Institute for SECURITY + TECHNOLOGY

Institute for Security and Technology
195 41st Street #11045
Oakland, CA 94611

March 15, 2025

Networking and Information Technology
Research and Development (NITRD) Program
Attn: Mr. Faisal D'Souza
2415 Eisenhower Ave
Alexandria, VA 22314

Subject: Comments in response to the OSTP's Request for Information on the AI Action Plan.

Dear Mr. D'Souza,

The Institute for Security and Technology (IST) appreciates the opportunity to submit input on the highest priority policy actions that should be included in the new AI Action Plan. IST is the critical action think tank uniting technology and policy leaders to create actionable solutions to emerging security challenges. Our goal is to assist national security policymakers in operating at the cutting edge, and we strongly feel AI is currently far outpacing our national security policy apparatus.

In our view, there is no greater national security priority for the United States than artificial intelligence. This is not hyperbole. We are convinced of the overriding national security imperative created by cutting edge AI developments and believe the impact from AI seen to date has been relatively insignificant compared to what is coming next.

IST regularly engages with a diverse range of stakeholders from across the AI ecosystem, including leading AI labs, to better understand the opportunities and emerging risks from cutting edge AI capabilities, develop technical and policy oriented risk reduction strategies, and drive forward powerful and yet responsible innovation. We remain convinced that it is possible to lead in AI development while also prioritizing safety and security.

IST has engaged with the technical AI community in depth since 2017. Our original focus was on the implications of AI for the battlefield,[1] and evolved to include developing recommendations for confidence building measures around AI and nuclear command, control, and communications (NC3) for the Department of State.[2] In more recent years, our efforts have broadened even further as AI has rapidly become more generally applicable. This work has included efforts focused on the implications of AI in cybersecurity and the offence-defense balance,[3] identifying the risks of "open source" AI,[4] investigating AI's impact on human cognition,[5] and deeper work on the role of AI in NC3.[6]

We are now deeply focused on addressing AI in a more comprehensive manner. Accordingly, our recommendations begin with one dominant idea: the United States needs an updated, comprehensive national security strategy for AI.

That strategy must take into consideration that artificial intelligence is now critical infrastructure, and that within the next few years, AI will underwrite if not absolutely revolutionize all elements of national power. A new national security strategy for AI must focus on (but not solely be limited to) the following strategic objectives:

- Strategic Objective #1: Achieve energy security and resilience
- Strategic Objective #2: Protect U.S. and allied technology
- Strategic Objective #3: Retain and expand world-class talent

---

[1] T4GS, "AI and Human Decision-Making: AI and the Battlefield", T4GS Reports, November 28, 2018, http://www.tech4gs.org/ai-and-human-decisionmaking.html

[2] Alexa Wehsener et al, "AI-NC3 Integration in an Adversarial Context: Strategic Stability Risks and Confidence Building Measures", Institute for Security and Technology, February 2023, https://securityandtechnology.org/virtual-library/reports/ai-nc3-integration-in-an-adversarial-context-strategic-stability-risks-and-confidence-building-measures/.

[3] Jennifer Tang, Tiffany Saade, and Steve Kelly, "The Implications of Artificial Intelligence in Cybersecurity: Shifting the Offense Defense Balance", Institute for Security and Technology, October 2024, https://securityandtechnology.org/wp-content/uploads/2024/10/The-Implications-of-Artificial-Intelligence-in-Cybersecurity.pdf.

[4] Zoë Brammer, "How Does Access Impact Risk? Assessing AI Foundation Model Risk Along a Gradient of Access", Institute for Security and Technology, December 2023, https://securityandtechnology.org/virtual-library/reports/how-does-access-impact-risk-assessing-ai-foundation-model-risk-along-a-gradient-of-access/.

[5] Gabrielle Tran and Eric Davis, "The Generative Identity Initiative: Exploring Generative AI's Impact on Cognition, Society, and the Future", Institute for Security and Technology, December 2024, https://securityandtechnology.org/wp-content/uploads/2025/01/The-Generative-Identity-Initiative.pdf

[6] IST Launching New Initiative with Support from Longview Philanthropy Focused on the Integration of AI into Nuclear Command, Control, and Communications, November 2024, Institute for Security and Technology, https://securityandtechnology.org/blog/ist-launching-new-initiative-with-support-from-longview-philanthropy-focused-on-the-integration-of-ai-into-nuclear-command-control-and-communications/

- Strategic Objective #4: Lead in AI development while prioritizing safety and security
- Strategic Objective #5: Press the multi-domain advantage
- Strategic Objective #6: Anticipate and shape artificial general intelligence (AGI)

We will now expand on these suggested objectives.

Achieve energy security and resilience

The data centers required to train and operate cutting edge AI require vast amounts of electricity, so much so that technology firms are building and purchasing dedicated energy resources. These include re-activated nuclear power plants, small modular nuclear reactors (SMRs), and solar farms. Such distributed energy resources, which are intended to be located closer to the point of use, can enhance resilience for critical facilities like data centers. But these connected technologies are also susceptible to attack and require additional cybersecurity focus. IST is convinced the United States can build the energy infrastructure needed to maintain a global AI advantage, but its security must be a top priority.

Protect U.S. and allied technology

The U.S. must move to secure frontier AI systems from espionage and sabotage risks; these become increasingly incentivized as both the AI systems themselves increase in economic value as well as the systems approach thresholds for automated AI research and development. We may be reaching this point as early as late this year and quite likely next year. The effects of covert small-scale sabotage at the time of automated AI R&D could compound to create a substantial setback, if not to a loss of the United States' lead in AI. Securing frontier AI development from nation-state attacks must be treated as a national security priority.

IST has long been a strong proponent of secure and resilient critical infrastructure. In line with this theme, IST is also currently leading an effort that is focused on what is needed for AI labs to reach Security Level 5 (SL5),[7] the highest current designation for AI security against nation-state attacks. The newly-established SL5 Task Force is an industry-focused

[7] Nevo et al, "Securing AI Model Weights: Preventing Theft and Misuse of Frontier Models", Rand Corporation, May 2024 https://www.rand.org/pubs/research_reports/RRA2849-1.html

multi-stakeholder working group with the mission to create the optionality for American frontier AI labs to deploy SL5 within one to three months of choosing to do so. We work on threat modelling, technical roadmapping for ML development-optimised security infrastructure, productivity cost estimates, and prototype critical components with input from frontier labs.

Along with making outsider attacks on U.S. frontier AI more expensive, investment in SL5 also helps mitigate some of the possible risks from agentic AI systems and malicious insiders. As the development of AI agents matures, we should expect them to begin taking autonomous actions at scale and to form novel threat vectors. Luckily, several of the SL5 interventions share infrastructural components with actions we expect to need to take to defend against these additional risks.

We propose a strategy of the U.S. government supporting efforts to attain said optionality, while resting the choice on whether to deploy with the frontier labs. With possibly only a very small time window of one to two years remaining until we may want to deploy, starting now allows for the necessary iteration to develop technical plans that minimize costs on productivity (as far as possible), as well as assures prerequisite steps are taken to allow for rapid deployment once the security and productivity tradeoff shifts to clearly favor deployment.

Further, we encourage the U.S. government to consider creating antitrust protections to allow frontier AI labs in coordinating and benefit-sharing with regards to AI security. Different companies in the AI industry will need to rapidly solve some very similar and resource-intensive issues in AI security (e.g., understanding how to build AI development-optimized secure facilities, how to build retrofittable large data centers, implementing stronger information compartmentalization without incurring productivity costs). Opportunities for carefully defined, limited exemptions from antitrust laws would allow such industry coordination on security standards and may facilitate much needed faster, cheaper, and more efficient adoption.

In addition to protecting the most advanced AI models from espionage and sabotage, we must also take steps to prevent adversarial nations from gaining access to essential components within the AI supply chain, like advanced microprocessors. Based on our team’s knowledge and experience from prior governmental roles, supplemented with more recent open source anecdotes, the People’s Republic of China is too often evading

our and likeminded nations' export controls, and we must do better. IST will soon kick-off a new year-long research effort to understand the root causes of this compliance failure and develop a comprehensive framework for an enhanced multi-agency AI chip export controls enforcement program within the U.S. national security apparatus.

Retain and expand world-class talent

The United States continues to lead the world in science, technology, engineering, and math higher education and attracts the best and brightest students from across the world. Given the insatiable market appetite for those who can design AI systems, the U.S. must augment its domestic workforce pipeline by selectively tapping into this pool of foreign talent through visa and permanent residency opportunities. At the same time, we must remain cognizant of the foreign intelligence threat, as this student body includes large numbers of foreign nationals from high-threat countries who, it is well understood, may be subject to tasking by their respective intelligence services. This risk is most immediately realized in the exposure of advanced research within the universities and proprietary business information through student job placements, but can persist and expand through work permits and U.S. employment gained post-graduation.

To ensure the United States is able to prevail in the global techno-industrial competition, IST recommends the Administration pursue a three-pronged AI workforce strategy:

1. *STEM Patriots* – Ensure the sufficiency of K-12 math, science, and computer technology education; encourage and incentivize U.S. students to pursue STEM higher education.
2. *Victory Visas* – As needed to achieve and maintain U.S. competitiveness, make streamlined work visas, a path to permanent residency, and even U.S. citizenship readily available to the best and brightest minds on AI and related emerging technologies.
3. *Wean & Lean* – In light of the foreign intelligence threat, better manage the number of foreign nationals from high-threat nations granted U.S. student visas; wean American universities from their dependency on foreign student tuition revenues from these locations.

Lead in AI development while prioritizing safety and security

AI's rapid advancement requires a parallel commitment to safety, security, and resilience. As AI systems become more autonomous and integrated into critical infrastructure, financial markets, and national security apparatuses, their vulnerabilities become strategic risks. Adversaries are already exploiting AI for disinformation, cybercrime, and automated attacks, underscoring the urgency of securing AI models, supply chains, and deployment pipelines against intrusion or manipulation. As such, AI builders—ranging from research labs, startups, and major tech firms—and AI users—spanning enterprises, governments, and infrastructure operators—must implement proportional safeguards, particularly in high-impact domains and applications.

IST recommends focusing on three critical risk categories: (1) the malicious use of AI, including fraud, disinformation, and attacks on critical infrastructure; (2) compliance failure, where AI systems fall short of regulatory and governance mandates; and (3) diminished human oversight, where increasing automation risks eroding essential human judgement in high-stakes decision making. In seeking to address these risks, IST submits for consideration elements of the following reports: "*A Lifecycle Approach to AI Risk Reduction*: *Tackling the Risk of Malicious Use Amid Implications of Openness*" (published in June 2024), *"Navigating AI Compliance: Tracing Failure Patterns in History*" (published in December 2024), and "*Navigating AI Compliance: Risk Mitigation Strategies for Safeguarding Against Future Failures*" (to be published in March 2025).[8,9]

In consultation with a working group of 20 stakeholders from leading AI labs, industry, academia, and civil society, IST has developed a comprehensive set of recommendations and best practices aimed at reducing AI-related risks and enhancing the development and deployment process for both AI builders and users. By embedding security, transparency, and accountability throughout the AI lifecycle, the U.S. can ensure AI development remains secure and resilient while continuing to drive innovation.

---

[8] Louie Kangeter, "A Lifecycle Approach to AI Risk Reduction," Institute for Security and Technology, June 2024, https://securityandtechnology.org/wp-content/uploads/2024/06/A-Lifecycle-Approach-to-AI-Risk-Reduction.pdf (p.7)

[9] Mariami Tkeshelashvili, Tiffany Saade, "Navigating AI Compliance, Part 1," Institute for Security and Technology, https://securityandtechnology.org/wp-content/uploads/2024/12/Navigating-AI-Compliance.pdf

IST considers that balancing AI innovation and risk management should be the cornerstone of the AI Action Plan or a broader new national security strategy for AI. Current geopolitical instability indicates that AI will increasingly shape how international tensions emerge and escalate, requiring enhanced U.S. government capacity to address these challenges. While continuing to invest in AI capabilities that enhance government effectiveness, the executive branch should not overlook implementing safe and secure AI deployment practices that protect human rights and user safety. By fostering an environment conducive to responsible innovation while prioritizing risk identification, assessment, and mitigation, the government can lead the way in harnessing AI's transformative potential while ensuring its development aligns with established security and safety protocols, in turn enhancing accountability and fostering public trust.

Press the multi-domain advantage

The term “strategic stability” is often only used in reference to nuclear weapons and their strategic implications. In the age of AI, that type of broad-based stability will only arise through the use of AI across all domains including the cyber and cognitive domains, as well as the more commonly discussed air, land, sea, and space. By doing so, the United States would maintain options for delivering an overwhelming response if and when a potential adversary oversteps certain thresholds of activity. But at the same time, these capabilities can be used during steady state to protect the homeland and American people by realizing the defender’s data advantage against malicious actors using AI systems.

The work we have done in the cyber and cognitive domains point to how AI is revolutionizing these issue areas. The offense-defence balance in cyber and AI is an issue area where IST has spent considerable time, and while we assess the balance may currently lean toward the defender, that will not last without significant, collective effort. The same is not the case in the cognitive domain: IST’s work as part of our Generative Identity Initiative (cited earlier) points to a reality where powerful AI tools can all too easily be used to manipulate and persuade large segments of the population in incredibly subtle and powerful ways. IST recommends the U.S. government consider these threats as part of any national security strategy for AI.

Anticipate and shape artificial general intelligence (AGI)

As we have already made clear at the outset of this letter, IST assesses there is no greater national security priority for the United States than AI. Within this assertion, we firmly believe the potential for artificial general intelligence (AGI) must be well understood as a national security challenge. The debate continues to evolve as to the probabilities and timelines associated with the creation of AGI, but what is undeniable is that there is a chance it becomes a reality within the next few years. Ignoring the possibility of such powerful tools would be at our national peril, both from the perspective of what they would mean for national power and also for how anyone realistically will be able to maintain control over such incredibly powerful capabilities.

The work currently underway to understand these implications is strong, but falls far short of the scale and scope of effort required to anticipate these potential impacts and challenges. IST is working closely with stakeholders and partners across the ecosystem to contribute to these discussions, research, and debates, but strongly recommends there be a much more robust national-level conversation regarding the potential implications of AGI. Leaving these discussions and developments solely in the hands of the private sector is no longer a realistic option, and in the vein of much of our other work, IST strongly advocates for new multi-stakeholder efforts to understand the nature of the technological developments around AGI and their national security implications.

We and the IST team welcome an opportunity to discuss our work and these comments with you. Thank you for considering them as you draft this essential AI Action Plan.

Regards,

Philip Reiner
Chief Executive Officer

Steve Kelly
Chief Trust Officer

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 AI Action Plan and associated documents without attribution.