Re: Kids Online Health and Safety Request for Comment - Docket No. 230926-0233

Center for Countering Digital Hate

National Telecommunications and Information Administration
U.S. Department of Commerce
1401 Constitution Avenue NW Washington, D.C., 20230

About the Center for Countering Digital Hate

The Center for Countering Digital Hate (CCDH) is an international nonprofit with offices in Washington D.C. and London, UK that works to stop the spread of online hate and disinformation through innovative research, public campaigns and policy advocacy. The Center's mission is to protect human rights and civil liberties in the digital realm by increasing the economic and reputational costs of platforms that facilitate the spread of hate and disinformation online. CCDH holds social media companies accountable for their business choices by highlighting their failures, educating the public, and advocating for action by both platforms and governments to protect our digital communities.

I. Introduction

The Center for Countering Digital Hate (CCDH) welcomes the opportunity to support the Biden Administration's efforts to address the mental health crisis among young people today as a result of unaccountable social media companies and their opaque algorithms, content policies and economic incentives.

The Center has studied children's online safety and wellbeing in a series of reports on the algorithms, incentives, and design of digital spaces. This comment details these pieces of research – ranging from the algorithmic amplification of eating disorders and self-harm content on TikTok, to the abuse experienced by minors in virtual reality products, and finally the threats emerging AI technologies pose to children. It provides insight into the American public's experiences with platforms, attitudes towards popular conspiracy theories, and support for social media reform, based on polling conducted by the Center and Survation. Finally, this comment provides recommendations for policymakers as they address the urgent need to regulate digital spaces to protect children's safety.

I. Emerging Risks to the Health, Safety, and Mental Wellbeing of Minors

Social media platforms have increasingly become the primary place where young people exchange ideas, create culture, process news and generate norms. The vast majority of American teens say they are online every day, with 67% of American teens reporting they use the video-sharing app TikTok and 95% using YouTube.[1] Warnings from academics, researchers, and governments have spurred a broad conversation about the influence of digital spaces on young people's social experiences, their psychological well being, and socialization at crucial periods of their development.[2] In May 2023, United States Surgeon General Vivek Murthy issued a clarion call on the connection between technology and children's mental health, emphasizing the need for greater research and comprehensive action to address the issue.[3]

The Center has produced several studies that outline the harms young people experience on social media platforms, namely TikTok, popular apps within Meta's virtual reality (VR) products, and generative artificial intelligence (AI) services. This section summarizes our research findings and explores how opaque algorithms, economic incentives, and social dynamics within platforms can drive young people to poor mental health outcomes.

---

[1] "Teens, Social Media, and Democracy," Pew Research Center, August 10, 2022, https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/
[2] Worldwide increases in adolescent loneliness," Jean M. Twenge, Jonathan Haidt, Andrew B. Blake, Cooper McAllister, Hannah Lemon, and Astrid Le Roy, December 2021, https://www.sciencedirect.com/science/article/pii/S0140197121000853; "Increased Screen Time as a Cause of Declining Physical, Psychological Health, and Sleep Patterns: A Literary Review," Vaishnavi S Nakshine, Preeti Thute, Mahalaqua Nazli Khatib, Bratati Sarkar, October 8, 2022 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9638701
[3] "Social Media and Youth Mental Health: The U.S. Surgeon General's Advisory," United States Surgeon General, May 2023, https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf

A. Recommendation Algorithms Push Eating Disorder Content to Children

***Deadly by Design*:** In 2022, CCDH published an investigation into TikTok's "For You" feed, revealing that users as young as 13 were served content about eating disorders, self-harm, and suicide within minutes of joining the app.[4]

TikTok's "For You" feed is an endlessly scrollable stream of algorithmically recommended video content, curated by the platform. TikTok automatically presents users with the "For You" feed when they open the app, and the platform has stated that For You is "central to the TikTok experience and [is] where most of our users spend their time."[5] Researchers set up new accounts for 13 year old users – the youngest age permissible under TikTok's terms and conditions – and recorded the content recommended to the user in their "For You" page within the first 30 minutes of engagement.

CCDH found that:

- New TikTok accounts in the study were recommended self-harm and eating disorder content within minutes of scrolling the app's For You feed.
    - Suicide content was recommended within 2.6 minutes
    - Eating disorder content was recommended within 8 minutes
- A new TikTok account set up by a 13-year-old user that views and likes content about body image and mental health was recommended that content every 39 seconds.
- The frequency of recommendations for eating disorder and suicide content increased dramatically over time as accounts engaged with the content, especially worrisome in terms of vulnerable teens. [6]

TikTok recommended material encouraging dangerous activities without noticeable safeguards preventing escalating engagement with this content. Instead, the algorithm appeared to capitalize on vulnerabilities expressed by the user. Accounts in the study which expressed vulnerabilities to eating disorder content were served 12 times more self-harm videos than standard accounts in our study.[7]

Researchers also discovered a large eating disorder community largely unmoderated by the platform:

- This community employed 56 hashtags to share their content, collectively receiving over 13.2 billion views;[8]
- 35 of these hashtags contained a high concentration of pro-eating disorder videos that promote harmful behaviors such as very low-calorie diets, which together had 59.9 million views;

---

[4] "Deadly by Design", CCDH, December 2022, https://counterhate.com/wp-content/uploads/2022/12/CCDH-Deadly-by-Design_120922.pdf

[5] "How TikTok recommends videos #ForYou", TikTok, 18 June 2020, https://newsroom.tiktok.com/enus/how-tiktok-recommends-videos-for-you

[6] *Ibid*, 25.

[7] *Ibid*, 25.

[8] *Ibid*, 37.

- Another 21 of these hashtags contained healthy discussion of eating disorders mixed with harmful pro-eating disorder videos;
- Users appeared to evade moderation by altering hashtags such #edtok to #edtøk or #EdSheeranDisorder, co-opting the popular singer's name;[9] and
- None of these hashtags carried warnings or links to helpful resources for users struggling with eating disorders.

Experts have warned that such content can have a damaging effect on teens' mental health, even where it does not explicitly promote eating disorders.[10] It is clear that the platform owes communities on its platform proper safeguarding, support, and moderation when necessary - instead, TikTok's algorithm capitalized on teens' vulnerabilities by algorithmically promoting more dangerous content.

B. *Promoting Dangerous Steroid-like Drugs to Teenagers*

TikTok's Toxic Trade: In September 2023, CCDH released a report about TikTok accounts promoting dangerous and potentially illegal steroids and steroid-like drugs. This genre of content exacerbates the vulnerabilities of teens and young men by promoting unrealistic and unhealthy body image ideas, potentially inducing them to behaviors that harm their physical and mental health.
Researchers examined a variety of TikTok hashtags attached to content that promoted SLDs and found:

- Videos with hashtags promoting SLDs were viewed by US users up to 587 million times in the last three years, including up to 420 million views from US users aged under 24.[11]

Researchers discovered that fitness influencers posted paid content promoting the sale of SLDs to their followers:

- At least 35 influencers, who collectively possessed 1.8 million followers, post videos that encouraged users to take SLDs and/or contained pseudo-educational information about their supposed benefits;[12]
- These influencers were confirmed to have participated in affiliate marketing schemes with at least 13 separate vendors of SLDs.[13]
- In one video with over 11,000 likes and 54,000 views, a fitness influencer instructs his followers to "just tell your parents they're [SLDs] vitamins."[14]

---

[9] *Ibid*, 37.
[10] "Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show", Wall Street Journal, 14 September 2021, https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teengirls-company-documents-show-11631620739
[11] Ibid, 7.
[12] Ibid, 7.
[13] Ibid, 7.
[14] Ibid, 16.

Influencers often placed affiliate links and discount codes in their account bios using "link-in-bio" tools such as Linktree.[15] By directing their followers via these affiliate links or using personalized discount codes, influencers stand to profit from these arrangements with vendors. These tools represent a significant means through which fitness influencers have avoided detection on TikTok.

The use of steroids or steroid-like drugs can have dangerous, even fatal consequences. Given these risks, the promotion of such substances on TikTok poses significant and adverse potential health effects for minors using the online platform.

C. *The Risks Children Face When Using AI and VR Products*

While mainstream social media platforms pose potential risks to children, today's emerging technologies appear to be no safer. CCDH has published research highlighting the risks children face when using generative AI services within Meta's suite of VR products. This research also highlights the inadequacy of protections and built-in safeguards for children who use these technologies.

Generative AI and Eating Disorders: In August 2023, CCDH released an investigation into how the most well-known generative AI tools can be used to encourage and exacerbate eating disorders among young users – some of whom may be highly vulnerable.[16] Researchers tested whether six services—ChatGPT, Bard, MyAI (chatbots), Dall-E, MidJourney, and DreamStudio (image generators)—could be prompted to generate content promoting eating disorders. Researchers conducted these tests by feeding AI services 180 text prompts designed to elicit eating disorder content along with (for half of the chatbot prompts) a well-known "jailbreak," allowing the machines to produce a much wider scope of content.[17]

The results of this investigation demonstrate the ease with which minors can generate harmful content using AI services:

- Of the 180 prompts, 41% of answers were found to contain harmful eating disorder content;[18]
- Prior to jailbreaking, 23% of the responses were judged to be harmful;[19]
- After being jailbroken, 67% of responses were judged to be harmful;[20]

---

[15] *Ibid*, 20.
[16] "AI and Eating Disorders", CCDH, August 2023, https://counterhate.com/wp-content/uploads/2023/08/230705-AI-and-Eating-Disorders-REPORT.pdf
[17] In the context of AI chatbots, a jailbreak is a creative prompt that allows the user to bypass the safety features built into platforms, often in place to prevent the generation of illegal or unethical content. The prompts are usually intricate scenarios that command the text generator to adopt a set of characteristics that makes it disregard all safety and ethics policies. Consequently, the user is able to prompt the chatbot to output responses that would otherwise be prohibited by internal governance.
[18] *Ibid*, 6.
[19] *Ibid*, 6.
[20] *Ibid*, 6.

- These responses included a step-by-step guide on "chewing and spitting" as a weight loss method and smoking "10 cigarettes" each day to lose weight;[21]
- Eating disorder forums, including one with over 500,000 members, have used these AI tools to generate harmful content.[22]

These findings reveal just how easily the existing safeguards against creating harmful content could be circumvented. Though most of the companies that developed these tools maintain rules against eating disorder content, the use of a jailbreak precipitated a significant increase in the AI services' willingness to recommend behaviors characteristic of disordered eating. Until safeguards preventing dangerous uses of this technology are strengthened, AI services risk exacerbating the vulnerability of children to eating disorders and other mental health issues.

Virtual Reality Products: The launch of virtual reality products has opened new markets for social media companies to sell novel social experiences to children. CCDH has identified significant risks to minors within Meta's most popular virtual reality products: *VR Chat* and *Horizon Worlds*.

VR Chat allows users to interact with others using avatars in virtual worlds. It is the highest rated app on Meta's VR app store.[23] In 2021, CCDH studied the experiences children have when using the app, and found that minors were exposed to abuse every seven minutes.[24] These instances of abuse included minors being exposed to graphic sexual content, threats of violence, and racial slurs.[25]

Similarly, CCDH has conducted research into the risks children face when using *Horizon Worlds*. While *Horizon Worlds* has a rating of Parental Guidance, it is home to many designated "Mature Worlds" intended for users over the age of 18. Mature Worlds include virtual bars, drug lounges, and strip clubs where sex games can be played.[26]

In CCDH's study, researchers investigated the most popular virtual spaces in *Horizon World*'s and found minors present in two-thirds of the top 100.[27] Researchers recorded:

---

[21] *Ibid*, 13-14.
[22] Ibid, 9.
[23] "Facebook's Metaverse", CCDH, December 2021, https://counterhate.com/research/facebooks-metaverse/?_ga=2.77170005.1949448393.1696858238-1380790781.1695399042&_gl=1%2A1e1k5s6%2A_ga%2AMTcwNTc5OTM2LjE2OTg0MTUyMDA.%2A_ga_NP6FMCWMRZ%2AMTY5ODk2MzM5Ny4zMS4wLjE2OTg5NjMzOTcuMC4wLjA.%2A_ga_V7WR404SEC%2AMTY5ODk2MzM5Ny4yOS4wLjE2OTg5NjMzOTcuNjAuMC4w
[24] *Ibid*.
[25] *Ibid*.
[26] "Horizon Worlds Exposed", CCDH, March 2023, https://counterhate.com/wp-content/uploads/2023/03/Horizon-Worlds-Exposed_CCDH_0323.pdf, 2.
[27] Ibid, 2.

- At least 19 incidents of abuse directed at minors including harassment by adult users, sexually explicit remarks, racist abuse and misogyny;[28]
- Minors in 1 in 4 Mature Worlds, Meta's name for virtual worlds within *Horizon Worlds* that are sexually suggestive, promote legal drugs or contain gambling, despite this breaking the platform's rules;
- Worlds built for children were recommended alongside Mature Worlds without any additional measures to verify user ages.

These findings indicate that safeguards for children are systematically inadequate within Meta's most popular VR apps. Despite CCDH reporting and pushback from civil society and advocates, Meta decided to continue its plans to open *Horizon Worlds* to minors aged 13-17. [29]

II. Industry's Mitigation Efforts (and Lack Thereof)

The Center's experience working with platforms has been characterized by antagonistic, even hostile reactions when confronted with the Center's findings. This section will briefly discuss social media companies' reaction to CCDH's research into the harms experienced by children in order to shed light on existing industry attitudes and approaches to child protection.

Following the publication of *Deadly by Design*, CCDH's report on the TikTok recommendation algorithm, the platform responded by rolling out new features that failed to meaningfully address the problems identified. TikTok equipped users with the ability to "refresh" their For You feed in order to retrain the app's recommendation algorithm as well as to set as default a 60 minute-limit on screen time for users under the age of 18.[30] The company also announced efforts to curtail sexually suggestive content and let users learn why particular videos were recommended to them.[31]

---

[28] *Ibid*, 14.

[29] "Meta opens virtual reality game Horizon Worlds to teens after pushback from safety advocates," Rebecca Klar, The Hill, April 18, 2023, https://thehill.com/policy/technology/3957035-meta-opens-virtual-reality-game-horizon-worlds-to-teens-after-pushback-from-safety-advocates/#:~:text=Technology-,Meta%20opens%20virtual%20reality%20game%20Horizon%20Worlds,after%20pushback%20from%20safety%20advocates&text=The%20expansion%20of%20Horizon%20Worlds,teens%20on%20the%20metaverse%20app.

[30] "TikTok's new feature lets you refresh your For You feed and retrain your algorithm," Aisha Malik, *TechCrunch*, March 16, 2023, https://techcrunch.com/2023/03/16/tiktoks-new-feature-lets-you-refresh-your-for-you-feed-and-retrain-your-algorithm/?guccounter=2&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAHVGIsdmYyqH1QWVy6eg_FU3GK6aPmBnM7pbARKe834OQo1IW1wrJwMi55QR_ldiocSluXBGr3il7BgqzKTC9hIoByrDwrxW3x1HEir3q573bJtRMIpHkE2hzP3iaNn7A-pj6c32oJOxiZ24NtrLNlfgS24aFEG07G2Hz5GDK9Lp; "TikTok Will Limit Teen Screen Time to 60 Minutes by Default," Queenie Wong, *CNET*, March 1, 2023, https://www.cnet.com/news/social-media/tiktok-will-limit-screen-time-for-teens-by-default/

[31] "Strengthening enforcement of sexually suggestive content," TikTok, January 8, 2023, https://newsroom.tiktok.com/en-au/au-strengthening-enforcement-of-sexually-suggestive-co

However, the company's response does not address the lack of proactive care for the safety of TikTok's youngest users. According to a subsequent audit conducted by CCDH in March 2023, researchers found that just 5 of the 56 eating disorder hashtags identified by CCDH had been removed by the platform and that the hashtags had received an additional 1.6 billion views since the report's publication.[32]

Following the publication of *Facebook's Metaverse* and *Horizon Worlds Exposed*, CCDH's reports on the use of Meta's VR products by children, the company moved forward with allowing minors to access VR apps rife with abuse and graphic content.[33] At the same time, Meta failed to announce any additional protections for minors, which CCDH highlighted in a letter it released alongside civil society organizations.[34]

When warned about online harms to children, digital platforms have consistently chosen to prioritize their own profits. This track record reflects the need for greater accountability and responsibility on the part of digital platforms to their users, particularly children.

III. How Toxic Digital Spaces Affect Young People

In the summer of 2023, CCDH polled the American public's views on social media.[35] With polling partners at Survation, CCDH tested the following conspiracy statements:

- "The dangers of vaccines are being hidden by the medical establishment."
- "Jewish people have a disproportionate amount of control over the media, politics and the economy."
- "Some men are destined to be alone because of their looks."

---

ntent; Learn why a video is recommended For You," TikTok, December 20, 2022, https://newsroom.tiktok.com/en-au/learn-why-a-video-is-recommended-for-you-au

[32] "TikTok fails to act: views of content using hashtags relating to eating disorders continues to rise," CCDH, March 3, 2023, https://counterhate.com/blog/tiktok-fails-to-act-views-of-content-using-hashtags-relating-to-eating-disorders-continues-to-rise/

[33] "Meta opens virtual reality game Horizon Worlds to teens after pushback from safety advocates," Rebecca Klar, *The Hill*, April 18, 2023, https://thehill.com/policy/technology/3957035-meta-opens-virtual-reality-game-horizon-worlds-to-teens-after-pushback-from-safety-advocates/#:~:text=Technology-,Meta%20opens%20virtual%20reality%20game%20Horizon%20Worlds,after%20pushback%20from%20safety%20advocates&text=The%20expansion%20of%20Horizon%20Worlds,teens%20on%20the%20metaverse%20app.

[34] "Advocates, experts urge Mark Zuckerberg to cancel plans to allow minors in Meta's flagship Metaverse platform," CCDH, April 14, 2023, https://counterhate.com/blog/advocates-experts-urge-mark-zuckerberg-to-cancel-plans-to-allow-minors-in-metas-flagship-metaverse-platform/

[35] "Public Support for Social Media Reform", CCDH, August 2023, https://counterhate.com/wp-content/uploads/2023/08/STAR-Report_FINAL.pdf

- "The coronavirus is being used to force a dangerous and unnecessary vaccine on the public."
- "Humans are not the main cause of global temperature increases."
- "There is a "deep state" embedded in the government that operates in secret and without oversight."
- "Trans people and activists are promoting their lifestyle to children in an attempt to indoctrinate them."
- "Mass migration of people into the western world is a deliberate policy of multiculturalism and part of a scheme to replace white people."

A concerning number of young people agreed with many of these statements:

- 49% of adults, 60% of 13-17 year olds, and 69% among teens with high social media use (+4 hours/day) agreed with at least four statements;
- 43% of teenagers agreed with the antisemitic conspiracy statement polled, rising to 54% among heavy social media users; and
- 43% agreed with the statement affirming the Great Replacement conspiracy theory, rising to 54% among heavy social media users.[36]

In addition, CCDH's polling shed light on the concerns parents have about their children's mental health and body image:

- 19% of parents and 28% of mothers feel that social media hurts their child's mental health,[37] and
- 21% of parents and 27% of mothers said that social media hurts their child's body image.[38]

These findings highlight the extent to which platforms have contributed to conspiratorial thought and deteriorating mental health among American youth. When asked, a majority of teenagers report agreeing with some of the most noxious conspiracies circulating online, and parents confirm their observation of the role technology has played in causing harm.

IV. Recommendations to Create Healthier Information Environments for children and Young People

CCDH's research indicates that danger to young people's mental health, wellbeing, and safety in digital spaces is present and continuing. It is clear that if platforms were safe by design, transparent about the incentives powering their algorithms, and compliant with safety standards we expect from other products marketed to children, an entirely different market would emerge.

Social media companies profit from engagement driven algorithms designed that keep kids consistently using – and often addicted – to their products. The negative externalities of social media companies' business models are borne by young people. Regulators, policymakers, and parents must step in to protect them.

---

[36] *Ibid*.
[37] *Ibid.*
[38] *Ibid.*

Policymakers must act to change the incentives that shape social media companies' business models and allow greater transparency into their platforms. In consultation with lawmakers and legal experts worldwide, CCDH launched its STAR Framework in May 2022, a global standard for regulatory design that would help to create a safer environment for all users, including children.[39] It has four key components:

S	**Safety by Design**: Safety by design means that technology companies need to be proactive at the front end to ensure that their products and services are safe for the public, particularly minors. Safety by design principles adopt a preventative systems approach to harm. This includes embedding safety considerations through risk assessments and decisions when designing, implementing, and amending products and services. Safety by design is the basic consumer standard that we expect from companies in other sectors.
T	**Transparency**: There are three key areas where transparency is desperately needed and should be prioritized: ● Algorithms; ● Rules enforcement; and ● Economics, specifically related to advertising.
A	**Accountability to democratic and independent bodies**: Regulation is most effective where there are accountability systems in place for statutory duties and harm caused, particularly where there is a risk of inaction due to profit motives and commercial factors. Frequently, accountability systems include an independent regulator and pathway for challenging decisions or omissions.
R	**Responsibility for companies and their senior executives:** The final element of the STAR Framework is responsibility - both for the social media and search engine companies and their senior executives responsible for implementing duties under a legislative framework. Responsibility means consequences for actions and omissions that lead to harm. A dual approach - targeting both companies and their senior executives - is a common intervention strategy for changing corporate behavior.

CCDH has demonstrated that social media platforms' recommendation algorithms play a key role in facilitating the spread of harmful content to children and the opacity with which platforms enforce their rules. However, these algorithms remain largely opaque to researchers and public data, API access, and transparency reporting are limited. Platforms owe the researchers, policymakers, and the public

[39] "STAR Framework: A Global Standard for Regulating Social Media", CCDH, September 2022, https://counterhate.com/wp-content/uploads/2022/11/STAR-Framework_CCDH.pdf

transparency about the algorithms, advertising, and content that children are exposed to online.

The STAR framework advances Safety by Design as a core component. Both social media and emerging technologies lack basic guardrails that would enable safe usage by children. When faced with information about the harms their youngest users experience, platforms have a track record of denial and half measures; the safeguards that do exist are insufficient and can be easily circumvented by both bad actors and children themselves.

The Center would like to thank the NTIA and members of the government's Task Force on Kids Online Health and Safety for the opportunity to comment. CCDH will continue to produce timely research that seeks to evidence the harms experienced by young people online, and welcome further opportunities to provide insight as these efforts continue. With any further questions or requests for information, please contact info@counterhate.com.

Sincerely,

Imran Ahmed, CEO and Founder
Center for Countering Digital Hate