PUBLIC SUBMISSION

Received: May 21, 2025
Tracking No. may-dz4c-4pqs
Comments Due: May 28, 2025 **Submission Type:** API

Docket: NSF-2025-OGC-0001
NITRD_FRDOC_0001

Comment On: NSF-2025-OGC-0001-0001
Request for Information: Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Document: NSF-2025-OGC-0001-DRAFT-0092
Comment on FR Doc # 2025-07332

Submitter Information

Organization: R Street Institute

General Comment

Comments of the R Street Institute in Request for Information on the Development of a 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan

Attachments

image

Comments of the R Street Institute in Request for Information on the Development of a 2025 National Artificial Intelligence Research and Development Strategic Plan

Comments of the R Street Institute in
Request for Information on the Development of a 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan

1411 K Street NW
Suite 900
Washington, D.C. 20005
202-525-5717

Free Markets. Real Solutions.
www.rstreet.org

May 29, 2025

Office of Science and Technology Policy
2415 Eisenhower Avenue
Alexandria, VA 22314
ostp-ai-rfostp-ai-rfi@nitrd.govi@nitrd.gov

Re: Request for Information on the Development of a 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan, Federal Register Number 2025-07332
Submitted Electronically
This Document is Approved for Public Dissemination.

Comments of the R Street Institute in
Request for Information on the Development of a 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan

I. OVERVIEW OF COMMENTS

Thank you for the opportunity to respond to the Request for Information (RFI) on the Development of the 2025 National Artificial Intelligence (AI) Research and Development (R&D) Strategic Plan.

As a nonpartisan, nonprofit public policy research organization headquartered in Washington, D.C., the R Street Institute (RSI) appreciates the thoughtful steps the Trump administration has taken to reinforce America's AI leadership through both this RFI and the earlier March 2025 RFI on the Development of an AI Action Plan.[1] These parallel processes reflect a deliberate, multistakeholder approach that fosters improved alignment between federal AI R&D priorities and the broader, shared strategic imperative of

[1] Adam Thierer, "Comments of the R Street Institute in Request for Information on the Development of an Artificial Intelligence (AI) Action Plan," R Street Institute, March 15, 2025. https://www.rstreet.org/outreach/comments-of-the-r-street-institute-in-request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan; Haiman Wong and Brandon Pugh, "Comments of the R Street Institute's Cybersecurity and Emerging Threats Team in Request for Information on the Development of an Artificial Intelligence (AI) Action Plan," R Street Institute, March 15, 2025. https://www.rstreet.org/outreach/comments-of-the-r-street-institutes-cybersecurity-and-emerging-threats-team-in-request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan.

advancing America's long-term leadership in technological innovation, national security, and economic competitiveness.

We commend the Trump administration's clear emphasis on removing regulatory obstacles to AI innovation and understanding that while our private sector remains well positioned to lead in nearer-term AI breakthroughs, the federal government also has a limited yet invaluable role to play in guiding our nation's AI innovation trajectory and investing in R&D where commercial incentives are low or payoffs are uncertain.[2] Strategic federal leadership and investment in continued high-risk, high-reward AI R&D, security-critical capabilities, and foundational AI domains ensures that America remains at the forefront of AI development—not only as an innovator but as the world's most trusted and resilient AI leader.

Over the past two years, RSI has brought together experts from academia, industry, civil society, and government to examine the intersection of emerging technologies, cybersecurity, and national security.[3] Our work has consistently underscored AI's growing role in both offensive and defensive cyber operations, as well as its broader potential to enhance America's competitive advantages in military, government, and national security applications.[4] However, the transformative benefits of AI requires a risk-based, security-centric approach to AI R&D that focuses on mitigating emerging threats through resilient design.[5]

In this spirit, we urge the development of the 2025 National AI R&D Strategic Plan to prioritize three key areas:

- Strengthening AI security and national security by advancing a scientific understanding of AI's capabilities and emerging risks;

- Leading and securing open-source AI development to promote resilience, transparency, and competitiveness against adversarial threats; and

---

[2] Adam Thierer, "Trump's New AI Executive Order Begins Undoing Biden's Bureaucratic Mess," R Street Institute, Jan. 23, 2025. https://www.rstreet.org/commentary/trumps-new-ai-executive-order-begins-undoing-bidens-bureaucratic-mess.

[3] "R Street Cybersecurity-Artificial Intelligence Working Group," R Street Institute, last accessed May 14, 2025. https://www.rstreet.org/home/our-issues/cybersecurity-and-emerging-threats/cyber-ai-working-group.

[4] Haiman Wong and Brandon Pugh, "Key Cybersecurity and AI Policy Priorities for Trump's Second Administration and the 119th Congress," R Street Institute, Jan. 6, 2025. https://www.rstreet.org/research/key-cybersecurity-and-ai-policy-priorities-for-trumps-second-administration-and-the-119th-congress.

[5] Haiman Wong et al., "Balancing Risk and Reward: AI Risk Tolerance in Cybersecurity," R Street Institute, April 15, 2024. https://www.rstreet.org/commentary/balancing-risk-and-reward-ai-risk-tolerance-in-cybersecurity; Haiman Wong, "Securing the Future of AI at the Edge: An Overview of AI Compute Security," R Street Institute, July 16, 2024. https://www.rstreet.org/research/securing-the-future-of-ai-at-the-edge-an-overview-of-ai-compute-security.

- Shaping the contours of agentic AI and the future of human-machine collaboration, including expanded efforts in explainability, compute integrity, and accountability mechanisms.

II. Advancing a Scientific Understanding of AI to Strengthen AI Security and National Security

At the 2024 AI Forum hosted by the Private and Civil Liberties Oversight Board (PCLOB), panelists from the National Institute of Standards and Technology (NIST) emphasized that a "comprehensive scientific understanding of AI's limitations and capabilities is vital for assessing its impact and countering adversarial threats".[6] Building a "scientific understanding" of AI could also equip researchers, developers, and end users with more reliable tools and robust methods to measure, test, and refine AI performance and decision-making processes over time—ensuring systems can operate securely and with greater trust and transparency.[7] Furthermore, developing a shared language to describe emerging AI concepts and behaviors is essential for driving consistent knowledge sharing and accelerating both continuous and iterative technological progress.[8]

Each of the following recommendations outlines key research areas where federal R&D leadership and investment can advance this scientific foundation:

1. *Develop Reliable AI Metrics:*
   The absence of universal, cross-sector AI security metrics creates inconsistencies in how organizations may evaluate AI systems, leading to fragmented practices and potential blind spots in risk management.[9] The 2025 National AI R&D Strategic Plan should prioritize the development of reliable metrics to assess data security, model protection, and resilience against emerging threats and vulnerabilities.[10] These AI metrics could also serve as a foundation for the development of more consistent AI audits across industry sectors based on their distinct sector-specific needs.[11]

2. *Advance Voluntary AI Auditing and Evaluation Standards:*
   Another ongoing R&D gap that persists in AI innovation is our ability to audit and evaluate AI system capabilities comprehensively.[12] To address this gap, the 2025 National AI R&D Strategic Plan should support the continuation of evaluation and standardization initiatives in AI security. Earlier efforts in this space, including those by the U.S. AI Safety Institute Consortium to

---

[6] Haiman Wong, "US May Be Losing the Race for Global AI Leadership," *Dark Reading*, Sept. 25, 2024. https://www.darkreading.com/cybersecurity-operations/us-losing-race-global-ai-leadership.
[7] Ibid.
[8] Ibid.
[9] Haiman Wong et al., "Assessing the Current State of AI-Cybersecurity Governance: Progress, Challenges, and Solutions," R Street Institute, May 21, 2024. https://www.rstreet.org/commentary/assessing-the-current-state-of-ai-cybersecurity-governance-progress-challenges-and-solutions.
[10] Ibid.
[11] Ibid.
[12] Ibid.

"[develop] guidelines for red-teaming, capability evaluations, risk management, safety and security, and watermarking synthetic content," should also be prioritized and promoted.[13]

3. *Expand R&D for Privacy-Enhancing Technologies (PETs):*
   Privacy-enhancing technologies (PETs), such as differential privacy, federated learning, and homomorphic encryption, are emerging technological solutions designed to allow organizations to extract insights from data while preserving privacy.[14] When integrated into AI systems, PETs can facilitate automated data anonymization or encrypted computation.[15] In doing so, PETs afford researchers and developers the opportunity to continue innovating while ensuring that sensitive data remains protected.[16] The 2025 National AI R&D Strategic Plan should expand R&D efforts for PETs to support its continued development and deployment across industry sectors.

By prioritizing these aforementioned gaps in AI innovation, the 2025 National AI R&D Strategic Plan can also help stakeholders across industry, academia, and the public sector build a clearer understanding of AI's real-world impact—particularly as AI capabilities are increasingly combined with traditional cybersecurity tools and legacy systems.[17] Advancing this scientific foundation will also enable more informed assessments of where legitimate AI risks persist, where concerns may be overstated, and how the United States can establish risk-tolerance parameters to guide AI implementation that maximizes its benefits while reinforcing our resilience.[18]

III. Securing Open-Source AI Ecosystems to Promote Resilience and Competitiveness

Open-source AI development has emerged as a cornerstone of AI innovation, fostering increased collaboration and democratizing access to advanced tools and capabilities by making model data, code, and even weights publicly available.[19] Importantly, this approach empowers independent researchers and developers, startups, and academic institutions to build on shared AI advancements, driving a dynamic innovation ecosystem where breakthroughs accelerate through collective contributions.[20] At the same time, however, the blurring line between open- and closed-source AI, coupled with the potential for adversarial nations to exploit open-source contributions and dependencies, introduces

---

[13] Ibid.
[14] Steven Ward, "Leveraging AI and Emerging Technology to Enhance Data Privacy and Security," R Street Institute, March 6, 2025. https://www.rstreet.org/research/leveraging-ai-and-emerging-technology-to-enhance-data-privacy-and-security/#offer-incentives-and-guidance-for-pet-development.
[15] Ibid.
[16] Ibid.
[17] Wong and Pugh. https://www.rstreet.org/research/key-cybersecurity-and-ai-policy-priorities-for-trumps-second-administration-and-the-119th-congress.
[18] Ibid.
[19] Haiman Wong, "Mapping the Open-Source AI Debate: Cybersecurity Implications and Policy Priorities," R Street Institute, April 17, 2025. https://www.rstreet.org/research/mapping-the-open-source-ai-debate-cybersecurity-implications-and-policy-priorities/#emerging-technological-solutions.
[20] Ibid.

significant cybersecurity and governance challenges.[21] The 2025 National AI R&D Strategic Plan must address these evolving risks to ensure open-source AI fulfills its potential to promote America's resilience and foster our competitiveness.

1. *Evaluate Security Risks in Open- and Closed-Source AI Systems:*
   The rapid integration of open-source AI development strategies into proprietary pipelines heightens the need to understand the unique limitations and security risks across both approaches. The 2025 National AI R&D Strategic Plan should support comparative evaluation initiatives aimed at assessing vulnerabilities in open- and closed-source AI systems, focusing on emerging threats like model tampering in public repositories.[22] These R&D efforts can inform robust security practices, ensuring open-source AI remains a driver of American innovation while proactively mitigating evolving security risks.

2. *Develop Automated Validation Tools and Embedded Provenance Tracking Systems for Open-Source AI:*
   Open-source AI's inherently collaborative and transparent nature risks bad-faith contributions and nefarious use cases that require advanced technological solutions that ensure security and accountability. The 2025 National AI R&D Strategic Plan should facilitate public-private partnerships or initiatives focused on developing automated validation tools for open-source repositories, datasets, models, libraries, and packages.[23] These R&D efforts could be modeled after the Defense Advanced Research Projects and Agency's existing AI and cybersecurity competitions, incentivizing collaboration between the public, private firms, government agencies, and academic institutions to engage in the creation of scalable validation frameworks and model-validation capabilities.[24]

   Moreover, researchers and developers should be encouraged to develop embedded provenance-tracking capabilities to enhance accountability, improve attribution, and deter malicious actors. These tools could employ cryptographic tagging or distributed ledger technologies to create an immutable record of any updates, contributors, and deployment contexts, thereby allowing the open-source AI community to verify, audit, and manage contributions in real time.[25]

3. *Examine Open-Source Software and AI Supply Chain Security:*
   Adversarial nations are increasingly targeting open-source software and AI supply chains—including security tools, code repositories, model dependencies, training datasets, and compute infrastructure—creating vulnerabilities that can be exploited to undermine U.S. national

[21] Ibid.
[22] Ibid.
[23] Ibid.
[24] Ibid.
[25] Ibid.

security.[26] For example, UNC5174, a Chinese state-sponsored espionage group, was recently observed releasing "open-source offensive security tools like VShell and WebSockets to [obscure] its target[ing] of Western governments, technology companies, research institutions, and think tanks".[27] Given these evolving campaigns, the 2025 National AI R&D Strategic Plan should prioritize research that examines how malicious threat actors may seek to compromise open-source software and AI supply chains, with a focus on identifying emerging techniques, vulnerabilities, and attack vectors. This work should also build on existing supply chain security guidance, including NIST's Secure Software Development Practices for Generative AI and Dual Use Foundation Models (SP 800-218A), and align with broader federal cybersecurity supply chain frameworks, such as NIST's Cybersecurity Supply Chain Risk Management Practices for Systems and Organizations (SP 800-161 Rev. 1).[28]

If secured, supported, and guided through strategic AI R&D, open-source AI ecosystems can serve as both an engine of innovation and a pillar of our national security.

IV. Securing Agentic AI and Shaping the Future of Human-Machine Collaboration

AI agents—those capable of autonomous decision-making, action, and adaptation—present emerging benefits and capabilities that challenge existing cybersecurity models, governance frameworks, and operational assumptions.[29] Agentic AI systems also have the potential to introduce risks like emergent multi-agent behaviors, model hijacking, and decision-making beyond human oversight, which often cuts across sectors, supply chains, and conventional ownership boundaries.[30] These cross-cutting risks are significant because they are unlikely to be fully addressed by private-sector innovation alone, thereby requiring dedicated federal R&D leadership and investment to advance security, accountability, and

---

[26] David Kirichenko, "Predictions for Open Source Security in 2025: AI, State Actors, and Supply Chains," Open Source Security Foundation, Jan. 23, 2025. https://openssf.org/blog/2025/01/23/predictions-for-open-source-security-in-2025-ai-state-actors-and-supply-chains; Haiman Wong, "DeepSeek's cybersecurity failures expose a bigger risk. Here's what we really should be watching," R Street Institute, Feb. 4, 2025. https://www.rstreet.org/commentary/deepseeks-cybersecurity-failures-expose-a-bigger-risk-heres-what-we-really-should-be-watching; Jai Vijayan, "China's Silk Typhoon APT Shifts to IT Supply Chain Attacks," *Dark Reading*, March 5, 2025. https://www.darkreading.com/remote-workforce/china-silk-typhoon-it-supply-chain-attacks; Andy Greenberg and Matt Burgess, "The Mystery of 'Jia Tan,' the XZ Backdoor Mastermind," *Wired*, April 3, 2024. https://www.wired.com/story/jia-tan-xz-backdoor; Jessica Lyons, "As nation-state hacking becomes 'more in your face,' are supply chains secure?", *The Register*, March 24, 2025. https://www.theregister.com/2025/03/24/nation_state_supply_chain_attack;

[27] Derek B. Johnson, "Chinese espionage group leans on open-source tools to mask intrusions," *CyberScoop*, April 15, 2025. https://cyberscoop.com/chinese-espionage-group-unc5174-open-source-tools.

[28] Harold Booth et al., "Secure Software Development Practices for Generative AI and Dual-Use Foundation Models: An SSDF Community Profile," National Institute of Standards and Technology, July 2024. https://csrc.nist.gov/pubs/sp/800/218/a/final; Jon Boyens et al., "Cybersecurity Supply Chain Risk Management Practices for Systems and Organizations," National Institute of Standards and Technology, May 2022. https://csrc.nist.gov/pubs/sp/800/161/r1/upd1/final.

[29] Kinza Yasar, "What are AI agents?", TechTarget, December 2024. https://www.techtarget.com/searchenterpriseai/definition/AI-agents.

[30] Anna Gutowska, "What are AI agents?", IBM, July 3, 2024. https://www.ibm.com/think/topics/ai-agents.

trust in agentic AI systems—especially as these capabilities are increasingly deployed in national security and critical infrastructure applications.[31]

1. *Expand R&D into Agentic AI Security and Adversarial Testing:*
   The 2025 National AI R&D Strategic Plan should prioritize foundational research that strengthens the security and resilience of agentic AI systems across their lifecycle. This includes investing in adversarial testing, agent-specific risk modeling, and resilience evaluations focused on architectural features like memory integrity, decision autonomy thresholds, and emergent behaviors.[32] Since AI agents are capable of independently learning and executing tasks, many of their emerging risks lack clear ownership and liability, making federal R&D leadership essential to ensure ongoing efforts are aligned with national security priorities, shared openly, and used to inform cross-sector coordination and best practices.[33] Public-private partnerships, interdisciplinary research hubs, and competitive grants can accelerate this work and ensure lessons learned are rapidly disseminated across the AI ecosystem.

2. *Develop Attribution, Interaction, and Response Capabilities for Agentic AI:*
   Agentic AI systems will require more sophisticated identification and traceability mechanisms than non-agentic AI systems or software supply chains. To facilitate meaningful oversight and promote accountability, the 2025 National AI R&D Strategic Plan should support research into persistent agent identifiers and dynamic logging systems capable of capturing the full lifecycle of agentic activity.[34] This includes tracking initial and ongoing data collection strategies, third-party dependencies, completed and pending tasks, reasoning logic streams, memory recall, and decision pathways.[35] Developing these attribution, interaction, and response capabilities for agentic AI systems would enable developers, end users, researchers, and cyber defenders to conduct real-time behavioral analysis, version-control monitoring, post-incident forensics, and attribution of specific actions to agent instances—providing critical visibility and explainability across multi-agent systems and high-stakes environments.[36] These efforts extend beyond existing provenance-tracking efforts by introducing agent-specific commitment devices, identity-binding mechanisms, and agent IDs to ensure that agent behavior remains consistent with the intended scope and objectives.[37]

---

[31] Phaedra Boinodiris and Jon Parker, "The evolving ethics and governance landscape of agentic AI," IBM, March 21, 2025. https://www.ibm.com/think/insights/ethics-governance-agentic-ai; Brandon Vigliarolo, "It begins: Pentagon to give AI agents a role in decision making, ops planning," *The Register*, March 5, 2025. https://www.theregister.com/2025/03/05/dod_taps_scale_to_bring.
[32] Ibid.
[33] Shomit Ghose, "The Next 'Next Big Thing': Agentic AI's Opportunities and Risks," UC Berkeley Sutardja Center for Entrepreneurship & Technology, Dec. 19, 2024. https://scet.berkeley.edu/the-next-next-big-thing-agentic-ais-opportunities-and-risks.
[34] Alan Chan et al., "Infrastructure for AI Agents," arXiv, Jan. 17, 2025. https://arxiv.org/html/2501.10114v1
[35] Ibid.
[36] Ibid.
[37] Ibid.

3. *Establish Agentic AI Testbeds and Human-Machine Collaboration Sandboxes:* AI agents are poised to transform business operations and workforce dynamics, with some experts suggesting that up to "80 percent of a worker's day-to-day tasks could be fully automated within the next five years".[38] Given the complex and still often unpredictable behaviors of agentic AI systems, the 2025 National AI R&D Strategic Plan should support the creation of secure, open-access testbeds and sandbox environments where these systems can be safely deployed and evaluated at scale. These environments would enable researchers, developers, and end users to stress-test agentic AI systems under simulated scenarios, allowing for the observation of emergent agentic behaviors, decision-making under uncertainty, and collaboration dynamics between humans and AI agents.[39] Beyond validating system performance, security, and resilience, these testbeds and sandboxes could serve as a "proving ground" for advancing strategies to manage the boundaries of agentic behavior, ensuring they stay aligned with their intended use cases and operate within defined, accountable limits.[40] Furthermore, these environments can also help refine approaches for clarifying permissible use cases and ensuring agentic AI systems augment human decision-making and talent, rather than undermining it or displacing it.[41]

As the private sector continues to accelerate the development of increasingly sophisticated agentic AI capabilities, the federal government has a crucial role to play in ensuring that our foundational understanding, oversight mechanisms, and agentic interaction infrastructure are ready to scale alongside these advancements.

V. Conclusion

AI security is not a constraint on innovation—it is a prerequisite for ensuring that America's AI advancements are scalable and resilient. As the United States crafts its 2025 National AI R&D Strategic Plan, it is imperative to recognize that securing evolving AI systems, agentic capabilities, and open-source ecosystems is foundational work that will sustain our momentum, competitiveness, and global leadership.

While adversaries may rush forward with fragile and opaque systems, America has always led by building cutting-edge technologies that the world can trust. That same approach must guide our AI R&D priorities, investments, and efforts—ensuring we lead the race to innovate on America's terms:

---

[38] Kate Whiting, "What is an AI agent and what will they do? Experts explain," World Economic Forum, July 24, 2024. https://www.weforum.org/stories/2024/07/what-is-an-ai-agent-experts-explain.

[39] Xuhui Zhou et al., "HAICOSYSTEM: An Ecosystem for Sandboxing Safety Risks in Human-AI Interactions," arXiv, Oct. 21, 2024. https://arxiv.org/abs/2409.16427; Daniel Weitekamp et al., "TutorGym: A Testbed for Evaluating AI Agents as Tutors and Students," arXiv, May 2, 2025. https://arxiv.org/abs/2505.01563; Jennifer Grannen et al., "Vocal Sandbox: Continual Learning and Adaptation for Situated Human-Robot Collaboration," arXiv, Nov. 4, 2024. https://arxiv.org/abs/2411.02599.

[40] Ibid.

[41] Shana Lynch, "Predictions for AI in 2025: Collaborative Agents, AI Skepticism, and New Risks," Stanford University Human-Centered Artificial Intelligence, Dec. 23, 2024. https://hai.stanford.edu/news/predictions-for-ai-in-2025-collaborative-agents-ai-skepticism-and-new-risks.

grounded in our values, committed to national security, and focused on fostering prosperity.

We are grateful for your attention to this national priority and stand ready to assist in advancing an AI R&D strategy that aligns with American values and delivers on its promise of innovation, resilience, and enduring leadership.

Respectfully submitted,

Haiman Wong, *Fellow, Cybersecurity and Emerging Threats Team*

This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in developing the 2025 National AI R&D Strategic Plan and associated documents without attribution.