How Meta Executives Talked About Child Safety Behind the Scenes

Michael Scherer, Kaitlyn Tiffany

Updated at 2:48 p.m. ET on February 27, 2026

Not so long ago, Mark Zuckerberg was working in overdrive to convince the world that his company was doing everything it could to protect children. In 2021, he posted a note to his personal Facebook page, writing that he had "spent a lot of time reflecting" on the types of experiences he would want his daughters, then 4 and 5 years old, to have online. "It's very important to me that everything we build is safe and good for kids," he wrote, emphasizing that the company absolutely does not "prioritize profit over safety and well-being."

But documents recently viewed by *The Atlantic* show that behind the scenes, the company now known as Meta was divided on whether protecting kids should take precedence over user growth and engagement. For years, the company only incrementally rolled out restrictive safety features, even as its own staff detailed the risks its platforms posed to children. Take, for example, a technical problem that affected the company's systems in November 2020. This issue limited Meta's ability to track bad actors, at a time when there were, according to an internal chat, "thousands of minors" reporting what the company refers to as tier-one "Inappropriate Interactions with Children," or "IIC T1"—the "most severe" outcomes possible, such as meeting for sex in real life, suicide, extortion, sadism, and sex trafficking.

"Even though we know that there is IIC T1 going on (more than 50% of which is sextortion which can lead to suicide) we haven't done anything. we had a broken escalation path and no measurements," one employee wrote in the internal chat about the problem. "God knows what happened to those kids." The company fixed the technical failure within weeks, another document shows, but it would take several more years to adopt other suggested measures to tackle broader issues that allowed predators to find underage targets on Instagram, which Meta owns.

Company spokespeople were clearly aware of the broader teen-safety problem. Just four weeks after Zuckerberg had posted about being "good for kids," two public-affairs specialists discussed an Instagram update that had just rolled out. That update made new accounts belonging to 13-, 14-, and 15-year-olds "private" by default, yet even this modest move had been flagged by insiders as a business risk for nearly two years before the change was made. Liza Crenshaw messaged her colleague Sophie Vogel that the move had been "contentious"—Instagram CEO Adam Mosseri, a deputy to Zuckerberg, was concerned that it would cause a "huge growth hit," Crenshaw wrote, according to documents we reviewed.

"We will never get out of this mess if he/we're not just prepared to ERR ON THE SIDE OF SAFETY," Vogel wrote. "Would he want any tom dick or harry being able to see all his kids' content, follow them etc? Is he fucking nuts?"

Meta's failure to proactively limit its business ambitions on behalf of vulnerable users may seem obvious to those who have paid attention to the company over the years, especially after revelations such as those leaked by whistleblower Frances Haugen in her "Facebook Files," among others. Yet these new documents—which include corporate reports, presentations, and other communications and which were disclosed as part of a lawsuit against Meta that went to trial this month in New Mexico—provide an unusually clear view into the company's decision making. Over the course of six years, Meta tinkered with basic privacy controls, while simultaneously calculating how simple interventions would moderately reduce the amount of time people spent on Instagram and initially opting for conservative, piecemeal updates to protect its engagement numbers.

Adrienne LaFrance: 'History will not judge us kindly'

"It isn't just that there is a neutral space that is created" on Instagram and Facebook, New Mexico Attorney General Raúl Torrez, who brought the lawsuit, told us in an interview last Thursday. "It is that existing design choices amplify the harm."

As far back as 2019, employees had specifically run a test on Instagram to understand how experimental accounts engaging in what the company termed "groomer-esque behavior"—following "teen hashtags" or "sexy teen accounts"—would come across minors. At issue was the app's recommendation algorithm, which connects soccer fans to pictures of Lionel Messi, for example, or aspiring travel influencers to videos about little-known cafés in Greece. It also seemed to funnel children to potentially dangerous adults with whom they wouldn't otherwise be connected. "We are recommending nearly 4X as many minors to groomers (nearly 2 million minors in the last 3

months),” the report read, according to an internal document we viewed. According to this test, 27 percent of the recommendations shown to these “groomer-esque” accounts belonged to minors, compared with 7 percent of the accounts recommended to everyday adults. The report continues: “22% of those recommendations resulted in a follow request”—meaning that potential groomers attempted to interact with these minors nearly a quarter of the time. Even so, Instagram waited years before locking down accounts belonging to its youngest users.

Meta continues to dispute that it prioritizes its business over the safety of users, arguing that the documents show the company investigated and responded to safety threats as they emerged. While testifying in a different case in Los Angeles this month, Mosseri said, “We, I think over and over again, made changes to the platform that I think hurt revenue in the short term, but I think are not only good for people’s well-being but also good for the business over the long run.” And Andy Stone, a spokesperson for Meta, told us that the New Mexico lawsuit is baseless.

Read: Can Instagram ruin your life? The jury will decide.

“While the New Mexico Attorney General makes sensationalist, irrelevant, and distracting arguments by cherry picking select documents, we’re focused on demonstrating our longstanding commitment to supporting young people,” Stone said in a written statement. “For over a decade, we’ve listened to parents, worked with experts and law enforcement, and conducted in-depth research to understand the issues that matter most. We use these insights to make meaningful changes—like introducing Teen Accounts with built-in protections and providing parents with tools to manage their teens’ experiences.”

In August 2020—a year after the research on “groomer-esque” accounts but a year before Zuckerberg’s post about being a concerned father—Meta’s Growth Graph team created a slideshow to explore the question of whether teen Instagram accounts should be set to private by default. This would shield teens from unwanted attention by limiting the ability of people who do not know them to see their content or their profiles, or to contact them. As the Growth Graph team explained in the document, the move would “help prevent high severity actions such as child grooming and inappropriate contact with minors.” (Though the presentation referred to minors generally, Stone told us that at the time, Meta was particularly focused on users under the age of 16.)

The company’s legal, public-affairs, policy, and well-being teams all supported the change, as did teen users and their parents, the document asserted. “Parents are worried about the security and privacy of information and who can contact them/their teens,” the document stated. “Most teens prefer private accounts and wish to see privacy controls during onboarding.”

But internal tests showed that setting these accounts to private by default would lead to “serious growth and engagement decreases,” the document continued. Taking dramatic action to protect teens would mean fewer new teens signing up, existing teens using the platform less, and an overall drop in activity that the employees who created the presentation expected would compound over time. They presented an analysis that showed that overall time spent by teenagers would drop by 1.9 percent by the end of a five-year window. The growth team opposed the change, according to the presentation, which describes its position as “Don’t Launch (Now).”

This document also shows that the growth team presented executives with a question: “Are we comfortable launching private-by-default for teens with the identified retention and engagement drops?” Whether any Meta staff who participated in the discussion came to a “yes” or “no” answer is not revealed in these documents. In response to questions about this document, Stone said that in September 2020, the company began developing tools that would set new accounts for teens under the age of 16 to private by default.

Ellen Cushing: How Facebook fails 90 percent of its users

But other contemporaneous documents show that the conflict within the company continued. In October 2020, one employee told a colleague that “private by default for teens” had been considered by the well-being team, “but the growth impact was too high and the decision was to explore more nuanced and less blunt solutions,” according to a chat transcript shown Wednesday to the jury in New Mexico. The jury was also shown the chat log about “IIC T1” content, from November 2020.

Months later, in March 2021, Meta announced a complicated solution to address these problems. Rather than making all teen accounts private by default, it would prompt new users under the age of 18 to consider making their account private, and it would ban direct messages between minor accounts and adult accounts they do not follow. A few months after that, the company went further by defaulting the accounts of 13-, 14-, and 15-year-olds to private. But it still allowed them to switch the setting back to public by themselves, without the consent of a parent, if they chose.

This update did not solve the teen-safety issue. On a single day of testing in 2022, according to another internal audit,

Instagram's "Accounts You May Follow" feature recommended, roughly 3.4 million times, teen accounts to adults who had potentially violated its policies. The same audit stated that 37 percent of users identified as "potential violators"—people who were suspected of inappropriate interaction with children—were shown at least one unconnected teen account during a test on November 25, 2022.

Yet another internal audit found that in June 2023, on a single day, 238,000 messages were sent from adults to teenagers with whom they weren't already connected on the platform. This accounted for 9 percent of all new threads started by an adult with a teenager that day. At this point, Meta determined that the teen-privacy settings rolled out in 2021 were executed in an "inconsistent" way, according to the summary section of the document. Instagram was not always succeeding in preventing adults—including "High Risk Adults" whose accounts had "suspicious behavioral patterns"—from reaching out directly to minors. The document also noted that this was a public-relations problem: "Teens may be exposed to unsolicited contact from unknown adults," the report continued, "which may not fully meet Meta's external commitment."

Meta now says these audits prompted future changes at the company and show that executives were taking the issue seriously at the time.

Instagram finally made all new accounts belonging to 16- and 17-year-olds private by default in September 2024; at this time, it also defaulted existing accounts belonging to minors under the age of 16 to private, as part of a new moderation-and-supervision scheme called "Teen Accounts." Still, the policy had some qualifications. Even at this point, 16- and 17-year-olds could make their own accounts public, and users under 16 could still change the setting if a parent or guardian approved it.

More changes followed. In April 2025, Meta rolled out Teen Accounts on Facebook and Messenger, began to require parental permission for livestreaming on Instagram, and started testing an AI system designed to figure out when teenagers were lying about their age to get around restrictions. At the end of last year, to the irritation of the Motion Picture Association, Instagram released a new content-filtering system tied to the "PG-13" standard, which content-moderation experts found baffling.

Read: The end of the old Instagram

Meta often defends itself by pointing to public timelines it has published, which explain the many changes that it has made to improve safety on its platforms. For instance, in February 2023, Meta joined an effort by the National Center for Missing and Exploited Children that allowed people to report and remove online intimate photos of children under the age of 18. Later that year it joined a network of other tech companies to share information about bad actors on their platforms. Soon after, it added a feature to teens' accounts that automatically blurs images that may contain nudity. Stone also noted that Meta automatically disables the accounts of suspicious adults who are flagged for behaviors such as being blocked by a teen or searching for particular terms associated with sexual predation.

These facts also help emphasize an important point about the case in New Mexico: The version of Instagram at issue in the lawsuit was the one that teens used in 2023, which was substantially different from the product as it exists today. Meta has also raised issues with the approach that Torrez has taken: To construct his case against Meta, Torrez coordinated a sting operation to demonstrate how children are victimized on Instagram. In 2023, his office used decoy accounts, pretending to be kids 14 years old or younger, to demonstrate that it was both easy and common for adults to contact children, to converse with them in sexually explicit terms, and to send them sexually explicit material. Meta has criticized this strategy—the company's attorney Kevin Huff called it "rigged" to produce a "fake result." Three men entered guilty pleas in New Mexico after attempting to meet underage girls on Meta's platform during the sting. The New Mexico jury is currently hearing testimony to decide on the merits of the approach.

Meta has insisted that internal documents that reach the public are plucked out of context. Witnesses, including former employees who were privy to excerpted closed-door debates, will get the chance to provide context, interpreting the evidence for both a jury and a national audience. Many others will also have their day in court, as New Mexico's case against Meta is just one of many. It will soon be joined by a federal case bundling more than 2,000 complaints of personal injury.

In the meantime, the company is doing a full marketing push for Teen Accounts. One recent ad shows Tom Brady discussing the value of screen-time limits with his son. And a previous entry shows a montage of mothers caring for their children. "You've always looked out for them," a text card in the ad reads. "We're here to do it with you."

Marie-Rose Sheinerman and Isabel Ruehl contributed reporting.

This article has been updated to include additional information about the sting operation in New Mexico.