed Reporting.** Even though the law does not specify precisely which fields must be included in reports, we have continuously iterated on our reports to make them more robust and actionable. We make use of the xml schema, adding tags such as 'viral,' 'infant' and 'potential meme' to help investigators prioritize the reports. Where we received feedback that certain data points are problematic, we've taken efforts to adjust reports accordingly.
- **Dedicated Support.** We've provided deep engagement and regular contact and access to subject matter specialists, and outreach liaisons. Some who work with your teams on a weekly, if not daily, basis. These channels have been extremely responsive, working to remediate issues as we are made aware of them and also alerting NCMEC when we discover an issue on our own.

Meta's Long History of Supporting NCMEC on Our Shared Goals

- **Case Management Tool.** Since 2019, we have helped fund and rebuild NCMEC's Case Management Tool to ensure international law enforcement can receive, triage, manage, and organize CyberTipline reports as part of the global response to child sexual exploitation.
- **International Relations.** We've also devoted our experienced internal teams to create pathways for collaboration between NCMEC and international law enforcement agencies who specialize in child exploitation. For example, following years of concern regarding India's absence from the VPN network, our engagements and training helped facilitate their onboarding to the CyberTipline pipeline in 2019 – and they have been an active participant ever since. We've similarly helped onboard other jurisdictions like Pakistan and Nigeria.
- **AMBER Alerts.** We launched AMBER Alerts on Facebook in 2015 alerting people in the designated search area of a missing child – providing a description, the location of the abduction, and any other pertinent, available information to help families and authorities successfully recover missing children. We have since expanded the program to 25 countries.

Protecting the Full Ecosystem

- **Understanding Offenders.** Understanding how and why people share child exploitative content is critically important for all those who are working to deploy effective comprehensive solutions to combat it. To help with that understanding, we conducted an in-depth analysis of the illegal child exploitative content we reported to NCMEC from Facebook and Instagram during two representative months, shared what we learned publicly, and used the information to deploy new tools, as well as new programs targeted at reducing the sharing of this abhorrent content across the ecosystem not just on our platform.

    - **Report it, Don't Share It.** Based on the learnings of the research outlined above, we launched a PSA campaign 'Report it, Don't Share it' in 19 countries reminding people that any sharing of this content, including to express outrage or get help, revictimizes a minor, and they should report it, not share it.
- **Sharing Our Technology**. Since August 2019, we have shared our own photo and video matching technology for free by open-sourcing two algorithms, PDQ and TMK+PDQF. These tools enable smaller developers and non-profits to more easily identify abusive content and share hashes — or digital fingerprints — of different types of harmful content. For those who already use their own or other content matching technology, these technologies are another layer of defense and allow hash-sharing systems to talk to each other, making the systems that much more powerful.
- **URL Lists.** We also work with our expert partners, such as the Internet Watch Foundation, to collect lists of external sites known for hosting child sexual exploitation material and block access to those sites from our platform.
- **Search Interventions**. We also use resources like Thorn's Keyword Hub, a list of known CSAM seeking terms, to prevent type-ahead for searches containing known child exploitation terms, and we display a pop-up warning when people attempt searches with these terms.
- **Amplifying Deterrence Organizations**. We point people toward deterrence organizations, and our partners, such as Stop it Now, have seen Meta's Help Center and other interventions drive an increasing amount of web traffic to their services aimed at preventing offending.
- **Child Safety Hackathons**. Since 2016, we've hosted annual hackathons with industry and NGOs to bring together engineers, data scientists, and designers from across the industry, as well as nonprofit partners, to code and prototype projects focused on making the internet a safer place for children and combatting child sexual exploitation.
- **Report Remove**. In 2020, we launched a pilot with the UK Government, IWF, and the NSPCC for minors to report sexually explicit, self-generated images through a partnership with the National Center for Missing and Exploited Children. The reporting channel, Report Remove, was launched formally in 2021.
- **Model National Response**. As part of the multi-stakeholder WePROTECT Global Alliance to end child sexual exploitation, we have developed a model national response to child exploitation that includes passing legislation against the sexual exploitation of children.
- **Tech Coalition Leadership**. For a decade we've been a driving force behind the work of industry in the Tech Coalition, including leading in the creation of Project Protect, a 15-year plan covering a range of initiatives targeted at better detection and action against child sexual exploitation online. Our investments in this organization have generated critical resources, mentorship and support to its now 26 members companies.
- **Aselo**. Since 2020, we have been supporting TechMatters' project to develop new technology to support child helplines around the world and make them more accessible to children in crisis – building a customizable, open-source contact-center platform that allows children and youth to reach out to helplines via voice, SMS, web chat, WhatsApp, and Facebook Messenger if they come across child-abuse material.

Specific Concerns & Meta's Approach To Resolving

We hope that this brief and not entirely complete delineation of our efforts serves as a reminder of our long history of productive collaboration and of our commitment to child protection and to supporting NCMEC's mission and helps us approach the work ahead with the same collaborative mindset. With that we turn to outlining the steps already taken to address the recent technical challenges and to next steps to develop more long-term solutions and put in place a collaborative working model designed to prevent and/or more quickly detect issues of this kind in the future.

NCMEC Concerns

1. **Compromised Accounts.** Users' accounts are sometimes compromised by bad actors who send CSAM. While the law doesn't specifically require it, Meta's reports generally include, among other things, registration information for the registered account holder. If law enforcement doesn't diligently investigate the tip, they may proceed against the compromised account holder rather than the bad actor who shared CSAM.
2. **Missing Upload IP.** While the law doesn't specifically require Meta to include upload IP information, it has been our regular practice to include it when we have it. Between December and February, a significant number of Meta's reports did not include Upload IP information.
3. **Non-violating Reports.** An incident with deployment of a new protocol of content matching technology had the unintentional effect of sending a significant number of non-violating content.

Meta's Response

Our paramount interest in the mission of protecting our platforms and children is to ensure that our technologies are deployed in the manner intended, on violating content and with reports containing accurate information about the associated accounts and persons responsible. We have put together an internal task force to investigate the specific issues documented in the letter and we look forward to an on-going dialogue in the spirit of remaining the best tech partner in the industry. The task force also will be reviewing all our work in this area to explore whether there are additional processes, systems or safeguards we believe should be put in place (e.g., alerting systems or brief holdback periods to help prevent erroneous reports).

Compromised Accounts

With respect to compromised accounts, our legal obligation remains in that we must report the incident to NCMEC, and indeed, a violation involving CSAM has apparently occurred – though the person likely responsible is the hacker as opposed to the registered user. This is not merely

a 'compliance exercise.' A serious violation against the person depicted in the image and the innocent account holder has occurred, and it is our important moral and legal duty to report it.

As we explained in our call, we have now globally rolled out default enhancements in our CyberTips where we include the signals that should help law enforcement be on alert that there may be a person other than the registered account holder in control of the account. Those fields flag when there's been a new phone number, email address, password change or addition of a new device to the account along with the 10 most recent IP addresses – which can be a tell-tale sign that someone took over the account and locked out the original user. We are looking at additional fields and signals that could contribute to a better understanding of possible compromise. We have collected feedback from members of your team on this work and will set up time with you to go over the plan and secure alignment and support from NCMEC.

We take this opportunity to underscore our understanding and expectation that law enforcement has a critical investigatory role to play in receiving cybertips from Meta and industry. Our cybertips include the standard language that all information in our reports must be independently investigated. While our efforts are focused on providing the best information that is reasonably available at the time we make the report, these reports are meant by Congress to be a starting point for law enforcement's investigations, not conclusive findings of fact or culpability. Indeed, it is settled law that cybertips are not admissible on their own as evidence in court (*see*, e.g. *U.S. v. Cameron*). It is particularly important that law enforcement appreciate the need to regularly rule out account compromise in their investigations because it is common knowledge that defendants routinely claim that their account was compromised as their defense. We will continue to engage with law enforcement to help train them on the type of signals to look for to determine likely account compromise – though this is an adversarial space and bad actors' methods will continue to evolve. As we have partnered with NCMEC for years on training law enforcement agents around the world, we offer to do the same here with respect to compromised accounts.

Missing Upload IP

For many years we have included Upload IPs in our reports. Due to an unrelated reconfiguration of an adjacent internal system, a technical issue prevented the Upload IP from feeding into the reports. Though our teams speak to NCMEC on nearly a daily basis, the unusual broad lack of upload IPs was not mentioned to us for nearly 3 months. As soon as we were informed we investigated the issue, identified the cause, and started work to remediate our systems, which is an ongoing process.

We continue to feel there is room to improve our collaboration here. Notably, we have previously raised with NCMEC the importance for industry having access to the output of the CyberTipline to be able to audit whether our reports are coming through in the manner we expect – such a feature would have helped us identify the issue much earlier. This was raised on multiple occasions as a feature request including but not limited to when Thorn first partnered with NCMEC in the scope of their partnership to revamp the CyberTipline, and during the May

2019 NCMEC Industry Roundtable event. While we are working to improve the auditing available on our end, in order to ensure a well-functioning ecosystem and prevent the proliferation of technical problems, we again recommend the building of an audit feature for industry into what our reports look like when they are received through the reporting API. We remain happy to partner with NCMEC to help implement such a feature.

Non-violating Reports

As we explained in the call, the batch of non-violating reports came through for a few days (mostly over a holiday weekend) when a new technology meant to improve the quality of proactive detection was improperly configured. Benign content that was part of a violating media, matched against identical benign content on non-violating media. Again, once we were made aware of the issue, we immediately resolved it and ceased matching against that content. This was a process error and we are working on fortifying our systems to include more safeguards to prevent these outcomes.

While the CyberTipline's original design never considered the possible scale at which it would be used today, it is critical that the systems be modernized to account for the way today's technology actually works at scale. We've previously raised with NCMEC the need to have a means to bulk-retract content that was erroneously reported. These types of tools are even more important as (1) more companies are scaling their work with automated systems which greatly magnify unintended mistakes, (2) proposed regulations would require all providers to apply these automated systems, thus increasing the risk for large-scale disruptions, and (3) new laws like the EU CSAM Derogation require companies to remediate CSAM detection mistakes at the earliest. Again, we stand ready to work with NCMEC on creating such a bulk-retract tool.

<u>Continued Collaboration & Innovation</u>

Finally, as mentioned during our call, we think it would be useful for our organizations to evaluate our current working model and processes to see if there are non-tech related improvements we can make to support our shared goals.

We also would like to kick off an effort with NCMEC on identifying systemic non-tech related challenges that could help stand up a more effective and efficient system focused in the first instance on prevention, where we can better measure impact and progress and ensure industry, NCMEC, law enforcement and other key stakeholders are best positioned to help safeguard kids. For example, working together we can advance more certainty and reduce false positive reports by supporting the development of specific legislative guidelines on how to determine "apparent CSAM." We also can identify and seek specific legislative allowances to not report certain unactionable reports that make it harder for investigators to hone in on the actionable reports. These examples are just the beginning.

We've worked together in the trenches for more than ten years to fight against these heinous acts. This work is hard, complex, but with rewarding safety outcomes. We appreciate the feedback that you've provided in the spirit of making all of us better equipped for this fight. We look forward to continuing to work together on our shared goals.

Antigone Davis
VP, Global Head of Safety

Guy Rosen
VP, Integrity