ANDREW R. GARBARINO, NEW YORK
CHAIRMAN

BENNIE G. THOMPSON, MISSISSIPPI
RANKING MEMBER

One Hundred Nineteenth Congress
Committee on Homeland Security
U.S. House of Representatives
Washington, DC 20515

July 31, 2026

Mr. Sam Altman
Chief Executive Officer
OpenAI
1455 3rd Street
San Francisco, CA 94158

Dear Mr. Altman:

We write to request a briefing regarding OpenAI models escaping a testing environment and autonomously attacking Hugging Face, an artificial intelligence (AI) platform, and accounts on other third-party services.[1] We are not only concerned that OpenAI's models overcame controls put in place to prevent the type of autonomous activity that ultimately occurred, but also that OpenAI seemingly failed to detect and disrupt this activity in real time.

Earlier this month, OpenAI revealed an "unprecedented cyber incident" during which a combination of its models – including a pre-release model – exploited a zero-day vulnerability in a third-party software to gain Internet access from an otherwise isolated environment during testing, identified ways to break into Hugging Face's systems, gained access to Hugging Face's systems, and stole data.[2] We understand that OpenAI's security team eventually discovered this activity,[3] and that Hugging Face's team had detected, disrupted, and begun investigating the incident by the time OpenAI connected with Hugging Face.[4] On July 28, OpenAI revealed that its models had compromised accounts on other third-party platforms in pursuit of gaining access to Hugging Face.[5]

This incident raises a number of serious questions regarding the rigor with which OpenAI secures and monitors its testing environments and the tools used to support them. More broadly, the incident presents difficult questions about the limitations of security guardrails intended to prevent

---

[1] OpenAI, *OpenAI and Hugging Face partner to Address Security Incident During Model Evaluation* (updated Jul. 28, 2026), https://openai.com/index/hugging-face-model-evaluation-security-incident/; Dell Cameron & Maxwell Zeff, *OpenAI's Rogue AI Agent Hacked More Than Just Hugging Face,* WIRED (Jul. 28, 2026), https://www.wired.com/story/openais-rogue-ai-agent-hacked-more-than-just-hugging-face/.
[2] *Id.*
[3] Raphael Satter, et al, *Its AI Agent Spent Days Hacking A Company, But Sources Say OpenAI Did Not Notice for a Week*, REUTERS (Jul. 25, 2026), https://www.reuters.com/business/its-ai-agent-spent-days-hacking-company-sources-say-openai-did-not-notice-week-2026-07-24/.
[4] OpenAI, *OpenAI and Hugging Face partner to Address Security Incident During Model Evaluation* (updated Jul. 28, 2026), https://openai.com/index/hugging-face-model-evaluation-security-incident/.
[5] *Id.;* Cameron & Zeff, *supra* note 1.

the misuse of AI models, both because they can be overcome and because they can also prevent network defenders from leveraging the models for security and forensic purposes.[6]

Safety, security, and innovation must coexist. We look forward to working with industry to better understand the limitations of security guardrails, how the security of model development and testing environments can be improved, and how to facilitate transparency regarding these processes to engender public trust.

Thank you for your prompt attention to this matter.

Sincerely,

ANDY OGLES
Chairman
Subcommittee on Cybersecurity
and Infrastructure Protection
Committee on Homeland Security

DELIA RAMIREZ
Ranking Member
Subcommittee on Cybersecurity
and Infrastructure Protection
Committee on Homeland Security

[6] *See* Georgia Hennessy, *The Latest OpenAI Drama Made Chinese AI the Hero*, BUSINESS INSIDER (Jul. 22, 2026), https://www.businessinsider.com/hugging-face-hack-openai-rogue-ai-china-cybersecurity-2026-7.