

Filed Under Seal

Exhibit A

Highly Confidential

**IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA**

PEOPLE OF THE STATE OF CALIFORNIA,
et al.,
Plaintiffs,

v.

META PLATFORMS, INC, Instagram, LLC,
Meta Payments, Inc., Meta Platforms
Technologies, LLC., et al.,
Defendants

IN RE: SOCIAL MEDIA ADOLESCENT
ADDICTION/PERSONAL INJURY
PRODUCTS LIABILITY LITIGATION

THIS DOCUMENT RELATES TO:
4:23-cv-05448

MDL No. 3047

Case No. 4:23-cv-05448-YGR

Judge: Hon. Yvonne Gonzalez Rogers

Magistrate Judge: Hon. Peter H. Kang

TRIAL REPORT OF ADAM L. ALTER, PH.D.
August 1, 2025

Highly Confidential

I. EXECUTIVE SUMMARY OF OPINIONS

1. This report is divided into two Parts: the first Part, labeled **1. DECEPTION**, examines how reasonable consumers would have understood various statements made by Meta personnel; the second Part, labeled **2. CHILD-DIRECTEDNESS OF META PLATFORMS**, examines the extent to which Meta platforms Facebook and Instagram, or portions of those platforms, were directed to children.

2. To understand the effect(s) that Meta’s public statements would have had on a reasonable consumer, in **Part 1** I have adopted the social influence analytic framework. This framework considers three relevant dimensions of influence: the context in which the statements were made; the tactics of persuasion used by the communicator and embedded within the statements; and the manner in which consumers process information. Based on this analysis of Meta’s public communications and internal documents, statements of Meta employees, as well as relevant academic literature, I have reached the following conclusions:

- A. Social media platforms are rapidly-evolving and opaque technology products about which consumers have less information than do the products’ manufacturers. Social media platforms are thus “credence goods,” meaning that it is difficult or impossible for consumers to evaluate such products or services even after those consumers have used them, because those consumers lack the knowledge or expertise necessary to conduct their own independent evaluations.
- B. Companies like Meta use public statements strategically to develop, protect, or repair consumer perceptions of their brands and products. When making public statements, Meta employed tactics of persuasion designed to influence consumer perception, including *drumbeat messaging*, *proof points*, *favorable metrics*, *agenda setting*, and *side-stepping criticism*.
- C. In complex decision-making environments, where, like here, the focal issues are highly technical and opaque, even highly motivated consumers are constrained in their ability to perform extensive assessments and to reach comprehensive conclusions about the product. In domains concerning health and safety, as in the present context, consumers reasonably tend to assume that a company’s communications are both truthful and complete. Furthermore, consumers may unconsciously rely on certain processing cues (or, *heuristics*) to make sense of company messaging. Consumers also tend to distill

Highly Confidential

central meanings, or “gists,” from a set of statements or claims. Reliance on such implicit assumptions and processing cues is characteristic of reasonable consumers.

- D. Meta was aware that consumers had major concerns about the effects of its platforms on the health and safety of the users of those platforms. Its knowledge of those concerns informed its communications strategy. Meta knew that uncertainty or concern about platform safety could hamper the growth of its platforms, engagement with those platforms, and, ultimately, Meta’s financial performance. Thus, to protect its brand and long-term viability as a company, Meta disseminated messages that were designed to shape consumer perceptions of the health and safety of its platforms in the company’s favor.
- E. Meta made a series of misleading public statements that advanced four distinct sets of claims that were relevant to individuals under age 13 (“U13”) and individuals of ages 13-17 (“teens”; collectively “young users”):
 - a. Claims that Meta prioritized user well-being, health, and safety, particularly for young users, over other interests, including its own business interests. Specifically, Meta claimed that consumer well-being was its “number one priority” or one of its top priorities, that it prioritized well-being over profit, and that it made significant and meaningful investments in well-being efforts.
 - b. Claims that Meta’s platforms are not addictive. Specifically, Meta claimed that its platforms do not operate to addict, were not designed to be addictive, and touted its time management tools such as Take a Break and other features as proof points to support this sentiment.
 - c. Claims that Meta’s platforms do not harm users. Specifically, Meta claimed that research shows that the overall effect of its platforms on user well-being is neutral or positive rather than negative; issued public transparency metrics that indicate harmful content and experiences on the platforms are rare; that its internal policies keep platforms safe for users; and that its platforms are suitable for young users.
 - d. Claims that Meta effectively prevents U13s from using its platforms. Specifically, Meta claimed that U13s are not allowed to use and do not use the platforms, and that it “especially” invests and engages in various efforts to inhibit U13 use in furtherance of its policy to prohibit U13s from using its platforms.
- F. Meta’s health and safety claims were meaningful and important to a significant number of consumers. Empirical evidence demonstrates that large numbers of parents and teens worry about the safety of Meta’s platforms, and the addictive nature of those platforms.
- G. In concluding that the above-referenced public statements by Meta are false, incomplete, or otherwise have the tendency to mislead, I compared the impression that the statements would give a reasonable consumer with assumptions that counsel directed me to make. I asked counsel to provide documents that would help me better understand the basis of these assumptions. There is a stark contrast between the impression that Meta’s public statements would give a reasonable consumer, on the one hand, and the facts associated

Highly Confidential

with the assumptions and documents I was given by counsel, on the other hand. Specifically:

- a. Directly inconsistent with its external statements, Meta has made numerous internal decisions that deprioritized well-being and safety efforts, often for reasons related to profits and growth. Meta has also repeatedly failed to invest and fund internally recommended initiatives related to well-being and safety (*see* Section X.A.vi). Accordingly, reasonable consumers would have been misled by Meta's claims that user well-being, health, and safety are its "number one priority" because such audiences would expect Meta to prioritize the welfare of its platforms' consumers over the company's growth and profit when making decisions about those platforms. Further, a reasonable consumer would expect Meta to make significant and meaningful investments to launch and promote products designed to improve well-being, whereas Meta did not in fact do so.
- b. Directly inconsistent with its external statements, Meta had internal knowledge that users reported addictive, problematic use of its platforms and internally acknowledged that problematic use consistent with addiction was occurring. In addition, Meta knew its time management tools such as Take a Break and related features were ineffective (*see* Section X.B.vi). As a result, reasonable consumers would have been misled by Meta's public statements disclaiming the addictiveness of its platforms. Such consumers would have understood those statements to mean that Meta was not aware that its products operate to induce problematic and/or addictive use and/or were designed to be addictive. Furthermore, such audiences would have been led to believe that the Take a Break and related features were effective at curbing addictive use, whereas those features were not effective at curbing addictive use.
- c. Directly inconsistent with its external statements, Meta had internal data and knowledge of the frequency of harms occurring and/or being experienced on its platform by users, including young users, and knew its policies and procedures were not fully effective in eradicating or preventing such harmful experiences, yet failed to disclose such information (*see* Section X.C.v). Thus, reasonable consumers would have been misled by Meta's claims regarding the capacity for its platforms to harm users. Such consumers would have been misled because they would have been led to believe that Meta had robust procedures for eradicating harmful content, and preventing such content from reaching its platforms, and that young users would have encountered such harmful content only in negligible quantities.
- d. Directly inconsistent with its external statements, Meta internally knew meaningful numbers of U13s were on its platform, its lax age-gating and age enforcement enabled underage use, it was underinvested in age assurance, and efforts related to age-gating, age verification, age-detection, and underage removal were ineffective and lagged behind industry standards (*see* Section X.D.iv). Thus, reasonable consumers would have been misled by Meta's claims regarding the company's approach to and effectiveness in preventing children

Highly Confidential

datapoint. If Meta had instead adopted a different methodology and measured prevalence by how many sessions included harmful content, or aggregated views across all harms, or measured by how many people were exposed to harm across a week, month, or year, the resulting harms ratio would have been much larger in an absolute sense.

284. A reasonable consumer would interpret such absolutely small percentages (*e.g.*, 0.14-0.15% or 0.05-0.06% prevalence of bullying and harassment for Facebook and Instagram in Statement C8) as signaling that exposure to such harmful content was similarly low in an absolute sense. A reasonable consumer would have neither the expertise nor the information to unpack Meta’s statistical and methodological approach to understand that low levels of prevalence according to the CSER methodology do not speak to the likelihood of exposure, or magnitude of harm, to individual users.

285. Meta’s decision to promote and publicize the CSER metric is in stark contrast to its decision *not* to disclose other internal metrics such as its Bad Experiences and Encounters Framework (“BEEF”), Tracking Reach of Integrity Problems Survey (“TRIPS”), or Negative Experiences Tracking Survey (“NETS”) data (further described below), which focused on individual users rather than page views, and produced statistics that were significantly larger and considerably more meaningful. As discussed below, Meta selectively presented consumers with a sanitized picture of its platforms in the form of CSER reports that omitted the alarming results of its internal research on the frequency of harms experienced by large numbers of users.

iii. Analysis of individual statements: *Policies*

286. Finally, as shown above in Table 10, Meta has made public statements conveying the position that its internal guidelines and policies successfully prevent users from viewing harmful or inappropriate content.

Highly Confidential

that the reach for witnessing bully and harassment as measured by TRIPS was 100 times that of policy-violating impressions (essentially the CSER prevalence metric).⁶⁹³

318. In response to the enormous disparity between Meta’s internal research and the external CSER data, certain Meta employees advocated for a new framework that eventually came to be known as the *Bad Experiences and Encounters Framework* (“BEEF”). Building on prior tracking surveys such as NETS and TRIPS, and adopting a number of the same survey questions, BEEF was designed to create a singular well-being metric that would better inform product design choices and reform Meta’s products.

319. A key proponent of BEEF was former Meta senior engineer and consultant, Arturo Béjar,⁶⁹⁴ who went to Meta’s top executives to explain the necessity for resourcing the BEEF survey. In a 2020 email to Chris Cox, Mr. Béjar wrote, “[t]oday we don’t understand well the relationship between bad experiences or harm from our user’s perspective... and prevalence (policy violating content).”⁶⁹⁵ Mr. Béjar noted that this gap in understanding was important because policy violations may fail to capture legitimate bad experiences like “mass harassment,” “non-credible or non-violent threats,” or “thinspiration.”⁶⁹⁶ To better understand this gap, he and several other staff on the well-being team proposed BEEF.⁶⁹⁷

320. The framework, developed by former Meta researcher Dr. Kyle Andrews with input from Mr. Béjar, surveyed and measured “bad experiences” that users had on Instagram.⁶⁹⁸ It aimed to

⁶⁹³ META3047MDL-019-00094148, “‘Bad Experiences’ Measurement: Plan for a 2021 plan,” November 19, 2020.

⁶⁹⁴ Deposition of Arturo Béjar, April 7, 2025, 36, 41, 43-48, 75.

⁶⁹⁵ Deposition of Arturo Béjar, April 7, 2025, 276-283, Exhibit 20.

⁶⁹⁶ Deposition of Arturo Béjar, April 7, 2025, Exhibit 19.

⁶⁹⁷ Deposition of Arturo Béjar, April 7, 2025, 282-283, Exhibits 19-20.

⁶⁹⁸ Deposition of Kyle Andrews, PhD, November 19, 2024, 179; Deposition of Arturo Béjar, April 7, 2025, 283.

Highly Confidential

take a broader and “more robust”⁶⁹⁹ assessment of user experience by including bad experiences that weren’t necessarily violating any platform rules, such as negative social comparison and unwanted advances.⁷⁰⁰ Like NETS and TRIPS before it, the survey also addressed bad experiences associated with violations of community standards such as bullying and self-harm.⁷⁰¹ And whereas the prevalence metric provided in CSERs captured the proportion of total views that were of violating content, the final BEEF survey results provided the percent of respondents who had experienced a bad experience in the previous seven days.⁷⁰²

321. Internally, Meta regarded the BEEF survey as one of its most significant sources of information for understanding user experience and exposure to harmful content. Lending to its importance was the fact that BEEF examined a “statistically significant” group of 237,923 users of Instagram, and thus its findings could be extrapolated across the platform as a whole.⁷⁰³ Employees of the company embraced the data upon their internal release, using BEEF results to inform team objectives and make product recommendations.⁷⁰⁴ Several teams at Meta adopted the first version of BEEF in the second half of 2021, including the Creators team (which used BEEF data to address the priorities of content creators on the platform), the Relevance Integrity team (which used BEEF data to validate and improve their models that found objectionable or reported content), the Equity team (which used BEEF data to inform goals for 2022), and the Youth team (which used BEEF data to inform policies on early teens).⁷⁰⁵

322. The results from the first version of BEEF echoed prior NETS and TRIPS results by

⁶⁹⁹ Deposition of Arturo Béjar, April 7, 2025, 283.

⁷⁰⁰ Deposition of Kyle Andrews, PhD, November 19, 2024, 187:1-6 and Exhibit 14.

⁷⁰¹ Deposition of Kyle Andrews, PhD, November 19, 2024, Exhibit 14.

⁷⁰² Deposition of Kyle Andrews, PhD, November 19, 2024, 223-242; *see also* Exhibit 14.

⁷⁰³ Deposition of Arturo Béjar, April 7, 2025, 295:21-296:19, Exhibit 21.

⁷⁰⁴ Deposition of Kyle Andrews, PhD, November 19, 2024, 257:19-258:9, Exhibit 17.

⁷⁰⁵ Deposition of Kyle Andrews, PhD, November 19, 2024, 257:19-258:9, Exhibit 19.

Highly Confidential

youth to engage with Meta’s platforms. NETS, TRIPS, and BEEF sought to understand harm from a user’s perspective, while CSERs measure harm on Meta’s platforms through narrow, mathematically deflated, incomplete proxies for bad experiences focused on views of content that violated Meta’s community standards.

331. Moreover, as explained by Mr. Béjar, these community standards were too narrowly defined.⁷²² In his deposition, Mr. Béjar explained that Meta didn’t “remove threats or violence or hate speech in a timely fashion. And they do allow attacks. And part of that is because of the way that these things are implemented, where what it takes to get a piece of content removed is very narrow.”⁷²³ However, given that bad experiences are subjective and context-dependent, the BEEF survey would have provided consumers with a more relevant piece of information: the rate at which users reported having bad experiences on the platform.

332. Meta’s decision to withhold this information from consumers, while at the same time pushing a highly favorable narrative based on the prevalence metric, has the tendency mislead consumers about the safety of its platforms. Mr. Béjar recognized that this omission was extremely important, stating that without BEEF data, “[y]ou are blindfolded as a parent.”⁷²⁴ He expanded on this observation during his deposition:

So, the internal data is saying 1 in 3 kids see bullying happening... in the last 7 days. And the external representation says out of 10,000 views, only 5 or 6, right, like a fraction of a percent. And so I think that this is a material or a critical gap of what teens are communicating they’ve experienced versus what the company is representing people see.⁷²⁵

⁷²² Deposition of Arturo Béjar, April 7, 2025, 204.

⁷²³ Deposition of Arturo Béjar, April 7, 2025, 215:21-216:1.

⁷²⁴ Deposition of Arturo Béjar, April 7, 2025, 330:12; *see also* Deposition of Kang-Xing Jin, November 8, 2024, 527:6-529:25.

⁷²⁵ Deposition of Arturo Béjar, April 7, 2025, 367:18-368:2.

Highly Confidential

consumers that U13s are not allowed to use, and do not use, Meta's platforms, and that Meta invests in and is engaged in effective age-gating, age-verification, age-enforcement, age-assurance, age-detection, and removal efforts to inhibit U13 use had the tendency to mislead. On the basis of Meta's statements, a reasonable consumer would be likely to believe that Meta was unaware that its platforms hosted substantial numbers of U13s, and that its lax age-gating efforts enabled such underage use. However, Meta did, in fact, have such knowledge. Similarly, a reasonable consumer would be likely to believe that Meta meaningfully invested in age-assurance, and would expect Meta's efforts at age-gating, age-verification, age-detection, age assurance and removal to be effective at keeping U13s from using its platforms. Again, Meta did not in fact prioritize investment in policing underage use and knew its efforts were not effective and behind industry standards. In general, Meta's statements regarding its purported efforts to keep U13s off its platforms imply that the company was proactive in addressing the issue, whereas internal evidence indicates that Meta was reactive, at best, and in many cases failed to act altogether. Individual flagged U13 accounts were removed to some extent, but the company did very little to dissuade and prevent U13s from joining and using the platforms in the first place. Meta's reactive posture is also reflected in its default practice of allowing U13s to join and remain on the platforms unless Meta could verify their underage status, rather than requiring users to affirmatively prove their age before being permitted to sign up.

384. As a result, because Meta's statements omit important and highly relevant information about the extent of improper underage usage of its platform and ineffectiveness of its age-enforcement efforts, and therefore provided an overall impression that U13 use is essentially a "non-issue," they had the tendency to mislead consumers who might have otherwise acted differently in monitoring their U13's use or ensuring disuse.

Highly Confidential

The undersigned hereby certifies their understanding that they owe a primary and overriding duty of candor and professional integrity to help the Court on matters within their expertise and in all submissions to, or testimony before, the Court. The undersigned further certifies that their report and opinions are not being presented for any improper purpose, such as to harass, cause unnecessary delay, or needlessly increase the cost of litigation.

A handwritten signature in dark ink, appearing to be 'A. Alter', is written over a horizontal line.

Adam L. Alter, Ph.D.

August 1, 2025

Date