The Honorable Donald J. Trump
President of the United States
The White House
1600 Pennsylvania Avenue
Washington, DC 20500

July 30, 2026

Dear Mr. President,

Over the last year, Americans have seen how the capabilities of frontier AI models can threaten national security, financial institutions, and cybersecurity. Last week, a new and foreseeable threat arose: an AI model capable of breaking out of a test environment without human direction and breaching a company's secure systems.

We could not have asked for a clearer warning shot. While there is little doubt that they will bring opportunities and benefits across a wide range of domains, AI models at today's frontier pose increasingly severe risks to our private sector, our national security, and the American public. Last week's incident, which occurred in a testing environment and was diligently reported by OpenAI, made clear that the United States government has a limited view into the advanced capabilities of frontier AI models and has not yet released a clear process for evaluating the safety concerns posed by advanced AI systems.

Given the severity of the incident and the broader vulnerabilities it revealed, we urge the Trump administration to open an investigation, with the support of independent auditors, into what occurred.

This investigation should determine how the breach occurred, assess whether existing safeguards and reporting mechanisms were adequate, and identify the steps necessary to prevent a similar incident from occurring again. The investigation's findings can also provide a foundation for the Trump administration to follow up on its June 2 executive order with a clear rules-based process for assessing the risks posed by advanced AI systems.

We unfortunately expected this moment would arrive. The release of Mythos earlier this year demonstrated the capabilities that frontier AI models have to break through traditional cybersecurity protocols within companies. In 2025, Anthropic revealed that in a controlled test setting, one of their models, Claude Opus 4, used blackmail and self-preservation techniques when faced with retraining or replacement. These and other similar incidents show that frontier models can behave in ways their developers do not intend and that existing safeguards may be insufficient to prevent.

A thorough investigation into this incident, supported by independent auditors, will help inform further executive action building on the Administration's June 2 executive order. Specifically, an investigation would help support executive action establishing clear rules requiring transparency into frontier internal capabilities, testing, and incidents; providing meaningful pre-deployment evaluation; and overseeing the release of models or systems that pose serious risks to the public, critical infrastructure, and our national security.

Thank you for your proactive work to address the risks posed by advanced AI. America still has an opportunity to institute protections that both preserve America's AI leadership and safeguard the people, institutions, and systems that are under threat by increasingly capable models. The administration should act before a warning shot becomes a preventable disaster.

Sincerely,

Brad Carson, President, Americans for Responsible Innovation
Brendan Steinhauser, CEO, The Alliance for Secure AI

J.B. Branch, Director of Federal AI Governance and Technology Policy, Public Citizen
Ryan Carrier, Executive Director, ForHumanity
Hamza Chaudhry, AI and National Security Lead, Future of Life Institute
Nicholas P. Garcia, Senior Policy Counsel, Public Knowledge
Vael Gates, Founder and Executive Director, Humans in Control
Adam Gleave, CEO and Co-Founder, FAR.AI
Ross Gruetzemacher, Executive Director, Transformative Futures Institute
Samuel Hammond, AI Policy Expert and Economist
Sid Hiregowdara, CEO and Co-Founder, CivAI
David Kasten, Head of Policy, Palisade Research
Colin McGlynn, AI Policy Advisor, Demand Progress Action
Ruby Scanlon, Tech and National Security Leader
Nate Soares, Author and AI Research and Policy Leader

Cc: The Honorable Todd Blanche, Acting Attorney General of the United States
The Honorable Howard Lutnick, Secretary of Commerce
The Honorable Markwayne Mullin, Secretary of Homeland Security
The Honorable Michael Kratsios, Director, White House Office of Science and Technology Policy
The Honorable Sean Cairncross, National Cyber Director