Exhibit M

Highly Confidential

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
OAKLAND DIVISION

IN RE: SOCIAL MEDIA ADOLESCENT ADDICTION/PERSONAL INJURY PRODUCTS LIABILITY LITIGATION THIS DOCUMENT RELATES TO: *People of the State of California, et al. v. Meta Platforms, Inc., et al.*	MDL No. 3047 Case Nos. 4:22-md-03047-YGR-PHK 4:23-cv-05448-YGR Judge Yvonne Gonzalez Rogers

EXPERT REBUTTAL REPORT OF MARY CATHERINE WIRTH

April 15th, 2026

companies, using in-house proprietary technologies and with little in the way of industry-shared vocabulary, benchmarking, agreed-upon standards, guidelines or external visibility.

38. Although you can read another company's public-facing terms of service and Community Guidelines, Trust & Safety program technical and operational functions deliberately aren't visible or obvious to end users, otherwise bad actors could easily evade a platform's protections. Because of this deliberate opacity, individual companies for many years had no choice but to "reinvent the wheel" – on content policy training, moderation and terms enforcement workflows, trust algorithm and classifier development, escalation processes and public reporting.

39. For example, when I began developing the Adobe Trust & Safety program around 2009, I had no industry organization to turn to, no industry "best practices" documents to cite – nothing to guide my advice to executives and product teams other than my own personal experience developing a similar program at my previous company. My ability to benchmark other companies' practices depended entirely on my own personal network of friends and former work colleagues who, like me, had left Yahoo! to go work at other large tech companies. Without those collaborative relationships, I would have been entirely in the dark.

40. In 2021, the Digital Trust & Safety Partnership (DTSP) was formed to redress this industry-wide problem.[22] The DTSP is an industry-led, voluntary collaboration among large technology companies focused on sharing and standardizing a broad variety of trust and safety practices — particularly in areas like child safety, age assurance, content moderation, terms enforcement and transparency reporting. DTSP does not impose new requirements or define minimum acceptable conduct or legal obligations. Rather, it shares examples of how mature Trust & Safety organizations approach common industry challenges at scale so that individual companies no longer need to "reinvent the wheel."

41. In September 2023, the DTSP published "Age Assurance Guiding Principles and Best Practices" (attached as Exhibit A).[23]

---

[22] Leading Technology Companies Launch Initiative to Promote a Safer and More Trustworthy Internet - Digital Trust & Safety Partnership (February 18, 2021 press release announcing the launch of the Digital Trust & Safety Partnership, "a first-of-its-kind initiative aimed at promoting a safer and more trustworthy internet.").

[23] Meta is a partner in DTSP, along with other tech companies like Google, Apple, TikTok, Microsoft, Snap, LinkedIn and Reddit. *See, About Us*, Digital Trust & Safety Partnership ("DTSP") (listing DTSP's "partners"), https://dtspartnership.org/#aboutus.

42. As the document's executive summary acknowledges, "every approach to age assurance presents trade-offs."[24] For example, more accurate methods may depend on the collection of additional or sensitive personal data and thus can be in tension with a service's privacy commitments to users, with a user's desire to speak anonymously, or with data minimization principles. Certain age assurance methods may create inequities among users or unfairly discriminate (for example, requiring a credit card as proof of adult status can unfairly exclude users without credit cards). And certain age assurance methods may not be economically feasible for smaller companies or app developers. Hence, there is no one-size-fits all solution.[25]

1. Overview of DTSP Guiding Principles

43. DTSP's age assurance best practices are organized around five guiding principles:

- Safety By Design Proportionate To Youth Risk: Identify, evaluate, and adjust for risks to youth to inform proportionate age assurance methods, as part of implementing safety-by-design.
- Data Minimization/Privacy Protective-Focus: Account for risks to user privacy and data protection as part of development, implementation, and ongoing assessment of age assurance approaches.
- Inclusivity & Equity: Ensure assurance approaches are broadly inclusive and accessible to all users, regardless of age, socioeconomic status, race, or other characteristics.
- Layered Investigation/Enforcement: Conduct layered enforcement operations to implement age assurance approaches.
- Transparency: Ensure that relevant age assurance policies and practices are transparent to the public, and report periodically to the public and other stakeholders regarding actions taken.

[24] *Age Assurance: Guiding Principles and Best Practices*, DTSP, 2 (September 2023), https://dtspartnership.org/wp-content/uploads/2023/09/DTSP_Age-Assurance-Best-Practices.pdf.

[25] *Id.*

44. DTSP emphasizes repeatedly that there is no one-size-fits-all "best practice" solution and that different platforms may reasonably make different design choices while still aligning with industry best practices.[22]

2. DTSP "Layered Enforcement"

45. DTSP defines "layered enforcement" as using multiple, independent age indicators, each of which either corroborates or escalates concerns about user age. No single method is expected to work alone; strength comes from combining some (or all) of them.

46. DTSP examples of layered enforcement practices include:

- Ensuring a neutral self-declared age gate at registration.
- Deploying an age assurance check if a user changes their self-declared age from one that was <18 to an age >18.
- Analyzing user behavior or signals for indicia that the user is under the allowed platform age (for example, on a social media platform that bars users under 13, a "Happy 11th Birthday!" message posted to a user's timeline).
- Training enforcement teams who review age verification escalations using reliable methods for verifying a misreported age (for example, by evaluating appearance or other signals indicating a younger age) and in appropriate methods for age assurance investigation follow up.
- Allow users to report other users who may have misreported their age.
- Implement technical measures that can help prevent users who have failed an age assurance test from circumventing controls (*e.g.*, by dropping a cookie that prevents a user from starting the process over again via a new registration).
- Potentially offering family accounts or parental verification where a young person's account may be attached to a parental account.[23]

IV. META'S AGE ASSURANCE PRACTICES MEET OR EXCEED INDUSTRY BEST PRACTICES

47. As explained above, experienced privacy and Trust & Safety practitioners look to two primary sources of guidance when developing, operating or evaluating the efficacy of an age assurance program. An age assurance program *must* do what is expressly required by the FTC

52. Both platforms likewise have well-defined processes followed by trained staff for investigating any circumstance where Meta obtains actual knowledge that a specific user is under the age of 13.[30]

B. Meta's Age Gating Program Fully Comports With DTSP Age Assurance Best Practices

53. Meta also takes an appropriately layered approach to age assurance for Facebook and Instagram, consistent with the DTSP age assurance best practices.

54. Meta's layered practices include:

- Deploying technical methods to help prevent users who have failed an age assurance test and have been deemed ineligible from circumventing the controls.[31]
- Employing machine-learning classifiers trained on reliable signals for predicting whether an account belongs to someone over or under the age of 18.[32]
- Directing accounts flagged as potentially underage through an automated process that determines whether the account should be escalated for human review or immediately actioned.[33]

---

age selector so that it was set at the date of registration when encountered. *See* Allison Hartnett Depo., Exh. 6 p. 25 *et. seq.*

[30] *See., e.g.,* META304 7MDL-012-00000003 (Meta trains thousands of human reviewers to review accounts that are reported by other users for potentially violating the Meta terms of service and content guidelines. In the course of reviewing these accounts in response to any type of report, a reviewer may notice certain details suggesting that the account holder may be a child under 13 (e.g., photos posted to the account reflecting age or text in the account bio stating an age under 13), in which case the reviewer will escalate the account for underage review. Meta may also identify underage users through other review processes including, for example, where Meta requires users to provide identification to access certain features (e.g., Meta Verified)). *See also* META3047MDL-146-00072824 (overview of age assurance workflows) and Allison Hartnett Deposition at p. 182 *et seq.*

[31] *See., e.g.*, META3047MDL-019-00129336 (describing 12 hour cooling off period to prevent additional registration attempts when user enters date of birth under 13) and META3047MDL-053-00012657 at 12662 (If a user tries to put in a U13 birthday, "After two consecutive attempts, the user will be prevented from creating an account and locked out of the registration process for a period of 12 hours.")

[32] *See* META3047MDL-012-00000005 *et seq.*; META3047MDL-146-00072880 (high-level overview of the age labels used for training and evaluating the age prediction models); Allison Hartnett Deposition at p. 175 *et. seq.*

[33] META3047MDL-012-00000003 *et. seq.* (Meta may learn that a user is underage directly from the user, if the user attempts to change the date of birth on their account to a date that would make them under 13. As of February 2021, the user will be automatically placed in an age checkpoint and must provide sufficient proof that they are old enough (i.e., 13 or over in the United States) to use Instagram before they will be permitted to continue to use the service.)

- Where classifiers are not confident in their prediction, flagging such accounts and escalating them for review by human moderators trained in investigating account age.[34]

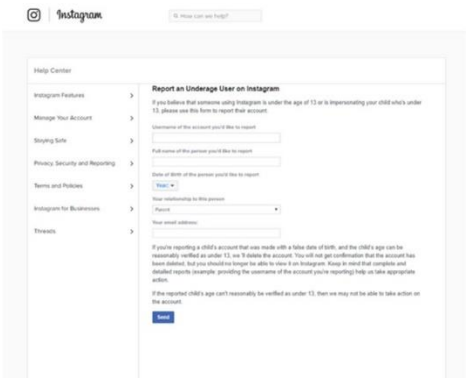

Figure 2 Instagram "Report an Underage User" Screen META3047MDL-040-00000322.

- Suspending accounts that have been checkpointed so that they are no longer accessible to the user or visible to other users during the evaluation process.[35]
- Flagging the data of any account that has been checkpointed as "not for business use."[36]
- Closing accounts that fail an age checkpoint and, once the period for user appeal has expired, erasing their data. Moreover, any accounts hard-linked to the failed account in Meta's Account Center (for example, a Facebook account a user has linked to an Instagram account) are likewise closed and deleted.[37]

---

[34] *See* META304 7MDL-012-00000001 at -0006-0008 (All potentially underage accounts that are manually reviewed are reviewed using Meta's content review platform, which is specifically designed for large-scale manual review efforts. The platform standardizes the underage review process and allows Meta to track and properly document decisions made. To ensure that the human review process is consistent and that reviewers evaluate each profile using the proper criteria, Meta's content review platform provides reviewers with a 'decision tree' or standardized flow that they must follow for each flagged account. Reviewers are also able to include notes and provide additional context for why they made a particular decision.)

[35] *See* META3047MDL-012-00000007.

[36] Lars Backstrom Deposition at p. 32.

[37] *See* META3047MDL-012-00000010. A "hard-linked" account is one where the user has affirmatively connected accounts together – *e.g.,* by logging in with their user name and password and choosing to link two accounts in the settings menu. Meta does *not* delete "soft matched" accounts (*i.e.* ones where data like device ID or IP address suggest they might belong to the same person) because the likelihood of erroneous account deactivation is simply too high. For example, deleting soft-matches would risk deleting the account of an older sibling or parent who logs into their account via a shared family desktop computer or the account of a co-worker who logs into their account via a common office wi-fi network IP address.

- Deploying an age assurance check via a third-party vendor if a user changes their self-declared age from one that was under 18 to an age over 18.[38]
- Automatically checkpointing accounts if users change their self-declared age from one that was 13 or older to an age under 13.[39]
- Allowing users (and non-users) to report directly to Meta that a particular user is under the age of 13.[40]
- Implementing "Teen Accounts" on both Instagram and Facebook, which default to protective settings and offer parents access to controls like approving changes to teen settings or permissions, monitoring usage, setting time limits and more.[41]

55. In my experience, this is everything I would expect a mature, responsible U.S.-based Internet or social media company's age assurance program to do — and more. Based upon my review and my professional experience, I am confident that Meta's program is at least comparable to (and in many instances likely exceeds) age assurance programs operated by DTSP members of similar size.

V. META IS NOT REQUIRED TO AGE VERIFY OR OBTAIN VERIFIABLE PARENTAL CONSENT

56. When evaluating Meta's practices, Mr. Estes ignores longstanding FTC guidance and industry-accepted principles that have guided U.S. age-gating development for more than two

[38] *See* META3047MDL-012-00000015 (When a user attempts to change their date of birth from an age under 18 years old to an age 18 years or older, Meta provides the user multiple options for verifying age, including the option to upload a self-portrait video. If a user elects this option, Meta will collect a video selfie from the user and share a selfie image with third-party service provider, Yoti, a leading age verification provider for several industries around the world, including social media, gaming, and age restricted ecommerce. Yoti's technology will estimate a user's age based on their facial features, and will share back to Meta the user's estimated age. Meta and Yoti delete the uploaded image after Yoti sends Meta the user's estimated age. If Yoti estimates that the user is under the age of 13, the user will be placed in an age checkpoint.).

[39] *See* META3047MDL-053-00012657 at 12664-65 (When a user changes their date of birth from an age 13 or older to an age under 13, their account is automatically placed in a checkpoint on both Instagram and Facebook, and the user is required to prove their age within the specified period to regain access to their account).

[40] *See, e.g.,* META3047MDL-012-00000002, *et. seq.* (Since at least January 1, 2019, anyone, including individuals who do not have an Instagram account, can report to Meta that a user is or appears to be under the age of 13 by filling out an online reporting form.); META3047MDL-053-00012709 at 12716 ("Anyone, including individuals who do not have a Facebook or Instagram account, can report to Meta that a user is or appears to be under the age of 13 by filling out a platform-specific online reporting form.").

[41] *See, e.g.,* Meta Family Center (https://familycenter.meta.com/).

motivation to speak anonymously: *i.e.,* whistleblowers, domestic violence or rape victims, LGBTQ youth, political dissidents, religious minorities and the like. It is industry best practice to develop Trust & Safety protocols "based on user feedback or other observed effects, including ensuring that the perspectives of minority and underrepresented communities are represented" and to "work with recognized third-party civil society groups and experts for input on policies."[54]

64. Given how highly the United States prioritizes free speech rights, it is no surprise that we, as a nation, have deliberately chosen *not* to mandate a "papers please" approach to online speech (or that U.S. social media services have declined to adopt this deeply disfavored approach voluntarily).

B. Privacy, Security, Competition and Public Opinion Considerations Also Prevent Voluntary Adoption of Universal Age Verification

65. There also are significant privacy, security, competitive and other practical considerations that prevent U.S. general-audience or social media platforms from adopting universal age verification.

66. The inconvenience, as well as the privacy and cybersecurity risks inherent in universal age verification, would inevitably deter many users from participating in online communities.[55] Age verification also inevitably produces false positives and erroneous exclusions. For example, adults without government-issued identification, users whose appearance does not match their identification documents, transgender users, users concerned about privacy, and users incorrectly flagged by facial age-estimation systems may have their right to speak (or access speech) wrongfully delayed or denied altogether.

67. More importantly, not a single U.S. social media company currently requires universal age verification — so how could that feature's absence from Facebook and Instagram render

---

media sites as an unconstitutional restriction on speech) https://www.aclu-wi.org/legislation/sb-758-social-media-age-verification/.

[54] *See* DTSP Best Practices Framework, Commitments 1 and 2, https://dtspartnership.org/best-practices/.

[55] *See, e.g., Internet Law Professors' Amicus Brief in Free Speech Coalition v. Paxton* (October 2024) at p. 2 ("Studies show that readers view age verification screens as inconvenient and not worth the hassle to access the desired content") and p. 4 (Age verification "introduces significant privacy and security risks to readers that will ultimately dissuade many adults from accessing constitutionally protected material.")

him on X, sharing your favorite recipes via Instagram, having a dialogue about the latest hair-loss treatments or wig vendors in a Facebook alopecia group, giving your two cents on the outcome of last night's sports game in a Reddit forum dedicated to the team. Not only is social media free, it is all about expression that, in the U.S., is entitled to the highest levels of protection. Erecting barriers to participation (by requiring every user to disclose a credit card, government identification or other sensitive personal information) would fundamentally alter the nature of those services and chill the very speech they are designed to facilitate. That is not the case for commercial gaming platforms.

VI. LACK OF TRUST & SAFETY OPERATIONS EXPERIENCE UNDERMINES THE ACCURACY OF BOTH EXPERTS' CONCLUSIONS

74. Dr. Sheatsley is a post-doctoral researcher in computer science and engineering at the University of Wisconsin, Madison. He has no Trust & Safety operations or privacy compliance experience. He has never built or overseen an age assurance program. He has never worked at a social media or technology company in any capacity. Indeed, he has never worked outside of academia at all.[65] Nonetheless, Dr. Sheatsley was hired by the plaintiffs to "evaluate and analyze the structured data Meta has produced in this litigation regarding its efforts to detect and remove users under the age of 13."[66]

75. Although Mr. Estes has practical, real-world experience at a technology company, he is a software developer. Like Dr. Sheatsley, Mr. Estes has no Trust & Safety operations or privacy compliance experience.

76. This lack of practical Trust & Safety experience leads both to repeatedly draw inaccurate conclusions from the data they have reviewed. What they each fail to recognize is that Trust & Safety systems cannot be designed solely to maximize the number of accounts disabled or the number of reports received. They must be designed to balance multiple competing considerations simultaneously, such as accuracy, reliability, the risk of false positives, the risk that reporting tools will be abused, adverse impact to rule-abiding users and the serious consequences of erroneously deleting an account and its contents.

[65] Sheatsley Report at ¶ 10, and at Appendix A.

[66] *Id* at ¶ 17.

A. Every Report or Match Does Not Represent A True Under-13 User

77. One consistent error both men make is assuming "more" reporting or "easier" reporting is always "better reporting." Or that every additional report, checkpoint, hard-link or soft-match represents a true under-13 account. This ignores the common factors Trust & Safety professionals have to take into consideration daily – *i.e.* the realities of false positives, abusive reporting, duplicate reports or the operational and reputational costs of wrongfully disabling rule-abiding users.

78. For example, both experts fault Meta's underage reporting workflow. Dr. Sheatsley criticizes Meta because it declined to roll out a Shortened Underage Reporting Flow (SURF) that would have enabled one-click reporting of underage users.[67] Mr. Estes just vaguely states (without support) that filing an underage report is "impossible."[68]

79. But, as reflected in discovery materials, Meta's existing under-13 reporting feature is "one of the most abused reporting flows at Meta."[69] It has been used millions of times to target and harass adult users, including political entities, countries, public officials, and celebrities (e.g., mass false reports directed at the Israeli Defense Forces, Kim Kardashian, Emmanuel Macron, and others).[70] With that context, any experienced Trust & Safety practitioner would intuitively predict that a frictionless, one-click reporting mechanism would be more likely to increase abusive reporting volume than meaningfully improve the detection of underage users.

80. Meta's SURF feature testing confirmed this. When the company rolled out the SURF feature to 1% of its user base, it resulted in a very large surge of U13 reports that were of exceedingly low quality – *i.e.,* the actionable rate of the reports received was fully 50% lower than those received via the existing longer form.[71] In Trust & Safety operations terms, a reporting feature that vastly increases the number of reports received while degrading the quality

[67] Sheatsley Report at ¶ 55 *et. seq.*

[68] Estes Report at ¶ 61.

[69] *See, e.g.,* Allison Hartnett Depo at Exh 6, p. 6 (describing it as "one of the most abused reporting flows at Meta.")

[70] *See* META3047MDL-208-00003249.

[71] *See* META3047MDL-111-00080128, *et. seq.*

84. Even though hard links are a far more reliable enforcement signal than soft matches, they still are imperfect –for example, when a parent links one of their own accounts to their child's account. The fact that Meta disables hard-linked accounts despite this imperfection, with no right of appeal, strikes what most Trust & Safety professionals would consider a responsible balance – *i.e.,* tolerance of some level of enforcement overbreadth, but not exceedingly high levels of overbreadth.

85. By contrast, soft-matching is probabilistic (and therefore, a far less reliable enforcement signal). A "soft match" is an inference that two accounts potentially *could* belong to the same user based on indirect signals like device ID, IP address, phone number, email address, or uploaded contacts. While Meta can infer that the accounts *may* belong to the same person, there is a far higher likelihood of error – for example, where everyone in a single household logs into their separate social media accounts using the same home desktop computer, where siblings share a single iPad, or where all students in a classroom log in using the same wi-fi network and shared IP address. Indeed, in Trust & Safety circles it is well-known that soft-matching enforcement has a disproportionately adverse impact on poorer families who are more likely to share devices rather than have their own.

86. Because Dr. Sheatsley mistakenly equates soft-match data with hard-link data, he concludes Meta should have treated every account soft-matched to an underage account as if it were definitively operated by the same user and deactivated them / deleted their data. Although he estimates that tens of thousands of additional accounts could have been disabled if Meta had propagated underage enforcement through all soft-matched accounts, he ignores what is obvious to any Trust & Safety professional – *i.e.,* responsible moderation ecosystems aren't designed just to maximize the number of accounts disabled without regard for whether such drastic action is warranted.

87. In the real world, signals used in Trust & Safety enforcement need to be carefully calibrated to the task at hand. It is perfectly reasonable to use a probabilistic signal like a soft-match when investigating a fraud claim or targeting advertising because the consequence of a false positive in these contexts is trivial – *i.e.,* you serve someone an irrelevant ad, or you recognize that other signals used in your fraud investigation contradict the soft-match signal and disregard it. By comparison, it would not be at all reasonable to use a soft-match as the sole

basis for automatically and permanently disabling user accounts and deleting their contents: the likelihood of error is far too high and the adverse impact on your user base is far too great.

C. Checkpointed Users' Data Does Not Remain Fully Accessible To Meta's Systems For Training

88. In his report, Dr. Sheatsley states that a "substantial number of underage user accounts remained in underage checkpoints for extended periods before being disabled, during which time their data continued to be fully accessible to Meta's systems—including for training and improving AI models."[75] Based upon evidence in the record, this is inaccurate.

89. This issue was discussed at length in the deposition of Meta Vice President of Engineering Lars Backstrom. Mr. Backstrom testified that, once an account is checkpointed, it is flagged by Meta as "not for business use" – which renders its data inaccessible to Meta's systems for training and other business purposes.[76] This is an industry standard approach to data management.

90. Moreover, Dr. Sheatsley ignores evidence in discovery that Meta's data deletion processes (among other things) are dictated by an existing federal consent decree and routinely audited by the FTC.[77] In other words, Meta's current deletion practices are the result of a joint agreement with a federal agency that has been approved by a federal court -- not something it (or its operations teams) can modify unilaterally.

[75] Sheatsley Report at ¶ 156.

[76] Lars Backstrom Deposition at p. 32. ("The process would begin when the account is check-pointed, or that is when the data would initially be flagged as not for business use.")

[77] *See* Allison Hartnett Deposition, Exh. 6, p. 15 (deletion processes routinely audited by FTC assessors in accordance with existing consent decree). *See also*, FTC Decision and Order Docket No. C-4365 at Section III (Deletion of Information) https://www.ftc.gov/system/files/documents/cases/c4365facebookmodifyingorder.pdf