Congress of the United States

Washington, DC 20515

May 13, 2026

The Honorable Sean Cairncross
National Cyber Director
Office of the National Cyber Director
The White House
1600 Pennsylvania Avenue NW
Washington, DC 20500

Dear Director Cairncross:

We write regarding the rapid development of frontier artificial intelligence systems that can discover, analyze, and exploit software vulnerabilities at a scale that existing public and private processes are not equipped to handle. We request that the Office of the National Cyber Director (ONCD) convene a federal-industry process to develop a plan to coordinate high volumes of vulnerability disclosures generated by advanced AI systems; expand controlled defensive access to these tools for trusted defenders; and recommend whether and how the U.S. government can support the American software ecosystem to validate, triage, and patch vulnerabilities in both commercial and government systems.

Anthropic's recent announcement of Claude Mythos Preview illustrates the issue. According to Anthropic, Mythos identified thousands of high-severity zero-day vulnerabilities in every major operating system and every major web browser, including vulnerabilities that had survived years of human review and automated testing. Anthropic also reported that Mythos was able to develop working exploits, chain vulnerabilities, and assist non-experts in finding serious software flaws.[1]

The overwhelming majority of these discoveries are not yet patched or public. As of their April 7 announcement, more than 99 percent of the vulnerabilities Anthropic discovered remained unpatched; for this reason, the company has chosen to withhold several details regarding its coordinated vulnerability disclosure process.[2] AI can empower defenders to discover a multitude of serious vulnerabilities, but the corresponding disclosure, validation, patching, and deployment efforts may struggle to keep pace.

The lesson from Mythos is not limited to cybersecurity or to just one company. Mythos' cyber capabilities exemplify broader advances in coding, reasoning, and agentic tool use. As

---

[1] Anthropic, *Project Glasswing: Securing Critical Software for the AI Era* (Apr. 7, 2026), https://www.anthropic.com/glasswing; Nicholas Carlini et al., *Assessing Claude Mythos Preview's Cybersecurity Capabilities*, Anthropic Frontier Red Team (Apr. 7, 2026, updated Apr. 9, 2026), https://red.anthropic.com/2026/mythos-preview/.

[2] Nicholas Carlini et al., *Assessing Claude Mythos Preview's Cybersecurity Capabilities*, Anthropic Frontier Red Team (Apr. 7, 2026, updated Apr. 9, 2026), https://red.anthropic.com/2026/mythos-preview/.

general-purpose models improve, they may acquire new capabilities in cyber, chemical, biological, radiological, and nuclear (CBRN)-relevant research, and AI research and development (R&D) itself before federal agencies and infrastructure owners have time to adjust. Regardless of how quickly one expects AI capabilities to advance, when important capabilities appear, federal agencies must be able to recognize them and respond quickly.

Recent developments reinforce the need for federal action. Prompted by the announcement of Mythos-level systems, financial regulators and allied governments have begun examining new cybersecurity threats to banks and critical software providers.[3] OpenAI recently published a cybersecurity action plan to democratize cyber defense, coordinate across government and industry, and strengthen the security of the models themselves.[4] The White House has also reportedly opposed Anthropic's proposal to expand access to Mythos.[5] These parallel developments highlight the need for clear, consistent criteria on how access should be expanded, limited, or delayed. As the Administration considers how to manage access to advanced AI systems, ONCD should consult with the Secretary of the Treasury on questions affecting financial and national security, and with the National Economic Council when decisions implicate U.S. competitiveness. Additionally, unauthorized users reportedly accessed Mythos through a third-party vendor environment.[6] While not model-weight theft, the incident is emblematic of a broader reality: soon after their capabilities are known, restricted frontier systems become targets. Our well-resourced adversaries, such as the People's Republic of China, will almost certainly try to obtain, copy, or steal comparable systems. As capabilities grow, so too does the importance of model security.

The April 2026 Office of Science and Technology Policy memorandum titled "Adversarial Distillation of American AI Models" (NSTM-4) addresses part of this threat. It warns that foreign entities, principally based in China, are already engaged in deliberate, industrial-scale campaigns to distill U.S. frontier AI systems through proxy accounts, jailbreaking techniques, and systematic extraction of proprietary capabilities.[7] These efforts make model-weight security, privacy-preserving monitoring, know-your-customer protocols, and rapid revocation important parts of any defensive access framework.

---

[3] Saeed Azhar, *Bessent, Powell Warned Bank CEOs About Anthropic Model Risks, Sources Say*, Reuters (Apr. 10, 2026), https://www.reuters.com/business/finance/bessent-powell-warn-bank-ceos-about-anthropic-model-risks-bloomberg-news-reports-2026-04-10/; Makiko Yamazaki, *Japan Launches Financial Task Force Amid AI Security Fears*, Reuters (Apr. 24, 2026, updated Apr. 27, 2026), https://www.reuters.com/sustainability/boards-policy-regulation/japan-launches-financial-task-force-amid-ai-security-fears-2026-04-24/.

[4] OpenAI, *Cybersecurity in the Intelligence Age* (Apr. 29, 2026), https://openai.com/index/cybersecurity-in-the-intelligence-age/.

[5] Robert McMillan & Amrith Ramkumar, *White House Opposes Anthropic's Plan to Expand Access to Mythos Model*, Wall St. J. (Apr. 30, 2026), https://www.wsj.com/tech/ai/white-house-opposes-anthropics-plan-to-expand-access-to-mythos-model-dc281ab5.

[6] *Anthropic's Mythos Model Accessed by Unauthorized Users, Bloomberg News Reports*, Reuters (Apr. 21, 2026), https://www.reuters.com/technology/anthropics-mythos-model-accessed-by-unauthorized-users-bloomberg-news-reports-2026-04-21/.

[7] Off. of Sci. & Tech. Pol'y, Exec. Off. of the President, *Adversarial Distillation of American AI Models (NSTM-4)* (Apr. 23, 2026), https://www.whitehouse.gov/wp-content/uploads/2026/04/NSTM-4.pdf.

The Administration's AI Action Plan also identifies relevant priorities, including guidance for remediating AI-specific vulnerabilities, consolidated sharing of known AI vulnerabilities, mature AI incident response capacity, national security evaluations of frontier models, and work on AI-related biosecurity.[8] The Cybersecurity and Infrastructure Security Agency (CISA) and the Joint Cyber Defense Collaborative (JCDC) have also published an AI Cybersecurity Collaboration Playbook to support voluntary information sharing across the AI ecosystem.[9] Those efforts should account not only for vulnerabilities in AI systems, but also for ordinary software vulnerabilities that advanced AI systems may discover or help weaponize. These efforts identify many of the correct participants, but AI-discovered vulnerabilities in everyday software will likely necessitate traditional cyber defense coordination beyond the AI-specific approaches named in the AI Action Plan and JCDC playbook.

We respectfully request that ONCD convene and coordinate an interagency and federal-industry process, with the Department of Homeland Security (DHS), acting through CISA, serving as the lead agency for vulnerability coordination and remediation planning; the Department of Commerce (DOC), acting through the National Institute of Standards and Technology (NIST) and the Center for AI Standards and Innovation (CAISI), supporting AI capability evaluation, technical standards, and trusted-access frameworks; and the Office of Management and Budget (OMB), acting through the Office of the Federal Chief Information Officer (OFCIO), supporting federal coordination and implementation, to take the following steps:

1. **Develop a plan to coordinate high volumes of vulnerability disclosures.** The plan ought to address how federal agencies, software vendors, open-source maintainers, and critical infrastructure partners will receive an influx of AI-discovered vulnerability reports; verify, deduplicate, and prioritize them; assess their severity; protect exploit code and other sensitive technical details; coordinate responsible disclosure as affected vendors and maintainers prepare fixes; and provide mitigation best practices to IT product and service providers for communicating with key customers when patches are not available or cannot be deployed quickly. The plan should describe what role, if any, the federal government should play to support the security of non-federal systems.

2. **Assess existing federal, industry, and open-source efforts to identify critical software dependencies; where gaps remain, develop a protected approach for prioritizing load-bearing software dependencies.** As appropriate, ONCD should work with CISA, CAISI, open-source security organizations, major IT and cybersecurity product vendors, product security engineering teams, and critical infrastructure operators to identify the open-source

---

[8] The White House, *America's AI Action Plan* (July 2025), https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

[9] Cybersecurity & Infrastructure Sec. Agency, *CISA, JCDC, Government and Industry Partners Publish AI Cybersecurity Collaboration Playbook* (Jan. 14, 2025), https://www.cisa.gov/news-events/news/cisa-jcdc-government-and-industry-partners-publish-ai-cybersecurity-collaboration-playbook.

and commercial software libraries most common and critical to networking, internet-facing, and monitoring products. Public criteria can help guide the effort, but the government should avoid publishing a target map for adversaries. Sensitive dependency information should only be shared with vetted stakeholders deemed essential.

3. **Find and fix vulnerabilities before our adversaries, and assist critical infrastructure in deploying and verifying fixes.**[10] First, the federal government should encourage frontier AI developers to provide early, free, or reduced-cost access to qualified open-source maintainers, nonprofit security organizations, and software providers that manage widely used systems. Second, the federal government should aid critical infrastructure owners and operators in deploying and verifying patches. In particular, the government should reinforce infrastructure where legacy systems, limited cyber staff, or operational constraints slow remediation. Where necessary, this should include support for patch testing, temporary mitigations, and technical assistance from CISA and the relevant sector risk management agencies. Third, for reported vulnerabilities, the federal government should identify whether R&D into automated validation, risk scoring, rapid legacy system updates, and mass distribution of fixes to trusted users is warranted, such as through the Defense Advanced Research Projects Agency (DARPA) or similar programs.

4. **Develop a framework for handling sensitive AI-generated dual-use findings.** The same information that helps defenders can also help attackers if released too broadly. In coordination with CAISI, the Secretary of the Treasury, the National Security Agency (NSA), U.S. Cyber Command, the Intelligence Community, and other relevant agencies, ONCD should develop a framework for deciding what can be disclosed publicly, what should be shared only with affected vendors or vetted defenders, and what should be handled through restricted government channels. This should include details on cyber exploits, proof-of-concept code, vulnerability chains, affected dependency lists, and sensitive model capability evaluations. It should also include CBRN-relevant findings, such as biological design, synthesis screening, or evaluation information that could be beneficial for science and public health if handled responsibly, but risky if disclosed without safeguards. For vulnerabilities implicating U.S. government operations or intelligence, the framework should clarify coordination with the U.S. computer network exploitation community and, where appropriate, whether the Vulnerabilities Equities Process needs to be adapted for an influx of AI-discovered vulnerabilities.

5. **Establish a voluntary framework for AI labs to provide early access to vetted defenders.** Several leading AI developers are already experimenting with trusted access for advanced cyber and life sciences models.[11] ONCD ought to build on those efforts, prioritizing access

---

[10] Miles Brundage, *Operation Patchlight*, Inst. for Progress (Aug. 11, 2025), https://ifp.org/operation-patchlight/.
[11] OpenAI, *Trusted Access for the Next Era of Cyber Defense* (Apr. 14, 2026), https://openai.com/index/scaling-trusted-access-for-cyber-defense/; OpenAI, *Introducing GPT-Rosalind for Life Sciences Research* (Apr. 16, 2026), https://openai.com/index/introducing-gpt-rosalind/; OpenAI, *Cybersecurity in the Intelligence Age* (Apr. 29, 2026),

for security engineering teams within major IT companies that build browsers, phones, mobile operating systems, and networking, authentication, and monitoring products; clarifying eligibility, vetting processes, permitted use cases, the appropriate oversight for high-risk capabilities, how third-party vendor environments are secured, responsible disclosure obligations, and how access is revoked if misused. Where appropriate, the framework should also address model-weight security, privacy-preserving abuse monitoring, customer verification protocols, privileged-access controls, and rapid revocation. This framework should address interagency prioritization to ensure access for CISA and other relevant federal agencies.[12]

6. **Establish a process for monitoring sudden frontier AI capability jumps.** Building on existing CAISI-led interagency evaluation efforts, ONCD should work with CAISI, CISA, DHS, DOC, the NSA, the Department of Energy (DOE), the Department of Defense (DOD), the Intelligence Community, the Secretary of the Treasury, and other relevant agencies to identify when models reach new thresholds in cyber exploitation, CBRN assistance, or AI research and development—including successor model development that could magnify model-weight security risks and materially shorten the time government and industry have to respond. For cyber, this process should determine what types of vulnerabilities a model can reliably discover or exploit; the token cost per useful vulnerability, exploit, or patch; and the effectiveness of post-training guardrails to prevent misuse. These findings should help inform whether a given model release or access expansion is more likely to reduce cybersecurity risk or increase offensive capability. This process should include ways for companies and evaluators to share sensitive results securely, including the sharing of classified or otherwise sensitive cyber, CBRN, and model-security evaluations, and should identify any legal or contractual restrictions constraining timely sharing. Additionally, the process ought to include a mechanism for timely congressional briefings when new capabilities affect national security or critical infrastructure. ONCD should also evaluate what secure government, industry, or public-private environments may be needed to run future systems more capable than Mythos, including the necessary infrastructure for defensive vulnerability discovery, validation, and patching before broader deployment.[13] Moreover, ONCD should plan for high-risk contingencies, such as confirmed or suspected model-weight theft; unauthorized access to a restricted frontier model; successful adversarial distillation of Mythos-class or superior capabilities; foreign-developed or open-weight models with comparable offensive cyber capabilities; or the release of a comparable U.S. model with limited, ineffective, or no safeguards.

---

https://openai.com/index/cybersecurity-in-the-intelligence-age/.

[12] Sam Sabin, *Scoop: CISA Lacks Access to Anthropic's Mythos*, Axios (Apr. 21, 2026), https://www.axios.com/2026/04/21/cisa-anthropic-mythos-ai-security.

[13] Sella Nevo, *A Sprint Toward Security Level 5*, Inst. for Progress (Aug. 11, 2025), https://ifp.org/a-sprint-toward-security-level-5/.

7. Identify barriers requiring congressional action. Please identify any legal, confidentiality, antitrust, liability, or other barriers limiting timely coordination between frontier AI developers, software vendors, open-source maintainers, critical infrastructure operators, cybersecurity researchers, and federal agencies. In particular, please note if Congress should consider additional statutory protections for sensitive cyber and CBRN information sharing among competitors, such as targeted updates to the Cybersecurity Information Sharing Act of 2015 or related liability, confidentiality, or antitrust protections; support for open-source maintainers; or federal accreditation of defenders who may receive access to advanced AI cyber tools. Please also identify whether existing legal authorities are sufficient where exploitation risk is imminent, ordinary patching is not feasible, and involuntary patching may be necessary. In accomplishing the above, ONCD should assess existing federal and industry efforts, identify any remaining gaps, and avoid duplicating effective ongoing work.

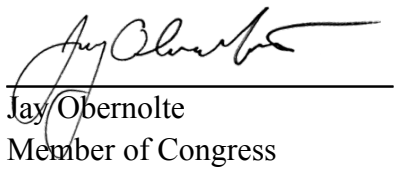
We ask that ONCD provide a staff-level briefing within 30 days on the status of these efforts and a written response within 45 days describing the Administration's current plan; the agencies responsible for implementation; an expected timeline for the initial federal-industry convening; and any additional statutory changes for Congress to consider.

We look forward to working with ONCD and other relevant agencies to preserve U.S. leadership in AI. We must ensure that this lead hardens American infrastructure before our adversaries use these tools against us.

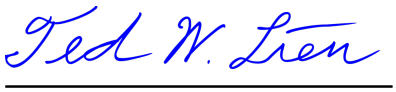
Sincerely,

Robert E. Latta
Member of Congress

Doris Matsui
Member of Congress



Jay Obernolte
Member of Congress



Ted W. Lieu
Member of Congress

Nathaniel Moran
Member of Congress

John R. Moolenaar
Member of Congress

Ro Khanna
Member of Congress

Gus M. Bilirakis
Member of Congress

Bill Foster
Member of Congress

Vince Fong
Member of Congress

Josh Gottheimer
Member of Congress

Nicholas J. Begich III
Congressman for All Alaska

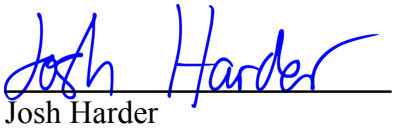
George Whitesides
Member of Congress

Mariannette J. Miller-Meeks, M.D.
Member of Congress

James R. Baird
Member of Congress

Michael A. Rulli
Member of Congress

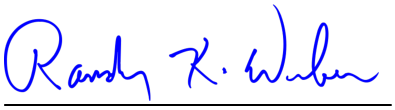
Kim Schrier, M.D.
Member of Congress



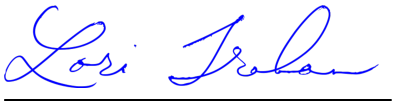
Josh Harder
Member of Congress

Pat Harrigan
Member of Congress

Earl L. "Buddy" Carter
Member of Congress

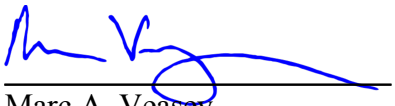


Randy K. Weber, Sr.
Member of Congress

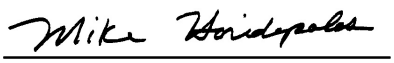


Lori Trahan
Member of Congress

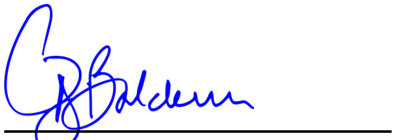
Erin Houchin
Member of Congress



Marc A. Veasey
Member of Congress



Mike Haridopolos
Member of Congress



Troy Balderson
Member of Congress

Donald S. Beyer Jr.
Member of Congress

Suzanne Bonamici
Member of Congress

Russell Fry
Member of Congress

Sara Jacobs
Member of Congress

Julie Fedorchak
Member of Congress

Debbie Dingell
Member of Congress

Laura Friedman
Member of Congress

Gary J. Palmer
Member of Congress

Raja Krishnamoorthi
Member of Congress

CC: The Honorable Howard W. Lutnick, Secretary of Commerce; The Honorable Markwayne Mullin, Secretary of Homeland Security; The Honorable Scott Bessent, Secretary of the Treasury; The Honorable Michael Kratsios, Director of the Office of Science and Technology Policy; The Honorable Tulsi Gabbard, Director of National Intelligence; The Honorable Pete Hegseth, Secretary of War; The Honorable Susie Wiles, White House Chief of Staff; The Honorable Kevin Hassett, Director of the National Economic Council; The Honorable Russell T. Vought, Director of the Office of Management and Budget