

AMENDED Exhibit 993

PLAINTIFFS' OMNIBUS OPPOSITION TO DEFENDANTS' MOTIONS FOR SUMMARY JUDGMENT

Case No.: 4:22-md-03047-YGR

MDL No. 3047

In Re: Social Media Adolescent Addiction/Personal Injury Products Liability Litigation

*In re: Social Media Adolescent Addiction/Personal Injury Products
Liability Litigation*

Expert Report of Minette Drumwright, Ph.D.

Submitted: June 13, 2025

A handwritten signature in cursive script, reading "Minette E. Drumwright".

Minette E. Drumwright, Ph.D.

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

156. Meta claims to practice corporate social responsibility through its own codes and the standards of care that it sets for itself as well as through public reassurances by its executives that its platforms are safe. However, an examination of Meta’s actual behavior and internal communications reveals that the company failed to meet its own stated standards of care as well as industry standards of care, and practiced pseudo corporate social responsibility. Moreover, through false assurances that it was meeting standards of care, teens, parents, schools, the U.S. Congress, and the American people at large were misled.

157. As I will elaborate below, Instagram violated Meta’s own codes and standards of care that boldly commit the company to build products responsibly *“to maximize the positive impact we have on people and society.”* Meta’s code commits employees to “design and build products that *prioritize safety . . . and include the highest protections for youth, such as age-appropriate features for youth audiences.*”³²² They must also *“anticipate and mitigate potential harms”*³²³ and *“[p]rovide appropriate warnings where necessary and articulate instructions for safe and responsible use of our products.”*³²⁴

158. Rather than warning kids, parents, schools, and the public about the harms of Meta’s platforms, Meta did just the opposite. In public forums such as subcommittees of the United

³²² <https://about.meta.com/code-of-conduct/>, p. 28 (emphasis mine).

³²³ <https://about.meta.com/code-of-conduct/>, p. 28 (emphasis mine).

³²⁴ <https://about.meta.com/code-of-conduct/>, p. 28 (emphasis mine).

States Congress, Meta’s executives boldly and repeatedly assured Congress and the American public that the company prioritizes safety and that their platforms are safe for teens. For example, when asked by Congresswoman Debbie Lesko in a Congressional hearing in March 2021 if Meta’s platform harms children, Meta CEO Mark Zuckerberg assured Congress that it does not and that designing products that improve people’s wellbeing is “very important” to the company:

LESKO: Do you believe that your platform harms children?

ZUCKERBERG: Congresswoman, I--I--I--***I don't believe so. This is something that we study and we care a lot about. Designing products that improve people's well-being is very important to us.***³²⁵

After Facebook Whistleblower Francis Haugen finished her testimony to Congress in early October of 2021, Zuckerberg professed in a Facebook post that “***it’s very important to me that everything that we build is safe and good for kids.***”³²⁶ He also asserted that Meta has “worked for years on industry-leading efforts to help people in moments of distress” that they experience on its platforms.³²⁷

159. In September 2021, Facebook’s Global Head of Safety Antigone Davis asserted in a Congressional hearing that Facebook “remains a leader in online safety” including on difficult issues such as combating bullying and child exploitation:

I work with internal teams and external stakeholders to ensure that Facebook remains a leader in online safety, including issues of bullying and combating child exploitation. This is some of the most important work that I have done in my career, and I am proud of the work that my team does every day.³²⁸

³²⁵ House Energy and Commerce Subcommittees on Communications and Technology and Consumer Protection Hold Joint Hearing on Misinformation and Disinformation on Online Platforms, March 25, 2021, p. 51 (emphasis mine).

³²⁶ Bejar Dep. Ex. 46.

³²⁷ *Id.*

³²⁸ Senate Committee on Science, Commerce, and Transportation Subcommittee on Consumer Protection, Product Safety, and Data Security, September 30, 2021, p. 3 (emphasis mine).

160. As another example, in December 2021, Instagram Head Adam Mosseri assured Congress of the primacy of youth safety at Meta:

We [Meta] all want teens to be safe online. *And from where I sit, there's no area more important than youth safety.*³²⁹

161. He also asserted that Meta invests more than its competitors—“more than anyone else”—in keeping teens safe by spending about \$5 billion dollars in 2021 alone and having more than 40,000 people working on safety and integrity:

*But we also invest, I believe more than anyone else, in keeping people, including teens, safe. We will spend around \$5 billion this year alone, and we have over 40,000 people working on safety and integrity at the company.*³³⁰

162. As noted above, numerous other standards of care stress companies' obligations to provide safe products and services. For example, the first right in the Consumer Bill of Rights is the right to safety, the American Marketing Association's first ethical norm is “[d]o no harm,” and the United Nations Guiding Principles on Business and Human Rights emphasizes that companies should “[s]eek to prevent or mitigate adverse human rights impacts that are directly linked to their operations, products or services by their business relationships, even if they have not contributed to those impacts.”³³¹ Certainly, warning consumers of potential harm in a timely manner is typically the most immediate way to prevent or mitigate adverse impacts.

163. Moreover, many codes recognize companies' special responsibilities to protect vulnerable groups such as children. For example, the American Marketing Association code asserts that marketers should recognize their “special commitments to vulnerable market

³²⁹ Senate Commerce, Science and Transportation Subcommittee on Consumer Protection, Product Safety and Data Security Holds Hearing on Social Media Reforms, December 8, 2021, p. 5 (emphasis mine).

³³⁰ *Id.*, p. 3 (emphasis mine).

³³¹ https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf at 15.

segments such as children,” and the International Chamber of Commerce Marketing and Advertising code admonishes businesspeople to avoid actions “that could have the effect of harming children and teens mentally, morally or physically.” Some codes, such as those set forth by the Children’s Online Privacy Protection Act and the Children’s Advertising Review Unit, focus exclusively on protecting children from harm.

164. As chronicled below, again and again Meta did not design for safety and failed to disclose the harms and warn parents, youth, and schools of the dangers of using its products. In fact, Meta leaders did just the opposite—they denied those harms existed and represented that their platforms were safe. As such, Meta misrepresented its actual behavior—the ultimate sign of pseudo corporate social responsibility. Leaders often rejected proposals that would have increased safety claiming either that they did not have the resources needed to design for safety or they could not accept the constraints on growth that safety measures would impose. Such claims ring hollow for a company with the substantial net income that Meta had during the relevant years: \$18.5 billion in 2019, \$29.1 billion in 2020, \$39.4 billion in 2021, \$23.2 billion in 2022, \$39.1 billion in 2023, and \$62 billion in 2024.³³²

165. Not only did Meta fail to meet standards of care and misrepresent what it was doing, Meta’s behavior contradicted the very premises upon which the private enterprise system and free markets are based. Meta broke the social contract between businesses and members of society that businesses will refrain from harming members of society and from systematically worsening the situation of a given societal group—in this case, teens and children. Meta failed to fulfill its duty

³³² SEC Form 10K for Fiscal Year Ended Dec. 31, 2021, p. 90; SEC Form 10K for Fiscal Year Ended Dec. 31, 2023, p. 90; SEC Form 10-K for Fiscal Year Ended Dec. 31, 2024.

of fidelity—to tell the truth and redress wrongful acts without delay. Likewise, Meta failed to fulfill its duty of nonmaleficence—the duty not to injure.

166. Meta’s actions did not pass the consumer sovereignty test on two grounds: 1) its young users did not have the *capability* to make safe and responsible decisions because of their developmental limitations, and 2) neither parents nor teens had the *information* that they needed to make safe and responsible decisions because of Meta’s failure to disclose and warn.

167. Moreover, Meta’s misrepresentations contradicted the premise of free markets—that marketing communications and information about product safety will be accurate, truthful, and not misleading. As such, as will be addressed below, Meta’s actions fell below the industry standards for corporate responsibility and were in violation of its own codes.

A. Meta Failed to Disclose the Harmful and Negative Experiences that Teens Encounter on Meta’s Platforms, and To the Contrary, Meta Stated that Its Platforms were Safe

168. Through its Community Standards, Meta assures the public that Instagram and Facebook are safe spaces that do not threaten the safety and dignity of its users. Specifically, Meta pledges that it does not allow nudity, illegal content, hate speech, bullying and abuse, glorification or encouragement of suicide, self-injury, or eating disorders, and graphic violence.³³³ For example, Meta assures users that it does not allow and will “*remove any content that encourages suicide, self-injury or eating disorders*” or “*that mocks victims or survivors of suicide, self-injury or eating disorders, as well as real time depictions of suicide or self-injury.*”³³⁴ As another example, Meta does not allow “*activity that sexually exploits or*

³³³ Bejar Dep. Ex. 14

³³⁴ <https://transparency.meta.com/policies/community-standards/suicide-self-injury/> (emphasis mine).

endangers children”³³⁵ and has “*zero tolerance when it comes to sharing sexual content involving minors.*”³³⁶ See Section VIII.C.2 for elaboration on Meta’s Community Standards.

169. Meta executives reinforced the Community Standards when they testified to Congress. For example, in 2024, Meta CEO Mark Zuckerberg testified, “Senator, my understanding is that we don’t allow sexually explicit content on the service for people of any age.”³³⁷ As another example, in 2021, Instagram Head Adam Mosseri testified to Congress, “[w]e don’t recommend an [sic] eating disorder-related content to people of any age.”³³⁸ Mosseri also testified, “[w]e don’t allow child or human trafficking of any kind.”³³⁹ Meta Global Head of Safety Antigone Davis also assured Congress that Meta does not allow the promotion of suicide or eating disorders:

Senator, we have been working with *suicide prevention* experts since 2006. We also work with *eating disorder* experts. *We don’t allow the promotion of either kind of content on our platform.*³⁴⁰

She asserted that Meta does not direct people to content that promotes eating disorders and uses AI to ensure that it does not exist on the platform:

Senator, *we do not direct people toward content that promotes eating disorders.* That actually violates our policies, and we remove that content when we become aware of it. *We actually use AI to find content like that and to remove it.*³⁴¹

³³⁵ <https://transparency.meta.com/policies/community-standards/child-sexual-exploitation-abuse-nudity/> (emphasis mine).

³³⁶ Jayakumar Dep. Ex. 27 META3047MDL-037-00266408; Jayakumar Dep. Ex. 28, p. 3 (emphasis mine).

³³⁷ U.S. Senate Judiciary Committee on Online Sexual Exploitation, Jan. 31, 2024, p. 22.

³³⁸ U.S. Senate Commerce, Science and Transportation Subcommittee on Consumer Protection, Product Safety and Data Security Hearing on Social Media Reforms, Dec. 8, 2021, p. 17.

³³⁹ *Id.* p. 24.

³⁴⁰ U.S. Senate Subcommittee on Consumer Protection, Product Safety, and Data of the Committee on Commerce, Science, and Transportation, Sept. 30, 2021, p. 26 (emphasis mine).

³⁴¹ *Id.* p. 29 (emphasis mine).

170. The assurances of Meta executives notwithstanding, Meta’s Community Standards provide a classic example of pseudo corporate social responsibility. As I will demonstrate in the following sections, teens routinely have negative experiences and suffer harms on Meta’s platforms that are seemingly at odds with Meta’s Community Standards and public assurances. While I reference Meta’s Community Standards and Meta’s leadership representations about content that it purportedly does not allow on its platforms, my opinions are not about content moderation or even the fact that bad content exists on these platforms. Rather, my opinions focus on how Meta has misrepresented the risks and harms that it is aware of and has failed to disclose and warn teens, parents, and the community at large of those risks and harms. In short, Meta has engaged in pseudo corporate responsibility by creating the appearance that Facebook and Instagram are safe spaces for teens when it possesses information indicating that millions of teens (and younger kids) are experiencing significant harms on those platforms on a daily basis.

171. Key to Meta’s misrepresentations and omissions regarding the safety of its platforms are the Community Standards Enforcement Report (CSER), which is published on Meta’s Transparency Center and has been published since 2018.³⁴² When Zuckerberg testified at a Congressional hearing in 2021, he emphasized the importance of CSER and referred to it as a “model” for other companies:

CONGRESSMAN LATTA: --Let me ask you quick now, where are those transparency reports being reported to and--and how often do you think those should be going out?

ZUCKERBERG: Well, Congressman, as--*as a model, Facebook has been doing something to this effect for every quarter, right, where we report on the prevalence of each category of harmful content and how effective our systems are at identifying that content and removing it in advance*, and--and I think the companies should be held accountable for having effective systems to do that broadly.³⁴³

³⁴² Bejar Dep., p. 288.

³⁴³ U.S. Senate Commerce, Science and Transportation Subcommittee on Consumer Protection, Product Safety and Data Security on Social Media Reforms, Dec. 8, 2021, p. 15 (emphasis mine).

172. Zuckerberg highlighted the supposed effectiveness of what he claimed is Meta’s “industry leading” AI systems:

You know our -- *our prevalence metrics suggest that, I think it's 99 percent or so of the content that we remove, we're able to identify automatically using AI system*. So I think that our efforts in this, while they're not perfect, *I think are industry leading*.³⁴⁴

173. Mosseri assured Congress that Meta is transparent about the amount of harm on its platform through its Community Standards Report:

Senator, I believe deeply in transparency. *I've spent an immense amount of time over the years not only trying to be transparent about how our algorithms work but also investing in making sure we're transparent about how much problematic content is on the platform. And I believe you can see that in our Community Standards Report*.³⁴⁵

174. Meta’s Global Head of Safety Antigone Davis also emphasized Meta’s supposed commitment to transparency through its CSER transparency report, its oversight board, and its human rights impact assessments when she testified to Congress:

Respectfully, Senator, *I think that our commitment to transparency in the last few years should be a very good indication of our commitment. We have launched a transparency report regularly*. We have set up an oversight board. We have human rights impact assessments. We are doing a tremendous amount to ensure transparency around our platform.³⁴⁶

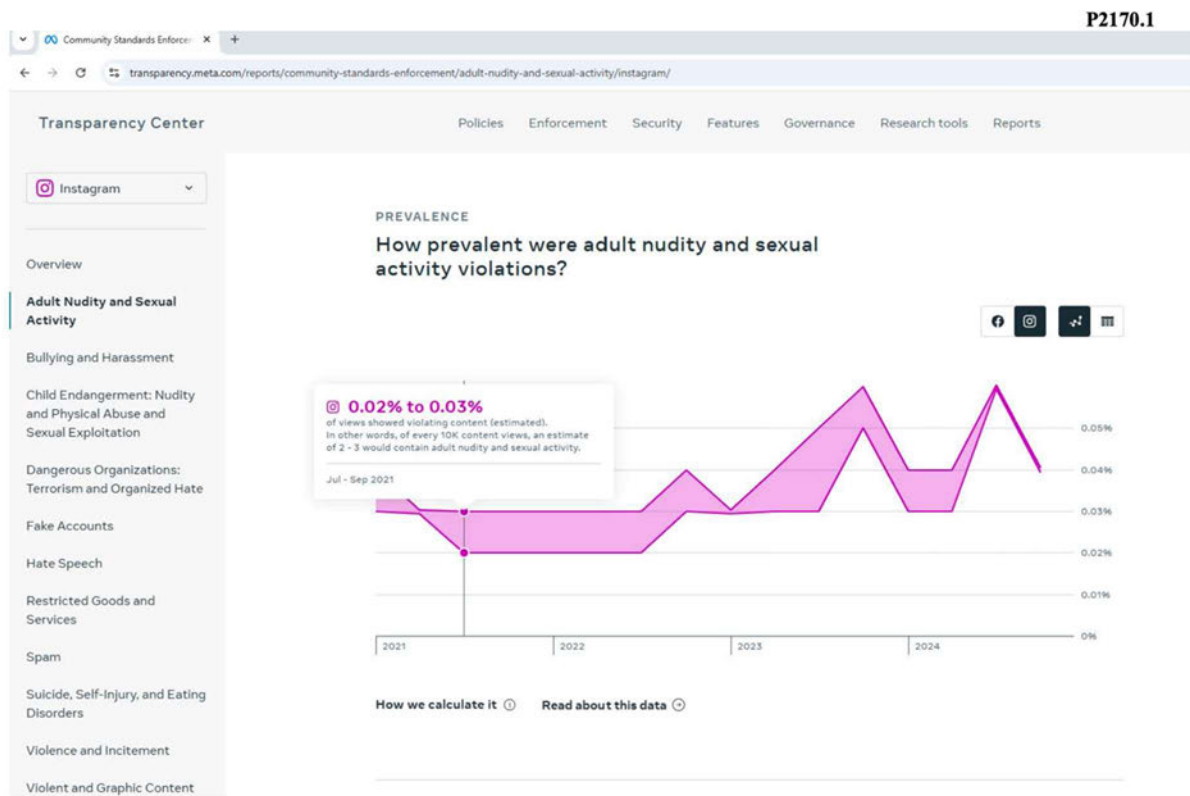
175. These CSER reports consistently represented that the “prevalence” of content violating Meta’s Community Standards on Instagram was extremely low; however, as demonstrated below, Meta had information from a variety of sources, including its own studies, that was far from the case. For example, in the third quarter of 2021, CSER reported that the

³⁴⁴ U.S. Senate Judiciary Hearing on Online Child Sexual Exploitation, Jan. 31, 2024, p. 22 (emphasis mine).

³⁴⁵ U.S. Senate Commerce, Science and Transportation Subcommittee on Consumer Protection, Product Safety and Data Security on Social Media Reforms, Dec. 8, 2021, p. 25 (emphasis mine).

³⁴⁶ U.S. Senate Subcommittee on Consumer Protection, Product Safety, and Data of the Committee on Commerce, Science, and Transportation, p. 26 (emphasis mine).

prevalence of adult nudity and sexual activity was between 0.02 and 0.03 percent, indicating that of 10,000 views, only two to three would contain adult nudity or sexual activity:³⁴⁷



176. In contrast, an extensive (n=237,923) and sophisticated internal Meta survey, the Bad Experiences and Encounters Framework (BEEF) conducted in June/July 2021, found that 16.3 percent of respondents had seen nudity *during the last seven days*.³⁴⁸ The percentages were even higher for teens—19.2 percent of 13- to 15-year-olds and 18.9 percent of 16- to 17-year-olds.³⁴⁹ Many of these users reported seeing nudity multiple times in the last seven days, some more than 10 times.³⁵⁰ In an earlier Instagram study conducted in October 2019, the Negative

³⁴⁷ Bejar Dep. Ex. 25.

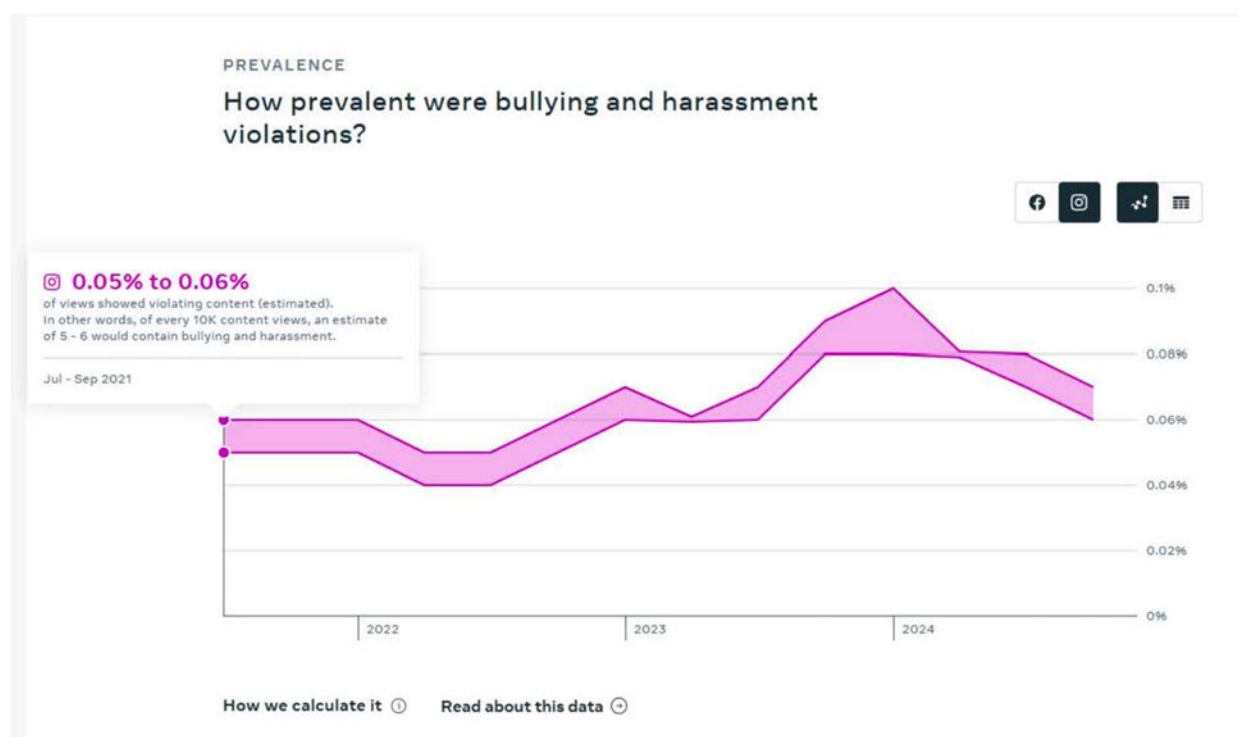
³⁴⁸ Bejar Dep. Ex. 21, p. 21, META3047MDL-004-00015049.

³⁴⁹ *Id.*

³⁵⁰ *Id.*, p. 11.

Experiences Survey (NES), revealed that 23 percent of all respondents and 26 percent of teen respondents had seen at least one example of nudity on Instagram during the last seven days.³⁵¹

177. This stark contrast between what Meta represents to the public versus what its internal data shows also exists for other harms that users experience on Instagram. As another example, Meta represented to the public that in the third quarter of 2021, the prevalence rate of bullying and harassment on Instagram was a miniscule 0.05 to 0.06 percent, indicating that of 10,000 views only five or six would involve bullying or harassment.³⁵²



178. The internal BEEF survey, however, showed that 28.3 percent of users reported that they had witnessed bullying on Instagram during the past seven days, and 8.1 percent reported that they had been the victim of bullying during the past seven days.³⁵³ Similarly, according to the

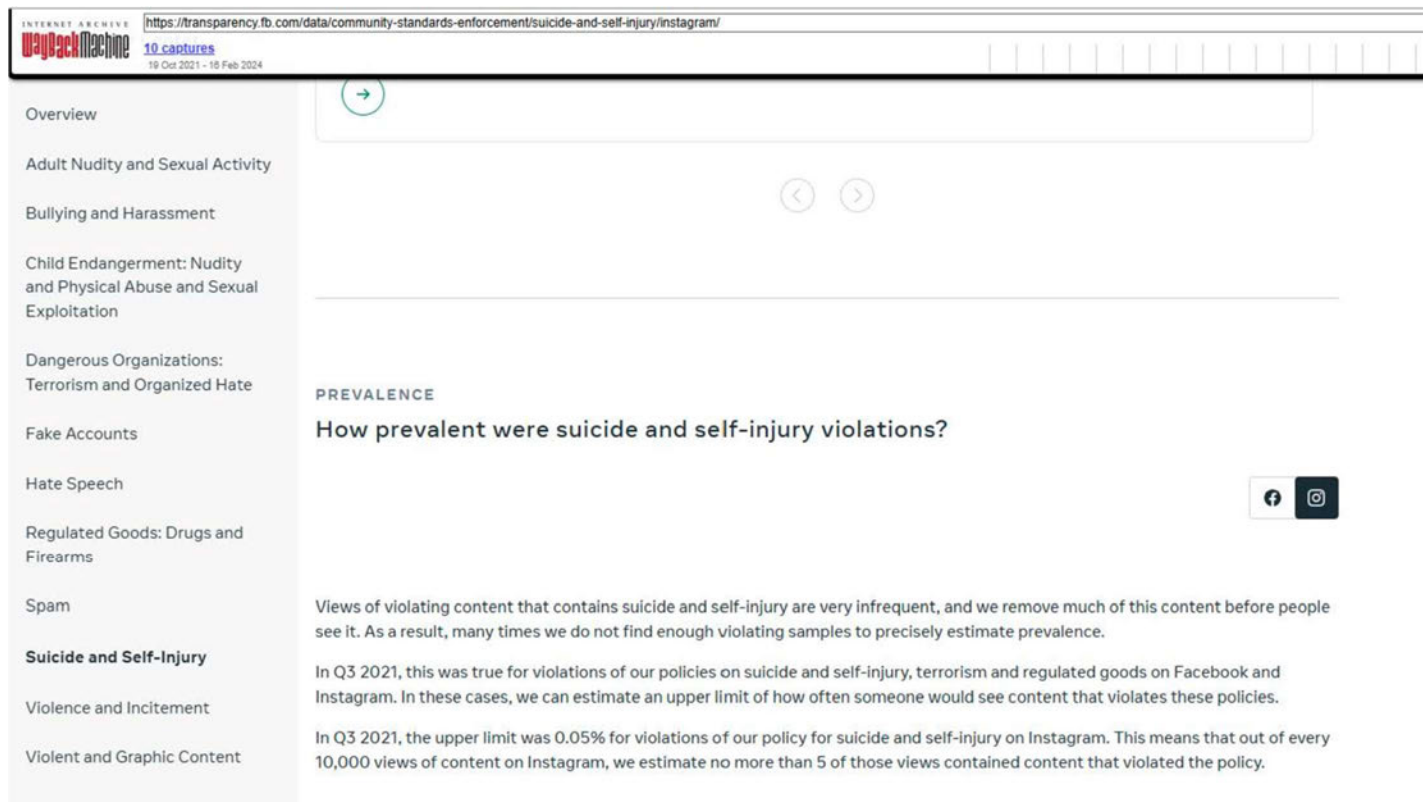
³⁵¹ Bejar Dep. Ex. 9, p. 7, META3047MDL-059-00000222.

³⁵² Bejar Dep. Ex. 26.

³⁵³ Bejar Dep. Ex. 21, p. 21, META3047MDL-004-00015049.

NES internal survey, 30 percent of teens surveyed in 2019 had seen bullying, and nine percent were victims of bullying on Instagram during the past seven days.³⁵⁴

179. As another example, in the third quarter of 2021, CSER represented that suicide and self-injury violations were so rare that they cannot very precisely estimate the prevalence.³⁵⁵ Meta then represented that the upper limit of such violations was only 0.05 percent, indicating that out of every 10,000 views, only five contained SSI violations:³⁵⁶



In contrast, the internal BEEF survey showed that 6.4 percent of the people surveyed had seen suicide and self-injury content on Instagram in the past seven days, and the percentages were even higher for teens (8.4% for 13-15-year-olds; 7.2% for 15–17-year-olds).³⁵⁷ According to NES, 17

³⁵⁴ Bejar Dep. Ex. 9, p. 7, META3047MDL-059-00000222.

³⁵⁵ Bejar Dep. Ex. 27.

³⁵⁶ *Id.*

³⁵⁷ Bejar Dep. Ex. 21, p. 21.

percent of teens (almost 1 in 5) reported having seen SSI content on Instagram during the past seven days.³⁵⁸

180. As another example, in September 2021, Meta represented that it “cannot estimate prevalence of child endangerment right now.”³⁵⁹ Conversely, the BEEF survey found that 13 percent of 13–15-year-olds had received an unwanted sexual advance in the last 7 days alone.³⁶⁰ For most users, this occurred more than once in the past seven days, and for over 8 percent, it occurred more than 10 times in the past seven days.³⁶¹ Nearly 94 percent reported that these unwanted sexual advances came from strangers that they did not know in real life or even online.³⁶²

181. Meta’s internal BEEF survey also showed that users were experiencing harms that Meta did not report at all in its public CSER reports. For instance, Meta’s Transparency Center does not even address the issue of negative social comparison experienced by users on its platforms.³⁶³ The BEEF survey, however, found that 21.4 percent—or one in five—kids aged 13–15 felt worse about themselves after using Instagram in the last 7 days.³⁶⁴

182. Alarming, BEEF results revealed that more than half of teens had one or more negative experiences in the past week. (Note that ages are listed vertically on the left):³⁶⁵

³⁵⁸ Bejar Dep. Ex. 9.

³⁵⁹ Bejar Dep., p. 368-69.

³⁶⁰ *Id.*

³⁶¹ Bejar Dep. Ex.21, p. 11, META3047MDL-004-00015039.

³⁶² *Id.* p. 13.

³⁶³ <https://transparency.meta.com/reports/community-standards-enforcement/>.

³⁶⁴ Bejar Dep. Ex. 21, p. 21, META3047MDL-004-00015049.

³⁶⁵ Bejar Dep. Ex. 21, p. 19, META3047MDL-004-00015047.

	None	1 or more (out of 5)	% of each age group experiencing one or more issues
13-15	5,283	6,215	54.1%
16-17	7,371	9,888	57.3%
18-21	16,033	21,164	56.9%
22-26	12,520	15,713	55.7%
27-34	13,019	13,426	50.8%
35-44	9,978	7,689	43.5%
45 or older	7,877	3,564	31.2%
Prefer not to say	5,647	5,229	48.1%
Total	77,727	82,888	51.6%

183. See the summary slide of BEEF results with ages on the top horizontally and negative experiences represented vertically:

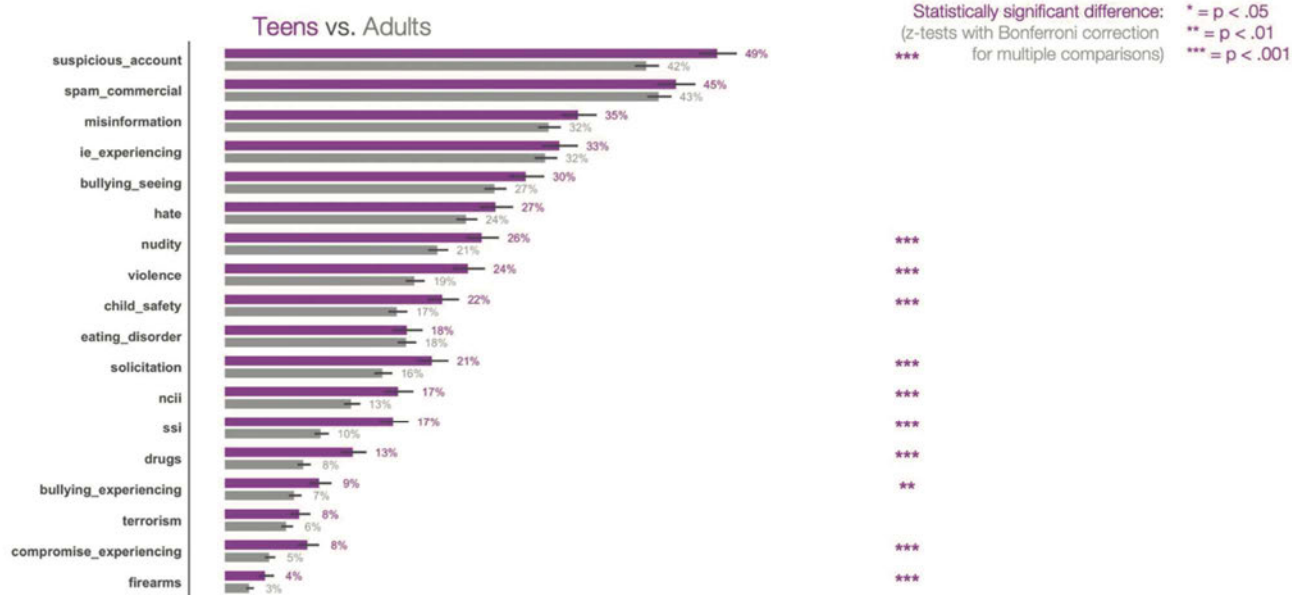
	Overall %	13-15	16-17	18-21	22-26	27-34	35-44	45+	Delta
		Column N %	Column N %	Column N %	Column N %	Column N %	Column N %	Column N %	Highest-lowest
commerciality	48.2%	37.7%a	42.5%b,f,g	47.6%c	51.2%d	50.5%d	44.8%b,e	39.6%a,f	13.50%
misinfo	30.3%	24.4%a	29.2%b	30.0%b,d	33.6%c	31.2%d	24.4%a	16.1%e	17.50%
bully witness	28.3%	27.2%a	29.4%b	29.9%b,c	31.1%c	27.4%a	19.8%d	10.6%e	20.50%
audience limitation	26.8%	28.2%a,b	29.3%a	28.1%a	26.7%b	24.9%c	20.3%d	15.9%e	13.40%
usability passive	25.5%	30.1%a	30.0%a	29.5%a	25.9%b	21.6%c	17.1%d	10.1%e	20.00%
hate witness	25.3%	26.0%a	28.5%b	27.9%b	27.9%b	22.5%c	15.7%d	9.0%e	19.50%
data privacy	24.4%	17.9%a	21.4%b	23.0%c	25.9%d	24.8%d,f	20.2%b	14.3%e	11.60%
perceived control	23.9%	18.4%a	20.0%a	24.8%b	26.7%c	24.7%b,c	20.9%a	14.2%d	12.50%
fake acct 1st	21.8%	15.8%a	21.2%b	24.8%c	23.4%c,d	22.9%b,d	17.4%a	13.2%e	11.60%
negative comparison	19.2%	21.4%a	19.5%b	20.1%a,b	19.4%b	17.5%c	13.4%d	6.8%e	14.60%
transparency	17.5%	14.9%a	19.1%b,f	19.0%b	17.1%c	15.1%a	12.0%d	8.4%e	10.70%
political posts	17.0%	14.3%a,d	16.3%a,b	17.3%b	18.0%b,c	15.2%a,f	12.4%d	8.7%e	9.30%
nudity	16.3%	19.2%a	18.4%a	18.2%a	16.7%b	13.6%c	11.3%d	6.2%e	13.00%
over enforcement	14.8%	14.1%a,c	16.8%b	16.3%b	14.8%a,f	12.9%c	9.1%d	5.9%e	10.90%
violence	12.8%	12.8%a,b	14.3%a	13.2%a	13.1%a	11.5%b	8.2%c	5.0%d	9.30%
unwanted advances	11.9%	13.0%a	14.1%a	14.2%a	11.0%b	8.0%c	7.7%c,d	6.9%d	7.30%
usability action	10.7%	11.2%a,e	11.1%a,e	11.1%a	10.3%a,b	9.6%b	8.5%c	6.6%d	4.60%
bully target	8.1%	10.8%a	9.7%a	8.1%b	7.1%b,c	6.5%c	4.9%d	3.4%e	7.40%
self harm	6.7%	8.4%a	7.2%a,b	7.5%a	5.8%b	3.5%c	2.5%c,d	1.9%d	6.50%
acct security	3.9%	5.2%a	3.2%b,c	4.2%a,b	3.2%c	2.1%d	1.6%d	1.3%d	3.90%
drugs	3.9%	3.7%a,b	3.5%a,b	3.8%a	2.7%a,b,c	3.6%a,b	2.3%b,c	1.7%c	2.10%
impersonation 1st	3.7%	5.5%a	3.5%a	5.0%a	3.1%a	1.2%b	0.7%b	0.3%b	5.20%

An expanded table with confidence intervals can be found [here](#).

184. See the NES summary results:

PERCEIVED REACH: TEENS & ADULTS

Overall, teens have significantly more negative experiences than adults. Perceived reach was higher among teens than adults for 11 out of the 18 problem areas that we asked about.



7

HIGHLY CONFIDENTIAL (COMPETITOR)

META3047MDL-059-00000222

185. Arturo Bejar is a former Meta employee that worked for years in the area of online safety for users, including kids.³⁶⁶ He has about 30 years of experience in the industry.³⁶⁷ At Meta, his work included assessing whether Meta was adequately protecting and preventing harms to kids and making recommendations to leadership about those issues.³⁶⁸ Bejar had two stints at Meta, first from 2009 to 2015 and later from 2019 to 2021.³⁶⁹ During his second stint, Bejar worked as a safety consultant for the Instagram wellbeing team.³⁷⁰ His work included overseeing the BEEF survey.³⁷¹

³⁶⁶ Bejar Dep., p. 12.

³⁶⁷ *Id.*, p. 17-18.

³⁶⁸ *Id.*

³⁶⁹ *Id.*, p. 36.

³⁷⁰ *Id.*, p. 254-255.

³⁷¹ *Id.*, p. 283.

186. Bejar testified that being transparent is very important for kids' safety.³⁷²

Transparency reduces harm to kids because the more that people understand, the more Meta can deploy effective ways of preventing harm.³⁷³ Bejar testified that it is absolutely essential for Meta to inform users and the public of the harms its platforms cause or contribute to cause so that people can make informed decisions.³⁷⁴ This is consistent with many of the professional codes and standards of care outlined above, which highlight the importance of transparency and the need for companies to disclose harms associated with their products.

187. Despite the need to be transparent, Meta engaged in pseudo corporate responsibility. Bejar characterized Meta's public statements on its Transparency Center as "profoundly misleading, dangerously misleading."³⁷⁵ Bejar explained that Meta only enforces and discloses violations of its Community Standards—which are narrowly defined—rather than preventing and disclosing many other actually harmful experiences.³⁷⁶ Further, the Transparency Center is deeply misleading as to the likelihood that a kid is going to experience a certain harm.³⁷⁷ For instance, the Community Standards and the Transparency Center make it seem like a kid will not receive unwanted sexual or eating disorder content, but that is not consistent with Bejar's experience or the BEEF survey results.³⁷⁸ There was a very large discrepancy between the information that Meta had and what Meta was telling the public about the prevalence and frequency of suicide and

³⁷² Bejar Dep., p. 53-54.

³⁷³ *Id.*

³⁷⁴ *Id.*, p. 69.

³⁷⁵ *Id.*, p. 422.

³⁷⁶ *Id.*, p. 204.

³⁷⁷ *Id.*

³⁷⁸ *Id.*, p. 229-31.

self-harm, nudity and sexual activity, violent and graphic content, and child endangerment occurring on its platforms.³⁷⁹

188. The 2019 NES survey results told Bejar that there is a meaningful amount of harm happening on Instagram every 7 days and that Meta had that data.³⁸⁰ The NES survey results also showed that these harms were happening to hundreds of millions of users on Instagram.³⁸¹ Bejar learned from the NES survey that people were experiencing harms at 100 to 400 times greater than what is disclosed publicly in the CSER reports.³⁸² Bejar testified that these NES survey results had not been published to anyone outside of the company, but he states that they should have been.³⁸³ If Meta was truly committed to transparency, as its top leadership represented in testimony to the public, it should have disclosed these findings.

189. The BEEF survey was more robust than prior surveys, and it had more respondents and covered more topics.³⁸⁴ It was large enough to be statistically significant and accurately represent the population of Instagram users.³⁸⁵ When Bejar received the BEEF survey results, he thought the data was staggering and that it was critical to bring it to the attention of Meta leadership as soon as possible.³⁸⁶ Bejar's expectation was that once this data was presented to Meta leadership, they would immediately start work to address these harms.³⁸⁷

³⁷⁹ *Id.*, p. 369-70.

³⁸⁰ *Id.*, p. 126.

³⁸¹ *Id.*, p. 127.

³⁸² *Id.*, p. 259.

³⁸³ *Id.*, p. 135.

³⁸⁴ Bejar Dep., p. 284-85.

³⁸⁵ *Id.*, p. 296.

³⁸⁶ *Id.*, p. 308.

³⁸⁷ *Id.*, p. 315.

190. Bejar testified that Meta's leadership had the ability to reduce these harms because many of the harms were occurring due to the product design of the platforms.³⁸⁸ For instance, the BEEF survey data showed that many of the harms that teens were experiencing came from strangers and interactions that Instagram's product design facilitated.³⁸⁹

191. The BEEF survey results became available around the same time that Meta was under external scrutiny after a whistleblower went public.³⁹⁰ Meta responded to the whistleblower incident by internally locking down the research around harms to users.³⁹¹ None of the BEEF results were disclosed to the public, but Bejar testified that they should have been.³⁹² And again, if Meta was truly committed to transparency, it would have made these survey results public. As stated by Bejar, parents need to know but are blindfolded without this data.³⁹³ Sharing research and being transparent is absolutely essential to help reduce harm to kids on Instagram, and Bejar testified that Meta's lack of transparency by failing to inform the public of its research, like BEEF, has contributed to harms to kids on Instagram.³⁹⁴

192. Not only did Meta fail to be transparent with the public about the results of its BEEF survey, but also it actively undermined the survey results. The BEEF survey also had asked respondents to identify what specific emotion they felt as a result of their negative experience and how long they felt that way.³⁹⁵ Bejar explained that this emotions data was important to understand how severe the bad experience was.³⁹⁶ But Kyle Andrews, who was responsible for

³⁸⁸ *Id.*, p. 315-320.

³⁸⁹ *Id.*, p. 318-319.

³⁹⁰ *Id.*, p. 327-28.

³⁹¹ *Id.*

³⁹² *Id.*, p. 330.

³⁹³ *Id.*

³⁹⁴ *Id.*, p. 330-31.

³⁹⁵ Bejar Dep. Ex.21, p. 6, META3047MDL-004-00015034.

³⁹⁶ Bejar Dep., pp. 292-95, 323.

running the BEEF survey, testified that he was told for “policy/legal reasons” to delete the emotions data and not analyze it.³⁹⁷ Andrews also stated that “we’re not allowed to ask about emotions in surveys anymore.”³⁹⁸ Bejar testified that deleting this data shows that Meta does not want to have information about the harm that people are experiencing on its platform and that it was ethically inappropriate to not analyze these data.³⁹⁹

193. Contrary to Zuckerberg’s testimony to Congress that Meta’s efforts around transparency and harm reduction are a model for other companies and industry leading, Bejar concluded that Meta is not an industry leader for transparency.⁴⁰⁰ As such, Instagram failed to meet its own and industry standards of care by creating the false impression that its platforms are safe but failing to disclose the significant risk of harm to users.

194. As I will elaborate below, Meta had many other ways to be aware that CSER was reporting highly inaccurate results in terms of how it reflected teens’ actual experiences on Instagram.

B. Meta Failed to Disclose the Risk of Problematic Use and Addiction

195. Meta’s approach to problematic use and addiction provides another example of pseudo corporate social responsibility. Meta leaders repeatedly denied in public that its platforms are addictive. For example, when Instagram Head Adam Mosseri was asked during a Congressional hearing if Instagram is addictive, he replied that the research does not suggest that Meta’s products are addictive and that Instagram helps more than it harms teens on 11 of 12 difficult issues that they face:

³⁹⁷ Andrews Dep., pp. 178, 365-73; Andrews Dep. Exs. 35, 36.

³⁹⁸ *Id.*

³⁹⁹ Bejar Dep., pp. 325-26.

⁴⁰⁰ *Id.*, p. 489.

Senator, respectfully, ***I don't believe the research suggests that our products are addictive.*** Research actually shows that on 11 of 12 difficult issues that teens face -- teens are struggling said Instagram helps more than harms.⁴⁰¹

196. Likewise, when Meta CEO Mark Zuckerberg was asked in a Congressional hearing if he thought too much time in front of screens passively consuming content is harmful to children's mental health, he replied, "I don't think that the research is conclusive on that."⁴⁰² When a congressman asked Zuckerberg, "So yes or no, do you agree that you make money off of creating an addiction to your platforms—Mr. Zuckerberg?" Zuckerberg replied, "Congressman, no. I don't agree with that."

197. Meta's Global Head of Safety Antigone Davis also denied that Meta's product was addictive when she testified to Congress:

So I disagree with calling our product addictive. I also think that is not how we build products.⁴⁰³

198. Internally, however, Meta employees recognized that overuse and addiction were significant forms of problematic use of Instagram, and they understood the drug-like properties of Instagram. For example, Instagram employee and senior researcher Shayli Jimenez referred to Instagram as a drug and noted how uncomfortable Instagram Head Adam Mosseri was when she talked about dopamine in a meeting:

Oh my gosh yall ***IG is a drug . . . We are causing Reward Deficit Disorder bc people are binging on IG so much they can't feel reward anymore . . .*** I know Adam [Mosseri] doesn't want to hear it – ***he freaked out when I talked about dopamine in my teen fundamentals leads review but it is undeniable. Its biological and psychological truth.***⁴⁰⁴

⁴⁰¹ U.S. Senate Commerce, Science and Transportation Subcommittee on Consumer Protection, Product Safety, and Data Security Hearing on Social Media Reforms, Dec. 8, 2021, p. 6.

⁴⁰² U.S. House of Representatives Energy and Commerce Subcommittee on Communications and Technology and Consumer Protection Joint Hearing on Misinformation and Disinformation on Online Platforms, March 25, 2021, p. 20.

⁴⁰³ U.S. Senate Subcommittee on Consumer Protection Product Safety, and Data Security of the Committee on Commerce, Science and Transportation, Sept. 30, 2021, p. 38 (emphasis mine).

⁴⁰⁴ Jimenez Dep. Ex. 17, META3047MDL-020-00342152.

Similarly, in May 2020, Jimenez acknowledged that “[t]een brains are much more sensitive to dopamine, one of the reasons that the risk of drug addiction is higher for adolescents and it’s the same thing that keeps them scrolling and scrolling. Due to the immature brain, they have a much harder time stopping even though they want to -- our own product foundation research has shown teens are unhappy with the amount of time they spend on our app.”⁴⁰⁵

199. [REDACTED], a member of Meta’s central youth team, acknowledged the addictive design of Meta’s platforms in that the algorithm is tuned to optimize engagement:

[REDACTED]

200. Former Meta employee and safety consultant, Arturo Bejar, characterized addiction as a “radioactive topic”⁴⁰⁷ at Meta and said that Meta employees were not supposed to use the term “addiction”; instead they were supposed to refer to “problematic use.”⁴⁰⁸ According to Kristin Hendrix, Head of Research for Instagram, Meta changed the term “addictive use” to “problematic use” because external advisors told them to do so.⁴⁰⁹

⁴⁰⁵ Jimenez Dep. Ex. 10, p. 9, META3047MDL-003-00191215.

⁴⁰⁶ [REDACTED] Dep., pp. 104-09; [REDACTED] Dep. Ex. 9, META3047MDL-047-00706686.

⁴⁰⁷ Bejar Dep., pp. 136-137, 555.

⁴⁰⁸ *Id.*

⁴⁰⁹ Hendrix Dep., p. 493.

201. When Safety and Wellbeing Lead Vaishnavi Jayakumar was asked to define problematic use, the first three issues that she named—“time spent online, addiction, and dependency”—were all relevant to social media addiction:

Problematic use was a category of behaviors that the research team had identified as being problematic ways in which people were using the platform. *That included topics like time spent online, addiction, dependency,* and social comparison.⁴¹⁰

202. Bejar states that Meta intentionally narrowly defined problematic use.⁴¹¹ Meta had redefined the issue so it was not studied or understood in good faith.⁴¹² According to Meta Researcher Jennifer Guadagno, up until late 2017, Meta had not done any research on “addiction.”⁴¹³

203. In December 2017, Meta Research Scientist Moira Burke posted a public blog on Meta’s website titled, “Hard Questions, Is Spending Time on Social Media Bad for Us.”⁴¹⁴ In her blog, Burke wrote, “[w]e know that people are concerned about how technology affects our attention spans and relationships, as well as how it affects children in the long run. We agree these are critically important questions, and we all have a lot more to learn.”⁴¹⁵ Burke acknowledged that at that time, there were a lot of open questions about how technology affects children and that additional research was needed.⁴¹⁶ Notwithstanding that Meta allowed children as young as 13 to use its services, Burke was not aware of any existing research that showed that Facebook or Instagram was safe for a 13-year-old kid.⁴¹⁷ Burke’s blog then announced that Meta had pledged

⁴¹⁰ Jayakumar Dep., p. 127-128.

⁴¹¹ Bejar Dep., p. 136-37.

⁴¹² *Id.*, p. 138.

⁴¹³ Guadagno Dep., p. 165.

⁴¹⁴ Burke Dep., pp. 40-41; Burke Dep. Ex. 4.

⁴¹⁵ *Id.* p. 48.

⁴¹⁶ *Id.* p. 48-49.

⁴¹⁷ *Id.* pp. 49-50.

\$1 million toward research to better understand the relationship between social media use and youth wellbeing, and made the commitment to “share our findings and improve our products.”⁴¹⁸

204. In mid-2018, Burke published an article titled “Problematic Facebook use: When people feel like Facebook negatively affects their life,” which presented the results from a survey of 20,000 U.S. Facebook users.⁴¹⁹ This survey defined problematic use as: “Serious problems with sleep, work, or relationships that [respondents] attribute to Facebook, AND concerns or preoccupations about how [users] use Facebook.” It found that “3.1% of Facebook users in the U.S. experience problematic use.”⁴²⁰ It also found that teenagers experience higher rates of problematic use than any other age group, and that “[p]eople who feel like their use of Facebook is problematic spend more time on the site and have more sessions than people who don’t experience problematic use. They also have a greater proportion of sessions late at night, and spend about 2 extra hours per month on Facebook late at night.”⁴²¹ Notably, the study stated that a broad definition of problematic use was used, so it represented the “upper bounds” of problematic use.⁴²²

205. Internally, however, Meta employees recognized that its 3.1 percent statistic did not represent the upper bounds of users experiencing problematic use or addiction. Burke posted an internal note on Workplace that described the same survey of 20,000 users.⁴²³ In that post, Burke acknowledged that there are “lots of ways to think about defining problematic use” and that “under the most liberal definition” 14 percent of users spend a lot more time on Facebook than

⁴¹⁸ *Id.* pp. 51-52.

⁴¹⁹ Burke Dep. Ex. 5, META3047MDL-039-00000058.

⁴²⁰ *Id.*

⁴²¹ *Id.*; Burke Dep., pp. 58-78.

⁴²² *Id.*; Burke Dep., p. 138.

⁴²³ META3047MDL-039-00000058.

they want to.⁴²⁴ In another internal document, Meta described the 3.1 percent statistic as Facebook users who experience “severe” problematic use, and acknowledged that more than half (55 percent) of Facebook users experience mild problematic use.⁴²⁵

206. In late 2019, in another internal Workplace note, Meta described Burke’s 3.1 percent statistic as representing “extreme problematic use” cases.⁴²⁶ This internal note states that based on how Meta was measuring problematic use in 2019, 12.5 percent of users were experiencing problematic use and “[a]nother 11 percent of users report negative life impact ‘very often’ but only ‘sometimes’ lacking control.”⁴²⁷ Notably, Burke conceded that these much higher problematic use statistics were never publicly disclosed.⁴²⁸

207. Meta employees also possessed information indicating that teens were particularly susceptible to overindulgence because of their developing brains. For example, a 2020 presentation entitled, “Teen Fundamentals,” examined teen brain chemistry and asserted that “teen brains are highly dependent on their temporal lobe where the reward system reigns supreme” and concluded that teens are “very vulnerable to overindulgence” because they do not have a “mature frontal cortex to help impose limits.”⁴²⁹ Instagram Senior Researcher Alison Lee noted the diversity of highly risky but rewarding experiences that could draw teens to Instagram:

. . . a short term reward and inexperience makes teens prone to risky behavior and there are plenty that present themselves online and on Instagram. This could be engaging with predators, consuming dark content, sharing nude photos or copycat self-harm.⁴³⁰

⁴²⁴ *Id.*

⁴²⁵ META3047MDL-003-00086233 at p. 41.

⁴²⁶ META3047MDL-039-00000400.

⁴²⁷ *Id.*

⁴²⁸ Burke Dep., pp. 175-76.

⁴²⁹ META3047MDL-003-000132740.

⁴³⁰ META3047MDL-003-00191215 (emphasis mine).

Lee asserted that whenever teens find something unexpected on Instagram, their brains deliver dopamine effects and that “[teens] are insatiable when it comes to ‘feel good’ dopamine effects.”⁴³¹ She also noted that “[d]ue to the immature brain, [teens] have a much harder time stopping even though they want to—our own product foundation research has shown teens are unhappy with the amount of time that they spend on our app.”⁴³²

208. As noted above, Meta employees had information showing that certain product features such as “likes,”—which Meta refused to make invisible—created rewards, and that too much reward equals addiction and too little reward equals depression.⁴³³

209. Other features such as notifications could prompt overuse and encourage dependency on Instagram among teens, as Jayakumar explained:

*... when you look at the developmental appropriateness of experiences for people between the age of 13 and 18, we generally acknowledge that we -- that notifications tend to draw them into the app repeatedly, and for more -- for more sessions and for longer time in session than we would otherwise want for young people of that age. And so notifications have been seen as a form of increasing the dependency of somebody on the app.*⁴³⁴

210. Nonetheless, Instagram’s growth team continued to focus on notifications to stimulate increased teen use. For example, in December of 2019, Meta highlighted notifications as a main strategy to shore up usage among millions of teens who were at risk of not being monthly or daily active users.⁴³⁵ Likewise, a focus on notifications was part of a strategy during the first half of 2024 to drive growth among teen daily users.⁴³⁶ KX Jin, one of Meta’s early

⁴³¹ META3047MDL-136-00009489; 490.

⁴³² META3047MDL-003-00191215.

⁴³³ META3047MDL-020-00342310.

⁴³⁴ Jayakumar Dep., p. 141-142 (emphasis mine).

⁴³⁵ META3047MDL-020-00535512-META3047MDL-020-00535514; Kilstein Dep., pp. 470-474.

⁴³⁶ Kilstein Dep. Ex. 13, pp. 2, 5.

engineers and later the Head of Health, agreed that features like late night push notifications, infinite scroll, auto-play, time bound content, and rankings may amplify problematic use.⁴³⁷

211. Jayakumar also acknowledged that recommendation algorithms such as those used in Reels and Explore contributed to overuse and addiction:

I think the design of the algorithm to promote sustained engagement by continuously showing appealing content and increasingly appealing content can play a role in drawing viewers into the experience for longer periods of time than they necessarily would want for themselves.⁴³⁸

212. These algorithms on parts of the platform such as Explore use infinite or endless scroll, which can cause users to succumb to “doom scrolling,” in which they view a long chain of the same topic,⁴³⁹ or to fall into “rabbit holes that can cause emotional spirals.” A slide in a presentation entitled “Social Comparison Convening Research” described how constant scrolling through appearance-related content—a rabbit hole—can have a negative effect on teens and leave them feeling “down and exhausted”:

Teens talk about cycles of constant scrolling and comparisons. They rabbit hole on appearance related content on social media and this cycle of seeing perfect images and comparisons leaves them feeling down and exhausted.⁴⁴⁰

213. Bejar also acknowledged that the design of Instagram causes or contributes to cause addiction, including the nature and frequency of notifications, never-ending infinite scroll of images and videos, and rewards such as likes and hearts.⁴⁴¹

214. Instagram had heavy users, whom it referred to as “power users” or “whales;” these were the “users with the most activity”⁴⁴² on Instagram, who would likely be prone to addiction.

⁴³⁷ Jin Dep., pp. 433-435; Jin Ex. 38 - META3047MDL-003-00166569.

⁴³⁸ Jayakumar Dep., p. 138 (emphasis mine).

⁴³⁹ Volichenko Dep., 92:7-95:24

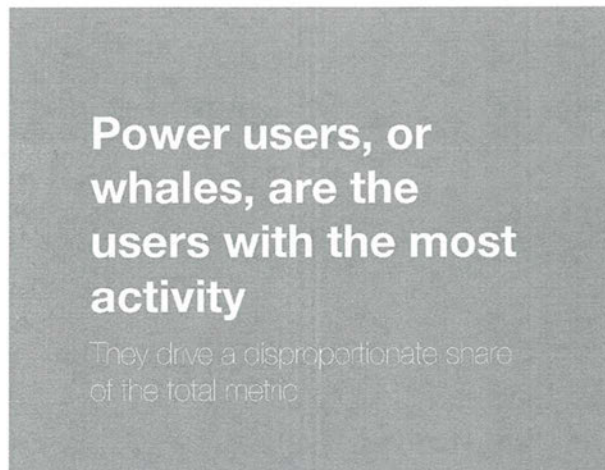
⁴⁴⁰ Bhutada Ex. 31 at 19, META3047MDL-044-00035618 at slide 17 (emphasis mine).

⁴⁴¹ Bejar Dep., pp. 139-40

⁴⁴² Kilstein Ex. 66, META3047MDL-047-00913004

In a presentation dated June of 2019, ██████████ reported that “whales,” who generally represent about five percent of Daily Active People (DAP), drive a “disproportionate share of the total metric.”⁴⁴³

DEFINITION



In this presentation, I use the top 5% of users, for the relevant metric and segment, to classify a user as a power user.

HIGHLY CONFIDENTIAL (COMPETITOR)
HIGHLY CONFIDENTIAL – ATTORNEYS' EYES ONLY

META3047MDL-047-00913004
METATNAG-019-01409913

215. ██████████ noted that many of Instagram’s “top line metrics,”⁴⁴⁴ which are the metrics that the engagement teams focus on most,⁴⁴⁵ are driven by whales or heavy users. For example, the top five percent of DAP drives 70 percent of “contribution to value,”⁴⁴⁶ which Kilstein defined as “the monetization of users” or how Instagram makes money from users.⁴⁴⁷ This metric is important because one of Instagram’s goals in June of 2019 was to “close the

⁴⁴³ Kilstein Dep. Ex. 66, META3047MDL-047-00913004.

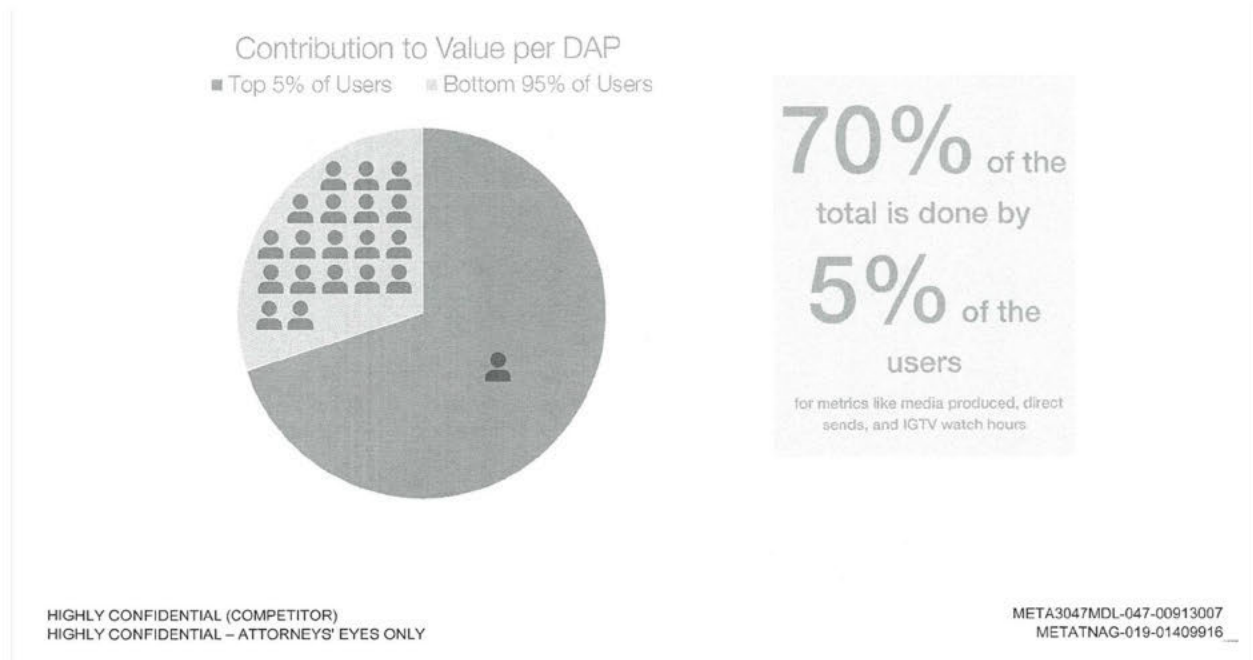
⁴⁴⁴ Kilstein Dep. Ex. 66, META3047MDL-047-00912999.

⁴⁴⁵ Kilstein Dep., p. 501.

⁴⁴⁶ Kilstein Dep. Ex. 66, META3047MDL-047-00913007

⁴⁴⁷ Kilstein Dep., p. 506.

monetization gap between Facebook and Instagram”⁴⁴⁸ or to make Instagram as financially viable as Facebook.



216. Notably, U.S. teens are power users or whales for important metrics such as overall time spent on Instagram.⁴⁴⁹ [REDACTED] head of analytics for Instagram, acknowledged that “we’ve known for years that the majority of top-line metrics movement we see (for teens and the overall population) is generated by a small subset of power users.”⁴⁵⁰ In fact, “70 percent of the total value is done 3 by 5 percent of users for metrics like media produced, direct sends, and IGTV watch hours.”⁴⁵¹ Specifically, from his own analyses, Data Scientist Darius Kilstein had information that 1,869,603 U.S. teens were heavy users or whales and were using Instagram 14 or

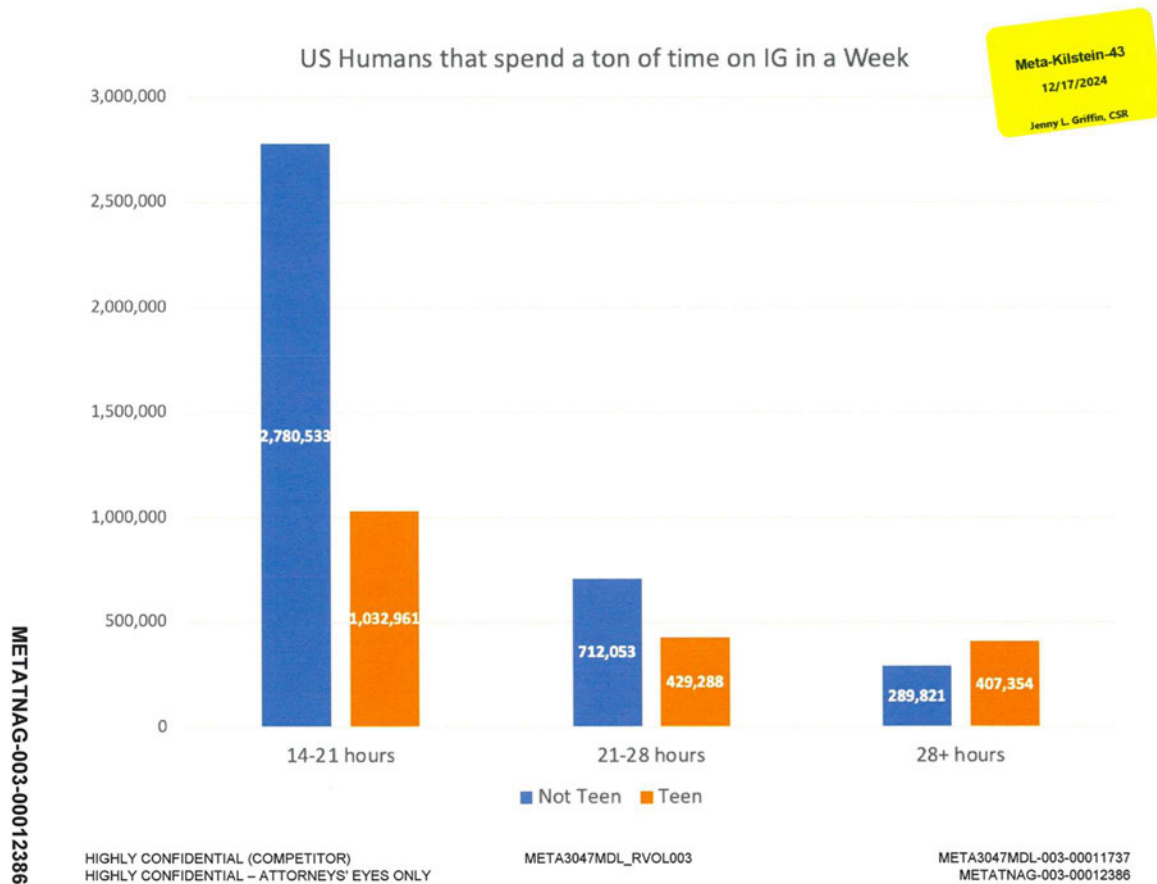
⁴⁴⁸ Kilstein Dep. Ex. 66, META3047MDL-047-00913005.

⁴⁴⁹ Kilstein Dep. Ex. 66, META3047MDL-047-00913012.

⁴⁵⁰ Kilstein Dep. Ex. 65, META3047MDL-040-00300434.

⁴⁵¹ Kilstein Dep. Ex. 66, p. 11, META3047MDL-047-00913007.

more hours a week as shown in the graph that he created:⁴⁵²



However, Kilstein did not address the problem of almost two million teens spending 14 or more hours a week on Instagram other than posting this chart on Workplace (Instagram’s Message Board) and mentioning it to his manager.⁴⁵³

217. Meta employees had information that youth were especially susceptible to social media addiction. The “Social Dilemma,” the Netflix documentary released in 2020, highlighted the dangers of social media addiction for them. For example, Kilstein noted to colleagues that the “Social Dilemma” says “*we’re creating a world of addicted monsters . . . A lot of it is probably*

⁴⁵² Kilstein Dep. Ex. 43, META3047MDL-003-00011737.

⁴⁵³ Kilstein Dep., p. 541.

true. Some is totally exaggerated. . . Worth watching. . .”⁴⁵⁴ In a message to Kilstein, Data Scientist [REDACTED] commented that recent media coverage of social media addiction made her “*feel like tech plays to humans’ inability to have self control lol* [laughing out loud].”⁴⁵⁵ Kilstein responded that he agreed that social media was playing to humans’ inability to have self-control and that he was worried most about young people and their developing brains:

Yeh that's exactly what the doco [documentary, “The Social Dilemma”] says. I think its true tbh [to be honest]. *Personally I think its most worrying for young people*. Like, as an adult, if you've got nothing better to do anyway, there's no major downside to spending time on social media *but I do worry what it does to young people who are still developing their brains and social skills, as well as being more susceptible [sic] to mean comments or lack of friends/feedback.*⁴⁵⁶

218. Also in 2020, Instagram employee [REDACTED] said that she did not think that there was a nefarious plot to addict people to Instagram, but that capitalism and its focus on endless growth fuels problems with addiction, along with “top down directives” from Meta’s leaders to make sure people keep coming back for more:

I don’t think there’s a nefarious plot to addict people but *I think that the nature of capitalism and endless growth sort of fuels it, regardless of intent*. . . . I think all of us (employees) work tirelessly to make our products a positive experience for people but the *top down directives drive it all towards making sure that people keep coming back for more.*⁴⁵⁷

219. However, by February 2021, the issue of social media addiction had faded from Meta’s agenda. For example, when Jayakumar asked a colleague on the safety policy team, [REDACTED], if “social media addiction and time spent online is still a companywide concern,”⁴⁵⁸ [REDACTED] said that it was not:

⁴⁵⁴ Kilstein Dep. Ex. 38, META3047MDL-003-00011718 (emphasis mine).

⁴⁵⁵ Kilstein Dep. Ex. 45, META3047MDL-003-00011760 (emphasis mine).

⁴⁵⁶ Kilstein Dep. Ex. 45, META3047MDL-003-00011760 (emphasis mine).

⁴⁵⁷ Jimenez Dep. Ex. 17, META3047MDL-020-00342153 (emphasis mine).

⁴⁵⁸ Jayakumar Dep. Ex. 12, META3047MDL-020-00277373.

Honestly, no. Given pandemic ***I don't think people are talking about time spent online anymore.*** Addiction too less so than last year.⁴⁵⁹

220. Jayakumar attributed Instagram's decreased concern about social media addiction to decreased external scrutiny:

We weren't getting as much external scrutiny around, you know, time spent or social media addiction, and as a result, internally there wasn't as much conversation or prioritization of those projects.⁴⁶⁰

She noted that Instagram was much more "sensitive to external critique" than other tech companies where she had worked:

Meta was my third tech company after Google and Twitter. And I think more so than my, you know, other tech experiences, ***Meta was much more sensitive to external critique*** and what – you know, ***the external conversation was around their products, and it really impacted how the company leaders prioritized internal projects.***⁴⁶¹

221. Burke testified that she also does not know whether Meta currently measures how many users experience problematic use on its platforms.⁴⁶²

222. Similarly, Bejar stated that while addiction and problematic use were problems for kids in 2021, Meta was not monitoring the issue or doing anything effective to prevent or substantially reduce addiction or problematic use.⁴⁶³ In November of 2021, Jayakumar acknowledged that social media addiction was still a significant problem and that Meta had not seized the opportunity to harvest a lot of "low hanging fruit":

Social media addiction is such a problem, and I optimistically think ***we have a lot of low-hanging fruit we can pick here.***⁴⁶⁴

⁴⁵⁹ *Id.* (emphasis mine).

⁴⁶⁰ Jayakumar Dep., p. 125 (emphasis mine).

⁴⁶¹ Jayakumar Dep., p. 126 (emphasis mine).

⁴⁶² Burke Dep., p. 174

⁴⁶³ Bejar Dep., pp. 433-34

⁴⁶⁴ Jayakumar Dep. Ex. 14, META3047MDL-040-00541333 (emphasis mine).

223. Meta’s own expert advisors, its Youth Advisors, informed Instagram employees of the addictive qualities of social media platforms. For example, in September of 2021, Pattie Gonsalves, one of Meta’s Youth Advisors, sent Jayakumar an op-ed that she wrote for a publication in India that warned of the harmful, addictive qualities of social media:

*By design, many social media platforms can be addictive, and their use has been associated with an increased risk of anxiety, depression, bullying, fostering negative body image, a fear of missing out. People share stories or post content on these with the hope of receiving positive feedback. The platforms are designed so that content shared has potential future rewards such as “Likes” or comments by other users, resulting in a recipe for constantly checking platforms.*⁴⁶⁵

Jayakumar acknowledged that Meta could have done more to address social media addiction.

224. Moreover, much of what Meta did do—using notifications and recommendation algorithms, delaying and weakening private-by-default,⁴⁶⁶ delaying the introduction of Quiet Mode⁴⁶⁷—contributed to overuse and addiction.

225. Despite internal and external knowledge of social media addiction, Jayakumar acknowledged that Meta never warned parents, youth, or schools that Instagram could be addictive during her time at the company,⁴⁶⁸ which ended in May 2023.⁴⁶⁹

226. Similarly, Bejar acknowledged that Meta did not disclose to the public or parents the risk of addiction or problematic use.⁴⁷⁰ Bejar also testified that Instagram did not do enough to prevent or substantially reduce addiction or problematic use for kids on the platform.⁴⁷¹

⁴⁶⁵ Jayakumar Dep. Ex. 41, META3047MDL-020-00294288 (emphasis mine).

⁴⁶⁶ Private-by-default was a feature introduced to make accounts for users under 18 private as the default setting to reduce unwanted attention and interactions.

⁴⁶⁷ Quiet mode was a feature introduced to enable users to turn off notifications at night so that they do not interrupt youth sleep.

⁴⁶⁸ Jayakumar Dep., p. 136.

⁴⁶⁹ Jayakumar Dep., p. 20.

⁴⁷⁰ Bejar Dep., pp. 144, 436.

⁴⁷¹ *Id.*

227. Meta’s actions fell below its own and industry standards of care in failing to address, disclose and warn about addiction. Meta’s leadership told the public that its products were not addictive. Meta made a public pledge to conduct research and to share its findings with the public. Meta only performed limited research on addiction, however, and euphemistically redefined it as problematic use. Meta only shared limited findings of its research, and then represented that the “upper bounds” of those who experience problematic use on its platforms was merely 3.1 percent. But internally, Meta described that as “severe” and “extreme” cases of problematic use and had information that over half of users experienced this problem in some form. Later research showed that the rate of problematic use was much higher, but Meta never disclosed that to the public, and there is no indication that Meta has even continued to measure users’ addiction to its platforms. Meta has failed to disclose its internal data showing the high risk of addiction and has thereby misrepresented the risk of addiction. In doing so, Meta has failed to meet standards of care, and practiced pseudo corporate social responsibility.

C. Meta Exacerbated and Failed to Disclose the Risk of Negative Social Comparison Among Teens

228. Negative social comparison “refers to the way that people feel about themselves when they compare their lives against the lives of other people around them.”⁴⁷² From a safety and public policy perspective, it represents another form of problematic use of Instagram and potential harm for users.⁴⁷³ Instagram’s safety and wellbeing and public policy employees possessed information indicating that negative social comparison was a problem for Instagram’s young users, especially girls: “[g]irls as young as 8 start to lose self-confidence because of

⁴⁷² Jayakumar Dep., p. 154.

⁴⁷³ Jayakumar Dep., p. 155.

negative self-image and social comparison. It's when they become aware of their bodies versus others.”⁴⁷⁴

229. A battery of Instagram studies have revealed that negative social comparison was indeed a particular problem on Instagram. For example, one study found that “51% of US teens feel there's pressure to be perfect on Instagram, significantly more than other social media apps” and that “[t]eens feel IG makes the problem worse.”⁴⁷⁵ Another study reported that 66 percent of teen girls on Instagram had experienced negative social comparison.⁴⁷⁶ A 2019 Instagram study had revealed that of the users who had experienced “a hard life moment” during the past 30 days, many said that the feeling to have to create “the perfect image, feeling not attractive and other negative feelings started on IG [Instagram].”⁴⁷⁷ In 2021, the BEEF survey results indicated that 21.4 percent of 13–15-year-olds and 19.5 percent of 15-17-year-olds had experienced negative social comparison on Instagram during the past seven days.⁴⁷⁸ As discussed above, these survey results were not disclosed to the public, but they leave no doubt that negative social comparison was a problematic issue that Instagram needed to address.

230. Meta's commitment to “design and build products that *prioritize safety* . . . and *include the highest protections for youth, such as age-appropriate features for youth audiences*”⁴⁷⁹ obligates the company to try to mitigate social comparison. Meta's platforms such as Instagram must also “*anticipate and mitigate potential harms*”⁴⁸⁰ resulting from social

⁴⁷⁴ Jayakumar Dep. Ex. 17, META3047MDL-040-00541685, a statement by Public Policy Employee Palak Sheth in a chat with Safety and Wellbeing Lead Vaishnavi Jayakumar.

⁴⁷⁵ META3047MDL-003-00156888 at 16.

⁴⁷⁶ META3047MDL-003-00156888.

⁴⁷⁷ META3047MDL-003-00000142.

⁴⁷⁸ Bejar Dep. Ex. 21, p. 21.

⁴⁷⁹ <https://about.meta.com/code-of-conduct/>, p. 28 (emphasis mine).

⁴⁸⁰ <https://about.meta.com/code-of-conduct/>, p. 28 (emphasis mine).

comparison and “[p]rovide appropriate warnings where necessary and articulate instructions for safe and responsible use of our products.”⁴⁸¹ Numerous other codes and standards of care reinforce Meta’s responsibilities. But, as I will be elaborate below, Meta violated these codes and failed to fulfill standards of care through a variety of actions: a) failing to remove or warn of the risk of beauty filters, b) failing to make “likes” invisible as a default setting, and c) failing to disclose the significant risk of eating disorders associated with Meta’s platforms and features.

1. Meta Failed to Remove or Disclose the Risk of Beauty Filters

231. Meta had information showing that negative social comparison is prompted and exacerbated by beauty filters, which are augmented reality effects that enable users to enhance their appearance often in the way that plastic surgery would (e.g., a different shaped nose).

Jayakumar defined beauty filters in the following way:

[Beauty filters are] *augmented reality effects, special effects, that you can apply to your image or your video that will alter your appearance in some way* . . . they’ll enhance your looks . . . different types of lips, eyes, makeup, teeth, and so on.⁴⁸²

232. In 2019, Instagram reached out to 18 experts on social comparison and beauty filters including psychologists, researchers, body image activists, and AR [Augmented Reality] professionals.⁴⁸³ They made sure to focus on “experts who could provide research insights and findings about the impact of social media and internet technologies on youth identity development, body image, including Body Dysmorphic Disorder (colloquially known as ‘Snapchat dysmorphia’), and mental health.”⁴⁸⁴ The experts concluded that beauty filters can be

⁴⁸¹ <https://about.meta.com/code-of-conduct/>, p. 28 (emphasis mine).

⁴⁸² Jayakumar Dep., p. 156 (emphasis mine).

⁴⁸³ Jayakumar Dep. Ex. 16, META3047MDL-040-00337136.

⁴⁸⁴ Jayakumar Dep. Ex. 16, META3047MDL-040-00337136.

severely damaging to both users and viewers especially if the users and viewers are children or otherwise vulnerable in some way:

- These extreme beauty effects can have severe impacts on both the individuals using the effects and those viewing the images.
- Children are particularly vulnerable, however many others are vulnerable as well: those with a history of mental health challenges, eating disorders, etc.⁴⁸⁵

233. A summary of a message sent to Andrew Bosworth, head of Instagram’s virtual reality team, confirmed that a “large majority” of the experts, “including individuals in the AR field, recommended prohibiting these filters, citing known impacts to body image and mental health of other forms of media that idealize unrealistic beauty standards.”⁴⁸⁶ The message also confirmed that “corrective actions that were less prohibitive” would not work because of the broad age range of people who were vulnerable and because research had demonstrated that labeling the effects as “photoshop” did not protect users from the negative effects:⁴⁸⁷

We discussed corrective actions that were less prohibitive, including limiting cosmetic surgery filters to adults and/or utilizing watermarks, however *these actions were not viewed as sufficient protections because (1) the universe of users who are vulnerable to negative impacts from such filters is much broader than those ages 13-17, and (2) past research on labeling photoshopped imagery found that they do not insulate viewers from its negative effects.* There was agreement in our conversations that we should be careful to limit cosmetic filters even when used in conjunction with a costume, animal features, fantasy effects, and accessories especially because these are often most popular with our youngest users.⁴⁸⁸

234. However, despite the strong warnings and recommendations of the experts, Jayakumar “received word that Boz [Andrew Bosworth] had overturned the initial decision” to remove the beauty filters.⁴⁸⁹ In explaining why Bosworth made the decision that he did,

⁴⁸⁵ Jayakumar Dep. Ex. 16, META3047MDL-040-00337135.

⁴⁸⁶ Jayakumar Dep. Ex. 16, META3047MDL-040-00337136.

⁴⁸⁷ Jayakumar Dep. Ex. 16, META3047MDL-040-00337136.

⁴⁸⁸ Jayakumar Dep. Ex. 16, META3047MDL-040-00337135 (emphasis mine).

⁴⁸⁹ Jayakumar Dep., p. 162.

Jayakumar referred to the popularity of the beauty filters and the competitive loss that Instagram would suffer vis a vis its competitors if it removed the beauty filters:

. . . these AR filters are incredibly popular. . . *They are also really popular on other platforms like Snap and TikTok . . . And I think banning them would have felt like a competitive loss . . .*⁴⁹⁰

235. Instagram went against the advice of experts and reversed the decision to remove beauty filters and, as such, failed to design for safety. Meta also failed to disclose the information it had received regarding the dangers of these filters and failed to proactively warn parents, youth, schools, and the community at large of the dangers of beauty filters. Meta’s actions were not consistent with its own and industry standards for corporate responsibility.

2. Meta Failed to Make “Likes” Invisible as a Default Setting

236. A “like” is a product feature that provides “a reaction that someone can make on the app to a piece of content that suggests that they saw, acknowledged, or liked this content in some way.”⁴⁹¹ It is a way to show appreciation and a measure of how engaged other users are on the platform. Researchers have determined that it is “numeric evaluative feedback” that is a form of social reward for adolescents.⁴⁹²

237. Jayakumar reported that Instagram had access to “a really compelling package of interviews and research” indicating that “making ‘likes’ invisible would be one way that Instagram could mitigate problems related to negative social comparison.”⁴⁹³

⁴⁹⁰ Jayakumar Dep., p. 165 (emphasis mine).

⁴⁹¹ Jayakumar Dep., pp. 165-166.

⁴⁹² Sanchez-Hernandez, M. Dolores, M. Carmen Herrera, and Francisca Exposito, “Does the Number of Likes Affect Adolescents’ Emotions? The Moderating Role of Social Comparison and Feedback Seeking on Instagram, *Journal of Psychology*, 2022, <https://doi.org/10.1080/00223980.2021.2024120>.

⁴⁹³ Jayakumar Dep., p. 166.

238. In 2019, Instagram Head Adam Mosseri announced that Instagram would release “a safety feature to combat visible ‘likes’ contributing to negative social comparison.”⁴⁹⁴ In an initiative referred to as Project Daisy, Instagram had tested “turning off the ‘like’ count in general for all users.”⁴⁹⁵ However, what Instagram actually launched was an “opt-in” option for users to go into the settings tab to turn off “likes” themselves. Jayakumar explained that “the version of Project Daisy that launched did not address the core issue of negative social comparison that the original version of the product was trying to target.”⁴⁹⁶ It was largely ineffective because “people didn’t really go in and adjust those settings in any significant quantities. And so . . . the goal of that feature, which was to reduce social comparison, didn’t really get achieved.”⁴⁹⁷

239. Note that when Project Daisy was opt-in, only 0.72 percent of users chose to opt-in and hide “like” counts as compared to 35 percent of users who left “like” counts hidden when Project Daisy was opt-out or a default setting.⁴⁹⁸ As explained by Bejar, the safety tools implemented by Meta were for the press and regulators, but were not effective.⁴⁹⁹ If a safety feature is opt-in, almost nobody will use it.⁵⁰⁰

240. Meta failed to take the step that would have been most protective of youth safety, and as such, failed to meet its obligations to design its products for safety. Meta failed to meet standards of care of numerous professional organizations regarding designing for safety. Meta also failed to disclose and warn youth users and their parents of the significant risks associated with the like feature.

⁴⁹⁴ Jayakumar Dep., p. 167.

⁴⁹⁵ Jayakumar Dep., pp. 167-168.

⁴⁹⁶ Jayakumar Dep., p. 168.

⁴⁹⁷ Jayakumar Dep., p. 169.

⁴⁹⁸ META3047MDL-004-00003255.

⁴⁹⁹ Bejar Dep., p. 166.

⁵⁰⁰ *Id.*, p. 167.

3. Meta Failed to Disclose and Warn of the Risk of Eating Disorders that Users of Its Platforms Could Experience

241. Contrary to public statements that it does not allow the promotion of eating disorders on its platform, Instagram actually *recommended and aggregated* content that promoted eating disorders among youth and never devoted the resources needed to address this problem. To be clear, the issue here is not the content itself, or that Instagram recommended and aggregated it; instead, the issue is that by telling users, parents and the public that it does not allow promotion of eating disorders on its platform—when in fact Instagram itself is doing just that—Meta failed to disclose its actions and created the highly misleading and dangerous perception that its platform is far safer than it actually is. Meta’s actions related to this issue are effectively the opposite of a warning, in that Meta created a false sense of safety and security about its platform.

242. As explained above, Meta’s Community Standards claim that it does not allow the promotion of eating disorders. Bejar testified that Meta’s Transparency Center makes it seem that there is no way that Meta is going to recommend eating disorder content to a kid, but that is not the case based on his experience.⁵⁰¹ As noted above, the NES study revealed that 18 percent of teens had seen someone promoting eating disorders on Instagram during the past seven days.⁵⁰² Bejar described these recommendations that are profoundly distressing to include recipes, lose weight fads, and body image issues.⁵⁰³ The Instagram wellbeing team had information about these user experiences and that they affected body image issues.⁵⁰⁴ “Thinspiration” was the label Instagram used for content that encourages being skinny and causes body image issues.⁵⁰⁵

⁵⁰¹ Bejar Dep., p. 231.

⁵⁰² Bejar Dep. Ex. 9, META3047MDL-059-00000222.

⁵⁰³ Bejar Dep., p. 178.

⁵⁰⁴ *Id.*, p. 268.

⁵⁰⁵ *Id.*, p. 267.

243. Meta also had internal research showing that Instagram frequently had content that promoted or encouraged body image issues and eating disorders.⁵⁰⁶ But Bejar reported that Meta did not disclose or warn parents or the public about issues relating to eating disorders and absolutely should have.⁵⁰⁷

244. Not only did Meta fail to disclose the frequency of this problem, but even worse, it actually was recommending it to young users. Algorithms, referred to as Instagram’s “recommendation surfaces”—e.g., Explore and Reels—provide users with aggregated content, much of which is “unconnected content” or content posted by people whom users do not know.⁵⁰⁸ The algorithms recommend such content based on “signals” or pieces of information about users’ interests that Instagram collects.⁵⁰⁹ Explore is a tab that provides aggregated images and video,⁵¹⁰ while Reels is a tab that provides short-form videos in a chained sequence so that the user can keep scrolling and get video after video.⁵¹¹

245. The algorithms that determine what to recommend to users based on their interests also supposedly determine what content to filter out because it violates community standards. However, problems occur when content is aggregated. Jayakumar explained that content that might not be a problem when a single image or piece of content is presented could create a problem for young users “when you’ve got dozens of these images or pieces of content in one

⁵⁰⁶ *Id.*, p. 270.

⁵⁰⁷ *Id.*, p. 271.

⁵⁰⁸ Jayakumar Dep., p. 179.

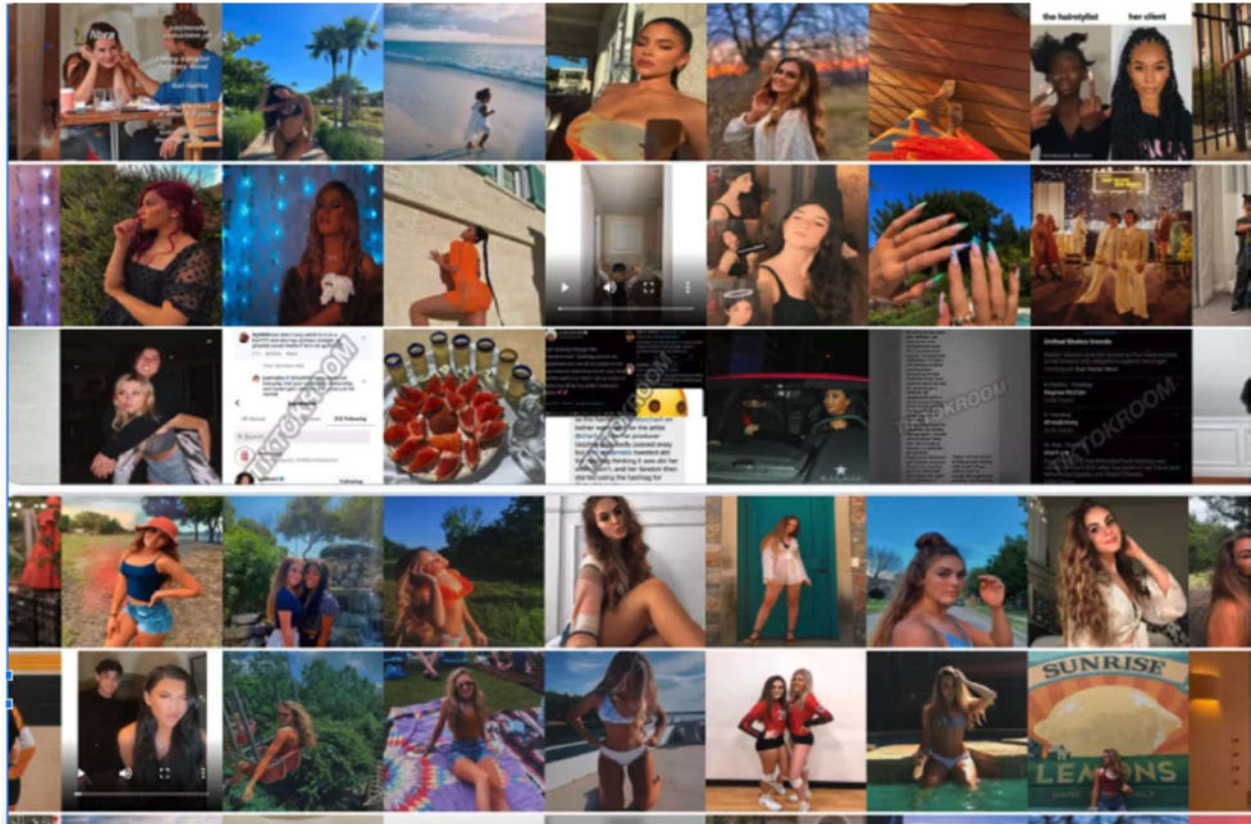
⁵⁰⁹ Jayakumar Dep., p. 172.

⁵¹⁰ Jayakumar Dep., pp. 169-170.

⁵¹¹ Jayakumar Dep., p. 171.

page on the Explore tab—that can be really overwhelming.”⁵¹² An example would be a teenage girl getting an entire page of diet tips on Explore.⁵¹³

246. See, as another example, the screenshot below of the Explore tab with many images of voluptuous and alluring teens, which was circulated in April 2021 by [REDACTED], an Instagram employee:⁵¹⁴



247. [REDACTED], an Instagram public policy employee, expressed concern about the problematic nature of such images on young users and pinned responsibility or “ownership” of the problem on Instagram since the images were “unconnected” content that Instagram had recommended to young users:

⁵¹² Jayakumar Dep., pp. 187-188.

⁵¹³ Jayakumar Dep., p. 191.

⁵¹⁴ Jayakumar Dep. Ex. 18, META3047MDL-145-00000003.

The images that █████ sent yesterday of an explore feed is really problematic, and *the ownership of that is also on us since it's an unconnected, recommended surface.*⁵¹⁵

248. Jayakumar elaborated on why the ownership of the problem and the responsibility for it were on Instagram rather than on young users:

. . . these are not accounts that . . . young people have chosen to follow. If they had chosen to follow them, you could argue that, you know, they made that choice to see that content. But in these cases, *we are recommending this content that's from people that you don't follow, and, therefore, the responsibility of making sure those are responsible recommendations falls to us.*⁵¹⁶

249. Jayakumar explained that “the difficulty with moderating . . . surfaces with aggregated content is that any one of those pieces of content is not going to violate any policy. It’s not even necessarily going to be *borderline violative* of any policies.”⁵¹⁷ However, in the aggregate such content can make “teens feel bad about themselves” because it prompts negative social comparison.⁵¹⁸

250. According to Jayakumar, one possible fix “was to expand the borderline policy to include more types of content,” but that was not done because “if you expand the borderline policy to include more types of content, then you could make an impact on recommendation surfaces.”⁵¹⁹ Presumably, such an impact on growth-related metrics would be negative in that it would restrict the unconnected content that could be recommended in the aggregate.

251. Moreover, Instagram never devoted the technical resources to fix this problem despite support for the changes among people working on its products as Jayakumar noted:

⁵¹⁵ Jayakumar Dep. Ex. 17, META3047MDL-040-00541685 (emphasis mine).

⁵¹⁶ Jayakumar Dep., p. 187 (emphasis mine).

⁵¹⁷ Jayakumar Dep., pp. 187-188 (emphasis mine).

⁵¹⁸ Jayakumar Dep. Ex. 18, META3047MDL-145-00000001.

⁵¹⁹ Jayakumar Dep., p. 198.

And we had very supportive product partners. *But due to just overall under-resourcing of this area, we certainly couldn't move as fast or as extensively as we would have liked,* given the solutions that were available to us.⁵²⁰

252. Instagram failed to devote the technical resources to fix the problematic recommendation surfaces, and they failed warn parents, youth, or schools of the problems with aggregated content on the recommendation surfaces on the platform.⁵²¹ Notably, in April 2021, Instagram Public Policy Employee [REDACTED] asserted that Instagram's failures to meet standards of care related to the aggregated content on its recommendation surfaces raised legal deceptiveness issues in addition to public policy health concerns:

. . . there are really meaty legal issues separate and apart of the policy and public health concerns. *Largely by way of deceptiveness* (which I think is something we should start to leverage more).⁵²²

253. Meta represented to the public that it does not allow the promotion of eating disorders on its platform, creating the false impression that the platforms are safe. Meta failed to disclose to parents, youth, schools, and the community at large about the actual frequency of this problem on its platform, and failed to disclose that it was actively recommending it to teens. Meta's actions did not comply with its own and industry standards of care for corporate responsibility.

D. Meta Failed to Disclose that Its Platforms were High Risk for Child Predation

254. As explained in detail above, Meta's Community Standards claim that *Meta does "not allow content or activity that sexually exploits or endangers children."*⁵²³ Meta further

⁵²⁰ Jayakumar Dep., p. 188.

⁵²¹ Jayakumar Dep., pp. 191-192.

⁵²² Jayakumar Dep. Ex. 17, META3047MDL-040-00541686 (emphasis mine).

⁵²³ <https://transparency.meta.com/policies/community-standards/child-sexual-exploitation-abuse-nudity/> (emphasis mine).

claims to have “*zero tolerance*” for material that sexually exploits minors.⁵²⁴ Meta’s strong public proclamations suggest that child sexual exploitation, solicitation, inappropriate interactions with children, exploitative intimate imagery and sextortion, sexualization of children, child nudity, and non-sexual child abuse do not frequently occur on its platforms. Notably, Meta’s Transparency Center lists “Child Endangerment: Nudity and Physical Abuse and Sexual Exploitation” as a harm that it measures, but it provides no estimate of its prevalence on the platform.⁵²⁵ Instead, Meta indicates that it is working on estimating prevalence and will provide such information when it is able to “confirm accuracy and meaningful data.”⁵²⁶

255. Meta’s former online safety consultant, Arturo Bejar, explained that the platform design and the way it promotes engagement encourages young girls to create racy content.⁵²⁷ Instagram creates an environment that facilitates grooming, sexual endangerment, child porn, and accounts selling child porn.⁵²⁸ Bejar also explained that Instagram creates an environment in which the worst kind of behaviors on the internet are rewarded via the application’s design.⁵²⁹ According to Bejar, Meta was aware of the problem with inappropriate interactions between adults and kids, but did not take appropriate actions to prevent it, and parents did not know that their kids are going to be recommended this type of content because Meta did not disclose or warn about it.⁵³⁰

⁵²⁴ Jayakumar Dep. Ex. 27 META3047MDL-037-00266408; Jayakumar Dep. Ex. 28, p. 3 (emphasis mine).

⁵²⁵ <https://transparency.meta.com/reports/community-standards-enforcement/child-nudity-and-sexual-exploitation/facebook/>.

⁵²⁶ *Id.*

⁵²⁷ Bejar Dep., p. 225.

⁵²⁸ *Id.*, p. 227.

⁵²⁹ *Id.*

⁵³⁰ *Id.*, pp. 221, 227-28.

1. Meta Failed to Create a Policy Against Minor Sexualization

256. Minor sexualization (MS) refers to “images or videos of minors behaving in a very sexualized way”—e.g., a skimpy outfit or a provocative dance.⁵³¹ Jayakumar said that MS “wouldn’t necessarily meet the threshold of being CEI [child explicit imagery or CSAM],” but “the closest analogy is minor sexualization is the borderline policy for CEI.”⁵³² When Jayakumar did her initial safety assessment of Instagram shortly after joining the company in January 2020, she remembered “flagging minor sexualization as a significant issue.”⁵³³ However, Instagram did not have a policy regarding MS, and it had not actively trained integrity classifiers to identify MS.⁵³⁴

257. An integrity classifier is “an algorithm that is a form of machine learning that is proactively scanning for potential violations to our community guidelines or our community standards.”⁵³⁵ Note that Meta also uses algorithms to tailor posts and ads to the preferences of individual users to provide personalized, engaging experiences for them. Jayakumar provided an example of how an integrity classifier works:

...an algorithm that is -- a form of machine learning that is *proactively scanning for potential violations to our community guidelines or our community standards*. And that could include searching for --*proactively searching for, you know, child sexual abuse material. Proactively searching for content that's encouraging suicide and self-harm. Proactively searching for hate speech or bullying and harassment.*⁵³⁶

⁵³¹ Jayakumar Dep., pp. 285-286.

⁵³² Jayakumar Dep., p. 286.

⁵³³ Jayakumar Dep., p. 284.

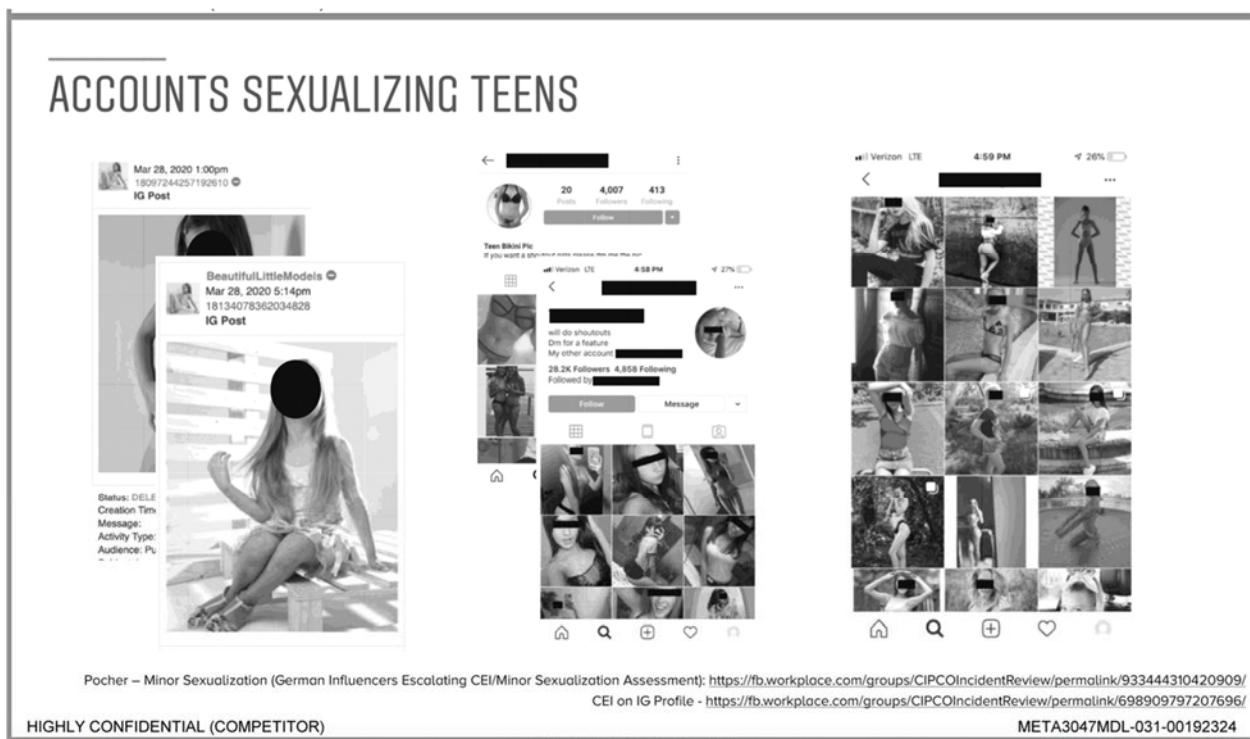
⁵³⁴ Jayakumar Dep., p. 284-285.

⁵³⁵ Jayakumar Dep., pp. 49-50.

⁵³⁶ Jayakumar Dep., pp. 49-50 (emphasis mine).

Meta claims to continuously invest in technology that enables its personnel to build sophisticated machine learning models—integrity classifiers—that can identify and flag these problems.⁵³⁷

258. Because of her concern about minor sexualization, Jayakumar was particularly alarmed when she saw the results of an Instagram test account that simulated “the seeking behavior of a bad actor”⁵³⁸ in March 2020. After only 24 hours, the test account’s Feed, Explore, and Recommended Accounts sections were “being filled almost exclusively with Minor Sexualization content/accounts.”⁵³⁹ Below is one example from a presentation:⁵⁴⁰



259. Jayakumar was particularly concerned because she was struggling to get Instagram to adopt an MS policy at the time:

⁵³⁷ <https://transparency.meta.com/enforcement/detecting-violations/how-enforcement-technology-works/>.

⁵³⁸ Jayakumar Dep. Ex. 33, META3047MDL-144-00000324.

⁵³⁹ Jayakumar Dep. Ex. 33, META3047MDL-144-00000324.

⁵⁴⁰ Jayakumar Dep. Ex. 33, META3047MDL-031-00192324.

I was also really concerned about the minor sexualization piece. *That was a policy that was in development at the time. And, you know, it was a policy that really, again, we were struggling to get across the finish line. And -- and seeing that these many pieces of content sexualizing minors had been associated with the account after just 24 hours suggested that it was a pretty significant issue for us to solve.*⁵⁴¹

260. As demonstrated in the Instagram presentation on bad actors, MS is problematic because it can enable bad actors to find minors, make inappropriate comments to them, and stalk them.⁵⁴² In November 2021, Instagram still did not have a policy on MS, so the National Center on Sexual Exploitation (NCSE) recommended that Instagram prohibit content sexualizing minors. NCSE noted that MS enables predators to identify and follow preteens,⁵⁴³ and it “is being used to signal to predators of available CSAM, trafficking victims, etc.”⁵⁴⁴

261. Instagram endangered youth not simply by failing to have a policy against MS, but also by failing to disclose to minors, parents and the public the significant risk of MS activities on Instagram, despite possessing information indicating that it was a significant issue on its platform. As such, it failed to meet standards of care for corporate responsibility.

2. Meta Endangered Youth Through Increased Interactions with Unconnected Adults

262. Meta possessed information indicating that unwanted interactions with “unconnected” adults was a problem for teens on Instagram from various sources. In 2021, the BEEF study confirmed this reality and gave an indication of the scale of the problem: 11.9 percent of the users surveyed had received unwanted sexual advances *during the past seven days*, and the unwanted advances were even more prevalent for teens: 13 percent for 13 to 15-year-olds

⁵⁴¹ Jayakumar Dep., pp. 283-284 (emphasis mine).

⁵⁴² Jayakumar Dep. Ex. 33, META3047MDL-031-00192335.

⁵⁴³ Jayakumar Dep. Ex. 35, META3047MDL-020-00651576.

⁵⁴⁴ Jayakumar Dep. Ex. 35, META3047MDL-020-00651579.

and 14.1 percent for 16 to 17-year-olds.⁵⁴⁵ The BEEF study also revealed that, remarkably, **93.8 percent** of the unwanted sexual advances were from *strangers* whom the user did not know in real life or online (i.e., unconnected adults):⁵⁴⁶

	I know them offline/ in real life	I know them, but only online	I don't know them
Negative comparison	15.5%	19.6%	64.9%
Bullying target	14.6%	16.0%	69.4%
Impersonation	9.9%	21.1%	69.0%
Drugs	9.7%	12.3%	78.0%
Self harm	9.5%	19.2%	71.3%
Political posts	8.9%	16.2%	75.0%
Bullying witness	6.4%	17.6%	76.0%
Violence	5.1%	13.7%	81.2%
Hate witness	4.6%	14.3%	81.1%
Misinfo	4.5%	9.7%	85.8%
Nudity	3.7%	9.1%	87.1%
Unwanted sexual advances	2.1%	4.2%	93.8%

263. Reels, which Instagram launched in August of 2020, endangered youth by exposing them to a “very wide range of audiences.”⁵⁴⁷ Reels is “a series of chained short-form videos that are recommended to users as they go through the Reels tab” on Instagram.⁵⁴⁸ Instagram’s Safety and Wellbeing Lead Vaishnavi Jayakumar described Reels as “Meta’s answer to TikTok,” and Instagram Senior Researcher Alison Lee implied that Reels was launched quickly as a competitive response and that Meta did not study safety until after its launch.⁵⁴⁹

264. As Jayakumar explained in a safety update to her colleagues, Reels can result in “unwanted interactions and unsafe experiences on the platform” by exposing young users to “unconnected adults” whom they do not know in real life or online:

⁵⁴⁵ Bejar Dep. Ex. 21, p. 21, META3047MDL-004-00015049.

⁵⁴⁶ Bejar Dep. Ex. 21, p. 13, META3047MDL-004-00015041.

⁵⁴⁷ Jayakumar Dep., p. 45.

⁵⁴⁸ Jayakumar Dep., p. 38.

⁵⁴⁹ Lee Dep., pp. 32, 42.

*Reels surfaces unconnected users and content to Instagram's audience, which changes the degree of virality and visibility an otherwise unknown account would get. This is great for discovering new creators, but can also come with more unwanted interactions and unsafe experiences on the platform.*⁵⁵⁰

She explained how Reels enables and incentivizes adult predators to contact and prey on teens:

But by introducing Reels, anyone, including people who never followed them, never knew that the account existed, would now know that the user existed, minor existed, and that they had these kind -- they had posted videos. So that could incentivize, for example, in the worst-case scenario, an adult predator. It could incentivize them to reach out to the minor, try to connect with them, befriend them, try, you know, send them messages, when they previously wouldn't have even known about the minor's existence on the platform.⁵⁵¹

265. Lee reported that there had been instances when Instagram delivered Reels content of young girls dancing to unconnected adults who likely included pedophiles,⁵⁵² and that Instagram research also suggested that young women were experiencing unwanted interactions because their content had been shown to people whom they did not want to see it.⁵⁵³

266. Early on, Jayakumar had recommended that minor accounts be filtered out from Reels altogether to mitigate the safety concerns, but that recommendation was rejected.⁵⁵⁴ She later substantially scaled back her recommendation and asked that the risk of Reels be studied, measured, and understood before the launch. Unfortunately, Jayakumar was unable to get this study done because of lack of resources: “[the] data science team just didn’t have the band width to handle it.”⁵⁵⁵ As such, Meta endangered youth and failed to meet standards of care related to designing for safety and protecting youth.

⁵⁵⁰ Jayakumar Dep. Ex. 3, META3047MDL-020-00271173 (emphasis mine).

⁵⁵¹ Jayakumar Dep., p. 46 (emphasis mine).

⁵⁵² Lee Dep., p. 58.

⁵⁵³ Lee Dep., p. 56.

⁵⁵⁴ Jayakumar Dep., p. 47.

⁵⁵⁵ Jayakumar Dep., p. 48.

267. In 2020, a study revealed that the number of minor victims engaging in sex talk with adults could be 38 times greater on Instagram's Direct Message than on Facebook's Messenger, a finding that was described as "significantly beyond previous expectations":

A one-time prevalence analysis indicated that inappropriate interactions with children could be *38X on IG Direct versus Facebook Messenger, which is significantly beyond previous expectations.*⁵⁵⁶

This statistic was particularly alarming because there were *seven times more minors on Instagram as on Facebook.*⁵⁵⁷

268. In 2021, the BEEF study confirmed the problems on Direct Message and revealed that 68.6 percent of unwanted sexual advances on Instagram occurred on Direct Message/Chat:⁵⁵⁸

In the last 7 days, where in Instagram did you see this? Please select all that apply.

	DM/ Chat	Feed/ Stories	Comment on my post	Explore/ Search	Hashtag/ Location Page	Reels	Someone's Profile	Somewhere else	I don't remember
fake accounts	73.1%	7.6%	10.5%	4.9%	4.2%	3.6%	14.6%	10.3%	16.0%
unwanted advances	68.6%	9.2%	7.4%	9.2%	5.5%	6.0%	16.7%	10.6%	12.9%
bully target	60.1%	17.6%	18.5%	12.1%	12.0%	10.2%	20.5%	15.2%	17.7%
impersonation 1st	26.0%	23.0%	21.9%	24.0%	21.3%	19.5%	35.4%	27.2%	35.8%
nudity	23.4%	26.4%	7.9%	37.6%	18.1%	22.7%	37.0%	12.8%	18.6%
usability action	20.8%	46.6%	11.1%	14.5%	8.9%	17.0%	16.4%	13.6%	30.6%
usability passive	20.7%	49.7%	7.5%	21.4%	7.6%	19.9%	15.6%	12.2%	29.1%
acct security	18.4%	24.0%	11.2%	15.1%	12.5%	12.6%	18.0%	16.2%	55.2%
misinfo	18.2%	37.3%	6.4%	38.3%	11.5%	17.4%	24.1%	13.3%	21.2%
bully witness	17.8%	30.6%	10.1%	24.4%	13.3%	16.7%	40.8%	17.2%	23.2%
self harm	14.9%	31.9%	12.6%	25.3%	14.4%	19.1%	29.4%	13.8%	26.9%
drugs	14.8%	34.8%	11.7%	22.5%	17.6%	16.4%	26.6%	13.5%	28.3%
perceived control	13.7%	37.9%	5.8%	37.8%	12.2%	25.9%	13.0%	9.6%	25.5%
data privacy	13.5%	26.3%	5.8%	21.7%	10.3%	11.9%	11.9%	14.3%	46.8%
violence	11.3%	34.0%	7.2%	34.5%	16.2%	20.2%	29.7%	11.2%	23.1%
transparency	10.1%	28.4%	7.5%	14.5%	10.4%	11.2%	35.6%	11.5%	34.5%
hate witness	9.5%	31.0%	9.9%	31.3%	17.5%	20.2%	39.4%	19.6%	21.9%
audience limitation	9.3%	44.4%	9.2%	13.0%	10.9%	9.9%	13.0%	7.3%	41.3%
negative comparison	8.7%	40.9%	6.4%	36.3%	12.3%	24.7%	34.8%	9.0%	22.4%
over enforcement	7.0%	35.7%	6.5%	12.1%	9.6%	11.6%	42.7%	11.2%	29.7%
commerciality	6.9%	64.7%	5.0%	20.6%	10.2%	17.1%	8.9%	7.4%	23.4%
political posts	6.2%	44.4%	6.6%	34.0%	16.3%	20.0%	23.7%	10.4%	27.8%

An expanded table with 95% confidence intervals can be found [here](#).

12

⁵⁵⁶ META3047MDL-019-00023968 (emphasis mine).

⁵⁵⁷ META3047MDL-034-00027362.

⁵⁵⁸ Bejar Dep. Ex. 21, p. 12, META3047MDL-004-00015040.

269. Instagram Safety and Wellbeing Lead Vaishnavi Jayakumar noted that she had been eager to alter Instagram to put in place the feature that Facebook already had in 2020 so that unconnected adults could not proactively contact minors on Instagram:

I was really keen on making sure that minors could not be proactively contacted by adults who had no connection to them.⁵⁵⁹

270. Likewise, [REDACTED] of Instagram's safety policy team wrote an email in which she also recommended that Meta "[b]lock reach of unconnected adults to IG [Instagram] minors."⁵⁶⁰ She noted that "grooming on IG [Instagram] occurs most likely ~2x as much as FB [Facebook]."⁵⁶¹

271. Despite the need and the obvious availability of a solution, Meta did not invest in making the necessary changes to Instagram. In explanation, there was a reference to "significant resource constraints,"⁵⁶² but [REDACTED] noted that allowing unconnected adults to contact youth was a "big growth bet" for Meta in that it would allow older adults on Facebook to reach out to their young relatives on Instagram—a rationale that she found unconvincing:

Currently, on MSGR [Facebook Messenger], we do not deliver messages initiated by unconnected adults to minors. In IGD [Instagram Direct Messages], we do not have this protection in place. For Interop [interoperability between Facebook Messenger and Instagram Direct Messages], product has committed to preserving this rule in MSGR, but not for cross-network messages sent by FB adults to IGD minors. Nor will they protect IGD minors from receiving messages Block reach of unconnected adults to IG minors from unconnected IGD adults. In light of the fact that grooming on IG occurs most likely ~2x as much as FB, their main reason not to build this protection for IGD minors is less than compelling: we do not want to block FB parents and older relatives from being able to reach out to their IGD children/relatives. They claim this is one of the main value props for Interop and a big growth bet.⁵⁶³

⁵⁵⁹ Jayakumar Dep., p. 452.

⁵⁶⁰ META3047MDL-019-00023968.

⁵⁶¹ *Id.*

⁵⁶² *Id.*

⁵⁶³ Jayakumar Dep. Ex. 45, META3047MDL-014-00206541 (emphasis mine).

272. Jayakumar confirmed that the reason Meta did not block unconnected adults on Facebook and Instagram from contacting youth on Instagram was not due to a lack of resources but because of the “big growth bet” and a “disagreement on the value of blocking unconnected adults.”⁵⁶⁴ She also noted that “there are a lot more young people using Instagram than, say, Facebook or Facebook Messenger” and that that young people have “far more discoverability on Instagram than on Facebook.”⁵⁶⁵ Alarming, as of November 2023, Meta had information that indicated there were “15.3 million high risk [Malicious Child Safety Actor] users on Messenger” who would be able to proactively contact teens on Instagram.⁵⁶⁶ According to Ravi Sinha, Meta’s safety policy manager, Meta’s policy in 2023 was extremely hard to defend.⁵⁶⁷

273. Meta placed its own growth ahead of safety for youth and failed to meet standards of care regarding designing for safety. Moreover, Meta failed to disclose these risks and warn youth, parents, schools, and the community at large.

3. Meta Failed to Address Deficiencies in its Processes for Identifying and Reporting Child Sexual Abuse Material and Failed to Disclose or Warn Users, Parents and the Public about the Risks of Exploitation of Minors on Its Platforms

274. Instagram’s failure to respond to deficiencies in its processes for identifying and reporting child sexual abuse material (CSAM) or child explicit imagery (CEI) is yet another example of pseudo corporate social responsibility.⁵⁶⁸

275. Recall that Meta’s 2020 and 2024 Community Guidelines declared its “zero tolerance” for sexual content involving minors:

⁵⁶⁴ Jayakumar Dep., p. 454.

⁵⁶⁵ Jayakumar Dep., p. 453.

⁵⁶⁶ META3047MDL-072-00046304.

⁵⁶⁷ META3047MDL-044-00067085.

⁵⁶⁸ Note that child sexual abuse material (CSAM) and child explicit imagery (CEI) are synonyms.

We have *zero tolerance when it comes to sharing sexual content involving minors* or threatening to post intimate images of others.⁵⁶⁹

And Meta documents showed the dire effects of child sex abuse on young victims—e.g., victims were 15 times more likely to develop schizophrenia or another psychotic disorder; four times more likely to abuse drugs; three times more likely to commit suicide; and more likely to suffer depression, PTSD, severe obesity diabetes, asthma, and myocardial infarction.⁵⁷⁰

276. Numerous documents between 2020 and 2024 confirmed that Instagram was falling far short of its own standards and that minors were continuing to encounter CSAM content on its platform. Yet Meta did not act to address these problems in a responsible manner and it neither disclosed to the public the extent of these problems on its platforms nor warned teens, parents, schools, or the larger community.

277. Shortly after Safety and Wellbeing Lead Vaishnavi Jayakumar joined Instagram in January 2020, she flagged problems related to “Accounts You May Know surfacing . . . teen accounts to users”⁵⁷¹ and Instagram’s lack of a reporting option for CSAM.⁵⁷² Shortly thereafter in March 2020, her concerns were heightened when [REDACTED], Instagram’s Latin American safety policy representative, wrote in a chat that she was receiving “a ton” of child abuse reports, which suggested that Instagram was not detecting CSAM with its integrity classifiers:

@silent Is anyone else receiving child abuse reports on platform? *In the last 24 hours I’ve received a ton :(I’ve seen terrible things, we’re not detecting it :(((*⁵⁷³

⁵⁶⁹ Jayakumar Dep. Exs. 27 and 28 (emphasis mine).

⁵⁷⁰ META3047MDL-046-00319359 at 18.

⁵⁷¹ Jayakumar Dep., p. 312.

⁵⁷² Jayakumar Dep., p. 276.

⁵⁷³ Jayakumar Dep. Ex. 29, META3047MDL-040-00541113 (emphasis mine).

278. Two days later, [REDACTED], head of public policy for Instagram in Europe, flagged another problem. When users tried to report a problem related to child sexual abuse content on Instagram, they received an immediate, pop-up message saying that Instagram could not review their report due to a shortage of personnel:⁵⁷⁴



We couldn't review your report.

We have fewer people available to review reports because of the coronavirus (COVID-19) pandemic, so we're only able to review content with the most potential for harm. If you don't want to see amyloves0916 on Instagram, you can unfollow, mute or block them to hide their posts and comments from your feed. Reports like yours are an important part of making Instagram a safe and welcoming place for everyone.

3m

279. As such, Jayakumar expressed concern that it appeared that Instagram's integrity classifiers—its machine learning algorithms that proactively look for content that could be violating its CSAM policy—were not working to identify CSAM. She recommended that Instagram put in place a specific reporting option for users to alert the company of CSAM. At that time, Instagram did not have a specific way for people to report CSAM⁵⁷⁵ even though it had options for reporting other types of material that violated its policies such as spam.⁵⁷⁶ In response to Jayakumar's recommendation, [REDACTED], a representative of Instagram's product policy group, expressed concern that creating the reporting system would be too much work and that users would abuse it:

⁵⁷⁴ Jayakumar Dep. Ex. 30, META3047MDL-050-00285737

⁵⁷⁵ Jayakumar Dep. Ex. 30, META3047MDL-050-00285735; Jayakumar Dep., p. 274

⁵⁷⁶ Jayakumar Dep., p. 276

I don't think updating our reporting flow options is minimal product work. And I'd be concerned that people would abuse this reporting option to report anything they want reviewed and we'd need to review it because it's been reported as CEI.⁵⁷⁷

280. Jayakumar testified that she had highlighted the need for a CSAM reporting option multiple times, but she always received “pushback” because “it would be significant product work and that people might abuse the reporting option.”⁵⁷⁸ In a chat in which her colleagues considered other options (e.g., having users report CSAM twice), Jayakumar asserted that the “best option is one where our classifier just doesn't get it wrong. . .”⁵⁷⁹

281. In June 2020, Instagram conducted its own test, which confirmed that its integrity classifiers were not identifying CSAM.⁵⁸⁰ A test account was created, and it followed about 70 accounts that were either sexualizing minors or talking about minors. Within 24 hours, the Feed, Explore, and Recommended accounts were “filled almost exclusively with Minor Sexualization content/accounts.” Most alarmingly, CSAM was recommended in the Explore section. That is, Instagram was actively recommending CSAM. The test account also had more than 20 follow requests.⁵⁸¹ At that point in time, Instagram did not have a policy on MS though Jayakumar had flagged it as a significant issue shortly after she joined Meta in January 2020.⁵⁸²

282. Instagram's deficiencies in identifying and removing CSAM were highlighted when compared to Facebook. In September 2020, Data Scientist [REDACTED] pointed out that Instagram removed 21 times fewer pieces of CSAM content than Facebook.⁵⁸³ She also reported

⁵⁷⁷ Jayakumar Dep. Ex. 30, META3047MDL-050-00285735.

⁵⁷⁸ Jayakumar Dep., p. 276.

⁵⁷⁹ Jayakumar Dep. Ex. 30, META3047MDL-050-00285735.

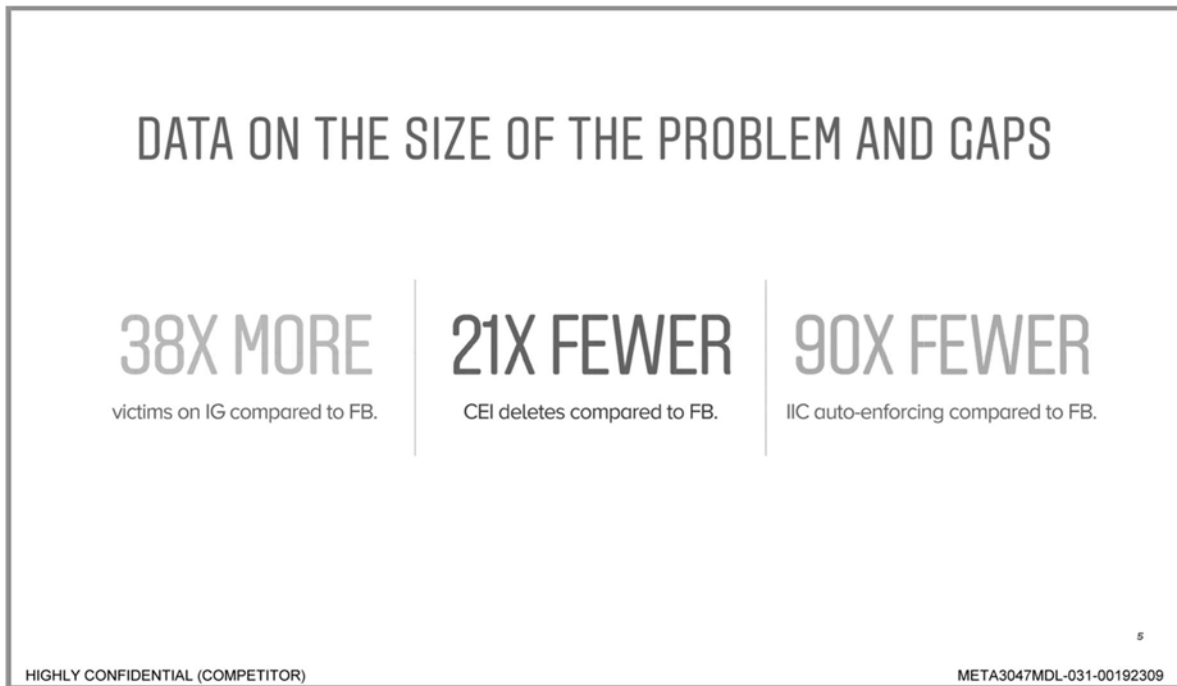
⁵⁸⁰ Jayakumar Dep. Ex. 31.

⁵⁸¹ Jayakumar Dep. Ex. 31, META3047MDL-050-00285735.

⁵⁸² Jayakumar Dep., p. 284.

⁵⁸³ Jayakumar Dep. Ex. 32, META3047MDL-047-01219676.

that Instagram was auto-enforcing “*90X fewer IIC [inappropriate interactions with children]*”⁵⁸⁴ *than Facebook*. The egregiousness of these differences is highlighted when one considers that there are *38 times more victims on Instagram than on Facebook*.⁵⁸⁵ See the summary slide below:⁵⁸⁶



283. Moreover, ██████ said that Instagram’s enforcement of its CSAM policy overall was not as good as Facebook’s and “prone to errors.”⁵⁸⁷ She attributed the discrepancy to “classifier recall”—that is Instagram’s integrity classifiers were not recognizing or capturing CSAM.⁵⁸⁸

⁵⁸⁴ Jayakumar Dep. Ex. 32, META3047MDL-047-01219676 (emphasis mine).

⁵⁸⁵ Jayakumar Dep. Ex. 33, META3047MDL-031-00192309.

⁵⁸⁶ Jayakumar Dep. Ex. 33, META3047MDL-031-00192309.

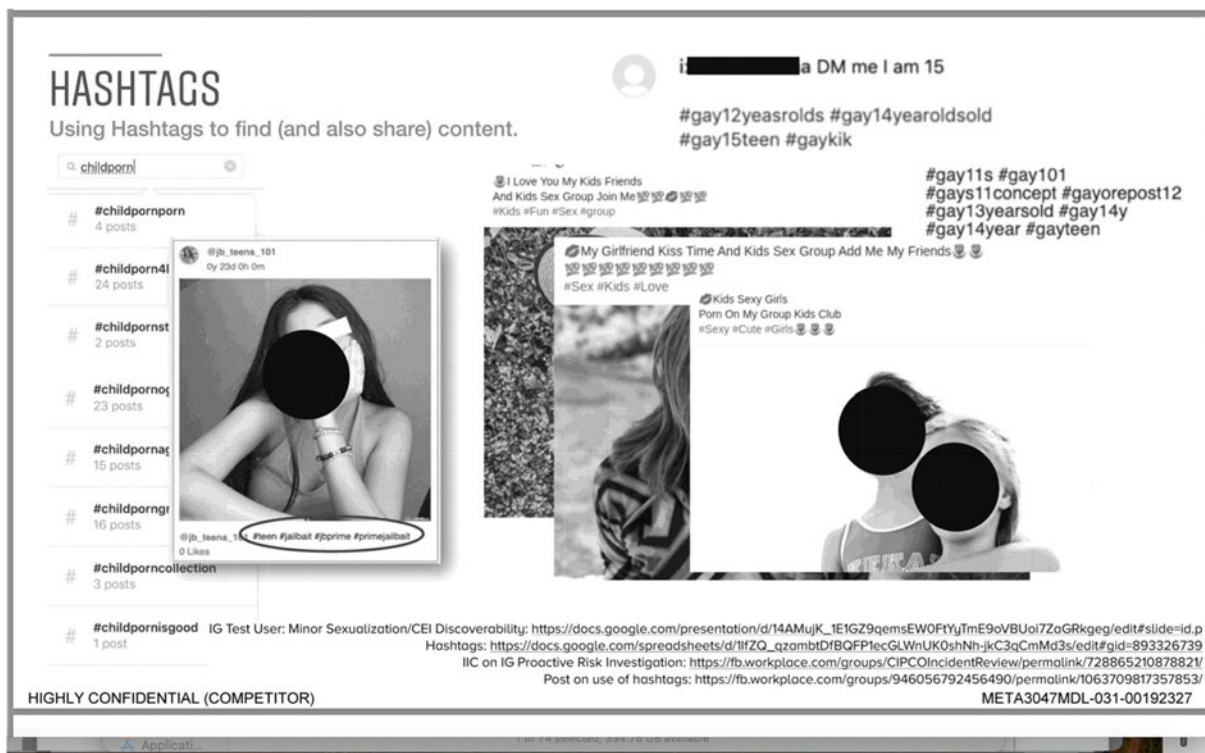
⁵⁸⁷ Jayakumar Dep. Ex. 32, META3047MDL-047-01219676.

⁵⁸⁸ *Id.*

284. Jayakumar explained that the reason Instagram's integrity classifiers were inferior to Facebook's was because they were trained on Facebook data, which was primarily text, rather than on images and videos, which dominated Instagram:

... the classifiers for child safety that ran across Instagram were not calibrated to Instagram. They were run centrally out of Facebook and they were trained on Facebook data. But Facebook data is mainly text and, you know, some images and video. Instagram is mostly images and videos. So the classifiers that ran on this content on IG [Instagram], just didn't have enough IG [Instagram] data to be able to correctly capture more IG [Instagram] content.⁵⁸⁹

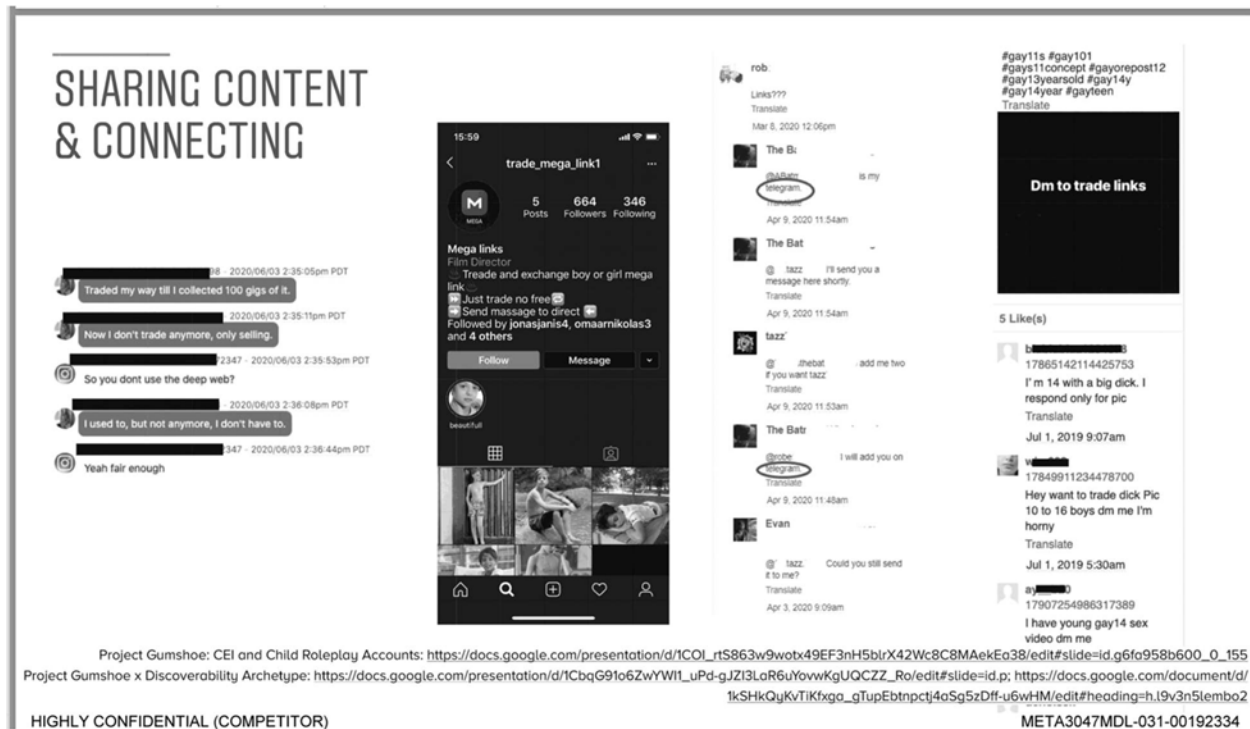
285. Another test account was created by Instagram to simulate bad actor behavior. A slide presentation documented that the test account surfaced not only sexualized images of children but also a number of hashtags related to finding and sharing child pornography⁵⁹⁰:



⁵⁸⁹ Jayakumar Dep., p. 290 (emphasis mine).

⁵⁹⁰ Jayakumar Dep. Ex. 33, META3047MDL-031-00192327.

286. The test account also produced evidence of what Jayakumar characterized as “child sexual abuse predators messaging one another, essentially trying to build community and trade images”⁵⁹¹ as documented in the slide below:⁵⁹²



287. The slide presentation documented another key deficiency of Instagram vis a vis Facebook. Facebook disrupted the ability of predators to discover minors while Instagram did not:

Facebook disrupts the ability for predators to discover minors (using People You May Know) to prevent unconnected adults from discovering or initiating conversations with minors. Instagram has no similar logic for Accounts You May Follow.⁵⁹³

288. Not only did Instagram fail to disrupt the ability of predators to discover minors, but prior to 2023, Instagram actually recommended minors to potentially suspicious adults and vice versa.⁵⁹⁴

⁵⁹¹ Jayakumar Dep., p. 308.

⁵⁹² Jayakumar Dep. Ex. 33, META3047MDL-031-00192334.

⁵⁹³ Jayakumar Dep. Ex. 33, META3047MDL-031-00192337 (emphasis mine).

⁵⁹⁴ Sinha Dep. Ex. 8, META3047MDL-073-00003522.

289. As if more documentation were needed, in November 2020, a Norwegian Broadcasting Corporation (NRK) story confirmed Instagram's deficiencies in identifying and removing CSAM or CEI (child explicit imagery). NRK found and reported 100 accounts with CEI to Instagram, but Instagram only removed 10 accounts or 1/10 in a week. NRK also reported that there was open buying and selling of CEI on Instagram—hundreds of accounts with photos of nude children, some as young as five years old:

NRK reported 100 accounts to Instagram that spread and sell CEI material. In one week, only ten of them have been removed, despite the fact that the accounts violate the company's own rules.

NRK recently revealed that there is open buying and selling of CEI on Instagram (first story published 5" Nov). Hundreds of accounts contain nude photos and highly sexualized material of children as young as 5 years old. The photos and videos are distributed, exchanged or sold to other users on the platform, despite the fact that it violates Instagram's own rules.

NRK has, via Instagram's own system, reported 100 of these accounts to the company. But in a week, only ten of them have been removed. . . ***Not only that: The accounts Instagram has not removed have in one week gained over 10,000 new followers.***⁵⁹⁵

290. NRK reported that Instagram responded quickly to their reports—apparently by automated responses—and told them that the accounts in question *did not violate company policy*:

More, not fewer photos

The accounts NRK reported were moderated quickly- so quickly that it was apparently done by machines. ***A number of auto-generated responses were returned, stating that the accounts did not violate the company's guidelines. Almost all the accounts reported were allowed to stay on the platform, despite the fact that they contain highly sexualized images, comments and icons about children as young as 5 years old.***⁵⁹⁶

291. [REDACTED], Facebook's head of communication in the Nordic region, acknowledged that Instagram had turned over too much decision power to technology:

Disregarding the obvious fix to the perception problem (introducing an artificial 1-hour delay), ***I think the reporters point to a real issue in our review system that mirrors the***

⁵⁹⁵ Jayakumar Dep. Ex. 34, META3047MDL-034-00333406 (emphasis mine).

⁵⁹⁶ Jayakumar Dep. Ex. 34, META3047MDL-034-00333406 (emphasis mine).

*sentiment we've been getting from press across several media stories: clearly we've handed over too much decision power to a technology that's not ready for that yet, and it's beginning to erode people's trust in our systems.*⁵⁹⁷

However, Jayakumar responded that Instagram's human review capacity was very low—approximately [REDACTED] of what was needed and was expected to fall to below [REDACTED] in the near future—despite the fact that Zuckerberg had announced that more full-time people would be assigned to review. The catch was that though the number of full-time workers had been increased, the number of contract workers had been reduced, so capacity would actually fall to [REDACTED] of what was needed:

The CO [Community Operations] *human review capacity for CEI is at [REDACTED]; CO Safety suspects this will continue to be true for many more months to come for large pools of queues with [REDACTED] projected supply for the near future.*⁵⁹⁸

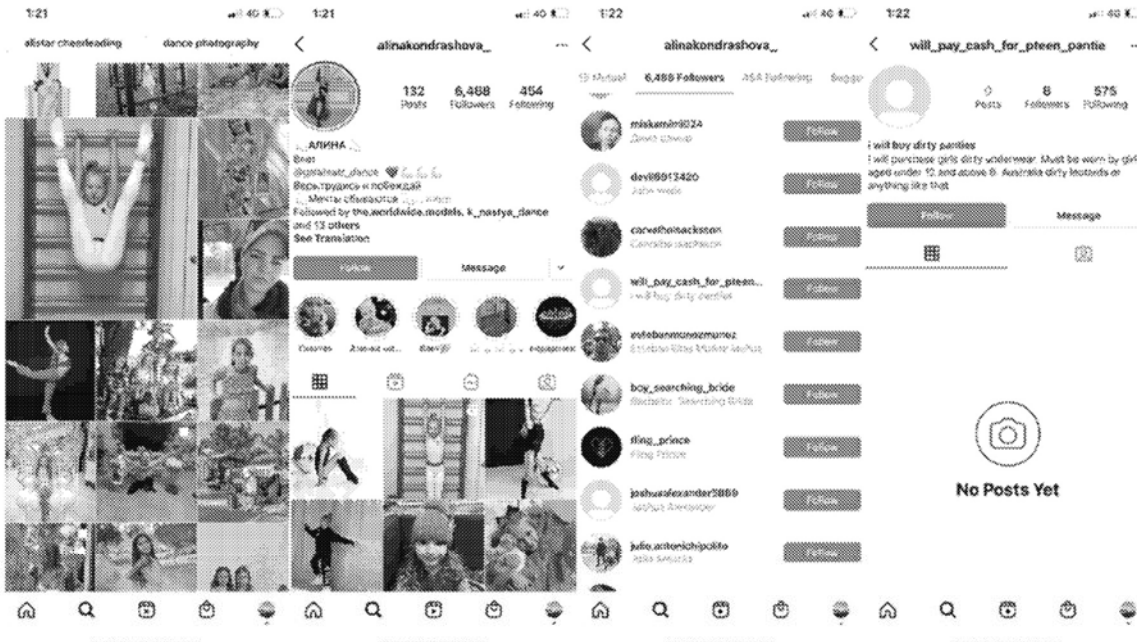
292. More evidence of Instagram's failure to meet standards of care was provided by Lena Nealon of the National Center on Sexual Exploitation (NCSE). In November of 2021, Nealon wrote a letter on behalf of NCSE in which she referred to Instagram's response to earlier criticisms as “almost lethargic considering the platform has been and continues to be cited as a top tool for a host of crimes and illegal behavior against children.”⁵⁹⁹ Among the problems documented in Nealon's letter was the fact that Instagram was still operating as a pedophile network that enabled predators to identify and contact their victims:⁶⁰⁰

⁵⁹⁷ Jayakumar Dep. Ex. 34, META3047MDL-034-00333403 (emphasis mine).

⁵⁹⁸ Jayakumar Dep. Ex. 34, META3047MDL-034-00333402 (emphasis mine).

⁵⁹⁹ Jayakumar Dep. Ex. 35, META3047MDL-020-00651573.

⁶⁰⁰ Jayakumar Dep. Ex. 35, META3047MDL-020-00651577.

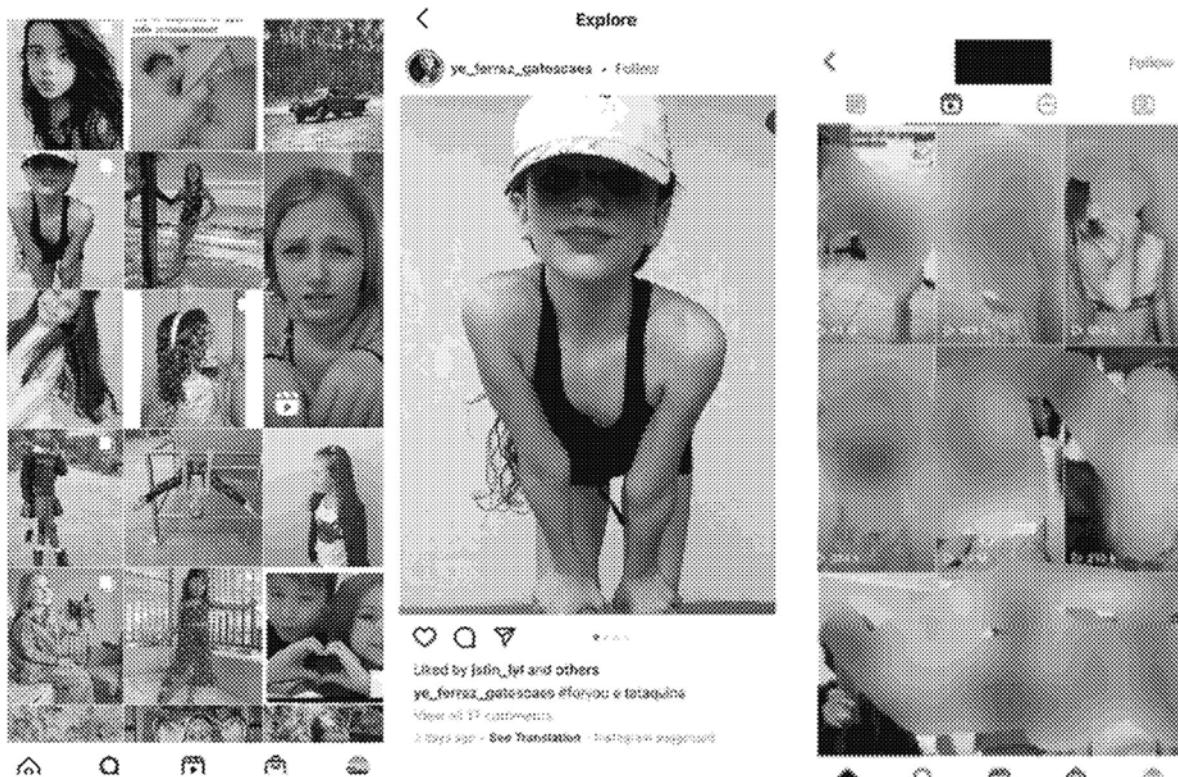


Accounts operating as pedophile networks



293. The letter also documented that “[m]inors, including preteens, are still being recommended via Explore, Suggested for You, and (particularly concerning) Reels, where preteens are essentially providing sexual entertainment for men.”⁶⁰¹

⁶⁰¹ Jayakumar Dep. Ex. 35, META3047MDL-020-00651576.



Pre-teen girls followed by/targets of approaches from predators.

294. Nealon noted that “[h]ashtags also remain a serious concern, given the ease with which they connect predators with children.”⁶⁰² She asserted that Instagram “was being used to signal to predators of available CSAM, trafficking victims, etc.”⁶⁰³ For example, the following warning notified users of CSAM but then, remarkably, enabled them to continue and click through to see the results:⁶⁰⁴

⁶⁰² Jayakumar Dep. Ex. 35, META3047MDL-020-00651578.

⁶⁰³ Jayakumar Dep. Ex. 35, META3047MDL-020-00651579.

⁶⁰⁴ Jayakumar Dep. Ex. 36, META3047MDL-020-00151907.

This May Be Associated with Child Sexual Abuse

This term is sometimes associated with sexual images or videos of children.

Child sexual abuse or viewing sexual imagery of children can lead to imprisonment and other severe personal consequences.

This abuse causes extreme harm to children and searching and viewing such materials adds to that harm.

If you see this type of activity here, see how to report it to us and to law enforcement in the link below.

[See How To Report](#)

[See results anyway](#)

295. Among the recommendations made by Nealon on behalf of NCSE were the very same recommendations that Jayakumar had been making since early 2020:

- 1) Invest in and expand use of existing technology to scan for sexually explicit images and posts and messages.
- 2) Prioritize using and/or creating tools to prevent, identify, and remove content, hashtags, and comments that sexualize children on minors' accounts, as well as on managed accounts for children 12 and under. Remove and block offenders.
- 3) Prohibit content sexualizing minors.⁶⁰⁵

296. As if this were not enough, in an article published in the *Wall Street Journal* in November 2023, Jeff Horowitz and Katherine Blunt reported that they had set up test accounts and that Instagram's system "served jarring doses of salacious content" to those test accounts,

⁶⁰⁵ Jayakumar Dep. Ex. 35, META3047MDL-020-00651575.

which included “risqué footage of children as well as overtly sexual adult videos—and ads for some of the biggest U.S. brands.”⁶⁰⁶

297. The journalists described their approach and the shocking results:

The 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. When the test accounts then followed some users who followed those same young people’s accounts, they yielded even more disturbing recommendations. *The platform served a mix of adult pornography and child-sexualizing material, such as a video of a clothed girl caressing her torso and another of a child pantomiming a sex act.*⁶⁰⁷

298. Despite the overwhelming evidence of CSAM on Instagram, Samantha Stetson, a Meta vice president, was quoted in the article, “Our systems are effective at reducing harmful content, and we’ve invested billions in safety, security and brand suitability solutions.”⁶⁰⁸

299. Repeated documentation from internal and external sources showed that Instagram’s systems were failing to identify and remove CSAM even though it touted a “zero tolerance” policy in its public-facing Community Guidelines. Instagram did not change its “zero tolerance policy” to reflect what was actually occurring on the platform. Meta also did not disclose to or warn parents, youth, or schools that its systems were grossly inadequate to enforce its purported policy and that child predation was rampant on its platforms.

300. Meta’s purported “zero tolerance” policy, coupled with its failures to disclose, is another example of Meta creating false assurances about the safety and security of its platforms, and it is another example of pseudo corporate social responsibility. Moreover, by not investing in

⁶⁰⁶ Jayakumar Dep. Ex. 38, p. 1.

⁶⁰⁷ Jayakumar Dep. Ex. 38, pp. 2-3 (emphasis mine).

⁶⁰⁸ Jayakumar Dep. Ex. 28, p. 2.

technology to enforce its purported standards, Meta did not design for safety. Meta’s actions were inconsistent with its own and industry standards for corporate responsibility.

4. Meta Failed to Disclose that It had Inadequate or Nonexistent Integrity Classifiers to Enforce its Community Standards

301. One of the key reasons that Meta had difficulty enforcing its Community Standards and fulfilling appropriate standards of care was because of gross inadequacies with its integrity classifiers. Meta’s website touts the effectiveness of its integrity classifiers and claims that “millions of violating posts and accounts” are removed automatically every day from Facebook and Instagram—“often before anyone sees it.”⁶⁰⁹

302. Meanwhile, Meta’s leaders asserted the efficacy of the company’s classifiers to Congress and the American people. For example, Meta CEO Mark Zuckerberg told Congress that Meta has invested billions of dollars in its integrity programs or integrity classifiers and that more than 1,000 engineers and 35,000 content reviewers review for problematic content on the platform:

*... we invested billions of dollars in--in our integrity programs [integrity classifiers], including having more than 1000 engineers working on this and 35,000 people doing content review across the company.*⁶¹⁰

303. As a specific example, Zuckerberg referred to Meta’s integrity classifiers dedicated to identifying and removing content related to child exploitation, which “thousands of people work on” and which he described as “generally pretty effective”:

⁶⁰⁹ <https://transparency.meta.com/enforcement/detecting-violations/technology-detects-violations/>.

⁶¹⁰ House Energy and Commerce Subcommittees on Communications and Technology and Consumer Protection Hold Joint Hearing on Misinformation and Disinformation on Online Platforms, March 25, 2021, p. 40 (emphasis mine).

In the--in the case of child exploitation content *we've been building systems for a long time that use AI and we have thousands of people working on being able to identify this content and remove it.* And I think our systems are generally pretty effective at this.⁶¹¹

304. As another example, Facebook Global Head of Safety Antigone Davis testified at a Congressional hearing that Meta had built AI or integrity classifiers to identify and rapidly respond to suicide-related content:

We've built *AI to identify suicide-related content* on our platform and rapidly respond with resources.⁶¹²

305. She also touted the ability of Meta's technology to identify suspicious adults and prevent them from having interactions with young people:

On Instagram, we prevent adults who have exhibited potentially suspicious behavior from interacting with young people's accounts. We won't show young people's accounts in Explore, Reels, or 'Accounts Suggested for You' to these adults. If they find young people's accounts by searching for their usernames, they won't be able to follow them. They also won't be able to see comments from young people on other people's posts, nor will they be able to leave comments on young people's posts.⁶¹³

306. Contrary to this representation, as noted above, Instagram actually recommended suspicious adults to minors and vice versa prior to 2023.⁶¹⁴

307. Meta also had no way to detect CSE (Child Sexual Exploitation) and CEI (Child Explicit Imagery) on Live Video: "CSE and CEI are Extremely High Severity abuses with real-world implications on Live Video. There is no detection to-date for these items, which is a major gap."⁶¹⁵ Similarly, Meta acknowledged that "CSE & CEI are the highest bucket of severity and we have no proactive detection."⁶¹⁶

⁶¹¹ *Id.*, p. 14 (emphasis mine).

⁶¹² Senate Committee on Science, Commerce, and Transportation Subcommittee on Consumer Protection, Product Safety, and Data Security, September 30, 2021, p. 8 (emphasis mine).

⁶¹³ *Id.*, p. 10 (emphasis mine).

⁶¹⁴ Sinha Dep. Ex. 8, META3047MDL-073-00003522.

⁶¹⁵ META3047MDL-050-00122454.

⁶¹⁶ *Id.*

308. Rather than warning the American public of the problems with their classifiers, Meta executives falsely reassured them. For example, Instagram Head Adam Mosseri represented to Congress that he personally believes deeply in transparency and has spent “an immense amount of time over the years” trying to be transparent about how Instagram’s algorithms work and about “how much problematic content is on the platform.”⁶¹⁷

309. Meta’s integrity classifiers, however, were far from adequate in enforcing Meta’s Community Standards. In some cases, they were even non-existent. Still, Meta executives did not disclose the inadequacies of their integrity classifiers, despite possessing information indicating that it was a significant issue on its platforms. Instead, they falsely reassured teens, parents, schools, and the American public that their platforms were safe. As such, Meta endangered youth and failed to meet the appropriate standards of care for corporate responsibility.

E. Meta Failed to Disclose that It had Ineffective Integrity Classifiers to Detect Problematic Content on Reels Short-Form Videos

310. Unbeknownst to Instagram’s teen users and their parents, Meta’s sophisticated technology was “very ineffective on the short-form video content”⁶¹⁸ of Reels. Instagram Safety and Wellbeing Lead Jayakumar explained that this was because the algorithms or models had been trained on image-based content, which dominated Instagram and Facebook at the time, not on videos.⁶¹⁹ She highlighted this fact in her initial analysis of Reels⁶²⁰ in a safety update email on July 27, 2020:

Our integrity classifiers are less effective on video and *not calibrated for short-form video in particular (i.e., all Reels content)*.⁶²¹

⁶¹⁷ Senate Committee on Science, Commerce, and Transportation Subcommittee on Consumer Protection, Product Safety, and Data Security, September 30, 2021, p. 25.

⁶¹⁸ Jayakumar Dep., p. 53.

⁶¹⁹ Jayakumar Dep., p. 49.

⁶²⁰ Jayakumar Dep., p. 49.

⁶²¹ Jayakumar Dep. Ex. 4, META3047MDL-034-00329534 (emphasis mine).

Jayakumar asked for a “mitigation”—a study of the effect of Reels on minors to understand the risk before the global launch—but she reported that she had “hit a wall with the Data Science team working on this due to limited bandwidth.”⁶²²

311. Reels was launched globally without any protections for youth or any disclosure to parents or teens about the inadequacies of its machine learning models and integrity classifiers and the risks that accompanied those deficiencies.⁶²³ Senior Researcher Alison Lee noted that the “guardrails” that Instagram put in place on Reels “were not sufficient to seriously consider the integrity impacts” and “that there were alternative strategies that they could have used.”⁶²⁴ As noted above, Reels was launched quickly as a competitive response to TikTok without studying safety issues.⁶²⁵

312. Nine months after it launched, Reels had more objectionable, supposedly non-recommendable, and violative content than any of the other Instagram spaces.⁶²⁶ For example, sexually suggestive content featuring minors was two to three times higher on Reels than on Explore.⁶²⁷

313. Lee explained that the Reels-related algorithms endangered girls by showing their content to people whom they did not intend:

Benign but still risky content featuring minors: Children dancing/posing unto itself is not problematic, but it becomes an issue if bad actors see it. *New approaches to protect children and teens (i.e., preventing visibility between known IIC adults and minors on other surfaces like Explore) should be ported to Reels as soon as possible.*⁶²⁸

⁶²² Jayakumar Dep. Ex. 4, META3047MDL-034-00329534.

⁶²³ Jayakumar Dep., pp. 56-57.

⁶²⁴ Lee Dep., pp. 51-53.

⁶²⁵ Lee Dep., p. 32, p.42.

⁶²⁶ Lee Dep. Ex. 4, p. 1; Lee Dep. Ex. 5, p. 11.

⁶²⁷ Lee Dep. Ex. 5, p. 11.

⁶²⁸ META3047MDL-074-00158945; Lee Dep., p. 108 (emphasis mine).

314. Lee recommended that the company “[i]nvest in age-specific engineering, research, and design to better protect and support teens in Reels.”⁶²⁹ However, because of *Wall Street Journal* leaks, Instagram’s communication team recommended that the company not proceed with such research on Reels safety. Lee testified that she believed that there was companywide concern that the research might make it into the public domain, and she was told not that it was not wise to proceed with research on teens and their negative experiences on Reels:

. . . given the timing and volatility of what was going on with the *Wall Street Journal*, ***they didn’t think it would be wise for us to proceed with any additional sensitive research that was on teens or their negative experiences on the platform.***⁶³⁰

315. Without integrity classifiers that were effective on the short-form videos of Reels, Meta could not live up to or enforce its Community Standards. Meta failed to disclose in its Community Standards or elsewhere that its purported standards did not conform to reality with respect to Reels⁶³¹ and that adequate safety guardrails did not exist. For example, Lee noted that to her knowledge, the information in the 2021 presentation, “Reels Integrity: Sizing the Challenges,” was not disclosed to the public.⁶³² Meta’s failure to disclose and warn users and the public about these safety risks, while touting safety in its Community Standards, once again sent a misleading and falsely reassuring message to minors and parents regarding the safety of Meta’s platform. Meta’s actions were inconsistent with the applicable standards of care for corporate responsibility.

⁶²⁹ Lee Dep. Ex. 4, p. 15.

⁶³⁰ Lee Dep., pp. 117-121 (emphasis mine).

⁶³¹ Jayakumar Dep., pp. 56-57.

⁶³² Lee Dep., p. 100.

F. Meta Failed to Disclose that It Had Ineffective Classifiers to Detect Inappropriate Interactions with Children

316. Instagram Safety and Wellbeing Lead Vaishnavi Jayakumar explained that a “set of classifiers” were supposed to be “monitoring and proactively detecting any inappropriate interactions” between adults and minors “so they could either protect the minor more or penalize the adult more, or some combination of both.”⁶³³ However, Meta’s interview-based research revealed that most of the teen girls interviewed had received what they characterized as weird, creepy, sexual, or otherwise inappropriate comments from adult men whom they did not know.⁶³⁴ Moreover, all of the teen girls interviewed said that they or one of their friends had “received inappropriate compliments or other unwanted sexualized interactions.”⁶³⁵

317. The problem was that integrity classifiers (i.e., “actor-level groomer classifiers”) were not operational on Instagram and fixing them was not a priority even though Safety Policy Team Member ██████████ characterized it as “low hanging fruit” that was “fully supported” by Instagram’s child safety and policy teams:

Product and Policy have aligned that *using the IIC [Inappropriate Interaction with Children] classifier is the best available safety mitigation for minors* in Disappearing Mode, planned for global launch in November. *However, the actor-level groomer classifier is not yet operational on IG, which means that we cannot use it to block groomers from accessing high-risk features like Disappearing Mode. We need about a week or two of CO [Community Operations] review capacity in order to calibrate the classifier for IG use. This is low hanging fruit* and fully supported by Child Safety CI and IG Policy. *We just need help getting CO to prioritize this.*⁶³⁶

318. Jayakumar explained that the IIC classifier “was trained on Facebook data and not really Instagram data,”⁶³⁷ so that is why it needed to be calibrated. That is, Facebook was text

⁶³³ Jayakumar Dep., p. 445.

⁶³⁴ META3047MDL-074-00164587 at 23-24.

⁶³⁵ *Id.*

⁶³⁶ Jayakumar Dep. Ex. 45; META3047MDL-014-00206542 (emphasis mine).

⁶³⁷ Jayakumar Dep., p. 455.

dominant, while Instagram was largely images. Even though the investment of time needed was modest—one week—the project had not received priority.

319. The fact that the project had not received priority is unfortunately consistent with other instances of Meta deprioritizing safety tools related to IIC, which was a problem of massive scale on its platform (by one estimate, *500,000 underage Instagram accounts receive IIC on a daily basis* in English markets alone).⁶³⁸

320. Parents and youth had no way of knowing the scope and scale of the risk of these inappropriate and dangerous interactions, or that Meta did not have effective safety tools to deal with them. Meta did not disclose these facts and, instead, provided safety assurances in its Community Standards that were misleading given Meta’s non-disclosures. Instagram’s failure to prioritize this “low hanging fruit” and failure to warn users of the problems with IIC classifiers were not consistent with the applicable standards of care. As such, Meta failed to meet standards of care related to warning users and their parents, designing for safety, and protecting youth.

G. Meta Failed to Disclose that It Had No Integrity Classifier to Detect Human Sex Trafficking

321. Instagram also had no integrity classifier to identify and flag instances of human sex trafficking. Instagram Safety Team Member [REDACTED] recommended that such a classifier be created, and in September 2020, she reported that it was “not ready; work has not been prioritized and more resources are needed.”⁶³⁹

We have one engineer dedicated to this work and we have no data science support to do some basic understand work to lower the strikes threshold for prostitution and sexual solicitation strikes (currently set at 17x, which is too high). This area is severely under-resourced given our SESTA-FOSTA liability.⁶⁴⁰

⁶³⁸ META3047MDL-014-00350154.

⁶³⁹ Jayakumar Dep. Ex. 45, META3047MDL-014-00206543.

⁶⁴⁰ Jayakumar Dep. Ex. 45, META3047MDL-014-00206542 (emphasis mine).

322. Instagram Safety and Wellbeing Lead Jayakumar explained that Instagram's strike rate of 17X meant that a user's account would not be removed from Instagram until the 17th violation for prostitution and sexual solicitation:

*17X, by any measure across the industry, is a very, very high strike threshold. That means that you could incur 16 violations for prostitution and sexual solicitation, and **upon the 17th violation, your account would be suspended.***⁶⁴¹

323. Instagram had no integrity classifier to detect human sex trafficking, and it failed to warn users that it was not quickly or reliably removing users who were soliciting prostitutes or promoting prostitution.⁶⁴² As such, Meta failed to meet applicable standards of care for corporate responsibility related to warning users and their parents, designing for safety, and protecting youth.

H. Meta Failed to Disclose that It Had Inadequate Age Management Technology and Processes

324. Meta represents to the public that it does not allow kids under 13 on its platforms and that it protects kids aged 13-17 from being exposed to inappropriate experiences. Meta fails to disclose that for many years it did not ask Instagram users to provide their age, that it did not have an effective system in place to identify underage users, that it made it difficult to report underage accounts, that when reports were made of underage accounts those reports went unaddressed, and that its age-gating of inappropriate experiences was actually very narrow.

325. Meta's stated policy is that only users 13 years and older are permitted to open an Instagram account and use the platform.⁶⁴³ Meta's executives have clearly created the impression that Meta enforces its age policy and prevents younger children from being on its platform. For

⁶⁴¹ Jayakumar Dep., p. 460 (emphasis mine).

⁶⁴² Jayakumar Dep., p. 461.

⁶⁴³ <https://www.meta.com/quest/parent-info/>; Jayakumar Dep., p. 81.

example, in 2021, Meta CEO Mark Zuckerberg assured Congress that Meta does not allow kids under 13 years old on its platform:

There is clearly a--a—a large number of people under the age of 13 who would want to use a service like Instagram. *We currently do not allow them to do that.*⁶⁴⁴

326. Instagram Head Adam Mosseri told Congress that Instagram invests heavily in identifying and removing users under the age of 13:

*We invest a lot to try to identify those under the age of 13. And whenever we find them, we remove [them].*⁶⁴⁵

327. Mosseri told Congress that Instagram uses classifiers to predict users' ages and then asks users to prove their age if they suspect they might be underage:

We build what we call classifiers, which try to predict age, and then, we ask people to prove their age if it looks like they might be too young.⁶⁴⁶

328. Meta's Global Head of Safety Antigone Davis also assured Congress that Meta removes the accounts of underaged users and consults with experts to create policies that prevent teens from seeing content that would be inappropriate for them ("age-gating content"):

When we learn that an underage user has created an account, we remove them from our platform. When it comes to those between 13 and 17, *we consult with experts to ensure that our policies properly account for their presence, for example, by age-gating content.*⁶⁴⁷

⁶⁴⁴ U.S. House of Representatives Energy and Commerce Subcommittees on Communications and Technology and Consumer Protection Hearing on Misinformation and Disinformation on Online Platforms, March 25, 2021, p. 17 (emphasis mine).

⁶⁴⁵ House Energy and Commerce Subcommittees on Communications and Technology and Consumer Protection Hold Joint Hearing on Misinformation and Disinformation on Online Platforms, March 25, 2021, p. 5 (emphasis mine).

⁶⁴⁶ *Id.* p. 14 (emphasis mine).

⁶⁴⁷ Senate Committee on Science, Commerce, and Transportation Subcommittee on Consumer Protection, Product Safety, and Data Security, September 30, 2021, p. 7 (emphasis mine).

329. Davis explained that Meta asks for users' birthdays when they register, allows anyone to report an underage account, has technology that can estimate people's ages, and is building technology to find and remove accounts of kids under 13 years old:

In addition to asking for people's date of birth when they register and allowing anyone to report a suspected underage account, *we have developed technology that allows us to estimate people's ages, for example, whether someone is younger or older than 18*, and we're building similar technology to find and remove accounts belonging to people under the age of 13. We train the technology using multiple signals.⁶⁴⁸

330. What Davis did not reveal is that although Facebook (now Meta) bought Instagram in April 2012,⁶⁴⁹ Instagram did not start even asking new users for their ages until more than seven years later in December 2019.⁶⁵⁰ Bejar testified that Meta executives are incorrect when they say that they do not allow kids under 13 on Instagram.⁶⁵¹ Bejar testified that during his first stint at Meta, Instagram did not ask kids for their age at sign up and there was likely millions of kids under 13 on Instagram.⁶⁵² At that time, Meta had a "don't ask, don't tell" policy around the age of users on Instagram.⁶⁵³

331. It was common knowledge within Meta that there were young kids using Instagram and, in fact, that was a significant reason behind why Meta wanted to acquire Instagram—Meta wanted very young kids on their apps.⁶⁵⁴ Even later, documents showed that Instagram only had the stated ages of 67 percent of users because Adam Mosseri would not support collecting ages

⁶⁴⁸ *Id.*, p. 9 (emphasis mine).

⁶⁴⁹ <https://www.britannica.com/money/Instagram>; Bejar Dep., pp. 183-84.

⁶⁵⁰ Jayakumar Dep., p. 82.

⁶⁵¹ Bejar Dep., p. 191.

⁶⁵² Bejar Dep., pp. 187-88.

⁶⁵³ Bejar Dep., pp. 183-84.

⁶⁵⁴ Bejar Dep., p. 185.

from existing users because he saw it as intrusive.⁶⁵⁵ As such, age models did not work well on Instagram.⁶⁵⁶

332. Even though Meta claims to remove the accounts of users who are under 13 years old when they find them or they are reported,⁶⁵⁷ Meta does not notify the children's parents when the accounts are removed.⁶⁵⁸ Moreover, Meta collected the same data from underaged users as from other users,⁶⁵⁹ but did not get the permission of their parents before collecting the data⁶⁶⁰ or give the parents opportunity to review the data.⁶⁶¹

333. Meta supposedly had integrity classifiers based on sophisticated technology and machine learning⁶⁶² along with 15,000 human reviewers to enforce its policies,⁶⁶³ so one would expect underaged users to be quickly identified and their accounts removed, but as detailed below, that was not the case.

1. Meta Did Not Build Integrity Classifiers for Detecting Underage Users and Failed to Review Accounts of Suspected Underage Users in a Timely Manner

334. In a 2020 document entitled, "Child Safety/State of Play," Antigone Davis described Instagram's vulnerability in preventing kids under 13 years of age from establishing and using an account. She identified the root of the problem as Meta leadership, who would not give priority to building an integrity classifier to detect minors under 13 years old:

Overall we are quite vulnerable when it comes to age management on our apps. To some that is because leadership has decided to live with certain risks at least in the near

⁶⁵⁵ Jayakumar Dep. Ex. 6, META3047MDL-034-00037247.

⁶⁵⁶ META3047MDL-034-00027362 at 6.

⁶⁵⁷ Jayakumar Dep., p. 81.

⁶⁵⁸ Jayakumar Dep., p. 82.

⁶⁵⁹ Jayakumar Dep., p. 80.

⁶⁶⁰ Jayakumar Dep., pp. 80-81.

⁶⁶¹ Jayakumar Dep., p. 81.

⁶⁶² <https://transparency.meta.com/enforcement/detecting-violations/how-enforcement-technology-works/>.

⁶⁶³ <https://transparency.meta.com/enforcement/detecting-violations/>.

term. For example in March, *we decided we would not build a classifier to detect minors under 13 years old.* Identity Integrity identified a PM [Product Manager] who will own age verification but *we will only prioritize this work when we face clear legal obligations to do so.*⁶⁶⁴

335. The inability to automate decisions to enforce against underage users was referred to as Instagram’s “biggest enforcement gap in relation to child safety.”⁶⁶⁵ Meta had information indicating that relying on users’ stated ages was problematic because roughly half of users were likely misrepresenting their ages.⁶⁶⁶

336. When a user was suspected of being underage, a review would be triggered—presumably by one of Meta’s more than 15,000 reviewers around the world who review potential violations on Facebook and Instagram in more than 80 languages.⁶⁶⁷ However, in 2020, Davis wrote that Instagram had a backlog of about 450,000 reports of potential users under age 13 waiting for review and that the backlog was “currently not resourced.”⁶⁶⁸ Note that while these accounts were waiting for review, “the users can continue to use the product normally.”⁶⁶⁹

337. Davis also noted that a number of children had Instagram accounts that were “hard-linked” to Facebook accounts of users who were known to be under 13 years of age. That is, the children had affirmatively connected their accounts in Meta’s systems, so Meta had information showing that the users were underaged. Davis noted that “basic COPPA [Children’s Online Privacy Protection Act] compliance” was at risk,⁶⁷⁰ so Identity Integrity agreed to cross-

⁶⁶⁴ Jayakumar Dep. Ex. 6, META3047MDL-034-00037245 (emphasis mine).

⁶⁶⁵ META3047MDL-034-00027362 at 6.

⁶⁶⁶ META3047MDL-014-00291926.

⁶⁶⁷ <https://transparency.meta.com/enforcement/detecting-violations/>.

⁶⁶⁸ Jayakumar Dep. Ex. 6, META3047MDL-034-00037245.

⁶⁶⁹ Jayakumar Dep., p. 445.

⁶⁷⁰ Jayakumar Dep. Ex. 6, META3047MDL-034-00037246.

enforcement [from Facebook to Instagram]. However, Davis feared that the project would not get resources because of constraints on the engineering team:

Ongoing: *Recently, Identity Integrity did agree to enable cross-enforcement of u13s [users under 13] on IG who have hard-linked FB accounts since basic COPPA [Children’s Online Privacy Protection Act] compliance is at risk when product does not prioritize checkpointing and disabling u13s. There is a chance, however, that this fails to get resourced because of engineering constraints on that team. However, there’s also a current u13 [under 13 years] backlog (~450K job backlog) that must be addressed via workload allocation and automation . . . This backlog is currently not resourced* and has been escalated to: Ops Resources. In the current state, we would not be able to reduce the backlog to zero for another year. The cross-enforcement of may also be at risk if the 2020 elections need more investment.⁶⁷¹

338. Davis’s acknowledgement was realized. In late September 2020, Jayakumar wrote in an email that “it was decided that we would *not* build a classifier to detect minors under 13 years old” but instead would pass the buck to industry partners.⁶⁷² She urged Instagram leadership to push the product team to clear the backlog of accounts waiting for review, disable the accounts of verified under 13 users, build a classifier, and support “the basic principle that if we’re using a signal to predict age for business [advertising] purposes, it should be used to enforce on age:”⁶⁷³

In March it was decided that we would not build a classifier to detect minors under 13 years old. Instead, we would push industry partners for device/OS level age verification and management solutions. . . . *We need leadership support to push product to act on knowledge we already have, i.e. clear our backlogs, disable verified u13s, build an u18 checkpoint to review minors who have been reported on 18+ products like FB Dating.* Against this backdrop, our enforcement is not consistent across all apps for a single known person. In addition, age models perform inconsistently across different markets. . . . *We also need leadership to support the basic principle that if we're using a signal to predict age for business purposes, it should be used to enforce on age.* For example, we need to use age prediction to gate minors from alcohol ads and other high-risk features where we currently rely solely on stated age.⁶⁷⁴ (emphasis mine)

339. Jayakumar elaborated on this “basic principle”:

⁶⁷¹ Jayakumar Dep. Ex. 6, META3047MDL-034-00037246 (emphasis mine).

⁶⁷² Jayakumar Dep. Ex. 45, META3047MDL-014-00206542 (emphasis mine).

⁶⁷³ Jayakumar Dep. Ex. 45, META3047MDL-014-00206542

⁶⁷⁴ Jayakumar Dep. Ex. 45, META3047MDL-014-00206542 (emphasis mine).

*... we sometimes use non-stated age signals ... to try and predict the age of that account, to decide what kind of ads might be most relevant to them, what the best type of promoted content would be for them. And if we are able to do that for business purposes, it says then, we should also be able to do that for integrity purposes.*⁶⁷⁵

340. In March 2021, [REDACTED], Child Safety Director for Messenger and Facebook, reported that Meta had a backlog of over 2.5 million age-related reports to review,⁶⁷⁶ so despite the pleas of the safety team, Instagram's leadership did not allocate the necessary resources. The gigantic backlog occurred despite the fact that it was very difficult for anyone to report underaged users on Instagram. As [REDACTED] reported, "[C]urrently there's no way for users to report under-13 users on Facebook in-app and it takes over seven screens for users to report under-13 users on Instagram."⁶⁷⁷ Jayakumar characterized Instagram's seven-screen reporting procedure as "a really complicated process."⁶⁷⁸ As explained by Bejar, it is easy to find under 13 kid accounts on Instagram but trying to report them is nearly impossible.⁶⁷⁹ Had it been easier to report underaged users, one would imagine that the backlog would have been even greater.

341. Having integrity classifiers that provide age-related data would have benefitted the development of safety-related products for minors as Jayakumar explained:

Minors who go viral on Instagram will likely experience increased interactions with unknown people. As their content and by extension their account becomes surfaced to more unconnected people, they would benefit from additional protections that they can turn on, like limiting who can send them private DMs or follow requests. ***However, with a lack of age data, we cannot effectively develop solutions at scale that address virality of minors on Instagram without also generating a large number of false positives/false negatives.***⁶⁸⁰

⁶⁷⁵ Jayakumar Dep., p. 464 (emphasis mine).

⁶⁷⁶ Jayakumar Dep. Ex. 9, META3047MDL-020-00278850.

⁶⁷⁷ Jayakumar Dep. Ex. 9, META3047MDL-020-00278850.

⁶⁷⁸ Jayakumar Dep., p. 100.

⁶⁷⁹ Bejar Dep., p. 193.

⁶⁸⁰ Jayakumar Dep., p. 62 (emphasis mine).

342. Jayakumar noted that having algorithms that infer age is also beneficial because many youth misstate their age:

There is also value to algorithmically inferred age sometimes. It's sometimes a more accurate predictor than stated age, especially in cases where people lie about their age.⁶⁸¹

343. Algorithms could also infer the ages of adults who misstate their age and pretend to be minors, which would circumvent many of Meta's protections for teens (e.g., adults are prevented from messaging teens).⁶⁸² As Ravi Sinha, Head of Child Safety Policy, said, "[N]othing prevents an adult from opening an IG or FB account and lying about their age today to say that they are a teen."⁶⁸³

344. Meta lacked adequate classifiers to identify underaged accounts, and its review system was horribly backlogged. Also, when Meta removed underaged accounts, it did not notify the users' parents. Moreover, Meta did not get parental permission before collecting information from under-13 users, despite the obligations of COPPA, and it did not let parents review the information it collected.⁶⁸⁴ Meta failed to warn parents, youth, schools, and the community at large that its systems were grossly inadequate in providing age-related data that would enable the company to detect underaged users. Moreover, by refusing to create a classifier to identify underaged users, Meta failed to meet standards of care regarding designing for safety. As such, Meta's actions were not consistent with its own and industry standards of care for corporate responsibility.

⁶⁸¹ Jayakumar Dep., p. 465.

⁶⁸² META3047MDL-046-00319359 at 52.

⁶⁸³ Sinha Dep., pp. 1445-46.

⁶⁸⁴ Jayakumar Dep., pp. 80-81.

2. Meta Failed to Disclose that It Had Ineffective Age Gating

345. When asked by her manager, Antigone Davis, what she saw as Instagram’s biggest gaps for youth safety in December 2021, Instagram’s Safety and Wellbeing Lead Vaishnavi Jayakumar highlighted Instagram’s inadequate and poorly enforced age-gating policies:

The lack of robust age-appropriate experiences, especially around, (a), the types of content, and (b), volume of content we surface to teens. We (mostly) assume that all content should be equally discoverable and viewable for the same duration of time by a 13-year-old and a 55-year-old. ***Our age-gating policies are currently very narrowly scoped, aren't all proactively enforced, and fall short of what one would reasonably consider to be age-appropriate.***⁶⁸⁵

346. Jayakumar explained that Instagram’s age-gating was limited and applied to a “very narrow scope of content” such as firearms and tobacco, and even then, these areas were “not proactively” enforced through machine learning and integrity classifiers:

So we had ***a very small set of policies that were age-gated***, meaning that you had to be 18-plus to view the content. So if you had, you know, logged -- if you had signed up as a 13 or 14-year-old, you would have gotten like a screen saying sorry, this, you know, content is age-restricted. But ***it applied to a very narrow scope of content . . . firearms, perhaps tobacco. . . And even when we had those policies, they weren't all proactively enforced. That means that we were waiting for reports of somebody posting this content or reports on this content to take action, rather than proactively scanning and, you know, age-gating that content.***⁶⁸⁶

347. Jayakumar wanted to extend Instagram’s age-gating to “borderline policies,” which “refer to content that doesn’t violate the community standards, but is something that we wouldn’t recommend.”⁶⁸⁷ She explained her logic and noted that this expansion would have been an easy product intervention that would have significantly increased what was age-gated:

. . . we say that if we aren't comfortable recommending content, then we should probably age-gate that content. ***So that would have expanded the amount of age-gated content very significantly. And it would have been a very easy product intervention.***⁶⁸⁸

⁶⁸⁵ Jayakumar Dep. Ex. 10, META3047MDL-014-00223594 (emphasis mine).

⁶⁸⁶ Jayakumar Dep., p. 114-115 (emphasis mine).

⁶⁸⁷ Jayakumar Dep., p. 115

⁶⁸⁸ Jayakumar Dep., pp. 115-116 (emphasis mine).

348. A year later in December of 2022, [REDACTED], Jayakumar's manager, confirmed that Instagram still did not have age-gating for a host of highly problematic areas including suicide and self-injury, eating disorders, graphic violence, and nudity:

We are testing age-gating borderline content for regulated goods [tobacco, firearms, alcohol] *but borderline content in other issue areas like SSI [suicide and self-injury], ED [eating disorders], graphic violence, nudity, still remains ungated and we do not yet have age-appropriate recommendation algorithms in most issue areas.*⁶⁸⁹

349. Alarming, in November 2023, Kilstein received a chat and a link from [REDACTED], a Meta Wellbeing staff member, that confirmed that Instagram had recommended child pornography to Wall Street Journal Reporter Jeff Horowitz, which is a violation of Meta's Community Standards. Since Meta lacked appropriate age-gating, this could be recommended to youth.⁶⁹⁰

350. By not committing the resources to age gate effectively, Meta failed to meet standards of care regarding designing for safety. Jayakumar confirmed that Meta had failed to warn parents that it did not have algorithms to detect and prevent youth from seeing harmful content in the same way that a 55-year-old user would.⁶⁹¹ As such, Meta also failed to meet standards of care regarding warning.

I. Meta Promoted What It Described as Safety Features that did Not Improve Safety

351. Over time, Meta has promoted a number of features that it claims improve safety on its platforms.⁶⁹² However, safety and wellbeing experts such as Arturo Bejar found the tools to be ineffective.⁶⁹³ Bejar testified that the safety tools were merely "placebo" that "sound good but

⁶⁸⁹ Jayakumar Dep. Ex. 11, META3047MDL-020-00331813 (emphasis mine).

⁶⁹⁰ Kilstein Dep. Ex. 68, META3047MDL-072-01293016; Ex 69 (video).

⁶⁹¹ Jayakumar Dep., pp. 120-121.

⁶⁹² See Mosseri Dep. Exs. 104 and 105 for a listing and a timeline of the introduction of these features.

⁶⁹³ See Expert Report of Dimitri Christakis, pp. 332-35, 343-53.

don't reduce real world harm," and that many were introduced defensively in response to external pressures such as public relations crises.⁶⁹⁴

352. Bejar attributed some of the ineffectiveness to Meta's decision to make the safety features "opt-in" rather than "opt-out."⁶⁹⁵ He said, "[i]f a feature is opt-in, almost no one will use it."⁶⁹⁶ Bejar estimated that the adoption rate for an opt-in safety tool would be "one to two percent of people if you are doing really well," which means 99% do not adopt the feature.⁶⁹⁷ Another reason why the safety tools were ineffective was that Meta made "no meaningful efforts to get people to use them."⁶⁹⁸ In addition, as discussed with some specific examples below, Meta delayed the introduction of some tools and introduced them in a weakened form, which some documents indicate was to blunt their potentially negative effects on growth and engagement.

1. Meta Delayed the Launch of Private By Default and Restricted It to New Users Under 16 Years

353. Instagram's private by default (PBD) feature, a safety feature for teens, is a prime example of pseudo corporate responsibility in which Instagram's actions went directly against Meta's many public declarations and commitments related to "building responsibly to keep people safe" and "providing the highest protections for youth."⁶⁹⁹

354. Documents indicate that Instagram was behind the industry in safety for minors and that some competitors had already taken important steps to reduce interactions between minors and adult strangers.⁷⁰⁰ The stated goal of PBD was to create a safer space by making teens'

⁶⁹⁴ Bejar Dep., pp. 166, 533-34.

⁶⁹⁵ Bejar Dep., p. 167.

⁶⁹⁶ *Id.*

⁶⁹⁷ Bejar Dep., pp. 545-46.

⁶⁹⁸ Bejar Dep., p. 167.

⁶⁹⁹ <https://about.meta.com/code-of-conduct/>, p. 28.

⁷⁰⁰ META3047MDL-031-00136981.

accounts visible only to the people whom they accepted as “friends” and gave permission to follow them.⁷⁰¹ Because PBD would limit the visibility of teens’ accounts to adult strangers,⁷⁰² it was presented that the feature would reduce teens’ bad experiences and mitigate high severity incidents.⁷⁰³ In 2019, in an electronic chat with his boss, [REDACTED], Instagram’s Director of Data Science Darius Kilstein said, “This [private by default] was never about safety. It was all for the PR wins.”⁷⁰⁴ To which [REDACTED] responded, “I know.”⁷⁰⁵

355. PBD was tested by Instagram in 2019, but it was not implemented until 2021. While Instagram’s policy, legal, communication, privacy, and wellbeing teams supported launching PBD earlier, the growth team objected because based on their test, it would result in a loss of 1.5 million active users.⁷⁰⁶ It also would likely result in lower user engagement.⁷⁰⁷ For example, PBD makes it harder for teens to find their friends,⁷⁰⁸ and “friending” is one of the main drivers of growth.⁷⁰⁹ The impact of PBD would translate into lower revenue for Instagram.⁷¹⁰

356. Vaishnavi Jayakumar reported that Instagram Leads had failed to approve the launch of PBD due to growth concerns:

we were initially supposed to launch private by default on September 2nd [2020] ***after approval from IG leads, but that has now been scrapped due to growth concerns***, and the product team has been given some time to come up with ways to mitigate the drop in growth.⁷¹¹

⁷⁰¹ META3047MDL-031-00136981.

⁷⁰² Jayakumar Dep., p. 68.

⁷⁰³ META3047MDL-031-00136996.

⁷⁰⁴ Kilstein Dep. Ex. 30, META3047MDL-020-00538138.

⁷⁰⁵ Kilstein Dep. Ex. 30, META3047MDL-020-00538138.

⁷⁰⁶ Kilstein Dep. Ex. 57, META3047MDL-031-00136981.

⁷⁰⁷ Jayakumar Dep., p. 70.

⁷⁰⁸ Kilstein Dep., p. 211.

⁷⁰⁹ Kilstein Dep., p. 119.

⁷¹⁰ Jayakumar Dep., p. 70.

⁷¹¹ META3047MDL-050-00285759, p. 69, Jayakumar Dep. Ex. 7 (emphasis mine).

The launch was delayed even though Instagram leaders acknowledged that Instagram was lagging the industry on teen safety:

Instagram is behind the industry in keeping our platform safe for minors. Other platforms, including TikTok, have taken bold moves to reduce interactions by strangers.⁷¹²

357. When PBD was finally implemented in 2021, documents indicate that it was designed to minimize its impact on teen growth metrics because, unlike the “strict version” that had been tested in 2019, it was limited to new users who had stated their age as 16 years old or under. Data Scientist [REDACTED] noted that only 6.5 percent of existing U.S. teen users would even see the notice about PBD’s availability, much less choose it:

BY the way I think the teen privacy stuff will have a much lower impact now that we are only talking about U16 and stated age. For example – only 11% of Existing US teen accounts have a stated age U16 – and of those 38% are already private. So only 60%*11% = 6.5% of us teens will even see the upsell - let alone go through with it.⁷¹³

358. The number of Instagram users identified as under age 16 was artificially low because Instagram did not verify ages,⁷¹⁴ and Instagram had information that many youth—e.g., about half of youth in the U.S.—misstated their ages.⁷¹⁵ Moreover, the default for PBD was not a typical default setting, which would require users to go into their account settings to opt-out. Instead, new users were presented with a screen asking them if they wanted a public or private account, and the private button was preselected.⁷¹⁶ Alarming, fewer young teens, arguably the most vulnerable users, selected private over public as compared to older cohorts.⁷¹⁷

⁷¹² Kilstein Dep. Ex. 57, META3047MDL-031-00136981 (emphasis mine).

⁷¹³ Kilstein Dep. Ex. 36, META3047MDL-003-00015911 (emphasis mine).

⁷¹⁴ Jayakumar Dep., p. 85.

⁷¹⁵ Jayakumar Dep. Ex. 8, META3047MDL-020-00298458; Bejar Dep., p. 194.

⁷¹⁶ Jayakumar Dep., p. 73

⁷¹⁷ META3047-020-00350419

359. If PBD had employed a typical default setting, Jayakumar stated that it would have resulted in more teens having private accounts because Instagram had data indicating that most teens do not change their settings.⁷¹⁸ Even more, PBD was not applied to all teen accounts until the end of 2024.⁷¹⁹

360. Meta delayed the launch of Private by Default and restricted it to new users under age 16. Documents and statements of Meta employees discussed this delay in terms of minimizing the impact of the feature on growth and engagement. To the extent that Meta prioritized growth over safety, it endangered youth and failed to meet its own and industry standards of care regarding designing for safety and protecting youth.

2. Meta Compromised the Effectiveness of PBD by Launching a Recommendation Feature for Teens that had the Opposite Effect of PBD

361. The potential negative impact of PBD on Instagram's growth was mitigated because its launch was accompanied by the launch of a new recommendation feature for teens that Kilstein described as designed to do the opposite of PBD—to “boost [teens] in suggested user recommendations”⁷²⁰

362. In discussing the recommendation feature, Data Engineer [REDACTED] expressed his confusion:

I'm still kinda confused about the goal [of the recommendation boost for teen accounts]. We want to make sure teens don't accidentally share more than they want, and avoid creepers, but then some of them totally want that and then *if we boost them, we are inviting people to see more teens in their suggestions*. Is it about decoupling clear privacy control from integrity issues, *or is it about saying we want to reduce integrity problems, but not wanting the consequence of less interactions and follows?*⁷²¹

Data Scientist [REDACTED] responded:

⁷¹⁸ Jayakumar Dep., p. 74

⁷¹⁹ Jayakumar Dep., p. 71

⁷²⁰ Kilstein Dep., p. 249

⁷²¹ Kilstein Dep. Ex. 37, META3047MDL-003-00016693 (emphasis mine).

*it is about looking good to regulators so that they don't block our under 13 year old IG version we are working on. Thats it.*⁷²²

363. Unfortunately, boosting suggested user recommendations for teens made them more vulnerable to contacts from adult strangers and less safe. This problem was further exacerbated by the fact that during this time Instagram was actually recommending suspicious adults to minors and vice versa.⁷²³ Thus, while Meta was publicly touting supposed safety features that it was implementing, it was actively working behind the curtain to undermine the effectiveness of those safety features. As such, Meta endangered youth and failed to meet standards of care related to designing for safety and protecting youth.

3. Meta Delayed the Introduction of Quiet Mode

364. Research demonstrated that problematic use was correlated with nighttime sessions on Instagram,⁷²⁴ so Quiet Mode was developed as a feature that would enable parents or youth to turn off notifications during the night so that they do not interrupt youth sleep. Notifications are messages that are external to Instagram that are designed to let users know when activity is taking place on Instagram, and as such, they encourage users to go back to Instagram to make sure that they are not missing out.⁷²⁵

365. Jayakumar explained the manner in which notifications can cause youth to spend more time on Instagram and become dependent on it—thus potentially playing a “contributing role” in social media addiction:⁷²⁶

⁷²² Kilstein Dep. Ex. 37, META3047MDL_RVOL003, META3047MDL-003-00016693 (emphasis mine).

⁷²³ Sinha Dep. Ex. 8, META3047MDL-073-00003522

⁷²⁴ Raychoudhury Dep. Ex. 9, at 24

⁷²⁵ Jayakumar Dep., p. 147

⁷²⁶ Jayakumar Dep., p. 141.

. . . notifications tend to draw *them [people 13-18 years] into the app repeatedly . . . for more sessions and for longer time in session than we would otherwise want for young people of that age*. And so notifications have been seen as a *form of increasing the dependency of somebody on the app*.⁷²⁷

366. Jayakumar said that upon joining Instagram she had “really wanted to pitch a number of sleep-promoting interventions to improve the well-being of young people on the platform.”⁷²⁸ In fact, she raised concerns about the impact of notifications on youth in March 2020⁷²⁹ and again in November 2021,⁷³⁰ but Quiet Mode was not launched until early 2023.⁷³¹ Even then, it was not a default option, so users had to go into their settings and opt for Quiet Mode.⁷³² Thus, Meta again undermined the effectiveness of its purported safety feature. Quiet Mode was not launched as a default setting for youth until Instagram introduced Teen Accounts in August 2024.⁷³³

367. As explained by Bejar, reducing notifications might help with addiction, but Instagram and Meta were not willing to substantively address the issue of reducing time that people spend or reducing notifications.⁷³⁴ In the interim, Meta did not warn youth or their parents about the problematic issues related to notifications and sleep.⁷³⁵

⁷²⁷ Jayakumar Dep., pp. 141-142 (emphasis mine).

⁷²⁸ Jayakumar Dep., p. 136.

⁷²⁹ Jayakumar Dep. Ex. 15, META3047MDL-020-00270223.

⁷³⁰ Jayakumar Dep. Ex. 14, META3047MDL-040-00541333.

⁷³¹ Jayakumar Dep., p. 148-149.

⁷³² Jayakumar Dep., p. 151.

⁷³³ Jayakumar Dep., p. 152.

⁷³⁴ Bejar Dep., p. 151.

⁷³⁵ Jayakumar Dep., p. 153.

4. Meta Failed to Make “Take a Break” a Default Setting to Combat Overuse and Addiction

368. Instagram had information indicating that long periods of use were a problem for teens.⁷³⁶ For example, in May 2020, Senior User Experience Researcher Shayli Jimenez acknowledged that because of their immature brains teens have a harder time stopping using Instagram than adults even when they are unhappy with the amount of time they spend on Instagram:

Due to the immature brain, [teens] have a much harder time stopping even though they want to—our own product foundation research has shown that teens are unhappy with the amount of time that they spend on our app.⁷³⁷

369. Seemingly in response, Instagram launched a feature called “Take a Break” that enabled teens to “opt-in” to receive a reminder to “take a break” after a session of a certain length of time that they could specify—i.e., 10, 20, or 30 minutes.⁷³⁸ Note that “opt-in” means that users have to 1) know about a feature, and 2) go into their apps and turn the feature on themselves. In contrast, “opt-out” means that the feature is automatically turned on for users, and if they want to disable it, they have to go into their settings and turn it off.⁷³⁹

370. When Instagram Head Adam Mosseri announced that Instagram would be testing the “Take A Break” feature, he professed Instagram’s concern for people’s well-being:

Supporting peoples' well-being is a priority for us - and giving people ways to manage their time in a way that's intentional and meaningful is one way we're thinking about that.⁷⁴⁰

⁷³⁶ Volichenko Dep. Ex. 2, META3047MDL-040-00258969.

⁷³⁷ META3047MDL-003-00191215.

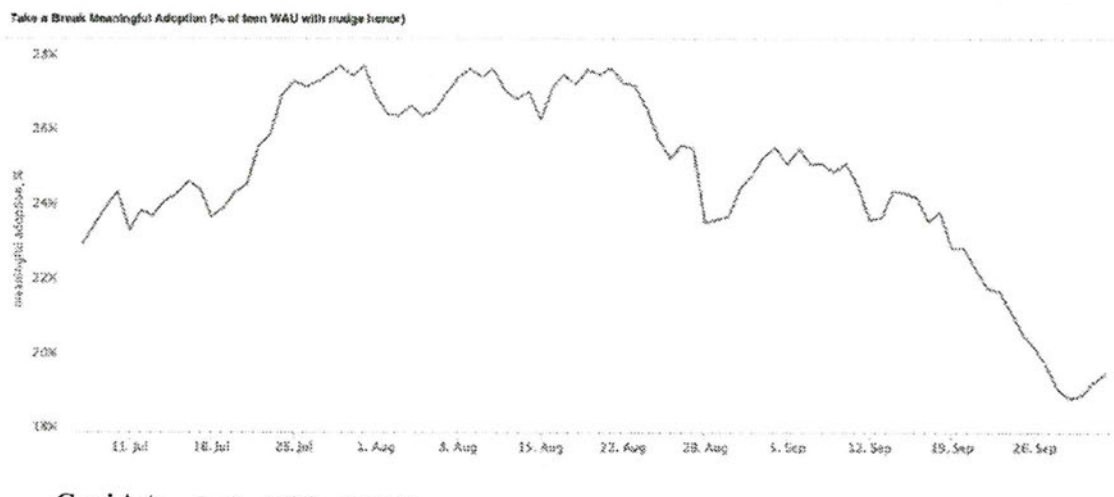
⁷³⁸ Volichenko Dep., pp. 92-93.

⁷³⁹ Volichenko Dep., p. 106.

⁷⁴⁰ Hutchinson, Andrew, “Instagram Tests New ‘Take a Break’ Feature to Encourage Users to Limit Time in the App,” Social Media Today, Nov. 10, 2021, <https://www.socialmediatoday.com/news/instagram-tests-new-take-a-break-feature-to-encourage-users-to-limit-time/609854/> (emphasis mine).

371. George Volichenko, a data scientist on the Teen Mental Health team, was assigned to assess the impact of “Take a Break.” His research revealed an adoption rate (users who attempted to opt-in) so low that the Teen Mental Health team asked to revise downward their “Take a Break” adoption goals from 0.26 percent to only 0.18 percent.⁷⁴¹ That is, if the team reached its goal, 99.82 percent of teens on Instagram would *not* be using the “Take a Break” feature. Volichenko stated that the reason adoption was so low was that “few people even knew about it and even fewer people enabled it.”⁷⁴²

372. Volichenko’s analysis also highlighted another issue with the “Take a Break” feature related to what he referred to as “meaningful adoption”—that is, teens who “actually honor the break reminder” and “end up taking a break.”⁷⁴³ Only a small percentage of the teens who opted-in to the “Take a Break” feature were actually taking a break when they were prompted to do so, and their numbers were dwindling as illustrated in the graph below:⁷⁴⁴



⁷⁴¹ Volichenko Dep. Ex. 3, META3047MDL-040-00399807.

⁷⁴² Volichenko Dep., p. 110.

⁷⁴³ Volichenko Dep., pp. 108-109.

⁷⁴⁴ Volichenko Dep. Ex. 4, META3047MDL-047-01170067.

373. Volichenko noted that a variety of things could make it easier or more difficult for the user to dismiss the reminder to take a break—e.g., “the font on the button, the size of the button, how prominent it is.”⁷⁴⁵ He noted that Instagram could make the reminder “in a way that it’s not dismissible at all. Like, you have to shut down the app.”⁷⁴⁶

374. To increase adoption and usage, Volichenko recommended that Instagram “revisit making this [Take a Break] an *opt-out* feature for teens.”⁷⁴⁷ As noted above, as an opt-out feature, “Take a Break” would automatically be turned on for all teens. Experts on Instagram’s own Youth Advisors group, who were established researchers and academics, also advised Meta to make the “Take A Break” feature a default setting. For example, [REDACTED], a youth development researcher, asserted that “Take a Break” should be a default setting because “younger people might not know that they need to take a break” and “relying on young people’s instincts is just not solid and that’s the challenge.”⁷⁴⁸

375. Nonetheless, “Take a Break” was not even tested as an opt-out feature due to Instagram’s “one-way door” practice, which meant once it was enabled, it would be difficult for Instagram to walk it back. Volichenko noted that the “one-way door” practice was especially sensitive for features related to the wellbeing space because to withdraw safety features “just looks really bad”.⁷⁴⁹

it just looks really bad, right? *Like the company is, like, doing something to protect the users, but then they’re, like, eh, we don’t want to protect users anymore. We’re going to turn it off.*⁷⁵⁰

⁷⁴⁵ Volichenko Dep., p. 110.

⁷⁴⁶ Volichenko Dep., pp. 110-111.

⁷⁴⁷ Volichenko Dep. Ex. 4, META3047MDL-047-01170067.

⁷⁴⁸ Jayakumar Dep. Ex. 39, META3047MDL-003-00021371.

⁷⁴⁹ Volichenko Dep., p. 184.

⁷⁵⁰ Volichenko Dep., p. 184 (emphasis mine).

376. Instagram leadership never made “Take a Break” an opt-out or default setting for teens. Volichenko’s perspective was that doing so would have negatively affected some of Instagram’s core metrics such as ad revenue.⁷⁵¹

377. As such, Instagram’s “Take a Break” feature was a classic form of pseudo corporate social responsibility in that it appeared to be a responsible safety feature and Instagram Head Adam Mosseri promoted it as such, but because of the manner in which Meta implemented it, the feature was ineffective and it actually had minimal impact on teen safety. Therefore, Meta’s actions did not meet standards of care regarding designing for safety.

J. The Above Actions Prioritized Growth Over Safety and were Reckless and Irresponsible

378. Meta claims to build “responsibly to keep people safe” and provide “the highest protections for youth,”⁷⁵² but in actuality, it does not. A primary indication was that concerns about growth—e.g., time spent on the platform—dominated concerns for wellbeing. However, Meta leaders resoundingly denied this in public forums. For example, Meta CEO Mark Zuckerberg denied that Meta set Instagram team goals related to increasing the amount of time that people spend on the platform:

I don’t give our newsfeed team, our Instagram team goals around increasing the amount of time that people spend.⁷⁵³

379. As noted in many places above, Instagram’s actions often prioritized growth over wellbeing. When asked who was prioritizing growth over safety, Jayakumar, Instagram’s first

⁷⁵¹ Volichenko Dep., p. 114.

⁷⁵² <https://about.meta.com/code-of-conduct/>, p. 28.

⁷⁵³ U.S. House of Representatives Energy and Commerce Subcommittees on Communications and Technology and Consumer Protection Hearing on Misinformation and Disinformation on Online Platforms, March 25, 2021, p. 21.

safety and wellbeing lead,⁷⁵⁴ said that it was Instagram’s leadership—company directors, VP’s and further up:

I think at the working level I met many great product managers, engineers, researchers, data scientists, policy folks, who were very committed to keeping children safe and protected on the platforms. *Where we typically run into challenges was at the leadership levels, when we took them for reviews to company directors, VPs, and further up.*⁷⁵⁵

While she acknowledged that in some instances Instagram leaders gave priority to safety, they “*didn’t prioritize safety over growth in many instances.*”⁷⁵⁶

380. Arturo Bejar testified that growth and engagement were the top priority and that safety was an afterthought.⁷⁵⁷ As noted above, one such instance involved PBD. Even though the wellbeing team, along with the legal, policy, communication, and privacy teams, supported the launch of PBD, the growth team objected over concerns about decreased user engagement. As a result, Instagram leadership delayed implementing PBD, and when it ultimately launched the feature, it was in a weakened form.⁷⁵⁸

381. Instagram Senior Researcher Alison Lee asserted that Instagram leaders’ priorities were on growing the product, growing the platform, and increasing engagement, while safety and wellbeing were third or fourth order priorities. As such, Lee asserted that concerns about engagement and growth superseded concerns about safety:

I think that the leaders at Instagram, their attention and their priorities were elsewhere. *Namely in growing the product, growing the platform, and increasing engagement. I would say that they perhaps cared as a third or fourth order priority about safety and well-being.* I don’t think that they did not care at all. *And I would say that those other*

⁷⁵⁴ Jayakumar Dep., p. 28.

⁷⁵⁵ Jayakumar Dep., p. 26 (emphasis mine).

⁷⁵⁶ Jayakumar Dep., p. 26 (emphasis mine).

⁷⁵⁷ Bejar Dep., p. 20.

⁷⁵⁸ Kilstein Dep. Ex. 57, META3047MDL-031-00136981.

concerns about engagement and growing the platform and product superceded [sic] those concerns about safety.⁷⁵⁹

382. Jayakumar said that the safety team “would need to be very careful not to suggest anything that could impact growth too significantly.”⁷⁶⁰

383. Jayakumar said that Instagram leadership had a higher benchmark for safety products than for other products—e.g., 100 percent likelihood that they product would improve safety—which placed additional burdens on safety-related products:

In some cases, we would see a very high benchmark for -- a very high standard required to get a safety product across the finish line that wasn't necessarily reflected in some of the benchmarks versus the other products. So we would have to have, you know, without a shadow of a doubt the evidence to show that there's a 100 -- really very, very high likelihood that this product will improve the safety. And if it was even just a high likelihood but not necessarily a very high likelihood, we faced challenges with getting that through.⁷⁶¹

384. Arturo Bejar also asserted that there was an unreasonably high bar for introducing a wellbeing product. He said that the wellbeing team had to demonstrate causality, which is difficult to do in a public health setting,⁷⁶² and that requirement made it incredibly difficult for the wellbeing team to introduce a new feature. In contrast, he said that the teams working on growth and engagement had ***no bar***—all they had to do was to have a good hypothesis:

And so the bar for a well-being feature and resourcing or working on a problematic area was high. And the bar to work on growth engagement, something that you would think would drive usage, was none, right. You could just -- you had to have a good thesis to say I believe this will increase usage by this much percentage or I believe this will -- this feature is going to be very successful and you were, like, off you go, right. And so -- and I noticed in my second stint, right, that the well-being team had an incredibly difficult job at trying to define and implement features that would make a difference.⁷⁶³

⁷⁵⁹ Lee Dep., pp. 460-46 (emphasis mine).

⁷⁶⁰ Jayakumar Dep., p. 32.

⁷⁶¹ Jayakumar Dep., p. 27 (emphasis mine).

⁷⁶² Bejar Dep., p. 142.

⁷⁶³ *Id.*, pp. 142-43 (emphasis mine).

385. Instagram appeared to make little effort to integrate wellbeing priorities with the initiatives of other teams such as the growth team. For example, Kilstein repeatedly claimed that risks to mental health and wellbeing “were not the growth team’s focus,”⁷⁶⁴ that “no one on the growth team was encouraged to think about mental health,”⁷⁶⁵ and that “the growth team’s sole focus was on increasing DAU [Daily Average Users] and MAU [Monthly Average Users].”⁷⁶⁶ Many of Instagram’s growth teams (e.g., acquisition, activation, graph) had engineers, data scientists, product managers, user experience experts, and design experts, but they did not incorporate safety or wellbeing experts.⁷⁶⁷

386. Instagram dedicated fewer human resources to safety than to product and growth—the wellbeing team was “tragically small.”⁷⁶⁸ For example, Instagram did not even have a dedicated, specialized safety officer prior to January 2020 when Jayakumar joined as the first safety officer. Previously, safety had been handled by public policy and product policy employees whom Jayakumar characterized as “[n]ot necessarily, you know, folks with a background in safety, per se.”⁷⁶⁹ Senior Researcher Alison Lee reported that each surface (e.g., Reels, Explore) had multiple product teams, but only one integrity team. For example, Reels had 15 or more teams working on engagement, but only one safety and wellbeing team.⁷⁷⁰

387. Jayakumar, who had had eight years of safety-related work experience, joined Instagram with the understanding that she would work directly with product managers to “evaluate the platform, understand what the risks and challenges were for children and teens, and

⁷⁶⁴ Kilstein Dep., p. 494.

⁷⁶⁵ Kilstein Dep., p. 494.

⁷⁶⁶ Kilstein Dep., p. 494.

⁷⁶⁷ Kilstein Dep., pp. 480-485.

⁷⁶⁸ Bejar Dep., pp. 148, 155.

⁷⁶⁹ Jayakumar Dep., p. 28.

⁷⁷⁰ Lee Dep., p. 163-164

then propose remediations, and then get the according budget and head count to do all that work.”⁷⁷¹ However, her work was thwarted because she had difficulty connecting with product-related employees and was not sponsored by Instagram leadership:

*What I found very quickly, though, was that it was very challenging to get connected to the product managers, the engineers, the folks who were doing the work who would benefit from the safety advice. And so I really had to virtually kind of chase down people and reach out to them, make sure they understood that I was there, that I had safety expertise that would improve the quality of the product, get my feedback, incorporate it into their work. And all that happened at the working level. **There was no kind of leadership sponsorship of that. Like my manager never – I don't think she ever even introduced me to the group that like I had joined the company. I don't think there was even, like, a post, you know, that I had joined the company. So that made it challenging.***⁷⁷²

Jayakumar said that her manager, Director of Safety Antigone Davis, did not give her enough support to do her job.⁷⁷³

388. By continually and repeatedly prioritizing growth over safety, Instagram failed to meet standards of care related to designing for safety and protecting its youngest and most vulnerable users. Further, Meta’s actions described above were reckless and irresponsible.

■ [REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

⁷⁷¹ Jayakumar Dep., p. 28.

⁷⁷² Jayakumar Dep., pp. 29-30 (emphasis mine).

⁷⁷³ Jayakumar Dep., p. 31.

390. Snap’s approach is characterized by what has been described in behavioral ethics as the “ostrich syndrome,” where one chooses to bury their head in the sand and ignore moral responsibilities and the related standards of care.⁷⁷⁴ The ostrich syndrome creates what scholars have labeled as “moral myopia,” in which one’s moral responsibilities do not come clearly into focus, and it can even result in moral blindness.⁷⁷⁵ Moreover, Snap failed to disclose the risks and warn teens, parents, schools, or the community at large.

⁷⁷⁴ Drumwright, Minette E. and Patrick E. Murphy, “How Advertising Practitioners View Ethics: Moral Muteness, Moral Myopia, and Moral Imagination,” *Journal of Advertising*, 33 (2): 7-24, 2004.

775 *Id.*

