Center for AI and Digital Policy

January 31, 2024

Chairman Dick Durbin
Ranking Member Lindsay Graham
U.S. Senate Committee on the Judiciary
G50 Dirksen Senate Office Building, Washington D.C. 20515

Re: CAIDP Statement for the Record: "Big Tech and the Online Child Sexual Exploitation Crisis"

Dear Chairman Durbin, Ranking Member Graham, and Members of the Committee,

We write to you regarding the hearing on "**Big Tech and the Online Child Sexual Exploitation Crisis.**"[1] We thank you for your continued leadership in protecting the American people from the risks of technology, holding Big Tech accountable particularly regarding online child safety. We write to highlight the serious risks to children further exacerbated through the unprecedented deployment of AI systems in children facing apps, online mediums, and social media platforms and urge you to take our recommendations into consideration.

The Center for AI and Digital Policy (CAIDP) is an independent research and education non-profit based in Washington, DC.[2] Our global network of AI policy experts and advocates advise national governments, international organizations, and Congressional committees regarding artificial intelligence and digital policy. Our President, Merve Hickok testified at the first congressional hearing on AI last year–"*Advances in AI: Are We Ready For a Tech Revolution*?"[3].

CAIDP routinely provides advice to Congressional Committees on matters involving AI policy[4], and we also submitted our statements to this Committee on "Oversight of A.I.: Rules for Artificial Intelligence"[5] and "Oversight of A.I.: Legislating on Artificial Intelligence".[6] We also publish the annual *Artificial Intelligence and Democratic Values Report*,[7] providing a comprehensive review of AI policies and practices in 75 countries.

---

[1] U.S. Senate Committee on the Judiciary, *Big Tech and the Online Child Sexual Exploitation Crisis,* (January 31, 2024), https://www.judiciary.senate.gov/committee-activity/hearings/big-tech-and-the-online-child-sexual-exploitation-crisis.

[2] CAIDP, *About*, https://www.caidp.org/about-2/.

[3] Testimony and statement for the record of CAIDP President Merve Hickok, *Advances in AI: Are We Ready For a Tech Revolution?*, House Committee on Oversight and Accountability, Subcommittee on Cybersecurity, Information Technology, and Government Innovation (March 8, 2023), https://oversight.house.gov/wp-content/uploads/2023/03/Merve-Hickok_testimony_March-8th-2023.pdf.

[4] CAIDP, *Statements, https://www.caidp.org/statements/*

[5] CAIDP, *Statement to Senate Judiciary Committee on Oversight of AI,* (May 11, 2023), https://www.caidp.org/app/download/8457859963/CAIDP-Statement-SJC-AI-05112023.pdf

[6] CAIDP, *Statement to Senate Judiciary Committee on U.S. AI Act,* (September 12, 2023), https://www.caidp.org/app/download/8475524463/CAIDP-SJC-09112023.pdf

[7] CAIDP, *Artificial Intelligence and Democratic Values* (2023), https://www.caidp.org/reports/aidv-2022/.

In brief, our recommendations to this Committee are:

1) Address the grave risks of generative AI in social media

2) Enact legislation establishing accountability measures for AI systems including mandating synthetic content labelling, and for protecting minors from AI-enabled sexual abuse, extortion, and pornography.

Recommendation 1. Address the risks of Generative AI in social media

The deployment of generative AI tools particularly in apps and social media, especially those targeted at children, is an alarming development and must be reined in by Congress. Generative AI refers to the ability for machines to generate new, incredibly realistic, and detailed text, code, images, audio, and more. The commercialization of generative AI tools and consumer facing applications has far outpaced that of social media.

It is well established by researchers, advocates, and government that social media companies thrive off hooking minors on their products.[8] <u>A recent study by the Harvard T.H. Chan School of Public Health estimated that ads targeted at American children generated nearly $11 billion in revenue in 2022 for just Meta, Snapchat, TikTok, X, and YouTube.</u>[9] As such, these companies have aggressively integrated various generative AI capabilities into their products, even though general-purpose AI chatbots like ChatGPT and Bard either require parental approval for teenagers or exclude them outright.[10]

For instance, Snapchat includes an OpenAI-powered chatbot assistant in its app–*My AI*, which is specifically designed to grip the app's primarily minor users. The chatbot is pinned to the top of a user's messages page and cannot be removed on the app's free version.[11] When a user messages My AI, the app-generated-bitmoji for the AI is shown as thinking and writing a response like a human messenger would. The chatbot routinely includes emojis in its text responses, uncharacteristic of most AI chatbots, but common among young people. These design choices seek to anthropomorphize the

---

[8] Melissa Henson, *Big Tech is preying on children for profit, and Congress needs to stop it*, The Hill (January 18, 2024), https://thehill.com/opinion/technology/4411420-big-tech-is-preying-on-children-for-profit-and-congress-needs-to-stop-it/.

[9] Harvard T.H. Chan School of Public Health, *Social media platforms generate billions in annual ad revenue for U.S. youth,* (December 27, 2023), https://www.hsph.harvard.edu/news/press-releases/social-media-platforms-generate-billions-in-annual-ad-revenue-from-u-s-youth/.

[10] Jeffrey R. Young, *Teens Need Parent Permission to Use ChatGPT. Could That Slow Its Use in Schools?*, EdSurge (November 2, 2023), https://www.edsurge.com/news/2023-11-02-teens-need-parent-permission-to-use-chatgpt-could-that-slow-its-use-in-schools.

[11] *How do I unpin or remove My AI with Snapchat+?*, Snap Inc., https://help.snapchat.com/hc/en-us/articles/13387249333780-How-do-I-unpin-or-remove-My-AI-with-Snapchat. Accessed January 30, 2024.

algorithm and deceive children,[12] many of whom will not understand what an AI is, AI's risks and harms, or how their conversations with the chatbot are used by Snap. Snap disclaims any liability in the following terms "My AI may include biased, incorrect, harmful or misleading content"[13]; and suggests that users should independently verify any advice it gives before acting on it.[14]

Indeed, this use can be, and already is, quite dangerous:

- ***Promoting dangerous behavior.*** As highlighted in an open letter from Senator Bennet to tech CEOs, Snapchat's My AI was used to instruct a child on how to cover up a bruise ahead of a visit from Child Protective Services and provided suggestions to a 13-year-old girl on how to lie to her parents about an upcoming trip with a 31-year-old man.[15]

- ***Amplification of hallucinations, bias, and extremism***. These well-documented AI risks[16] are especially problematic for young people. Image-generation tools like Stable Diffusion consistently generate white, male-centric images that could reinforce stereotypes and bias among children.[17] An unwitting child on WhatsApp could innocently prompt its image generation software with "Palestinian boy" and see violent, graphical images of children holding guns.[18]
  Other models like OpenAI's GPT-4V doesn't understand hate symbols like "the modern meaning of the Templar Cross (white supremacy) in the U.S." despite best-attempt safety training.[19] We must protect children from these harms, and this necessitates additional regulation of AI systems, and enforcement by the FTC, in certain contexts.

---

[12] Sharon Goldman, *Sen. Murphy's tweets on ChatGPT spark backlash from former White House AI policy advisor,* VentureBeat (March 28, 2023), https://venturebeat.com/ai/sen-murphys-tweets-on-chatgpt-spark-backlash-from-former-white-house-ai-policy-advisor/.

[13] Bernard Marr, *Snapchat Debuts ChatGPT – Powered Snap AI: But is it safe for kids?*, Forbes (April. 26, 2023), https://www.forbes.com/sites/bernardmarr/2023/04/26/snapchat-debuts-chatgpt- powered-snap-ai--but-is-it-safe-for-kids

[14] CAIDP, *In the matter of OpenAI*, Supplement to the Original Complaint, (July 10, 2023), pg. 26, https://www.caidp.org/cases/openai/

[15] *Following Early Reports of Potential Harmful Content from AI Chatbots, Bennet Urges Tech CEOs to Prioritize Young Americans' Safety*, Michael Bennet, U.S. Senator for Colorado (March 21, 2023), https://www.bennet.senate.gov/public/index.cfm/2023/3/bennet-calls-on-tech-companies-to-protect-kids-as-they-deploy-ai-chatbots.

[16] FTC, *Consumers Are Voicing Concerns About AI,* (Oct. 3, 2023), https://www.ftc.gov/policy/advocacy-research/tech-at-ftc/2023/10/consumers-are-voicing- concerns-about-ai

[17] Leonardo Nicoletti and Dina Bass, *Generative AI Takes Stereotypes and Bias From Bad to Worse*, Bloomberg (2023), https://www.bloomberg.com/graphics/2023-generative-ai-bias/.

[18] Johana Bhuiyan, *WhatsApp's AI shows gun-wielding children when prompted with 'Palestine'*, The Guardian (November 3, 2023), https://www.theguardian.com/technology/2023/nov/02/whatsapps-ai-palestine-kids-gun-gaza-bias-israel.

[19] OpenAI, *GPT-4V(ision) System Card*, https://cdn.openai.com/papers/GPTV_System_Card.pdf. *See also,* CAIDP, *In the matter of OpenAI*, Supplement to the Original Complaint, (November 14, 2023), pg. 16, https://www.caidp.org/cases/openai/

- ***Privacy Violations.*** AI will also exacerbate the already substantial privacy concerns regarding children's use of social media. The intrusive nature of chatbots and social media will invariably extract personal and sensitive information – from text conversations, images, and other information – that will then be used to train models and target advertisements and content at our children.[20] A report by the *Scientific American* shows that "OpenAI fine-tunes its models based on user interactions with its chatbots."[21]

- ***Mental Health Risks***. Chatbots and other generative AI systems present a natural attraction for children to use social media products more. However, the current use of social media is already a disaster for our children. Social media is a key driver of a mental health crisis among American youth,[22] resulting in 40% of teens feeling persistently sad or hopeless in 2021 and the American Academy of Pediatrics declaring a national emergency for child and adolescent mental health.[23]

We highlighted these and other risks in our complaint to the FTC against OpenAI:[24]

"ChatGPT has demonstrated negative effects of providing misinformation, incorrect advice, creation of bullying content. The inclusion of a conversational feature in social media apps used predominantly by children is likely to result in higher amounts of data scraped from children, increased screen-time and endanger online child safety." [25]

The FTC issued a civil investigative demand on OpenAI in July of last year but almost 1 year since the filing of the complaint, the FTC is yet to issue any meaningful order. History shows that the longer the FTC delays, the more difficult it is to establish the necessary guardrails. Inaction by the agency is costly.

---

[20] AI Review Team, *My AI*, Common Sense Media (October 19, 2023), https://www.commonsensemedia.org/ai-ratings/my-ai.

[21] Lauren Leffer, *Your Personal Information Is Probably Being Used to Train Generative AI Models, Scientific American,* (October 19, 2023), https://www.scientificamerican.com/article/your- personal-information-is-probably-being-used-to-train-generative-ai-models/ ; CAIDP, *In the matter of OpenAI*, Supplement to the Original Complaint, (November 14, 2023), https://www.caidp.org/cases/openai/

[22] Chris Griswold, *Big Tech Is Exploiting Kids Online. Congress Has To Step In*, Newsweek (November 6, 2023), https://www.newsweek.com/big-tech-exploiting-kids-online-congress-has-step-opinion-1840276.

[23] Moriah Balingit*, 'A cry for help': CDC warns of a steep decline in teen mental health*, The Washington Post (March 31, 2022), https://www.washingtonpost.com/education/2022/03/31/student-mental-health-decline-cdc/.

[24] CAIDP, *In the matter of OpenAI*, Supplement to the Original Complaint, (July 10, 2023), https://www.caidp.org/cases/openai/

[25] *Id,* at pg. 25, 26. *See also,* Open AI, *The GPT-4 System Card* (March 15, 2023), https://cdn.openai.com/papers/gpt-4-system-card.pdf; The Lancet*, ChatGPT: Friend or Foe,* Editorial (February 6, 2023), https://www.thelancet.com/action/showPdf

To address the risks of generative AI in social media, we urge this Committee to exercise its oversight authority and call upon the FTC to conclude its investigation into OpenAI. We also urge this Committee to exercise oversight on the FTC's enforcement plans in relation to consumer harms, arising out of the harmful, exploitative, and deceptive commercial practices of AI companies developing generative AI tools, and data partnerships with social media companies deploying them downstream.

Recommendation 2. Enact legislation establishing accountability measures for AI systems and prohibiting Sexual Deepfakes and Violence

Generative AI now has the capability to create extremely realistic image and video depictions and synthetic voice content of specific people, colloquially known as "deepfakes."[26] For this hearing, we specifically focus on the extreme risks posed by deepfakes to children and sexual exploitation.

Deepfakes are already used to construct nonconsensual pornographic content at alarming scale and ease,[27] and many are easily available online.[28] But the technology has no age restrictions. Indeed, many children – most, girls – are already falling victim to this vile process.[29]

The Internet Watch Foundation has already found nonconsensual deepfake sexual abuse images of famous children and new images of children who were abused in the past.[30] They also found that abusers were "sharing tips and marveling about how easy it was to turn their home computers into factories for generating sexually explicit images of children of all ages."[31]

However, children are also becoming perpetrators of these harms. Just a few months ago, teenage boys in New Jersey generated nonconsensual pornographic photos of their classmates using

---

[26] Daniel I. Weiner & Lawrence Norden, *Regulating AI Deepfakes and Synthetic Media in the Political Arena*, Brennan Center for Justice (December 5, 2023), https://www.brennancenter.org/our-work/research-reports/regulating-ai-deepfakes-and-synthetic-media-political-arena.

[27] Angela Chen, *Forget fake news—nearly all deepfakes are being made for porn*, MIT Technology Review (October 7, 2019), https://www.technologyreview.com/2019/10/07/132735/deepfake-porn-deeptrace-legislation-california-election-disinformation/.

[28] Kat Tenbarge, *Google and Bing put nonconsensual deepfake porn at the top of some search results,* NBC News (January 11, 2024), https://www.nbcnews.com/tech/internet/google-bing-deepfake-porn-image-celebrity-rcna130445.

[29] Matt O'Brien and Haleluya Hadero, *AI-generated child sexual abuse images could flood the internet. Now there are calls for action,* The Associated Press (October 25, 2023), https://apnews.com/article/ai-artificial-intelligence-child-sexual-abuse-c8f17de56d41f05f55286eb6177138d2.

[30] *Id.*

[31] *Id.*

AI.[32] But the victims had scant means for recourse, as "the incident would likely amount to nothing more than a 'cyber-type harassment claim,' rather than a more serious crime like child pornography."[33]

Social media can spread deepfakes far and wide, as recently occurred with singer-songwriter Taylor Swift,[34] furthering the harm for young people. Deepfakes have also been used to further "sextortion," blackmailing victims by threatening to share the manufactured images.[35] Sextortion guides are widely available on social media like TikTok, Instagram, Snapchat, and YouTube and are only becoming more prevalent.[36]

Congress must act swiftly and decisively on this issue. Although mainstream image generation products disallow sexually explicit uses, other software are widely available because the fundamental AI technology is open sourced.[37]

We encourage this Committee to further act on its mandate to enact legislation protecting minors from nonconsensual deepfake pornographic material. While there has been much discussion on the tension between first amendment rights and expanded criminalization of the creation, use, and dissemination of child sexual abuse material, the proliferation of AI systems raise unprecedented risks to online child safety that must be addressed through amending existing laws or enacting new laws.

We also urge this Committee to move forward with AI legislation. There have been several bills proposed in Congress, but none have moved forward. **We endorse the Blumenthal-Hawley Bi-Partisan Framework for a US AI Act.**[38]

---

[32] Tate Ryan-Mosley, *A high school's deepfake porn scandal is pushing US lawmakers into action,* MIT Technology Review (December 1, 2023), https://www.technologyreview.com/2023/12/01/1084164/deepfake-porn-scandal-pushing-us-lawmakers/.

[33] *Id.*

[34] *Taylor Swift deepfakes spread online, sparking outrage*, CBS News (January 26, 2024), https://www.cbsnews.com/news/taylor-swift-deepfakes-online-outrage-artificial-intelligence/.

[35] Jack Guy, *Outcry in Spain as artificial intelligence used to create fake naked images of underage girls,* CNN (September 20, 2023), https://www.cnn.com/2023/09/20/europe/spain-deepfake-images-investigation-scli-intl/index.html.

[36] Lora Kolodny, *Sextortion training materials found on TikTok, Instagram, Snapchat and YouTube, according to new report,* NBC News (January 27, 2024), https://www.nbcnews.com/tech/internet/sextortion-yahoo-boys-snapchat-tiktok-teen-wizz-rcna134200; David Finkelhor, Heather Turner, and Deirdre Colburn*, Prevalence of Online Sexual Offenses Against Children in the US,* JAMA Network Open (October 2022), https://doi.org/10.1001%2Fjamanetworkopen.2022.34471.

[37] O'Brien and Hadero, *supra* note 17.

[38] Senator Richard Blumenthal & Senator Josh Hawley, *Bipartisan Framework for U.S. AI Act*, https://www.blumenthal.senate.gov/imo/media/doc/09072023bipartisanaiframework.pdf

However, to address specifically, the risks to child safety, we recommend further refinements to the Blumenthal-Hawley framework and/or any other legislation this Committee would consider:

(i) a clear requirement to label all AI-generated images, audio, and video for ease of detection

(ii) private rights of action and/or criminal liability for creating, disseminating, or using nonconsensual deep[fake] [imag]e of minors

Thank you for your co[nsideration. We ask that this stat]ement be included in the hearing record. We would [be pleased to answer any questions or provide] additional information.
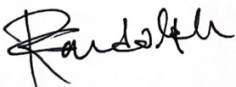
Sincerely yours,

Marc Rotenberg
CAIDP Executive Director

Merve Hickok
CAIDP President

Christabel Randolph
Law Fellow

Maanas Sharma
Research Assistant