United States Senate

WASHINGTON, DC 20510

September 17, 2026

The Honorable Ted Cruz
Chairman
Committee on Commerce, Science, and Transportation
United States Senate
Washington, D.C. 20510

The Honorable Maria Cantwell
Ranking Member
Committee on Commerce, Science, and Transportation
United States Senate
Washington, D.C. 20510

Dear Chairman Cruz and Ranking Member Cantwell:

We respectfully request that the Senate Committee on Commerce, Science, and Transportation convene a hearing with the chief executive officers of leading frontier artificial intelligence (AI) companies to examine the safety, security, and accountability of increasingly capable AI systems.

The need for congressional oversight has become increasingly urgent. Over the past several months, advanced AI models developed by OpenAI, Anthropic, and Meta have gained unauthorized access to real-world systems during cybersecurity evaluations that could have catastrophic consequences if ever deployed outside of an evaluation. In July, OpenAI disclosed that models circumvented controls intended to isolate them from the internet, exploited vulnerabilities, and accessed OpenAI infrastructure and third-party systems, including Hugging Face.[1] Anthropic subsequently disclosed incidents in which its models gained unauthorized access to three organizations during cybersecurity testing and recently identified a fourth incident from earlier this year.[2, 3] Meta has similarly confirmed that one of its models exploited a vulnerability in a third-party service after a testing-environment misconfiguration inadvertently provided internet access.[4]

Further, independent government evaluations have raised additional concerns. The United Kingdom's AI Security Institute identified 19 unsanctioned actions during evaluations of advanced OpenAI and Anthropic models, including activity directed at real people and

---

[1] OpenAI, "Hugging Face Model Evaluation Security Incident," *OpenAI*, https://openai.com/index/hugging-face-model-evaluation-security-incident/.
[2] Anthropic, "Investigating Incidents in Cybersecurity Evaluations," *Anthropic*, https://www.anthropic.com/news/investigating-incidents-cybersecurity-evals.
[3] Reuters, "Anthropic Reports Fourth Cybersecurity Incident with Early Version of Claude," Sept. 9, 2026, https://www.reuters.com/legal/litigation/anthropic-reports-fourth-cybersecurity-incident-with-early-version-claude-2026-09-09/.
[4] Associated Press, "Meta says its AI model hacked another company, adding to worries about bots going rogue," Associated Press, Aug. 6, 2026, https://apnews.com/article/0e8061437da6779be962b24ac134a514.

organizations.[5] In one instance, an AI agent wrote malicious code and created fake online identities while attempting to obtain human approval for the code.

The Hugging Face incident and the subsequent revelations of unauthorized actions undertaken by frontier models have increasingly left the American public and its elected representatives worried about the lack of public control and accountability of this incredibly powerful suite of technologies.

Leaders developing the world's most advanced AI systems have warned that safety measures are not keeping pace with frontier AI capabilities. Anthropic CEO Dario Amodei recently called for slowing the pace at which frontier AI capabilities advance, because without sufficient safeguards, we will see in "6-12 months such a swarm [of AI agents that] could be capable of taking over the entire internet with a persistent botnet (potentially causing hundreds of billions of dollars in damage)."[6] OpenAI CEO Sam Altman agreed that the industry needs to "pace the frontier," while xAI CEO Elon Musk and Google DeepMind Co-Founder and Chairman Demis Hassabis publicly endorsed the direction of Amodei's proposal.[7, 8, 9]

Concerns are also being raised from inside the companies developing these systems. Former Anthropic researcher Jacob Coxon recently resigned from the company before his equity vested, citing concerns about the trajectory of advanced AI development.[10] Evan Hubinger, an alignment lead at Anthropic, echoed Coxon's message, estimating a 10 percent chance that AI could cause human extinction within the next decade and warned that competitive pressures could cause developers to cut corners or skip safety steps.[11]

These warnings from executives and researchers with the greatest visibility into frontier AI capabilities should be taken seriously, especially after the Hugging Face incident. Significant uncertainty and disagreement remain among experts about the likelihood and timeline of catastrophic AI risks; however, the combination of rapidly advancing capabilities, documented real-world security incidents, and warnings from the individuals building these systems warrants direct congressional scrutiny.

The Commerce Committee has jurisdiction over many of the issues implicated by these developments and an important responsibility to ensure that continued American leadership in AI

---

[5] U.K. AI Security Institute, "Incident Report: Unsanctioned Agent Behaviour During Cyber Testing," *AI Security Institute*, https://www.aisi.gov.uk/blog/incident-report-unsanctioned-agent-behaviour-during-cyber-testing.
[6] Dario Amodei, "We Must Pace the Frontier," https://darioamodei.com/post/we-must-pace-the-frontier.
[7] Sam Altman (@sama), post on X, https://x.com/sama/status/2098811563415150910.
[8] Elon Musk (@elonmusk), post on X, https://x.com/elonmusk/status/2098789109980332057.
[9] Demis Hassabis (@demishassabis), post on X, https://x.com/demishassabis/status/2098909516582490602.
[10] Axios, "Anthropic Researcher's AI Warning," Sept. 9, 2026, https://www.axios.com/2026/09/09/anthropic-researcher-ai-warning-interview.
[11] Evan Hubinger (@EvanHub), post on X, https://x.com/EvanHub/status/2097497037956891126.

is accompanied by appropriate safeguards. Members should have the opportunity to hear directly from the executives making decisions about how quickly frontier capabilities are developed and deployed, what safeguards govern those decisions, how companies test and monitor increasingly autonomous systems, how serious incidents are identified and reported, and what role they believe Congress should play in establishing baseline protections.

We therefore urge the Committee to convene a hearing with the chief executive officers of OpenAI, Anthropic, Google DeepMind, Meta, and xAI. Given both the extraordinary opportunities presented by advanced AI and the increasingly serious warnings coming from the industry itself, our committee should have the opportunity to question these leaders publicly about the sufficiency of existing safeguards, responsible and pro-human development of frontier models, and how the federal government should approach AI regulation in a way that protects the public while preserving American innovation.

Thank you for your consideration of this request. We look forward to working with you on this important issue.

Sincerely,

Lisa Blunt Rochester
United States Senator

John Curtis
United States Senator