An Open Letter to Members of the United States Congress
Congress Must Investigate the OpenAI–Hugging Face
Security Incident and Strengthen Federal AI Oversight

July 31, 2026

Dear Members of Congress:

We write to urge Congress to recognize the recent OpenAI–Hugging Face security incident as a historic inflection point in the development of artificial intelligence (AI) and a clear warning that the continued absence of meaningful federal oversight of frontier AI companies has left Americans, and people around the world, in unnecessary danger as increasingly powerful AI systems are being built and deployed by Big Tech without legally binding standards for safety, security, transparency, or accountability. **Accordingly, Congress should immediately open a Congressional investigation into the incident and examine whether the development and testing of frontier AI systems require stronger statutory safeguards and independent oversight.**

For years, a wide range of experts - academics, civil society organizations, and AI labs - have called for improved regulation of AI models, citing empirical evidence of improved capabilities to enable bad actors and fundamental flaws in the models that perpetuate harmful outcomes for citizens. Yet despite these repeated warnings, Congress has been slow to establish a comprehensive federal framework to govern the development and deployment of frontier AI systems. The recent OpenAI model attack on Hugging Face's infrastructure—a leading open-source AI platform used by millions of developers, researchers, companies, and governments to build and deploy artificial intelligence systems—marks a watershed moment in AI governance and underscores that the future experts have warned about has finally arrived.

According to the public accounts, an advanced OpenAI system, while undergoing internal testing, escaped its intended containment environment by exploiting a previously unknown software vulnerability. After obtaining broader access, it autonomously identified, planned, and executed a multi-stage cyber intrusion into Hugging Face's infrastructure to accomplish its assigned objective which went unnoticed by OpenAI for a week. Subsequent public reporting and analyses revealed the model remained active outside its intended testing environment for approximately four days, carrying out thousands of autonomous actions across the open internet. This incident also affected Modal Labs' customer account, suggesting the model's cyber operations extended beyond a single target. The incident represents an unprecedented demonstration that frontier AI systems need to be better tested and secured so we can proactively prevent them from carrying out sustained unanticipated operations that cross organizational boundaries, risking critical infrastructure, personal information, and digital integrity.

This incident arose from decisions OpenAI made during the development and evaluation of its frontier AI system. OpenAI determined the system's capabilities, designed the testing

environment, established the objectives it was expected to pursue, and controlled the safeguards intended to limit its operation. The resulting security incident therefore underscores the risks that can arise when private companies are permitted to conduct consequential real-world evaluations of frontier systems without legally enforceable standards governing safety, security, containment, independent oversight, or accountability. As AI developers continue to expand the capabilities of these systems, Congress should ensure that the risks created by frontier AI development are not borne primarily by the public, downstream organizations, or critical infrastructure.

The OpenAI–Hugging Face incident should not be viewed as an isolated anomaly. Last year, reporting indicated that threat actors used Anthropic's Claude models in connection with the GTIG cyber campaign, further demonstrating that AI systems are already being leveraged in real-world cyber operations across multiple platforms. These incidents preview the kinds of AI-related security risks that frontier AI systems may pose if left without appropriate safeguards. While these incidents were contained before more significant harm occurred, we may not be as fortunate next time. Congress has a narrowing window to act before a more consequential AI-related disaster forces action in the wake of irreversible harm. Rather than waiting for that moment, lawmakers should seize this opportunity to establish thoughtful, risk-based guardrails that protect the public while preserving the authority of states to respond to emerging AI risks and adopt stronger protections as technology continues to evolve.

While the formation of the Open Secure AI Alliance reflects a growing recognition within industry that AI security demands greater attention, it also underscores that companies developing these systems acknowledge the risks are significant enough to warrant coordinated action. But this voluntary initiative is not the same as responsible governance. Voluntary alliances cannot compel participation, establish uniform safety standards, ensure independent evaluation, or provide meaningful accountability when failures occur. Congress should view this initiative not as further evidence that the era of relying exclusively on voluntary commitments has reached its limit. The public deserves safeguards grounded in law, not promises grounded in goodwill.

The events disclosed by OpenAI and Hugging Face demonstrate that existing voluntary commitments are not, by themselves, sufficient to address frontier AI systems. The United States' greatest strength is both its capacity to innovate and its willingness to govern innovation responsibly. As the global leader in artificial intelligence, our responsibility extends beyond winning the race for market share, investment, or technological supremacy. True leadership is measured by whether we have the wisdom and courage to ensure that these extraordinary technologies serve humanity safely and ethically.

Congress can ensure that America is remembered not only as the nation that built the most powerful AI systems, but as the nation that had the foresight to govern them responsibly. We respectfully urge Congress to seize this historic moment and establish the safeguards necessary

to ensure that American leadership in artificial intelligence is defined by accountability, responsibility, and moral leadership worthy of generations to come.

Sincerely,

Public Citizen
Tech Oversight Project
Americans for Responsible Innovation
350 Conejo / San Fernando Valley
The Alliance for Secure AI
Animals Are Sentient Beings, Inc.
Autistic Women & Nonbinary Network
Black Voters Matter Fund
Cascadia Climate Action Now
Center for Progressive Reform
Clean Elections Texas
Climate Defenders
Common Cause
Consumer Action
Demand Progress Economic Security Project Action Education Fund
Enabled Emissions Campaign
EncodeAI
Future of Life Institute
Friend of the Earth U.S.
Green America
Humane Intelligence Public Benefit Corporation
Indivisible
Long Beach Alliance for Clean Energy
MARBE SA
Memphis Community Against Pollution
Mid-Ohio Valley Climate Action
Mpower Change
National Association of Voice Actors (NAVA)
National Coalition Against Cryptomining
Novadaur
Oil and Gas Action Network
Our Subscription to Addiction
People Power United
Physicians for Social Responsibility Pennsylvania
Progressives for Climate
Public Knowledge
ReThink Citizens
Seneca Lake Guardian
Stand.earth
The Value Alliance
Third Act SoCal
United for Respect

Mark Linder, Professor, Syracuse University
Maroussia Lévesque, Assistant Professor in Law & AI, Queen's University
Eli Pariser, author / activist *The Filter Bubble: What the Internet Is Hiding from You*
Allison Stanger, Distinguished Endowed Professor, Middlebury College and Santa Fe Institute
Shenja van der Graaf, Associate Professor, Communication Science, University of Twente (NL)
Moshe Vardi, University Professor and George Distinguished Service Professor in Computational Engineering, Rice University
Salil Vadhan, Vicky Joseph Professor of Computer Science & Applied Mathematics, Harvard University
Shlomit Wagman, Former Director-General of the Israel Money Laundering and Terror Financing Prohibition Authority, Fellow at Harvard Kennedy School;
Jim Waldo, Gordon McKay Professor of the Practice of Computer Science and Technology Policy, Harvard University
Baobao Zhang, Associate Professor, Political Science, Syracuse University