AMERICANS *for* RESPONSIBLE INNOVATION

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

May 11, 2026

Dear Mr. President:

I write today to commend the Administration's reported intent to screen frontier AI models before they are released to the public. ARI has long advocated for evaluating advanced generative AI systems for dangerous capabilities and for supporting existing federal agencies in their oversight of AI deployments in regulated use cases.[1]

The arrival of Claude Mythos brings into specific relief the broad national security risks of frontier models. In order to safeguard our nation, this step change in capabilities requires an evolution from the current voluntary, developer-led oversight regime towards structured pre-deployment assurance.[2] The recent agreements that frontier labs signed with the Center for AI Standards and Innovation (CAISI) to conduct pre-deployment evaluations signal that the industry itself is coming to terms with the fact that its models are becoming too consequential for national security to rely on developer discretion alone.[3]

This Administration has an historic opportunity to establish clear guardrails for frontier labs that simultaneously ensure American AI development remains the most innovative in the world, while protecting the public from its largest risks. The Administration's reported approach is welcomed and timely, and I encourage you to consider the following policy recommendations to ensure any new framework remains maximally effective and resilient.

RECOMMENDATIONS

Which developers are reviewed: As the Administration considers which models should be subjected to review, ARI recommends targeting developers as a proxy for regulating frontier models. This is preferable to a regime targeting individual models, since today's frontier systems derive their capabilities from a combination of baseline models, scaffolding, and compute; therefore any

[1] "Our Priorities - Americans for Responsible Innovation." 2025. Americans for Responsible Innovation. July 18, 2025. https://ari.us/our-priorities/.
[2] Haykel, Iskandar. 2026. "After Mythos." Americans for Responsible Innovation. April 22, 2026. https://ari.us/policy-bytes/after-mythos/.
[3] "CAISI Signs Agreements Regarding Frontier AI National Security Testing with Google DeepMind, Microsoft and XAI | NIST." 2026. NIST. May 5, 2026. https://www.nist.gov/news-events/news/2026/05/caisi-signs-agreements-regarding-frontier-ai-national-security-testing.

threshold defined by model characteristics alone may become obsolete as the field rapidly advances.[4] To this end, the Administration should consider developers with $100 million or more in aggregate compute expenditure on frontier model training within the preceding 12 months, or frontier model developers with $500 million or more in annual revenue from AI products and services, with affiliate aggregation to prevent evasion through corporate restructuring. Because frontier AI capabilities increasingly derive from scaffolding and compute alongside the underlying models, dollar-denominated thresholds better capture the full scope of frontier investment than compute-based metrics alone.[5] Using a threshold like this also ensures that "small tech" developers and entrepreneurs will not be encumbered by any oversight requirements.

Which domains are reviewed: The current reported focus on cybersecurity capabilities is warranted and timely. I propose adding, at a minimum, Chemical, Biological, Radiological, Nuclear, and Explosive (CBRNE) as an additional priority risk domain. A recent study revealed critical and widespread safety gaps in this domain across ten Large Language Models (LLMs), highlighting an urgent need for standardized pre-deployment evaluations.[6] Additional risk domains such as autonomous harmful conduct; oversight subversion and deceptive alignment; autonomous self-replication and resource acquisition; autonomous AI R&D and high impact internal deployment; and manipulation and influence operations should be considered as additional areas of interest and potential review before deployment is approved.

Who reviews: The Trump Administration has made critical investments in CAISI to position the agency as industry's primary point of contact within the U.S. government to facilitate testing and collaborative research. Until a dedicated agency that possesses the right technical talent and resources can be established by an act of Congress, ARI believes the Administration should utilize CAISI's technical expertise and accumulated evaluation record. As CAISI already signed agreements with frontier labs, it should play, at a minimum, a technical-advisory role, specifically to provide deep evaluation support, methodology guidance, and input to standards derivation. Any working group that is constituted should also consider what role, if any, the National Labs might play in a federal review process. Particularly with respect to cybersecurity issues, many of the labs already possess the talent and actively conduct research that could contribute to this mission.

Enforcement: In order to make this oversight regime immediately enforceable, the Administration should consider these specific mechanisms: (i) directing NIST/CAISI to develop evaluation methodologies, benchmarks, and disclosure formats; (ii) conditioning federal contracts on having completed the established pre-deployment evaluation based on these NIST/CAISI best practices and receiving a greenlight for release; and (iii) requiring labs to report under Title VII of the Defense Production Act specifically for cybersecurity and CBRNE risk domains that have national defense

---

[4] Ball, Dean W, and Ketan Ramakrishnan. 2025. "Entity-Based Regulation in Frontier AI Governance." Carnegie Endowment for International Peace. July 7, 2025. https://carnegieendowment.org/russia-eurasia/research/2025/07/artificial-intelligence-regulation-united-states.

[5] ibid.

[6] Kumar, Divyanshu, Nitin Aravind Birur, Tanay Baswa, Sahil Agarwal, and Prashanth Harshangi. 2025. "Quantifying CBRN Risk in Frontier Models." Arxiv.org. October 2025. https://arxiv.org/html/2510.21133v1.

relevance. In addition to these first steps, I encourage you to work with Congress to establish a durable enforcement body within the Department of Commerce. This office should be vested with civil penalty authority for administrative non-compliance and statutory authority to refer representational or willful criminal violations to the Federal Trade Commission or Department of Justice, respectively.

As this Administration looks to protect American citizens and infrastructure, I encourage you to consider incident reporting as a companion and complement to pre-deployment measures. As the case of unauthorized access to Mythos showed, major security incidents can occur irrespective of whether a model is released or not, making pre-release gates necessary but insufficient.[7] Adding a deployment-agnostic incident reporting companion to pre-release review would create a more robust framework and help identify "near misses" before they become active threats to the public.

I additionally commend the Administration's reported desire to design a robust, balanced, and transparent oversight regime and encourage you to consider the following methods to best inoculate this system against possible abuses or overt politicization:

Utilize Unbiased Technical Experts: The Administration's reported plan to prioritize technical experts from leading firms in the proposed executive action is sensible, since they most intimately understand the technology. However, I encourage the Administration to additionally include independent technical research organizations, as well as academic AI researchers. Including independent, third-party technical voices would add public legitimacy and help ensure the resulting oversight regime reflects a balance of perspectives rather than relying exclusively on the companies whose products it governs for outside expertise. In support of this effort, I encourage you to direct the Secretary of Commerce to begin exploring the requirements for creating an ecosystem or marketplace that would support this kind of third-party verification.

Minimize Conflicts of Interest: The Administration should take steps to ensure independence when entering into a commercial dispute with a covered company whose core product is reviewed by this pre-deployment regime. This could include, at a minimum, recusal rules for specific pre-deployment reviewers from a government entity when that entity is in active litigation, a contract dispute, or enforcement action against a covered company.

Maintain Public Confidence: The Administration should pursue transparency and measures to help ensure accountability and public acceptance of any review process. This could include maintaining a public catalogue of the evaluation methodology and findings to enable systematic comparison across developers within each risk domain with confidentiality carve-outs for trade secrets and material posing specific harms to public safety. This could also include periodic classified reporting to

[7] Metz, Rachel. 2026. "Anthropic's Mythos Model Is Being Accessed by Unauthorized Users." Bloomberg.com. Bloomberg. April 21, 2026. https://www.bloomberg.com/news/articles/2026-04-21/anthropic-s-mythos-model-is-being-accessed-by-unauthorized-users.

Congress, annual public reports at appropriate levels of detail, and GAO audit authority as an existing and effective mechanism for surfacing problems.

While the anticipated Executive Order is a necessary first step toward establishing any kind of pre-deployment security review, this vital work will require more than a temporary order that can be easily reversed by the next administration. In order to ensure this Administration's commitment to help safeguard the American people, I urge you to work with Congress to codify in statute the regime described in this letter. This should include design choices for the regime itself, such as scoping and covered domains, as well as the safeguards to bolster this system against gaming and abuse.

As the Administration continues to weigh options and explore next steps, ARI stands ready to engage as a constructive partner in shaping an oversight regime that protects both American innovation and the public interest. ARI has developed several policy proposals that support the Administration's current ambitions and are consistent with my recommendations throughout this letter. ARI would welcome the opportunity to engage further with the Administration to discuss any of these proposals to support your goals of mitigating AI risks while enabling American leadership in AI innovation.

Sincerely,

Brad Carson
President
Americans for Responsible Innovation

Cc: The Honorable JD Vance, Vice President of the United States
The Honorable Scott Bessent, Secretary of the Treasury
The Honorable Howard Lutnick, Secretary of Commerce
The Honorable Michael Kratsios, Director, Office of Science and Technology Policy
The Honorable Kevin Hassett, Director, National Economic Council