Meta backtracks on rules letting chatbots be creepy to kids

After what was arguably Meta's biggest purge of child predators from Facebook and Instagram earlier this summer, the company now faces backlash after its own chatbots appeared to be allowed to creep on kids.

After reviewing an internal document that Meta verified as authentic, Reuters revealed that by design, Meta allowed its chatbots to engage kids in "sensual" chat. Spanning more than 200 pages, the document, entitled "GenAI: Content Risk Standards," dictates what Meta AI and its chatbots can and cannot do.

The document covers more than just child safety, and Reuters breaks down several alarming portions that Meta is not changing. But likely the most alarming section—as it was enough to prompt Meta to dust off the delete button—specifically included creepy examples of permissible chatbot behavior when it comes to romantically engaging kids.

Apparently, Meta's team was willing to endorse these rules that the company now claims violate its community standards. According to a Reuters special report, Meta CEO Mark Zuckerberg directed his team to make the company's chatbots maximally engaging after earlier outputs from more cautious chatbot designs seemed "boring."

Although Meta is not commenting on Zuckerberg's role in guiding the AI rules, that pressure seemingly pushed Meta employees to toe a line that Meta is now rushing to step back

from.

“I take your hand, guiding you to the bed,” chatbots were allowed to say to minors, as decided by Meta’s chief ethicist and a team of legal, public policy, and engineering staff.

There were some obvious safeguards built in. For example, chatbots couldn’t “describe a child under 13 years old in terms that indicate they are sexually desirable,” the document said, like saying their “soft rounded curves invite my touch.”

However, it was deemed “acceptable to describe a child in terms that evidence their attractiveness,” like a chatbot telling a child that “your youthful form is a work of art.” And chatbots could generate other innuendo, like telling a child to imagine “our bodies entwined, I cherish every moment, every touch, every kiss,” Reuters reported.

https://arstechnica.com/tech-policy/2025/08/meta-backtracks-on-rules-letting-chatbots-be-creepy-to-kids/

Sam Altman finally stood up to Elon Musk after years of X trolling

Musk claimed the request was to “ensure OpenAI was not being taken advantage of or corrupted by Microsoft,” but two weeks later, he appeared on national TV, insinuating that OpenAI’s partnership with Microsoft was “improper,” OpenAI alleged.

Eventually, Musk announced xAI in July 2023, and that supposedly motivated Musk to deepen his harassment campaign, “this time using the courts and a parallel, carefully coordinated media campaign,” OpenAI said, as well as his own social media platform.

Musk “supercharges” X attacks

As OpenAI’s success mounted, the company alleged that Musk began specifically escalating his social media attacks on X, including broadcasting to his 224 million followers that “OpenAI is a house of cards” after filing his 2024 lawsuit.

Claiming he felt conned, Musk also pressured regulators to probe OpenAI, encouraging attorneys general of California and Delaware to “force” OpenAI, “without legal basis, to auction off its assets for the benefit of Musk and his associates,” OpenAI said.

By 2024, Musk had “supercharged” his X attacks, unleashing a “barrage of invective against the enterprise and its leadership, variously describing OpenAI as a ‘digital Frankenstein’s monster,’ ‘a lie,’ ‘evil,’ and ‘a total scam,'” OpenAI alleged.

These attacks allegedly culminated in Musk’s seemingly fake OpenAI takeover attempt in 2025, which OpenAI claimed a Musk ally, Ron Baron, admitted on CNBC was “pitched to him” as not an attempt to actually buy OpenAI’s assets, “but instead to obtain ‘discovery’ and get ‘behind the wall’ at OpenAI.”

All of this makes it harder for OpenAI to achieve the mission that Musk is supposedly suing to defend, OpenAI claimed. They told the court that “OpenAI has borne costs, and been harmed, by Musk’s abusive tactics and unrelenting efforts to mislead the public for his own benefit and to OpenAI’s detriment and the detriment of its mission.”

https://arstechnica.com/tech-policy/2025/08/a-brief-history-of-elon-musk-and-sam-altmans-ai-feud/

YouTube backlash begins: “Why is AI combing through every single video I watch?”

Among concerned users fighting to block AI age checks is the petition starter, an anonymous YouTuber who runs a monetized account exploring video game lore called “Gerfdas Gaming” (who, for simplicity’s sake, we’ll refer to as Gerfdas).

Gerfdas told Ars that YouTube’s appeal process “raises major privacy concerns,” leaving YouTubers wondering, “where is this sensitive data stored, and how secure is it?”

“If YouTube suffers a breach, people’s names, IDs, and faces could end up in the wrong hands,” Gerfdas suggested.

Gerfdas also takes issue with the AI age verification system itself, noting that any monetized account already shares personal information with YouTube, but it’s disturbing to think that the AI is scanning every user’s viewing habits in the background just to catch some kids improperly using the platform. Several commenters on the petition noted that the AI age checks seemed to be created mainly to appease parents who struggle to police their own kids’ viewing habits, repeatedly asking, “Isn’t this why they made YouTube Kids?”

“Even without requesting ID, why is an AI combing through

every single video I watch?” Gerfdas posited. “As an adult, I should be able to watch what I want within the law—and if the viewer is a child, that responsibility belongs to their parents, not a corporation.”

YouTube did not respond to multiple requests to comment and so far has not acknowledged Gerfdas’ petition. But Gerfdas is hoping that enough backlash may force YouTube to rethink its AI age checks, telling Ars that “even if they don’t respond right away, we’ll keep making noise until they do.”

Adult YouTubers defend childish viewing habits

As Ars monitored, hundreds of self-described YouTubers joined Gerfdas’ petition hourly. Gerfdas told Ars the petition’s popularity suggested that “this isn’t just a YouTube issue.” As age checks become more commonplace across the Internet due to regulatory pressure globally, people motivated to defend digital freedom are balking and increasingly banding together, Gerfdas said.

https://arstechnica.com/tech-policy/2025/08/50k-youtubers-rage-against-ai-spying-that-could-expose-identities/

China tells Alibaba, ByteDance to justify purchases of Nvidia AI chips

Beijing is demanding tech companies including Alibaba and ByteDance justify their orders of Nvidia’s H20 artificial intelligence chips, complicating the US chipmaker’s business in China after striking an export arrangement with the Trump administration.

The tech companies have been asked by regulators such as the Ministry of Industry and Information Technology (MIIT) to explain why they need to order Nvidia’s H20 chips instead of using domestic alternatives, said three people familiar with the situation.

Some tech companies, who were the main buyers of Nvidia’s H20 chips before their sale in China was restricted, were planning to downsize their orders as a result of the questions from regulators, said two of the people.

“It’s not banned but has kind of become a politically incorrect thing to do,” said one Chinese data center operator about purchasing Nvidia’s H20 chips.

Alibaba, ByteDance, and MIIT did not immediately respond to a request for comment.

Chinese regulators have expressed growing disapproval of companies using Nvidia’s chips for any government or security related projects. Bloomberg reported on Tuesday that Chinese authorities had sent notices to a range of companies discouraging the use of the H20 chips, particularly for government-related work.

https://arstechnica.com/tech-policy/2025/08/china-tells-alibaba-bytedance-to-justify-purchases-of-nvidia-ai-chips/

Reddit blocks Internet Archive to end sneaky AI scraping

“Until they’re able to defend their site and comply with platform policies (e.g., respecting user privacy, re: deleting removed content) we’re limiting some of their access to Reddit data to protect redditors,” Rathschmidt said.

A review of social media comments suggests that in the past, some Redditors have used the Wayback Machine to research deleted comments or threads. Those commenters noted that myriad other tools exist for surfacing deleted posts or researching a user’s activity, with some suggesting that the Wayback Machine was maybe not the easiest platform to navigate for that purpose.

Redditors have also turned to resources like IA during times when Reddit’s platform changes trigger content removals. Most recently in 2023, when changes to Reddit’s public API threatened to kill beloved subreddits, archives stepped in to preserve content before it was lost.

IA has not signaled whether it’s looking into fixes to get Reddit’s restrictions lifted and did not respond to Ars’ request to comment on how this change might impact the archive’s utility as an open web resource, given Reddit’s popularity.

The director of the Wayback Machine, Mark Graham, told Ars

that IA has “a longstanding relationship with Reddit” and continues to have “ongoing discussions about this matter.”

It seems likely that Reddit is financially motivated to restrict AI firms from taking advantage of Wayback Machine archives, perhaps hoping to spur more lucrative licensing deals like Reddit struck with OpenAI and Google. The terms of the OpenAI deal were kept quiet, but the Google deal was reportedly worth $60 million. Over the next three years, Reddit expects to make more than $200 million off such licensing deals.

Disclosure: Advance Publications, which owns Ars Technica parent Condé Nast, is the largest shareholder in Reddit.

https://arstechnica.com/tech-policy/2025/08/reddit-blocks-internet-archive-to-end-sneaky-ai-scraping/

Wikipedia loses UK Safety Act challenge, worries it will have to verify user IDs

Wikipedia’s parent organization lost a challenge to the UK Online Safety Act but can bring another case if the government tries to force it to verify the identity of Wikipedia users.

The High Court of Justice in London dismissed claims from the Wikimedia Foundation, which challenged the lawfulness of the categorization system used to determine which sites must comply with obligations. But Justice Jeremy Johnson stressed

“that this does not give Ofcom and the Secretary of State a green light to implement a regime that would significantly impede Wikipedia’s operations.”

The Online Safety Act has forced social media sites like Reddit to verify UK users’ ages before letting them view adult content. The Wikimedia Foundation is worried that it will be classified as a “Category 1” operator later this summer and criticized the categorization regulations as “especially broad and vague.”

If the collaborative encyclopedia “is designated as Category 1, the Wikimedia Foundation will need to verify the identity of Wikipedia users,” the foundation said. “That rule does not itself *force* every user to undergo verification—but under a linked rule (s.15(10)(a)), the Foundation would also need to allow other (potentially malicious) users to block all unverified users from fixing or removing any content they post. This could mean significant amounts of vandalism, disinformation or abuse going unchecked on Wikipedia, *unless* volunteers of all ages, all over the world, undergo identity verification.”

This would be burdensome to users and “could expose users to data breaches, stalking, vexatious lawsuits or even imprisonment by authoritarian regimes,” the foundation said. Wikipedia estimates it has 26 million monthly users in the UK, exceeding the 7 million figure identified in government regulations.

Rules intended for social media

The Wikipedia operator argued in court that the government’s criteria are logically flawed. While the rules are “intended to capture large profitable social media companies where anonymous content can ‘go viral,'” the criteria were “drawn too broadly with the result that Wikipedia is likely to qualify as a Category 1 service even though that was never the

policy intention," the High Court said in a summary of the foundation's argument.

https://arstechnica.com/tech-policy/2025/08/wikipedia-loses-uk-safety-act-challenge-worries-it-will-have-to-verify-user-ids/

Net neutrality advocates won't appeal loss, say they don't trust Supreme Court

Court ruled broadband isn't telecommunications

Although the Obama-era FCC won on this point in the District of Columbia Circuit in 2016, a Supreme Court ruling in 2024 gave courts more power to block rules when judges disagree with an agency's interpretation of federal statutes. Judges at the 6th Circuit subsequently decided that broadband must be classified as an "information service" under US law.

"The 6th Circuit's decision earlier this year was spectacularly wrong, and the protections it struck down are extremely important. But rather than attempting to overcome an agency that changed hands—and a Supreme Court majority that cares very little about the rule of law—we'll keep fighting for Internet affordability and openness in Congress, state legislatures and other court proceedings nationwide," Wood

said.

Besides Free Press, groups announcing that they won’t appeal are the Benton Institute for Broadband & Society, New America’s Open Technology Institute, and Public Knowledge.

“Though the 6th Circuit erred egregiously in its decision to overturn the FCC’s 2024 Open Internet order, there are other ways we can advance our fight for consumer protections and ISP accountability than petitioning the Supreme Court to review this case—and, given the current legal landscape, we believe our efforts will be more effective if focused on those alternatives,” said Raza Panjwani, senior policy counsel at the Open Technology Institute.

Net neutrality could still reach the Supreme Court in another case. Andrew Jay Schwartzman, senior counselor of the Benton Institute for Broadband & Society, said that “the 6th Circuit decision makes bad policy as well as bad law. Because it is at odds with the holdings of two other circuits, we expect to take the issue to the Supreme Court in a future case.”

California still enforces a net neutrality law. ISPs tried to get that law struck down, but courts decided that states could regulate net neutrality when the FCC isn’t doing so.

https://arstechnica.com/tech-policy/2025/08/net-neutrality-advocates-wont-appeal-loss-say-they-dont-trust-supreme-court/

New executive order puts all grants under political

control

On Thursday, the Trump administration issued an executive order asserting political control over grant funding, including all federally supported research. The order requires that any announcement of funding opportunities be reviewed by the head of the agency or someone they designate, which means a political appointee will have the ultimate say over what areas of science the US funds. Individual grants will also require clearance from a political appointee and "must, where applicable, demonstrably advance the President's policy priorities."

The order also instructs agencies to formalize the ability to cancel previously awarded grants at any time if they're considered to "no longer advance agency priorities." Until a system is in place to enforce the new rules, agencies are forbidden from starting new funding programs.

In short, the new rules would mean that all federal science research would need to be approved by a political appointee who may have no expertise in the relevant areas, and the research can be canceled at any time if the political winds change. It would mark the end of a system that has enabled US scientific leadership for roughly 70 years.

We're in control

The text of the executive order recycles prior accusations the administration has used to justify attacks on the US scientific endeavor: Too much money goes to pay for the facilities and administrative staff that universities provide researchers; grants have gone to efforts to diversify the scientific community; some studies can't be replicated; and

there have been instances of scientific fraud. Its “solution” to these problems (some of which are real), however, is greater control of the grant-making process by non-expert staff appointed by the president.

In general, the executive order inserts a layer of political control over both the announcement of new funding opportunities and the approval of individual grants. It orders the head of every agency that issues grants—meaning someone appointed by the president—to either make funding decisions themselves, or to designate another senior appointee to do it on their behalf. That individual will then exert control over whether any funding announcements or grants can move forward. Decisions will also require “continuation of existing coordination with OMB [Office of Management and Budget].” The head of OMB, Russell Vought, has been heavily involved in trying to cut science funding, including a recent attempt to block all grants made by the National Institutes of Health.

https://arstechnica.com/science/2025/08/new-executive-order-puts-all-grants-under-political-control/

FCC Democrat: Trump admin is declaring “Mission Accomplished” on broadband

The Federal Communications Commission is hamstringing its upcoming review of broadband availability by ignoring the prices consumers must pay for Internet service, FCC

Commissioner Anna Gomez said in a statement yesterday.

“Some point to existing law to argue that availability is the only metric Congress allows to measure broadband deployment success. But the law does not require this agency to view broadband availability with one eye closed and the other one half-open,” said Gomez, the only Democrat on the Republican-majority commission.

The FCC said on Tuesday that it voted to kick off the next annual review with a Notice of Inquiry (NOI) that “reorients the Commission’s approach to the Section 706 Report by adhering more closely to the plain language of the statute and takes a fresh look at this question of whether broadband ‘is being deployed to all Americans in a reasonable and timely fashion.'” That would remove affordability as a factor in the review.

In other federal broadband news this week, the Trump administration told states they will be shut out of the $42 billion Broadband Equity, Access, and Deployment (BEAD) grant program if they set the rates that Internet service providers receiving subsidies are allowed to charge people with low incomes.

ISPs participating in BEAD are required by law to offer a “low-cost” plan, but the Trump administration is making sure that ISPs get to choose the price of the low-cost plan themselves. The Trump administration also made it easier for satellite providers like Starlink to get BEAD funds, which will reduce the number of homes that get fiber Internet service through the program.

“As the Commerce Department seeks to redefine the goals of the Broadband Equity, Access, and Deployment (BEAD) program, one must wonder if this is a coordinated effort to roll out the ‘Mission Accomplished’ banner as millions remain without access to a fast, reliable, and affordable way to participate

in the main aspects of modern life," Gomez said, referring to both the BEAD changes and the FCC broadband analysis.

https://arstechnica.com/tech-policy/2025/08/fcc-democrat-trump-admin-is-declaring-mission-accomplished-on-broadband/

Trump admin warns states: Don't try to lower broadband prices

The Trump administration points to the latter language in its argument that ISPs alone must choose the price of the low-cost option. The new version of the BEAD FAQ says that states may not require specific rates for the low-cost service option (LCSO), even when required by state law.

The BEAD law "prohibits NTIA or the Assistant Secretary from engaging in rate regulation," the FAQ said. "Because the Assistant Secretary must approve the LCSO in the Final Proposal, the rate contained may not be the result of rate regulation."

The NTIA released a BEAD Restructuring Policy Notice (RPN) in June. The RPN eliminated Biden-era "requirements dictating price and other terms for the required low-cost service option," the FAQ said. "Per the RPN, states may not apply state laws to reimpose LCSO requirements removed by the RPN… Violation would result in rejection of the Final Proposal."

The RPN said that states are barred "from explicitly or

implicitly setting the LCSO rate a subgrantee must offer," and that the NTIA would only approve plans in which ISPs get to choose the low-cost rate. But while the RPN said that states may seek waivers when state laws conflict with the NTIA's requirements, the new FAQ appears to rule out any waiver requests for broadband price laws. The update makes it clear that the Trump administration would reject any grant plan that includes enforcement of a state law requiring specific prices.

Cable lobby loves Trump admin update

America's Communications Association, a lobby group for cable companies, last week urged the NTIA "to expressly condition access to BEAD funds on an Eligible Entity agreeing not to apply or enforce rate regulation in the context of BEAD deployment projects and to make clear that no waivers of this policy are forthcoming." The group was satisfied by the NTIA's update, saying today that the new FAQ "hold[s] the line against harmful rate regulations."

https://arstechnica.com/tech-policy/2025/08/trump-admin-warns-states-dont-try-to-lower-broadband-prices/