

Filed Under Seal

Exhibit D

**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.,
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 COLIN GRAY
May 16, 2025

Shilton²⁹ demonstrate, ethical sensitivity is a key component of ethical practice, defined as the ability to recognize, understand, and judge ethical issues that arise during design.

71. A technologist's engagement with ethics is often influenced by factors such as organizational culture, economic pressure, lack of knowledge, or decision-making outside of the control of the individual technologist that may not align with the principles outlined in codes of ethics.³⁰ For example, companies can create financial incentives that reduce the ability of technologists to abide by ethical principles by measuring key performance indicators according to attention and engagement without consideration for user well-being.³¹ Such incentives create potential barriers to action that can result in legitimate ethical concerns being rejected by organizations. These barriers can include the inability of individual contributors to influence decision-making that is in user or societal best interest.³²

d. Ethical and Professional Standards Were Repeatedly Ignored Across Meta's Platforms

72. There is evidence that professional standards of ethical technology design were either disregarded or actively circumvented by Meta. Meta's platforms reflect a pattern of exploiting user vulnerabilities, particularly among adolescents, through engagement-maximizing systems that failed to prioritize user autonomy, safety, or well-being. This pattern reveals that technologists at Meta failed to uphold the ethical responsibilities that computing professionals are expected to uphold.

73. While each element of the ACM Code of Ethics and ABET standards is broadly defined, each aligns with decisions that Meta should have taken in order to support user autonomy, agency, and well-being. These principles and standards require not only that computing professionals minimize foreseeable harm and protect vulnerable populations, but also that they take a proactive role in assessing potential risks and a speedily reactive role in

²⁹ Boyd, K. L., & Shilton, K. (2021).

³⁰ Gray, C. M., & Chivukula, S. S. (2019); Watkins, C. R., Gray, C. M., Toombs, A. L., & Parsons, P. (2020, July); Wong, R. Y. (2021).

³¹ Gray, C. M., & Chivukula, S. S. (2019); Watkins, C. R., Gray, C. M., Toombs, A. L., & Parsons, P. (2020, July); Wong, R. Y. (2021).

³² Watkins et al. (2020); see also Wong (2021).

responding to emerging evidence of harm. Ethical design practice is not just about preventing known dangers—it is about embedding care, accountability, and responsiveness into the entire lifecycle of a system.

74. In the case of Meta’s platforms, that did not happen. Meta had the ability to anticipate the impact of their design decisions. It had information showing that large numbers of adolescents were using their platforms and that their functionality would create harms—especially for young users still developing self-regulation and digital literacy. And yet, Meta chose not to act, even in cases where its own evidence clearly indicated the harms these platforms were promulgating.

75. These decisions could have leveraged existing resources to take the complexity of digital systems into account. For example, Meta could have used a range of ethics-focused methods (Chivukula et al., 2024; Chivukula & Gray, 2025) or traditional user studies to better understand the potential impacts of its design decisions on user behavior and well-being. In cases where information on potential harms or risks to users already existed (such as through the outcomes of user studies that were performed or utilization of external research), Meta could have turned this into action—making design decisions to minimize further harm to users.

76. Yet in each case, Meta designed (and continued to re-design) systems that failed to meet the principles and standards set out by computing professional societies. None of the platforms implemented sufficient protective defaults for adolescent users, and in many cases, defaults actively placed youth at risk by making their accounts public, enabling contact from strangers, and/or exposing them to addictive content feeds. Notification systems across all platforms employed recapture strategies and nagging, overriding user intent to disengage. Account deletion and supervision flows created additional friction, deterring users from seeking to escape addictive or other problematic use on Meta’s platforms.

77. These harms were foreseeable. Meta had access to internal user research, external studies, and regulatory scrutiny indicating that adolescents were particularly vulnerable to the types of engagement loops and feedback systems these platforms engineered. For instance, Meta

from a parent or guardian over some engagement features (e.g., time limits, safety settings) this functionality suffers from many of the issues discussed above, providing a façade of protection rather than a substantive safeguard.

b. Meta Could Have Designed Instagram to Be Less Harmful

126. As discussed above, Meta’s design incorporated dark and attention-capture patterns that exploited known psychological and social vulnerabilities. These practices violated well-established professional ethical standards in computing and design, particularly those codified by the ACM Code of Ethics and ABET accreditation standards, which emphasize risk assessment, harm avoidance, and the responsibility to consider vulnerable populations. Importantly, less harmful designs were readily available. Meta possessed the technical capacity to implement more socially responsible and ethical alternatives. For example:

127. **More Robust Age Gating:** Meta could have implemented additional layers of verification—such as facial age estimation, device-based inferences, or multi-step confirmations involving parents or IDs—to substantially reduce access by underage users. As early as 2012, Meta considered such verification requirements as applied to Facebook, noting they could be launched within approximately six weeks. (META3047MDL-034-00385870). Instead, Meta used easily bypassed self-report mechanisms that encouraged misrepresentation.

128. **Adding Friction during Onboarding:** Rather than immediately immersing users in addictive content and social engagement, Meta could have introduced onboarding sequences that prioritized safety, social media literacy, and patterns of intentional use. This might include prompts about content types, default privacy settings, or reflective pauses before following recommendations or turning on notifications—ideally supervised by parents. Again, as early as 2012, in the context of Facebook, Meta considered requiring parent verification prior to providing children under 13 access to the platform. (META3047MDL-034-00385870).

129. **Privacy-Respectful Defaults:** Meta chose to default teen accounts to public in varying degrees until September 17, 2024, exposing young users to unwanted and potentially harmful interactions. The system could, however, have been designed with privacy by default

and more robust explanations of the risks of visibility, again in consultation with the teen's parents, from the very beginning. More generally, Meta did not begin implementing privacy-respectful defaults for teens and children until March 2021, even though, at least as early as July 2019, Meta researchers recommended doing so. (META3047MDL-065-00124111 at 119; Castaneda Dep. Ex. 14; Kilstein Dep. Ex. 57).

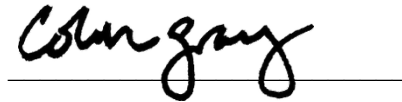
130. **Meaningful Parental Controls:** Meta could have developed parental controls much earlier in the platform's history. And even with the controls it did eventually roll out in 2022 and 2023, the supervisory mechanisms could have been durable and verifiable, including mandatory adult approval for sensitive settings or notification changes, safeguards against alternative account creation, and reliable communication channels that inform adults of critical actions. For instance, guidance from academic research in 2017 on developing self-regulation strategies between parents and children regarding technology tools could have been adopted (Hiniker et al., 2017). Instead, controls were easily disabled and poorly communicated.

131. **Streamlined Exit Options:** For users who sought to delete their accounts, Meta could have offered a clear, linear path to deletion—respecting user agency and mental health. Instead, it created a labyrinth of **forced action** and **obstruction**, making account exit much more difficult than account creation. Design alternatives—such as one-click deletion, shorter and clearly communicated timelines, or deletion flows that prioritized a user's choice—were technically feasible but deprioritized in favor of user retention, as Mark Zuckerberg himself testified (Zuckerberg (Vol. 1) Dep. 207:1-208:23). Meta failed to adopt less harmful alternatives—and took actions to increase friction and the use of dark patterns over time instead of reducing this friction (Seyson & Willett, 2025). It did so despite both information about the problem and available solutions, illustrating ethically problematic deployment of dark patterns to retain users who prefer to leave.

132. **Human-Centered Notification Systems:** Rather than exploiting user anxiety through **recapture notifications** and **social engineering** tactics, Meta could have encouraged users to set boundaries around communication, offering bulk notification control, opt-out

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.

Dated: May 16, 2025

A handwritten signature in black ink, reading "Colin Gray", written over a horizontal line.

Colin M. Gray, PhD