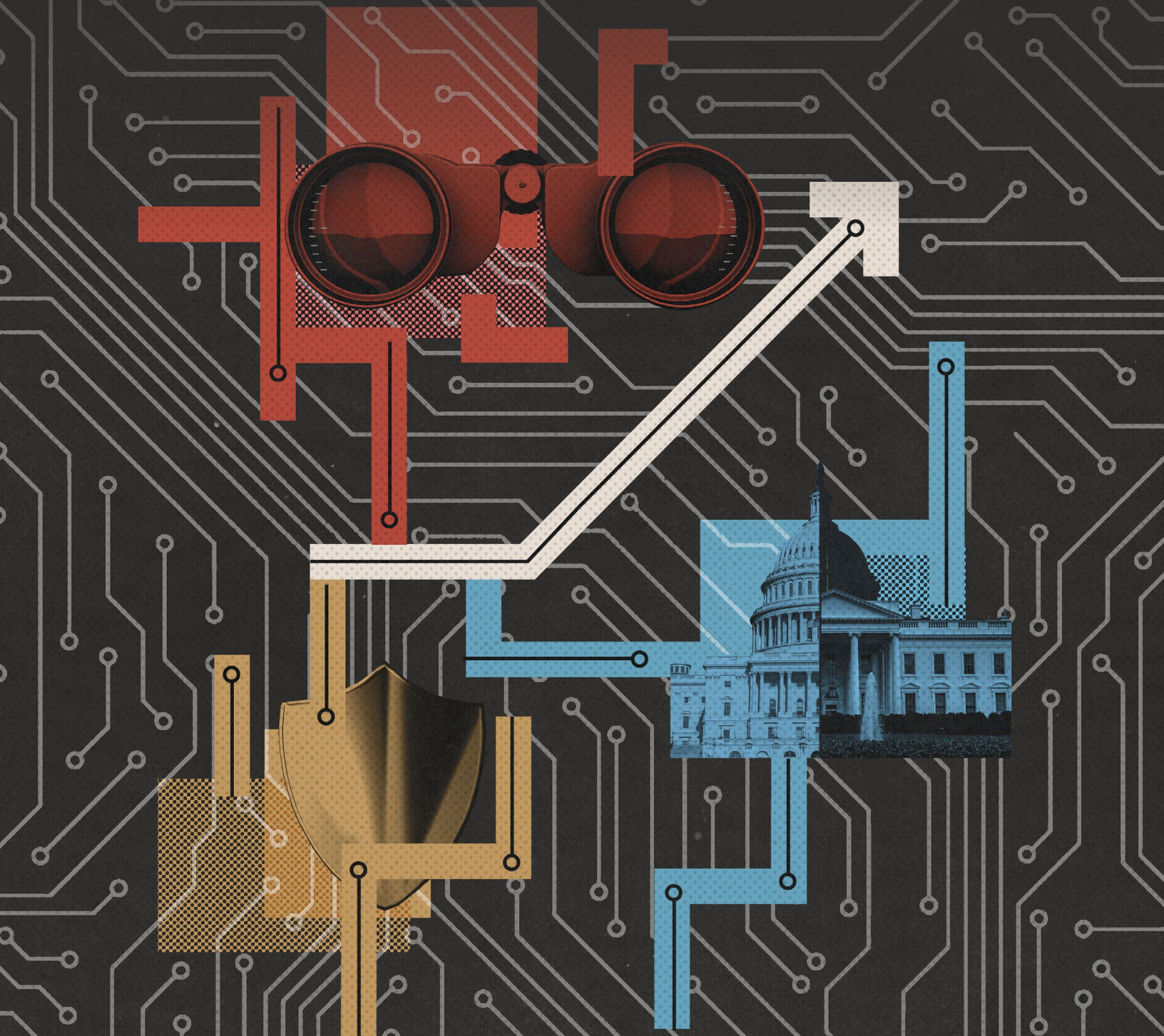
CNAS | AMERICA'S EDGE

PREPARED, NOT PARALYZED

Managing AI Risks to Drive American Leadership

Janet Egan, Spencer Michaels, and Caleb Withers



About the Authors

Janet Egan is the deputy director and senior fellow with the Technology and National Security Program at the Center for a New American Security (CNAS). Her research focuses on the national security implications of artificial intelligence (AI) and other emerging technologies, including how compute policy can be used to manage the risks and opportunities of advanced AI systems. Prior to joining CNAS, Egan was a director in the Australian Government Department of the Prime Minister and Cabinet. She has applied experience working on policy at the intersection of national security, economics, and international relations. Her work has spanned issues such as 5G security, cybersecurity, countering foreign interference, foreign investment and trade, and critical infrastructure regulations. She was a member of the prime minister's task force on critical technologies and the inaugural director of policy in Australia's Office of Supply Chain Resilience. Egan holds a master's degree in public policy from the Harvard Kennedy School and a bachelor of arts from Monash University in Australia.

Spencer Michaels is an independent researcher working with the Technology and National Security Program at CNAS. His work focuses on the national security risks and geopolitical implications of emerging technologies. Michaels was formerly a Joseph S. Nye, Jr. National Security intern at CNAS. He holds a BA from Amherst College with a double major in Russian and LJST (Law, Jurisprudence, and Social Thought). His undergraduate research focused on Russian geopolitics, AI, philosophy, and contemporary warfare. He speaks Russian fluently.

Caleb Withers is a research associate for the Technology and National Security Program at CNAS. He focuses on frontier AI and national security, including emerging AI capabilities, their impacts in the biological and cyber domains, and compute policy. Before CNAS, Withers worked as a policy analyst for a variety of New Zealand government departments. He has an MA in security studies from Georgetown University, concentrating in technology and security, and a bachelor of commerce from Victoria University of Wellington, New Zealand, majoring in economics and information systems.

About the Technology and National Security Program

The CNAS Technology and National Security Program produces cutting-edge policy research to secure America's edge in emerging technologies, while managing potential risks to security and democratic values. The program produces bold, actionable recommendations to drive U.S. and allied leadership in responsible technology innovation, adoption, and governance. The Technology and National Security Program focuses on three high-impact technology areas: AI, biotechnology, and quantum information sciences. It also conducts cross-cutting research to strengthen U.S. technology statecraft to promote secure, resilient, and rights-respecting digital infrastructure and ecosystems abroad. A focus of the program is convening the technology and policy communities to bridge gaps and develop solutions.

Acknowledgments

The authors are grateful to Cole Salvador, Vivek Chilukuri, Paul Scharre, Keegan McBride, Charlie Bullock, Emily Kilkrease, Carson Ezell, Brian McGrail, Muhammad Mustafa Nasim Ul Ghani, Joe O'Brien, Michelle Nie, and Ryan Beane for their valuable feedback and suggestions on earlier drafts of this report. The report also would not have been possible without the research, editorial, and design contributions of CNAS colleagues Caroline Steel, Melody Cook, Maura McCarthy, and Alina Spatz. This report was made possible with the generous support of Founders Pledge.

As a research and policy institution committed to the highest standards of organizational, intellectual, and personal integrity, CNAS maintains strict intellectual independence and sole editorial direction and control over its ideas, projects, publications, events, and other research activities. CNAS does not take institutional positions on policy issues and the content of CNAS publications reflects the views of their authors alone. In keeping with its mission and values, CNAS does not engage in lobbying activity and complies fully with all applicable federal, state, and local laws. CNAS will not engage in any representational activities or advocacy on behalf of any entities or interests and, to the extent that the Center accepts funding from non-U.S. sources, its activities will be limited to bona fide scholastic, academic, and research-related activities, consistent with applicable federal law. The Center publicly acknowledges on its website annually all donors who contribute.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

THE TRUMP ADMINISTRATION has embraced a pro-innovation approach to artificial intelligence (AI) policy. Its AI Action Plan, released July 2025, underscores the private sector's central role in advancing AI breakthroughs and positioning the United States as the world's leading AI power.[1] At the Paris AI Action Summit in February 2025, Vice President JD Vance cautioned that an overly restrictive approach to AI development "would mean paralyzing one of the most promising technologies we have seen in generations."[2]

Yet this emphasis on innovation does not diminish the government's critical role in ensuring national security. On the contrary, AI advances will yield significant threats alongside unprecedented potential in this domain. Experts warn of advanced AI introducing more autonomous cyber weapons, bestowing a broader pool of actors with the know-how to develop biological weapons, and potentially malfunctioning in ways that cause massive damage.[3] Private and public sector leaders alike have echoed these concerns.[4] The urgent task for policymakers is to ensure that the federal government can anticipate and manage the national security implications of AI with advanced capabilities—without resorting to blunt, ill-targeted, or burdensome regulation that would undermine America's innovative edge. In other words, the government must prepare at once for potential risks from rapidly advancing AI without imposing onerous regulations that unduly stifle the technology's vast potential for good.

The status quo is insufficient: Technical expertise in advanced AI remains concentrated in a handful of companies, and the government is playing catch-up. Existing voluntary information-sharing commitments between AI labs and the federal government already face hurdles and likely will prove insufficient over time as the costs of providing transparency increase. Meanwhile, the private sector lacks both the national security expertise and the commercial incentives to manage these risks to the national interest. The United States cannot afford policies built solely on speculative fears. Yet in the face of real and rapid progress in national security–relevant capabilities, neither can it risk allowing an AI-driven disaster or a regulatory vacuum to derail technological progress.

While much of the policy debate rightly focuses on innovation and accelerating adoption, this report concentrates on a less developed but equally vital counterpart: managing the risks that could undermine those ambitions without stifling AI's innovative potential. Effective risk management is not a brake on progress but a prerequisite for it, playing an essential role in sustaining public trust, preventing setbacks, shaping global standards, and ensuring that American leadership in AI endures over the long term.

Yet the pace of AI progress is accelerating, exacerbating the difficulties of developing evidence-based policy and being responsive to emerging risks and opportunities. The federal government needs to strengthen its ability to manage AI risks without

overregulating. It can do this through building three interconnected capacities:

- **Situational awareness** to detect, analyze, and communicate emerging AI risks and opportunities;
- **Agile policymaking** that can adapt and scale proportionately to evolving threats; and
- **Incident response and readiness** to manage and contain significant AI-related incidents should they arise.

The AI Action Plan provides an ambitious foundation for these capacities. But it remains a high-level blueprint, leaving gaps in coverage and open questions around implementation and authorities.

This report makes the case for robust, proactive federal government engagement in AI risk management. It examines the current state of U.S. preparedness, assesses the AI Action Plan's contributions, identifies persistent shortcomings and gaps, and offers solutions to address them. The report advances the following recommendations for U.S. policymakers:

To establish AI situational awareness in government:

1. Empower and equip the Center for AI Standards and Innovation (CAISI) as the federal government's center of technical AI expertise and evaluation.
    1.1. Designate CAISI as the federal government's interagency lead for AI risks.
    1.2. Fund CAISI sufficiently to execute its critical role.
2. Strengthen information flows from frontier AI developers to government.
    2.1. Congress should pass a bill enacting greater protections for AI whistleblowers.
    2.2. The Office of Science and Technology Policy (OSTP) should work with CAISI and other relevant agencies to develop a plan for mandating information sharing and testing for dangerous capabilities, in case voluntary mechanisms prove inadequate.

To bolster policy agility:

3. Establish an interagency AI National Security working group, co-led by the OSTP and the National Security Council, to strengthen intra-government coordination on AI national security risks.
4. Prepare contingency planning for AI risk scenarios to allow expedited policy action.
5. Establish regular congressional reports by the AI National Security interagency working group to ensure Congress is aware of emerging risks and policy options.
6. Work with allies and partners to harmonize policy approaches to identified AI risks.

To strengthen incident response capacity:

7. Build stronger interconnectivity between the range of agencies and stakeholders that would need to coordinate a response to an incident.
    7.1. Engage AI companies and experts in updating the Cybersecurity and Infrastructure Security Agency incident response playbooks.
    7.2. Ensure the AI Information Sharing and Analysis Center also includes representatives from the AI industry.
    7.3. Conduct regular tabletop exercises including government, private sector, and nonprofit representatives to bolster connectivity between incident responders across public and private sectors.
8. Establish a mechanism for post-incident review and lesson learning.
9. Engage with international partners, including adversaries, on best practices for real-time AI incident response.

INTRODUCTION

THE RAPID ADVANCEMENT OF artificial intelligence (AI) presents policymakers with the challenge of governing a transformative technology that is both critical to national security and primarily driven by private innovation. The platitude that societies should harness the benefits of AI while managing its risks fails to address the central question: How?[5] As AI capabilities continue to advance rapidly, developing a sophisticated answer has become essential for maintaining America's technological leadership.

The Trump administration has signaled a clear commitment to American AI dominance, championing innovation over restrictive regulation.[6] However, this innovation-first approach does not negate the need for risk assessment and preparedness.[7] The potential consequences of advanced AI systems—from automated cyberattacks to exacerbated biosecurity risks to potential loss of control—demand serious attention from policymakers and national security experts. Leading researchers have identified scenarios where AI capabilities, if left unmanaged, could pose significant threats to national security, economic stability, and public safety.[8] Yet the significant uncertainty about these scenarios necessitates a nimble approach that emphasizes increasing policymaker awareness and speedy response.

Recognizing both the opportunity and urgency of this moment, the White House released the AI Action Plan in July 2025, outlining an ambitious agenda for adopting AI and managing the uncertainty inherent in any rapidly evolving sector. As President Donald Trump outlined at the AI Action Plan's launch, "This technology brings the potential for bad as well as for good, for peril as well as for progress. . . . We want to have rules, but they have to be smart."[9]

The key to smart rules lies in developing evidence-based policies that can adapt to emerging capabilities without stifling innovation. Yet here lies a fundamental tension: AI is evolving far faster than the traditional policy cycle, and most early regulatory proposals inevitably will rely on projections rather than robust evidence. Policymakers must balance the imperative to ground decisions in data with the reality that waiting for comprehensive evidence risks leaving society vulnerable to rapidly emerging threats. For many AI-enabled risks, defaulting to a wait-and-see approach will be woefully inadequate, because effective mitigations will take time to develop. Moreover, waiting for threats to fully materialize before implementing safeguards could result in a catastrophic

> **"This technology brings the potential for bad as well as for good, for peril as well as for progress. . . . We want to have rules, but they have to be smart."**
>
> —PRESIDENT DONALD TRUMP

incident that triggers public backlash and undermines the future of American AI progress—much like how nuclear accidents at Three Mile Island and Chernobyl derailed nuclear energy development for decades.[10] The stakes extend beyond domestic security and innovation. If the United States lags in shaping international AI standards and norms, it risks ceding strategic ground to competitors and allowing adversaries to shape the rules of the road for this critical technology.

Prioritizing three key capabilities can help the federal government to balance the tension between evidence and speed and chart a path that promotes both U.S. innovation and security. First, the government needs situational awareness—the ability to monitor emerging AI capabilities and risks and understand their implications for national security and society. Second, the government should increase its policy agility—the ability to adapt and change policy settings quickly, if risks materialize that warrant new approaches. Finally, the government needs robust incident response—the ability to effectively contain damage from AI incidents. Ideally, these capabilities will be mutually reinforcing: Situational awareness of emerging capabilities allows the government to more agilely develop targeted policies, including those that bolster incident preparedness. Lessons learned from incident response can help the government update its situational awareness and inform whether policies and rules need urgent revision.

AI has the potential to revolutionize national defense, accelerate scientific breakthroughs, and strengthen American competitiveness for generations. Realizing this vision requires an environment where innovation can proceed with public confidence. Effective government preparedness plays a role by establishing the conditions under which development can flourish. When policymakers have visibility into emerging capabilities, they can better target evidence-based measures to address genuine risks without imposing sweeping restrictions. When the government can adapt rules agilely, it can better respond to opportunities and risks as technology quickly evolves. When the government has well-practiced, robust incident response frameworks, individual setbacks won't trigger the kind of panic that could set American AI back years. The three capabilities outlined in this report—situational awareness, policy agility, and incident preparedness—form the foundation of this enabling framework, securing American leadership by creating a stable, trusted environment in which innovation can continue at pace and at scale.

The AI Action Plan outlined important first steps for AI preparedness, establishing a solid blueprint for bolstering the federal government's ability to understand, analyze, and respond to emerging capabilities and risks. Building on this foundation, the next phase of implementation requires translating the AI Action Plan's strategic vision into operational reality. Many of the plan's forward-looking initiatives now need additional authorities, detailed rulemaking, and dedicated resourcing to move from concept to execution. More work is needed to translate the vision of the AI Action Plan into action, plug residual gaps, and position the United States to responsibly lead the world through the AI transition.

This report addresses these challenges and charts a pathway forward to bolster federal government AI preparedness. It outlines why government involvement in AI risk preparedness is critical to national security and U.S. global leadership. The analysis focuses on the three critical capacities the government needs to address AI risks effectively: situational awareness, policy agility, and incident preparedness. Finally, the report highlights gaps in the U.S. AI preparedness ecosystem and recommends measures to address these shortfalls.

AI RISK MANAGEMENT: A PUBLIC PROBLEM IN PRIVATE HANDS

THE FEDERAL GOVERNMENT is confronting a technology that is vital to national security yet primarily developed and understood by the private sector. America's private sector has propelled the United States to global AI leadership through unprecedented investment and breakthroughs. Yet this success creates a dilemma: The dual-use nature of AI systems means the same growing capabilities that offer societal benefits also present national security risks. These risks increasingly demand government oversight, even as much of the expertise needed to understand them remains concentrated in private hands. The challenge lies in striking the right balance—too aggressive an approach risks stifling the innovation that underpins American leadership, while inaction leaves critical national security vulnerabilities unaddressed. This section explores the tension between innovation and national security and makes the case for a stronger government role in AI risk management.

The Private Sector Drives American AI Leadership

Within days of his second inauguration, President Trump signaled his administration's AI priorities through the executive order on "Removing Barriers to American Leadership in Artificial Intelligence."[11] The order credits America's AI dominance to "the strength of our free markets, world-class research institutions, and entrepreneurial spirit," beginning a policy framework that positions private sector innovation as the cornerstone of national AI leadership.[12] Unlike the development of nuclear technology, where government labs such as Los Alamos led the charge, or the early internet, which emerged from Defense Advanced Research Projects Agency (DARPA) research, private companies have dominated the development and deployment of AI since the emergence of the deep learning paradigm.[13]

Investments from the U.S. private sector dwarf government AI spending. Meta, Amazon, Google, and Microsoft alone are expected to invest as much as $320 billion into AI development and infrastructure in 2025.[14] OpenAI's flagship Stargate program is expected to drive $500 billion of investment in AI infrastructure in the United States over the next four years, with a similarly large investment project planned for the United Arab Emirates and smaller initiatives in Norway and the United Kingdom (UK).[15]

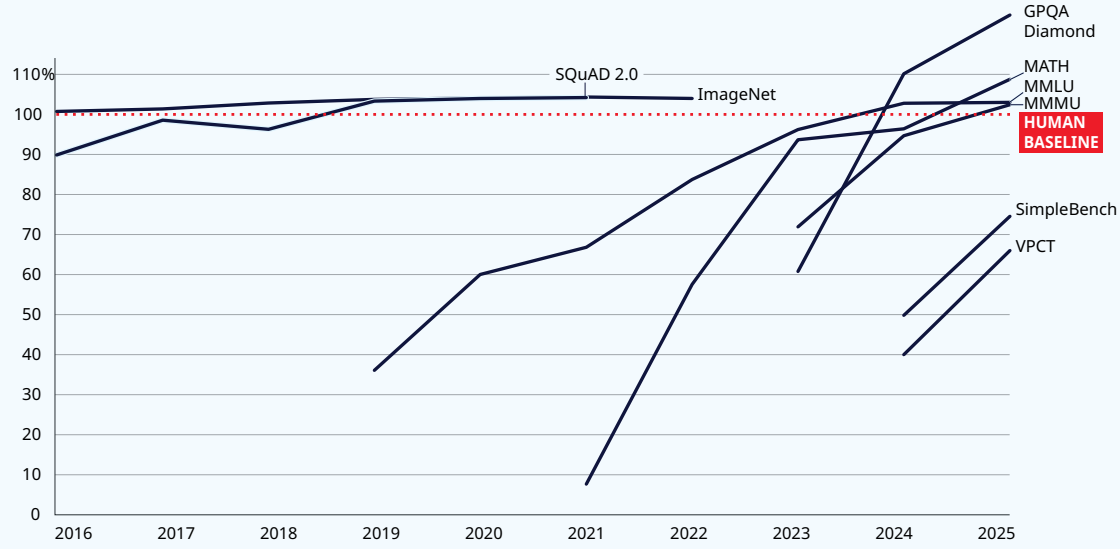
America's private sector has established the United States as the global leader in AI development. In 2024, U.S. companies released almost twice as many notable models as China (defined as models that made a state-of-the-art improvement on a benchmark) as well as models that were highly cited, historically relevant, or widely used.[16] Also in 2024, U.S. private investment in AI was $109.1 billion, almost 12 times

more than the equivalent in China.[17] The United States houses 74.4 percent of the world's compute, almost 94 percent of which the private sector owns.[18]

This surge of private sector investment has propelled remarkable growth in AI capabilities, leading to astronomical progress that shows few signs of stopping. AI models' performance on benchmarks measuring a range of capabilities has surpassed human baselines (Figure 1). Frontier AI models are increasingly completing lengthy and complex tasks across diverse fields. Where OpenAI's GPT-4, released in March 2023, could complete software engineering tasks that take humans five minutes on average, GPT-5, released just two and a half years later, can complete tasks that take humans an average of two hours and 17 minutes.[19] Another notable metric is the ability of AI models to outperform human experts in answering PhD-level questions in various subject areas. OpenAI's GPT-4, released in 2023, achieved only 36 percent on a key PhD-level benchmark; the next year, its o1 reasoning model scored 77 percent, surpassing human experts for the first time. By the release of GPT-5 in 2025, the model had reached 85 percent.[20] This growth in AI capability is happening faster than even experts anticipated. In 2022, a group of domain experts and forecasters had placed the likelihood of an AI system officially earning a gold medal at the 2025 International Math Olympiad, a prestigious mathematics competition for high school students, at between 2.3 percent and 8.6 percent.[21] Yet in July 2025, Google's Gemini 2.5 Deep Think system achieved this milestone.[22] In short, AI systems are approaching and exceeding human-level performance on complex reasoning tasks at a startling rate.

AI capabilities also have been improving in domains directly relevant to national security. Consider the dramatic acceleration in AI's dual-use expertise in virology—the science of viruses. In July 2023, GPT-4 Turbo outperformed 43 percent of experts on the Virology Capabilities Test, a benchmark measuring AI's ability to troubleshoot advanced virology laboratory situations with explicit dual-use potential. In a survey of expert virologists, most predicted that an AI model would not surpass a top virology team until after 2030. Yet by 2025, OpenAI's o3 model surpassed 94 percent of experts and matched the performance of top-tier virologists.[23]

Figure 1 | AI Performance on Benchmarks Relative to Human Baseline[24]



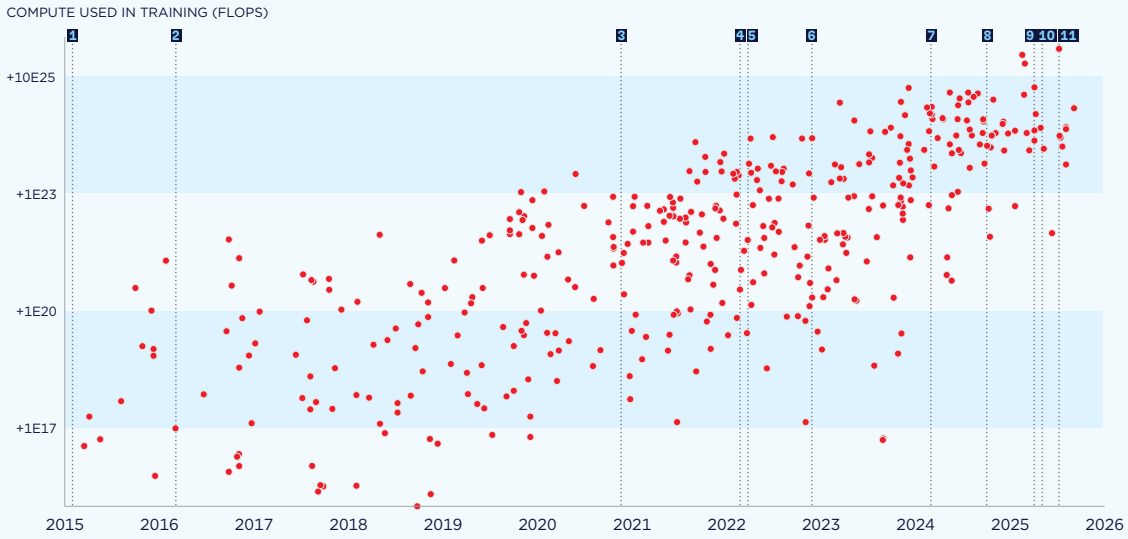
Key | AI Benchmarks

- Stanford Question Answering Dataset (SQuAD) 2.0
- ImageNet Top-5
- Massive Multitask Language Understanding (MMLU)
- Mathematics Aptitude Test of Heuristics (MATH)
- Graduate-Level Google-Proof Q&A (GPQA) Diamond
- Massive Multi-discipline Multimodal Understanding (MMMU)
- SimpleBench
- Visual Physics Comprehension Test (VPCT)

AI systems are rapidly approaching—and in some cases exceeding—human performance across multiple domains. This graph compares AI capabilities against human baselines and shows the accelerating pace of improvement.

Source: This graphic builds on analysis from Nestor Maslej et al., Artificial Intelligence Index Report 2025 (Stanford, CA: Stanford University, April 2025), https://hai-production.s3.amazonaws.com/files/hai_ai_index_report_2025.pdf.

Figure 2 | AI Training Compute and Capabilities[25]



1. February 6, 2015
Deep learning systems beat humans at image classification tasks.

2. March 15, 2016
AlphaGo beats world Go champion Lee Sedol.

3. November 30, 2020
AlphaFold successfully predicts protein structures, a process that usually requires painstaking and expensive experiments.

4. February 24, 2022
The first AI-designed drug begins human clinical trials.

5. March 7, 2022
Researchers use AI to identify 40,000 potentially lethal molecules in under six hours.

6. November 30, 2022
OpenAI releases ChatGPT.

7. February 14, 2024
OpenAI detects and disrupts state-affiliated hackers using AI to refine malware.

8. September 12, 2024
OpenAI's o1 model surpasses expert performance on PhD-level science questions.

9. April 23, 2025
Anthropic reports that a user with limited technical skill was able to develop malware that typically requires higher expertise using Claude.

April 29, 2025
AI models beat expert virologists at dual-use tasks.

10. May 22, 2025
Anthropic activates AI Safety Level 3 safeguards, increasing security measures.

11. July 14, 2025
The Department of Defense awards leading AI companies $200 million contracts to advance national security.

July 17, 2025
OpenAI assesses its ChatGPT Agent system to have high-risk biological and chemical capabilities.

July 21, 2025
Gemini Deep Think receives a gold medal at the International

Since the advent of deep learning in 2016, AI capabilities have shown a consistent trend: As the amount of compute used to train models grows, so too do their capabilities and risks.

Source: "Notable AI Models," in Data on AI Models, Epoch AI, accessed October 29, 2025, https://epoch.ai/data/ai-models.

Critical Risks Demand Government Action

There are three key reasons why stronger federal action is essential as national security capabilities emerge from advanced AI models. First, private sector incentives alone are insufficient to mitigate national security risks. Second, government expertise is needed to understand and contextualize the relevance of emerging capabilities within the national security landscape. Third, the government needs to understand and have visibility of emerging capabilities to ensure it can deploy them rapidly, responsibly, and effectively in support of national defense. This section will first present growing evidence of AI's expanding role in national security, before outlining the imperative for more active government involvement.

While private sector innovation fuels AI progress, risks are emerging that sit squarely within the government's purview. "The improper use of AI systems," said the Office of Science and Technology Policy (OSTP) Director Michael Kratsios at a September 2025 United Nations meeting, "can erode deterrence, create destabilizing effects, and reinforce systems of political control and social engineering."[26] Industry leaders and national security experts also have warned that advanced AI systems soon could enable serious threats to U.S. national security. During a 2023 Senate Judiciary Committee hearing, Dario Amodei, CEO of Anthropic, testified that AI "represents a grave threat to U.S. national security" by "enabling many more actors to carry out large-scale biological attacks."[27] OpenAI CEO Sam Altman warned in 2015 that AI poses "probably the greatest threat to the continued existence of humanity," acknowledging in 2025 congressional testimony that "it feels like a sort of new era of human history" requiring "humility and some caution."[28] Palisade Research has developed proofs of concept demonstrating how today's AI systems already can be used for increasingly autonomous hacking.[29] AI today is the worst it will ever be. The next generation of models will be even more capable.[30]

AI's dual-use risks are no longer theoretical. A 2025 Anthropic report revealed that AI models are now being used to carry out sophisticated cyberattacks, including by actors with few technical skills.[31] In October 2025, Microsoft researchers demonstrated that AI protein design tools could redesign dangerous toxins to evade biosecurity screening systems used by DNA synthesis companies.[32] Using openly available AI tools, the team generated variants of controlled highly toxic proteins, such as ricin, that were able to bypass commercial screening software. While Microsoft worked with industry and government to develop patches, some fixes remain incomplete as of November 1, 2025, with one industry executive who coauthored the report warning, "We're in something of an arms race."[33]

Private incentives alone are not well aligned with national security priorities, and a purely market-driven approach will soon prove inadequate. The five leading U.S. AI labs—Anthropic, Google DeepMind, Meta, OpenAI, and xAI—have published frameworks outlining their approaches to potential risks in their models.[34] But these are nonbinding voluntary commitments. The costs of adhering to those commitments also are increasing, as AI capabilities that previously were only theoretical are now within reach. OpenAI's and Anthropic's preparedness frameworks have spurred these companies to place enhanced biosecurity safeguards on their most recent systems.[35] Future systems may demand more resource-intensive mitigations or deployment delays during an intensely competitive race for AI dominance.

AI today is the worst it will ever be. The next generation of models will be even more capable.

Some labs also have come under criticism for delaying and minimizing their adherence to voluntary safety commitments. For example, at the 2024 Seoul Frontier AI Safety Summit, Google agreed to publicly report system capabilities and disclose how external groups, including governments, were involved in assessing model risks.[36] However, when Google DeepMind released Gemini 2.5 Pro just over a year later, it delayed publishing such documentation—releasing a brief model card three weeks after launch and not providing more comprehensive safety evaluation until over a month after the model was released.[37] This prompted strong criticism from lawmakers and

experts, including a cross-party group of 60 UK parliamentarians.[38] Competitive pressures between labs, and as part of Sino-American technology competition, likely will incentivize labs to reduce transparency and robust safeguards, despite growing risks.

Government expertise is necessary to properly evaluate dangers. While AI labs possess technical talent, the U.S. government holds unique information and expertise in managing national security risks in domains such as cybersecurity, nuclear security, and biosecurity. Since early 2024, for example, Anthropic has partnered with the Department of Energy's National Nuclear Security Administration to help assess their models for nuclear national security risks.[39] Such analysis cannot take place without strong government expertise. Private sector companies developing AI systems are not the right actors to fully understand and account for national security externalities—the potential for AI capabilities to affect critical infrastructure, diplomatic relations, or economic security in ways that extend far beyond individual company interests.

AI also has potentially revolutionary applications to military, intelligence, and defense—areas over which the federal government has jurisdiction. Advanced AI systems could transform intelligence analysis, improve battlefield decision-making, and enable autonomous defense systems. If realized, these capabilities could fundamentally shift the military landscape and determine strategic superiority for decades. China's civil-military fusion strategy ensures that AI breakthroughs flow directly into defensive applications. Without efforts to rapidly evaluate and deploy robust AI innovations for defense purposes, America risks ceding technological superiority.[40]

Ultimately, AI risks represent a public problem in private hands: The government has the mandate to protect national security, but the private sector has the technical AI expertise. The deepest technical knowledge about how models operate, what they are capable of, and what their vulnerabilities are, resides primarily within the companies developing these systems. This is a critical national security gap: The United States cannot afford to have the government cut out of the most transformative technology of the era. Despite their deep technical expertise, private companies lack the incentives and legal authority necessary to appropriately monitor and manage national security risks and opportunities.

Poorly Targeted Regulation Could Undermine the U.S. AI Lead

The Trump administration has called out burdensome and poorly targeted regulation as a threat to U.S. AI leadership. The risks of regulatory overreach are particularly acute given the global nature of AI competition. While the United States maintains its current leadership position, competitors like China are making substantial investments in AI development.[41] The Chinese Communist Party's (CCP's) lack of democratic checks and balances enables it to exercise greater central direction and cut through domestic regulations that block progress.

AI risks represent a public problem in private hands: The government has the mandate to protect national security, but the private sector has the technical AI expertise.

If the United States enacts unilateral and heavy-handed regulation that responds only to speculative risks, it could constrain American companies while leaving foreign competitors unencumbered. At his September 2025 address to the UN Security Council, OSTP Director Michael Kratsios said that "broad overregulation incentivizes centralization, stifles innovation, and increases the danger that these tools will be used for tyranny and conquest."[42]

Federal Inaction Leaves a Vacuum for States to Fill

At the same time, an absence of trust that the U.S. federal government is managing AI risks can undermine U.S. competitiveness. In the vacuum of federal regulations, states have moved ahead, with commentators and industry representatives counting up to 700 AI-related proposed bills in 2024 alone.[43] While critics argue this overstates the magnitude by taking an expansive definition of "AI related," the point

CASE STUDY

The European Union: When Regulation Hampers Innovation

The European Union (EU) offers a cautionary tale of how well-intended regulation can undermine competitiveness. Over recent decades, Brussels has erected an increasingly complex regulatory environment that has hampered investment and innovation in the technology sector. Key regulatory frameworks shaping this environment include the General Data Protection Regulation (GDPR), the Digital Services Act, the Digital Markets Act, and now, the EU AI Act.[44] Although Brussels justified each measure as essential for protecting citizens and mitigating risks, their cumulative effect has handicapped European AI competitiveness.

While regulatory frameworks alone do not determine technological competitiveness, government oversight can amplify existing disadvantages and dampen innovation. The *2024 State of European Tech* report found that, despite Europe's ambitious founders and successful companies, regulation remains a key barrier preventing European tech from reaching its full potential.[45]

The AI sector illustrates these impacts clearly. As recently as 2021, European organizations contributed to just under a fifth of notable AI models identified by Epoch AI.[46] By 2024, their share had fallen to just 5.2 percent. Even Mistral, Europe's AI champion, now lags: On a key benchmark, GPQA Diamond, Mistral's 3 model scores 60 percent versus GPT-5's 85 percent.[47]

The International Monetary Fund estimates the AI Act alone could reduce AI-driven productivity gains by 15 percent, rising to 30 percent when combined with other European regulations.[48] Some American companies already have begun limiting new feature releases to EU customers due to liability concerns.[49]

These burdens hit smaller firms hardest. A 2024 report commissioned by the European Commission warned that "innovative companies that want to scale up in Europe are hindered at every stage by inconsistent and restrictive regulations."[50] This assessment is backed by industry experience: A 2024 survey of EU tech stakeholders found that more than half said the GDPR and the AI Act have harmed conditions for founding and scaling companies.[51] In 2024, the world gained 203 new companies valued at over $1 billion, known as "unicorns," but only three came from the EU.[52] Between 2008 and 2021, nearly a third of European unicorns relocated overseas.[53] In early 2025, Dutch AI startup Bird announced its departure from Europe, citing that the continent "lacks the environment we need to innovate in an AI-first era."[54] Europe's talent hemorrhage compounds these challenges. Despite having a higher concentration of AI researchers per capita than either the United States or China, the continent continues bleeding talent to higher-paying U.S. opportunities.[55]

Ironically, this environment undermines the EU AI Act's central goal. Europe's shrinking pool of competitive AI firms and persistent talent drain mean it lacks the capacity required to understand, test, and address emerging AI risks. Rather than expanding resilience, these rules leave Europe more dependent on foreign suppliers and less able to develop its own safeguards—an outcome that amplifies, rather than reduces, national security risks.

Some European officials now have publicly acknowledged the negative effect of onerous AI regulations.[56] In late 2024, French President Emmanuel Macron said, "We are overregulating and underinvesting. In the two to three years to come, if we follow our classical agenda, we will be out of the [AI] market."[57] That same month, European Central Bank President Christine Lagarde warned that the United States "is developing artificial intelligence very rapidly and is already starting to see a number of major champions. Meanwhile, Europe not only has no major champions, but it is a pioneer in the regulation of artificial intelligence."[58] In June, Swedish Prime Minister Ulf Kristersson criticized the EU AI Act for being "confusing" and called for the regulations to be paused.[59] Nevertheless, as of November 1, 2025, the European Commission remains committed to the EU AI Act.[60]

stands that states increasingly may seek to fill the regulatory vacuum.[61]

California has been one of the most active states in regulating AI, enacting Senate Bill 53 (SB 53), also known as the Transparency in Frontier Artificial Intelligence Act, in late September 2025.[62] The law requires companies developing frontier AI models (defined by the amount of compute used to train the model) to publish safety frameworks, report critical incidents to state emergency services, and protect whistleblowers.[63] SB 53 has been commended for increasing transparency, with ex–OSTP official Dean Ball calling SB 53 "ultimately a kind of victory for sane, reasonable voices."[64] However, monitoring for national security risks sits within the purview of the federal government and requires national security expertise. SB 53 may contain sensible elements, but it could be the first of many overlapping state policies. As Director Kratsios described in July 2025, "Having a patchwork of regulations across the entire country just doesn't make sense. It is not pro-innovation."[65] For companies, especially startups, navigating a maze of inconsistent state-level rules could impose additional compliance costs that could slow innovation and adoption.[66]

Simply halting state-level regulation via broad federal preemption would bring its own dangers. Without a robust federal framework to replace state action, such a measure would leave critical gaps in national security. It also would prevent states from fulfilling their traditional role as "laboratories of democracy," where diverse governance approaches can be tested and refined.[67] Durable progress requires not just limiting state action but establishing clear, enforceable, and forward-looking federal leadership.

Responsible Governance Can Benefit Innovation

Targeted government involvement in promoting safety can strengthen innovation by building the public trust essential for widespread adoption. To unlock the full economic and national security potential of AI, people, businesses, and organizations must feel comfortable embracing it. Yet the U.S. public remains distrustful of AI: A 2024 Pew survey revealed that twice as many Americans anticipate a negative impact from AI on the United States over the next two decades compared to those who expect a positive impact.[68] Without trust, even the most advanced technologies cannot achieve their promise.

This trust deficit could undermine effective AI diffusion and adoption in the United States and abroad. Lasting competitive advantage from a technology depends on broad adoption across the economy and society.[69] The government needs to play an active role in understanding the potential of AI and removing barriers to beneficial use cases. At the same time, the potential for large-scale incidents should be taken seriously, lest a high-profile incident trigger public backlash and overly restrictive regulation that stifles progress—paralleling the trajectory of nuclear energy.

After the accidents at Three Mile Island and Chernobyl, U.S. approval for new reactors halted for more than 30 years, and support for new nuclear plants went from a majority to a minority position.[70] America's nuclear energy expansion ended, despite the technology's clear benefits, and only now are we just beginning to pay down that debt with nuclear buildout.[71]

The costs of that lost opportunity are staggering. Extrapolating from historical energy data, a fully nuclear-powered world in 2024 would have seen roughly 5,300 deaths from accidents and pollution. This same amount of power, if generated purely by coal, would have caused 4.4 million deaths.[72] In reality, the Three Mile Island incident resulted in zero deaths and minimal property damage.[73] The reactor's containment systems functioned as designed, preventing any significant release of radiation. Chernobyl, which occurred seven years later, was a fundamentally different disaster resulting from a flawed Soviet-era design that never could have been built under U.S. safety standards.[74]

To unlock the full economic and national security potential of AI, people, businesses, and organizations must feel comfortable embracing it.

The public reaction to both events was not proportional to the actual harm, but nevertheless, public fear curtailed nuclear energy's vast potential. AI faces a similar risk. Without credible safeguards and responsible governance, one major incident could derail transformative progress for decades. Proactive regulation that ensures safety while fostering trust is therefore not a brake on innovation—it is the foundation for sustainable American innovation in AI.[75] Ensuring regulations remain responsive to emerging insights and robust technical analysis will be key to finding the balance.

Exporting U.S. Frameworks and Technologies

In the absence of federal U.S. rules, global frameworks to manage AI risks are being set by other nations. In July 2025, major U.S. AI companies signed the European Union's General-Purpose AI Code of Practice, which requires comprehensive model documentation, safety evaluations, and formal governance structures.[76] While engagement with these international commitments remains uneven—Meta, for example, has refused to sign, and xAI only signed one chapter—the existence of international frameworks will begin to shape global norms and standards.[77] Absent U.S. federal frameworks, there is a risk that

CASE STUDY

The Aviation Industry: When Regulation Drove Trust

The aviation industry provides an example of how government regulation can foster innovation by strengthening public trust. In the early 20th century, commercial aviation leaders sought federal regulation, recognizing that official safety standards would make flying more credible with the public.[78] As Herbert Hoover observed in 1921, "This is the only industry that favors having itself regulated by government."[79]

Before the Federal Aviation Administration (FAA) was established in 1958, aviation oversight was fragmented and ineffective. Responsibilities were split between multiple agencies and the military, leading to bureaucratic conflict, overlapping priorities, and weak enforcement.[80] This fragmented approach undermined both safety and public confidence in commercial aviation. The creation of a unified federal framework through the FAA transformed the industry. Systematic safety oversight reassured the public, dramatically reduced risks, and enabled aviation's extraordinary expansion.[81]

This analogy does not perfectly apply to AI governance. Aircraft are highly specialized technologies with clear and observable catastrophic failure modes. AI, by contrast, is a general-purpose technology with more speculative and varied risk profiles. These different sectors have different challenges.

Nevertheless, this example underscores a critical lesson applicable to AI: Government regulation, when designed to provide clarity and consistency, can be an engine of progress rather than a brake. In aviation, clear rules and credible enforcement built the trust necessary for growth and innovation. Political and individual tolerance for incidents that are highly visible, cause massive damage, and are out of individuals' control is remarkably low. After the January 2025 midair collision in Washington, D.C., the deadliest aviation disaster in America since 2001, public confidence in air safety dropped.[82] But clear rules and credible enforcement—such as those the FAA provides—guards against knee-jerk overregulation even after disasters. Even critics of regulatory overreach acknowledge the FAA's "legitimate role of assuring minimum standards of passenger safety."[83] The Airline Deregulation Act of 1978 abolished the Civil Aeronautics Board, which had tightly controlled airline routes, fares, and market entry, stifling competition and keeping prices artificially high.[84] This dramatically lowered costs and expanded access to air travel.[85] Yet throughout this transformation, the FAA's safety mandate remained and continued to function, demonstrating that technical safety standards can coexist with, and even enable, dynamic market competition. In this context, the FAA has been explored as an inspiration for common-sense AI governance.[86]

other nations will shape the de facto global standards for AI governance, leaving the United States to follow, rather than lead, in responsible AI development.

China has moved aggressively to exploit the void left by American regulatory hesitancy. In July 2025, Beijing announced the World AI Cooperation Organization, headquartered in Shanghai, to set global AI standards and facilitate multilateral cooperation with the Global South.[87] Other initiatives, such as the AI Plus International Cooperation Initiative and AI Capacity Building Action Plan for Good and for All, are further steps China has taken in recent years to set AI rules and standards across the world.[88] Following the UN General Assembly in September 2025, Chinese Deputy Minister of Foreign Affairs Ma Zhaoxu said that "China is firmly committed to being a source of international public goods," supporting the UN's role in AI governance and "ready to work with all sides to strengthen alignment and coordination on development strategies, governance rules and technical standards."[89] Without credible U.S. governance frameworks that inspire confidence through transparency and demonstrable safety, the United States risks ceding global AI influence to the People's Republic of China.

In the absence of federal U.S. rules, global frameworks to manage AI risks are being set by other nations.

Clear U.S. governance frameworks and standards can help strengthen the global uptake of American AI technologies. Promoting the U.S. AI stack has been a clear focus of the Trump administration, featured prominently in both the AI Action Plan and a dedicated executive order.[90] At an Asia-Pacific Economic Cooperation meeting in August 2025, Director Kratsios articulated this vision: "We believe that by packaging the American AI stack and making it available to you, we can strengthen our friendships, empower each of

our nations' AI innovation, and secure a peaceful future of shared prosperity."[91] However, the United States' current lead in AI is heavily shaped by its dominance in closed-source models, whose underlying code and weights remain undisclosed and proprietary. When governments and businesses around the world rely on these models for critical infrastructure, sensitive data, or even routine operations, they are placing their trust in the technology itself, the American companies that build and deploy it, and the governance systems that underpin it. If the United States can demonstrate a mature approach to governance and credibly signal that U.S. companies are anticipating, preventing, and managing AI-related risks effectively, this could strengthen trust in American AI offerings.

How can Washington strike the right balance? Too much or poorly targeted regulation risks suffocating the innovative ecosystem that underpins U.S. AI leadership. But a failure to prepare adequately would risk ceding national security leadership or watching progress slow as a result of major AI incidents or a patchwork of state regulations. History underscores the importance of calibration. For example, in early 2023, more than 1,000 technology leaders, including Elon Musk, signed a public letter calling for pausing the training of models larger than GPT-4.[92] Had policymakers and executives halted AI development based on these early speculative concerns, the United States would have foregone transformative breakthroughs.

The trajectory of AI remains highly uncertain but continues at pace. While the precise contours of effective regulation are not yet clear, the opportunity exists now to lay the groundwork for the "smart rules" that President Trump called for at the Winning the AI Race summit.[93] Achieving this will require the federal government to build the institutional agility and technical expertise to understand and manage the AI transition effectively. This report outlines three core capacities that can help the U.S. government govern effectively in the AI age.

Three Critical Government Capacities for the AI Age

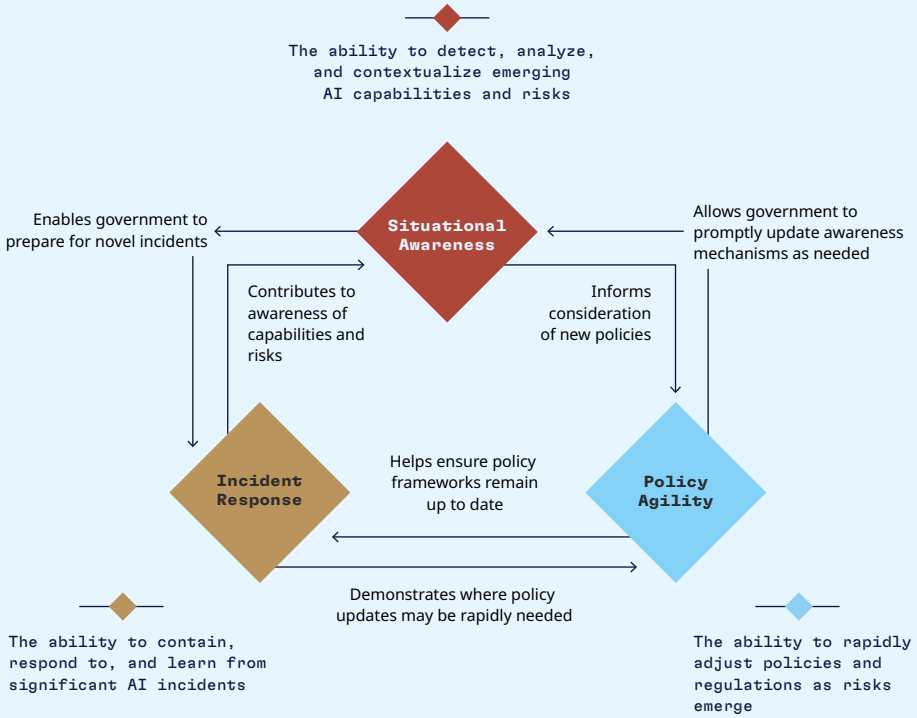
To craft evidence-based, proportionate, and timely AI policy, the U.S. government must strengthen three interconnected capacities to effectively manage AI national security risks and opportunities.

- **Situational awareness** includes the ability to detect, analyze, and understand emerging AI capabilities, risks, and opportunities. Achieving this requires robust information sharing between AI developers and the government, as well as the analytical capacity within government to contextualize this information for its broader national security and societal implications.

- **Policy agility** requires the capacity to adapt quickly to use or adjust policy levers, if and when required. The slow pace of traditional bureaucratic processes threatens to leave policy consistently behind AI development and unable to address novel national security risks. Developing agility to ensure that policy and regulatory approaches match AI's speed and evolving risks will be essential to manage both risks and opportunities effectively.

- **Incident response** encompasses the government's ability to contain and respond to major AI-related incidents. AI-related crises could affect critical infrastructure, financial systems, information environments, and national defense simultaneously in ways that transcend existing agency jurisdictions and response frameworks. A credible response system will require updated coordination mechanisms that unite federal agencies, state authorities, private sector stakeholders, and, where relevant, international partners.

Figure 3 | Three Capacities for Government Preparedness



The three capabilities necessary for a robust incident preparedness regime—situational awareness, policy agility, and incident response—are mutually reinforcing.

Established effectively, these three capabilities will be mutually reinforcing (Figure 3). Situational awareness provides the intelligence foundation that informs both proactive policy adjustments and incident preparedness planning. Policy agility ensures that insights from monitoring AI developments can translate into protective measures before threats materialize, while also providing the legal and operational frameworks necessary for effective incident response. Incident response capabilities not only manage acute crises but generate critical lessons, which enhance situational awareness and reveal gaps in policy frameworks that may spur policy change.

Successfully navigating this challenge—enabling private sector innovation while ensuring the government can fulfill its national security responsibilities—requires a careful reimagining of how the government operates in the AI era. The following sections outline each of the proposed capacities in greater detail and make recommendations to guide their implementation to bolster AI preparedness and responsiveness in the federal government.

KEY CAPACITY: SITUATIONAL AWARENESS

EFFECTIVE GOVERNANCE of AI's national security implications requires situational awareness—having visibility and understanding of emerging AI capabilities and their implications for national security. The government needs to be able to understand what these capabilities are as they emerge to respond swiftly to national security capabilities and risks arising from the most advanced AI models. The pace of AI development means that windows for effective government engagement may be narrower than in previous technology transitions, making proactive preparation essential.

Frontier AI development is dominated by private developers with resources that dwarf government capabilities.[94] The government does not by default have visibility into the frontier of AI development, nor the technical capacity to assess emerging risks.

The government cannot effectively prepare for and protect against threats it cannot see or understand. When private companies develop AI capabilities that could enable autonomous cyber weapons, novel biological weapons, or other national security risks, the government's reactive posture means threats may be realized before officials can assess and respond to their implications comprehensively.

Many of the most important responses to AI-enabled risks will not be in the AI domain itself, and these interventions require time to scale up and implement. Situational awareness can help provide adequate lead time for proactive preparation, such as:

- Increasing critical infrastructure defenses in light of novel or scaled-up AI-enabled cyberattacks
- Scaling up biodefense efforts and safeguards if AI lowers barriers to biological weapons development
- Understanding priorities for AI investment in national security applications to maintain America's competitive advantage
- Identifying gaps in government authorities that may need updating to address groundbreaking new AI capabilities

As of November 1, 2025, the government has relied largely on voluntary information sharing from companies and voluntary early access to advanced models. In 2024, OpenAI and Anthropic signed voluntary memorandums of understanding (MOUs) with the U.S. AI Safety Institute (now the Center for AI Standards and Innovation, or CAISI) that allowed access to their models prior to public release for testing.[95] In their responses to the OSTP's call for comments on the new AI Action Plan, both companies supported voluntary information sharing between private developers and the government.[96]

These voluntary arrangements and agreements have worked so far, but engagement is uneven across the sector and could end.[97] OpenAI and Anthropic have routinely engaged with CAISI and the UK AI Security Institute (UK AISI) by providing early

access to their flagship models.[98] This is not uniform, however: Google's and xAI's most recent model cards (Gemini 2.5 and Grok 4) do not explicitly mention information sharing with the government. While xAI did sign an MOU with the U.S. government under the Biden administration, and Google includes government engagement as a part of its "AI Principles," it is unclear what level of information sharing this has entailed.[99]

Moreover, while AI developers currently may be transparent with the government, evolving market dynamics and regulatory uncertainty may cause this to change. Paradoxically, efforts by developers to surface risks in their models proactively can make their models *appear* riskier than the models of competitors who remain silent about similar issues. For example, Anthropic attracted criticism after disclosing that safety evaluations of its Claude Opus 4 model showed concerning behavior: When given prompts encouraging initiative taking, the model often would attempt to contact external parties, including law enforcement and journalists, to report scenarios involving serious user misconduct.[100] However, subsequent efforts to test other models revealed that Claude was not unique in this behavior. Multiple leading models also sometimes exhibited similar whistleblowing tendencies in comparable scenarios.[101] Separately, a frontier developer reportedly received legal advice against collaborating with the government on chemical, biological, radiological, nuclear, and explosive (CBRNE) assessments due to potential risks around liability and regulation—essentially discouraging the very cooperation that safety requires.[102] These factors collectively create perverse incentives that may significantly increase the likelihood of AI developers undersharing in the future, potentially undermining effective government oversight of AI risks.

RECOMMENDATION 1

Empower and equip the Center for AI Standards and Innovation as the federal government's center of technical AI expertise and evaluation.

Information on emerging AI capabilities is only useful if it can be interpreted and contextualized properly. The federal government therefore needs robust technical capacity to assess emerging risks and opportunities and to act accordingly. As Director Kratsios noted in July 2025, the government is uniquely positioned for this role because "we have experts in these spaces" with deep domain knowledge of CBRNE risks, making federal agencies "very well equipped to be able to supply the subject matter experts to run these evals and create the testing harnesses within places like DOE [the Department of Energy]."[103] While all agencies will need to expand their AI expertise, staying abreast of the most advanced capabilities and the national security implications they may have demands concentrated, specialized knowledge.

The United States is not starting from a blank slate on AI risk management. The National Institute for Standards and Technology (NIST) and the Cybersecurity and Infrastructure Security Agency (CISA) have spun up several efforts in this area, from guidance documents such as the *AI Risk Management Framework* and *AI Cybersecurity Collaboration Playbook* to broader engagement with private industry stakeholders.[104] Sector regulators such as the Federal Aviation Administration and Food and Drug Administration have similarly developed procedures.[105] The Chief Digital and Artificial Intelligence Office in the Department of War (DoW) aims to accelerate adoption of AI.[106] The DOE published an *Artificial Intelligence Strategy* in October 2025.[107]

But across all of these efforts, deep technical expertise in AI is required to be effective. Designating a central hub of technical AI expertise to lead collaboration on AI risk management efforts will help solidify existing efforts and fill gaps in the current approach.

RECOMMENDATION 1.1

Designate CAISI as the federal government's interagency lead for AI risks.

CAISI has emerged as the natural hub for this expertise. Founded in November 2023 as the AI Safety Institute, it was restructured into CAISI in June 2025 to align with the Trump administration's pro-innovation AI strategy.[108] The AI Action Plan reinforces CAISI's role in technical AI analysis by designating it as the lead agency for initiatives requiring deep expertise and sustained engagement with frontier

AI companies.[109] In particular, the AI Action Plan gives CAISI the mandate of working with relevant agencies to evaluate frontier AI systems for national security risks, including CBRNE capabilities, as well as novel security risks.[110] These are areas that can be thoroughly assessed only with government expertise and classified data. As the lead on interagency AI risk management, CAISI would complement, rather than replace, other agencies' domain specific expertise and work collaboratively across government.

This approach yields multiple strategic benefits. Evaluating U.S. AI models, particularly predeployment, offers a preview of the future AI security landscape beyond just the capabilities of individual systems. While predicting exactly when AI models will achieve specific capabilities remains challenging, history shows a consistent pattern: Once frontier models demonstrate a capability, that same capability reliably appears in foreign, cheaper, and/or open-weight models within a short period of time.[111] Understanding the current frontier provides insights into the capabilities that will likely be available to a broader range of actors in the near future, allowing the government to take proactive actions for societal resilience when appropriate.[112]

With all information sharing from the private sector to the government currently being voluntary, CAISI's success depends on maintaining positive and collaborative relationships with frontier AI developers—something the organization already is working to foster. In September 2025, both OpenAI and Anthropic released blog posts with detailed accounts of their collaborations with CAISI and UK AISI.[113] These accounts reveal that this close engagement has gone further than red teaming and evaluating models. For example, both companies reported that the collaboration helped them identify vulnerabilities in their jailbreak defense systems, and Anthropic reported that CAISI and UK AISI helped strengthen its broader security, risk monitoring, and response protocols.[114]

CAISI's collaboration with international AI institutes also has been formalized through bilateral agreements. The September 2025 Tech Prosperity Deal between the United States and the United Kingdom committed both nations to "advancing pro-innovation AI policy frameworks" while "advancing the partnership between the U.S. Center for AI Standards and Innovation and the UK AI Security Institute towards a shared mission to promote secure AI innovation, including through working towards best practices in metrology and standards development for AI models."[115]

Beyond conducting evaluations and collaborating with AI companies, the AI Action Plan gives CAISI a clear role in assessing Chinese models for censorship and CCP ideology, analyzing potential risks from using models from adversarial countries, and developing technical standards for highly secure AI datacenters.[116] CAISI already has taken action in this space. In late September 2025, CAISI published an evaluation comparing leading U.S. models with those of DeepSeek—a leading Chinese AI lab.[117] Using a mix of public and self-developed benchmarks, CAISI found that DeepSeek's models were more expensive and susceptible to certain adversarial attacks, and that they advanced CCP narratives.

CAISI's collaboration with DARPA and the National Science Foundation (NSF) on a development program "to advance AI interpretability, AI control

Office of Science and Technology Policy Director Michael Kratsios and Chair of the President's Council of Advisors on Science and Technology David Sacks join President Donald Trump at the Winning the AI Race event commemorating the release of the AI Action Plan. (Chip Somodevilla/Getty Images)

Ensuring that models remain controllable and robust against manipulation will be vital if the United States is to safely deploy AI in critical systems.

systems, and adversarial robustness" will further strengthen its technical expertise.[118] Even as AI systems improve, they fundamentally remain black boxes; even leading AI researchers cannot fully explain why a model produces a certain output.[119] This literal blind spot is one of the largest inhibitors to creating provably safe and secure AI models.[120] Similarly, ensuring that models remain controllable and robust against manipulation will be vital if the United States is to safely deploy AI in critical systems. While the private sector and researchers have made some progress in these areas, they remain unsolved technical problems germane to national security interests.

To carry out these responsibilities, CAISI must be equipped with the capacity to rigorously interrogate and evaluate AI models and to deliver trusted technical advice across government. Given the breadth of AI capabilities, CAISI should play a key role in coordinating with a broad range of stakeholders and distributing situational awareness while preserving agencies' abilities to develop specialized AI capabilities within their own missions. Properly resourced, empowered, and working closely with relevant agencies, CAISI can serve as the federal government's focal point for AI situational awareness and play a decisive role in safeguarding U.S. security while advancing innovation.

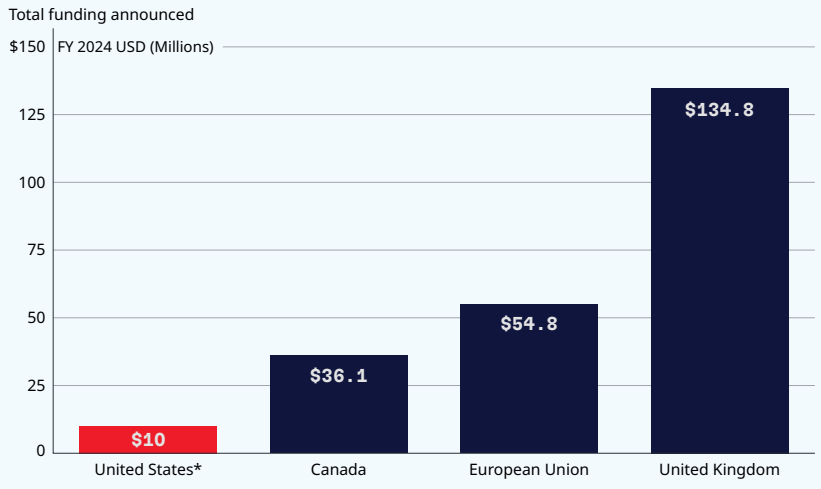
RECOMMENDATION 1.2

Fund CAISI sufficiently to execute its critical role.

The AI Action Plan's ambitious goals for CAISI stand at odds with fiscal realities. NIST historically has faced financial pressures, lacking the personnel and equipment necessary to fulfill its mission.[121] In fiscal year (FY) 2025, NIST received just under $1.1 billion, of which only $10–20 million went to CAISI.[122] Current budget proposals do not increase CAISI's funding for FY 2026. This funding envelope pales in comparison to the AI institutes of international counterparts (Figure 4). After an initial $134 million, the UK AISI will receive $88 million for FY 2026.[123] The EU AI Office was provided with $54.2 million to begin operations and now employs more than 100 experts.[124]

The U.S. government should fund CAISI to the tune of $59.2 million per year (see the appendix). This is critical if the United States wants to shape AI and be able to set "smart rules."[126] AI is inherently dual use, and its national security implications are a matter of *when*, not *if*. Converting the strength of America's private AI sector into strategic advantage requires that the government be able to understand frontier capabilities and

Figure 4 | Committed Funding for AI Safety Institutes in 2024[125]



The United States has committed significantly less funding to its AI institute compared to international peers. While funding structures differ—the European Union and Canada provided one-off investments while the United Kingdom offers ongoing support—the disparity remains stark.

*FY 2024 funding for the Center for AI Standards and Innovation (CAISI) is unclear. Appropriations documents indicate $10 million, but the actual committed level may be closer to $6 million.

ensure they serve the interests of the United States, not those of its adversaries. Providing CAISI with $59.2 million in funding still would equate to less than half a percent of total private investment in AI in 2024.[127]

This level of resourcing is also a critical precondition for attracting high quality talent. CAISI cannot execute the AI Action Plan's goals without access to leading AI experts. While the AI Action Plan orders agencies to prioritize the recruitment of leading AI researchers, they struggle to attract the talent necessary to upholding this directive.[128] The challenge of competing with industry for talent is particularly acute for the evaluation of frontier AI models: The same expertise necessary for designing and conducting evaluations is needed for the more lucrative work of training and refining state-of-the-art models.[129]

This struggle stems in part from the government's inability to match private sector salaries. In NIST's FY 2025 congressional budget submission, the highest-paying role under the AI evaluations budget umbrella, which included the former iteration of CAISI, was $164k per year.[130] As of November 1, 2025, a role on OpenAI's evaluation team pays $200–$370k, plus equity.[131]

The AI Action Plan seeks to address this challenge through collaborative approaches that tap into external expertise. For example, the plan recommends convening the DoW, DOE, CAISI, Department of Homeland Security (DHS), and NSF with academic partners to "solicit the best and brightest from U.S. academia" to test and evaluate AI systems.[132] This crowdsourcing model provides valuable insights and competitive incentives, channeling them to the relevant agencies.

Federally funded research and development centers (FFRDCs) offer another promising avenue for accessing specialized expertise. These institutions operate under a unique model designed to address complex, long-term research challenges that neither government staff nor traditional contractors can adequately handle.[133] What makes FFRDCs particularly valuable is their privileged access to government data, facilities, and personnel, including sensitive and proprietary information that would be typically unavailable to contractors. This access enables deeper integration with government operations while maintaining technical independence. As of November 1, 2025, there are 44 such institutions.[134] The Department of Commerce's existing partnership with the MITRE Corporation to operate the National Cybersecurity Center of Excellence demonstrates how this structure can work for emerging technology challenges.[135] Enlisting technical talent through this FFRDC could help unblock talent bottlenecks.

Still, these efforts must complement, rather than replace, internal government expertise. With the classified nature of many national security risks and capabilities, the government will need to monitor and limit what can be done externally. Ultimately, CAISI's funding and resourcing will determine whether the U.S. government will develop the capacity to understand and shape the trajectory of AI or risk being sidelined from the most transformative technology of our time.

RECOMMENDATION 2

Strengthen information flows from frontier AI developers to government.

The U.S. government should enhance transparency by improving channels for information sharing between frontier AI companies and the government. This requires bolstering whistleblower protections to ensure employees can safely raise concerns about risks, misconduct, or unreported vulnerabilities. Ideally, this will help ensure strong adherence to voluntary reporting. But if evidence emerges that voluntary measures are proving insufficient, the government should be prepared to establish mandatory reporting requirements to guarantee timely access to critical information on emerging AI national security capabilities and risks.

RECOMMENDATION 2.1

Congress should pass a bill enacting greater protections for AI whistleblowers.

When an AI system develops novel dual-use capabilities or when safety safeguards break down, those working most closely with the technology are often the first to recognize the risks. Yet these individuals—who represent vital sources of intelligence—face significant legal and professional risks when raising AI-specific concerns. For a government that seeks to

If evidence emerges that voluntary measures are proving insufficient, the government should be prepared to establish mandatory reporting requirements to guarantee timely access to critical information on emerging AI national security capabilities and risks.

rely on voluntary reporting to maintain situational awareness in a rapidly evolving technological landscape, robust whistleblower protections are essential to national security preparedness.

Existing safeguards for blowing the whistle on financial fraud and corporate misconduct broadly apply to AI companies, but such general protections focus on illegal actions and are insufficient to address AI's national security risks.[136] In other critical industries, such as finance and pharmaceuticals, specific whistleblower protections are assured legislatively, and some states have passed AI-specific protections.[137]

The threat of legal retaliation against whistleblowers could prove a significant barrier to unearthing evidence regarding national security threats. High-profile firings at Google, Meta, and OpenAI have demonstrated tech companies' willingness to retaliate against those who raise concerns about internal practices.[138] Some companies have used contractual mechanisms to suppress potential whistleblowing or unsanctioned communication with the government.[139] In 2024, a group of current and former OpenAI and Google employees signed a public letter calling for more whistleblower protections.[140] In 2024, anonymous OpenAI employees sent a letter to the Securities and Exchange Commission alleging that the company's hiring contracts and exit paperwork "forced employees to waive their rights to whistleblower incentives and compensation, and required employees to notify the company of communication with government regulators."[141] While OpenAI released former employees from these agreements after public backlash, such incidents illustrate how the lack of statutory protections lead to whistleblowers being disincentivized from reporting potential dangers.[142]

Congress has begun to respond to this critical gap. In May 2025, Senator Chuck Grassley and Representative Jay Obernolte introduced the bipartisan AI Whistleblower Protection Act (AI WPA).[143] The legislation would extend protections to those who report "any failure to appropriately respond to a substantial and specific danger that the development, deployment, or use of artificial intelligence may pose to public safety, public health, or national security."[144] The AI WPA drafts were referred to the relevant committees in May, but as of November 1, 2025, no further action has been taken.

Whistleblower protections have the potential to bolster government's situational awareness of emerging AI risks. By creating legal safeguards for reporting AI-specific risks and prohibiting retaliation, robust protections would increase the likelihood the government receives early warning and awareness of emerging threats. Ensuring patriotic workers at frontier AI companies feel empowered to speak out, if national security is at stake, also will help incentivize adherence to voluntary information-sharing commitments.

RECOMMENDATION 2.2

The OSTP should work with CAISI and other relevant agencies to develop a plan for mandating information sharing and testing for dangerous capabilities, in case voluntary mechanisms prove inadequate.

While voluntary standards can be implemented quickly and with minimal friction, they may fail to deliver the level of transparency needed for national security. The U.S. government must be prepared to act decisively to secure access to critical information if evidence emerges that voluntary measures are breaking down. This plan should assess both existing government powers and authorities to mandate information sharing and assess whether additional authorities may be required.

One of the government's most potent tools for accessing information from private industry is the Defense Production Act (DPA). First enacted during the Korean War, the DPA provides the executive branch with powers to influence domestic industry for national defense interests.[145] Apart from a handful of very specific provisions, the DPA must be periodically

reauthorized by Congress with the last authorization being in the National Defense Authorization Act 2019.[146] This most recent authorization lapsed on September 30, 2025, and as of November 1, 2025, Congress has not reauthorized the Act.[147]

Among its provisions, section 705 of the DPA allows the president to order industrial assessments and subpoena information from industry if necessary for national defense. Between 2018 and 2024, 17 such assessments were conducted.[148] In October 2023, President Joe Biden ordered the invocation of these DPA authorities through section 4.2(a) of his executive order on AI, mandating regular reporting requirements on a wide range of issues relating to the development of dual-use foundation models.[149] Critics objected that this invocation stretched DPA authorities beyond their proper scope for military and national defense purposes into peacetime regulation of private industry and did not stem from an urgent need or lack of congressional attention.[150] The Trump administration's rescission of the Biden administration's AI executive order in January 2025 discontinued this approach.[151]

However, should AI national security capabilities and risks be realized, the targeted use of such authorities may be more clearly justified. As part of its contingency planning for mandatory information sharing, the government should evaluate whether the DPA is the most appropriate authority for this purpose, or whether alternative legal frameworks would be more effective. As part of this, the OSTP could explore how other tools, such as liability protections, could incentivize information sharing.

Beyond information sharing, the government also should prepare mechanisms to require testing and evaluation of AI systems for dangerous capabilities before deployment, from both companies and third parties (including government in areas that require national security expertise). A mandatory testing regime would ensure that developers cannot deploy systems with dual-use capabilities without first demonstrating that adequate safeguards are in place. This approach would formalize current voluntary commitments, while providing the government with direct visibility into the most concerning capability thresholds before they enter widespread use.[152]

These recommendations will complement other existing initiatives that contribute more broadly to situational awareness. One key example is the National AI Research Resource (NAIRR) pilot, which expands access to compute and other resources for researchers across U.S. universities.[153] By enabling a broader pool of researchers to experiment with frontier AI capabilities, NAIRR ultimately increases the scientific understanding of AI systems in a way that could prove useful to situational awareness. This expanded network of evaluators and experimenters can uncover novel applications or overlooked risks that government and private companies otherwise may not prioritize. Another relevant mechanism is the establishment of regulatory sandboxes or AI Centers of Excellence, where innovators will be able to deploy and test new tools in more relaxed regulatory environments in exchange for open sharing of data and results.[154]

These initiatives strengthen situational awareness by ensuring the federal government receives critical information from the private sector and has the capacity to evaluate it. But awareness alone is not enough. Without the ability to act quickly and decisively, even the best intelligence and insights will fall short.

KEY CAPACITY: POLICY AGILITY

EQUIPPED WITH situational awareness about emerging AI capabilities and risks, the government also must be able to act quickly when developments demand immediate policy action. While situational awareness reveals what threats and capabilities are on the horizon, agility determines whether the government can respond in time to manage them. The stakes are too high to rely solely on slow-moving congressional processes when gaps in existing national security protections are identified. Without this capacity, even perfect situational awareness is meaningless. Awareness of emerging threats is necessary, but not sufficient, to deliver the kind of "smart rules" President Trump has called for.[155] To succeed, the government requires institutional mechanisms that allow it to act before risks are realized.

Government agility in the AI era means having clear authorities and streamlined decision-making processes that enable policies and strategies to be rapidly adjusted. Frontier AI capabilities may emerge suddenly and unpredictably, often catching even technical experts off guard. The government must be equipped to pivot quickly, updating its policies and actions to keep pace with technological change. The challenge extends beyond simply having the right tools available; it demands a cultural shift within government. Traditional bureaucratic processes that prioritize extensive deliberation over speed may prove inadequate when AI capabilities can emerge and proliferate within weeks and months rather than years.

Yet agility must not come at the expense of democratic accountability. Success will require equipping America's democratic institutions to operate at the pace that emerging technologies demand, rather than circumventing them. The AI Action Plan provides important foundations for developing skills in this domain, particularly in improving information flow and interagency coordination. Translating these recommendations into operational reality will require sustained effort across the executive branch and Congress. The government must move beyond treating AI policy as a purely technical challenge requiring only expert analysis and recognize it as an institutional challenge requiring new forms of coordination, decision-making, and rapid response.

While situational awareness reveals what threats and capabilities are on the horizon, agility determines whether the government can respond in time to manage them.

RECOMMENDATION 3

Establish an interagency AI National Security working group, co-led by the OSTP and the NSC, to strengthen intragovernment coordination on AI national security risks.

The complexity and pace of modern AI development requires unprecedented coordination across the federal government. No single agency can independently manage the full range of national security risks and opportunities AI presents. While CAISI can serve as the government's hub of technical expertise, effective preparedness requires strong collaboration across intelligence, defense, homeland security, law enforcement, and regulatory agencies.

History underscores the dangers of poor government coordination.[156] National security failures often arise not from a lack of intelligence, but from failures to integrate and disseminate it. This challenge is particularly acute with classified information, where strict "need-to-know" protocols can lead to overcaution and undersharing. Following the September 11, 2001, terror attacks, an independent commission concluded that the government already possessed the intelligence necessary to anticipate the attacks but that information was siloed across agencies, preventing a coherent picture from emerging.[157] The AI Action Plan acknowledges this challenge by directing agencies to "prioritize, collect, and distribute intelligence on foreign frontier AI projects that may have national security implications."[158] It explicitly calls for a greater information sharing, bringing together expertise from the intelligence community, the DOE, CAISI, the NSC, and the OSTP.

The federal government already maintains interagency coordination mechanisms that could serve as models. The Cyber Response Group (CRG), established by presidential memorandum in 2016, brings together senior representatives from across the national security community to coordinate policy development and incident response for significant cyber threats.[159] Critically, the CRG operates on an ongoing basis, not just during crises. It meets regularly to develop policies and strategies, receive updates from federal cybersecurity centers, and resolve coordination issues before they become emergencies. When significant incidents do occur, the CRG can activate enhanced coordination procedures and stand up a Cyber Unified Coordination Group for operational response, such as during the 2020 SolarWinds breach.[160] An AI-focused coordination body could adopt similar structures while addressing the distinct technical and policy challenges posed by advanced AI systems.

The AI Action Plan also directs the federal government to strengthen coordination mechanisms more broadly. For example, it formalizes the Chief Artificial Intelligence Officer Council (CAIOC) as "the primary venue for interagency coordination and collaboration on AI adoption" within the government, which will allow agencies to harmonize adoption rules and standards.[161]

Yet, adoption-focused coordination is not sufficient for national security preparedness. The government also needs a dedicated AI and National Security Group that focuses explicitly on identifying and managing national security risks and opportunities. This group should complement, but remain distinct from, the CAIOC, since the expertise required for national security policymaking differs from that needed for internal adoption and regulatory compliance. Convened and overseen by the OSTP and the NSC—with close collaboration from CAISI and participation from the national security community and other relevant agencies—this group would ensure that the right actors are informed, connected, and engaged in shaping timely responses to AI-driven risks.

The AI Action Plan further outlines pathways to improve information flows between technical assessment bodies and policy agencies, strengthen connectivity with the private sector, and create new links between the intelligence community and the Bureau of Industry and Security for export controls.[162] These measures represent important building blocks for agility, ensuring situational awareness can quickly reach the agencies best positioned to act.

However, the AI Action Plan still falls short. Its reforms are largely limited to improving information sharing and coordination, without addressing what happens when situational awareness reveals that current policy settings are dangerously inadequate. Missing are mechanisms for rapid policy changes when AI models cross critical thresholds or exhibit national security–relevant capabilities.

RECOMMENDATION 4

Prepare contingency planning for AI risk scenarios to allow expedited policy action.

Even with strong coordination, current regulatory processes likely will be too slow to keep pace with AI development. Standard federal rulemaking typically includes 60-day comment periods and extensive interagency review processes that can stretch policy implementation across multiple years.[163] Major legislation moves even more slowly. The CHIPS Act took over a year to pass through Congress. Typical regulatory rulemaking can take two to three years from proposal to final implementation.[164] While executive orders can direct agencies to act more quickly, the policies they commission still require time to develop and implement. In the 180 days between President Trump's January 2025 executive order and the delivery of the July 2025 AI Action Plan it commissioned, the time horizon of software engineering tasks AI models were capable of completing nearly tripled. During that same period, the leading model's performance on SimpleBench—a benchmark for basic reasoning tasks where unspecialized humans can outperform AI—jumped from 41.7 percent to 62.4 percent.[165]

The government cannot afford to remain reactive—crafting policy only after high-consequence capabilities already have emerged and proliferated. Once systems with serious national security implications are deployed, retroactive restrictions become technically, economically, and politically far more difficult. For example, if a model capable of automating catastrophic cyberattacks has its model weights openly released, clawing back such capabilities would be virtually impossible. The costs of delayed planning are evident in existing response frameworks: The *National Cyber Incident Response Plan*, first published in 2016, went eight years without substantive revision despite dramatic changes in the threat landscape.[166] The 2023 *National Cybersecurity Strategy* explicitly called for an update to address this gap, but the delay demonstrates how planning documents can become outdated when not regularly maintained.[167] A draft update was published for comment in late 2024, but no permanent updates have been made as of November 1, 2025.[168]

To enhance preparedness, the AI and National Security Group (as previously proposed) informed by CAISI's evaluations, should develop and update *policy* playbooks proactively—distinct from *incident response* playbooks (see the incident response section). These playbooks should map out the authorities and levers available to the government to respond to various emerging AI risks and capabilities. They also should include detail on escalation pathways to ensure key decision makers are kept informed of emerging risks. Where needed, these playbooks could incorporate privileged and classified information, ensuring holistic planning for high-consequence scenarios. While some of this work already may be occurring across government, centralizing and formalizing the process would ensure a coordinated, whole-of-government approach and enable rapid action when required.

Once systems with serious national security implications are deployed, retroactive restrictions become technically, economically, and politically far more difficult.

RECOMMENDATION 5

Establish regular congressional reports by the AI National Security interagency working group to ensure Congress is aware of emerging risks and policy options.

Executive branch agencies require congressional authorization to secure new powers or advance legislation in support of AI national security. As OSTP Director Michael Kratsios underscored in a September 2025 hearing, "The administration can only promote and protect America's position as the global AI standard setter with the legislative branch's support."[169] For this reason, Congress must remain informed about the AI risk landscape, the statutory limits agencies face, and the authorities they may need to respond effectively.

Office of Science and Technology Policy Director Michael Kratsios testifies during a Senate hearing on the AI Action Plan on September 10, 2025. (Andrew Caballero-Reynolds/AFP via Getty Images)

A proven mechanism for this is congressionally mandated reporting. Congress frequently directs agencies to provide reports to exercise oversight and inform legislative decisions.[170] These can take several forms: notification requirements that alert Congress to specific actions, descriptive reports providing factual information about agency activities and program operations, strategic plans, and studies or evaluations that address forward-looking concerns and emerging issues, often including recommendations for legislative action.[171]

Congressionally mandated reports often are one-time analyses of specific issues or situations. The National Defense Authorization Act for FY 2019, for example, ordered the formation of a National Security Commission on Artificial Intelligence that produced an almost 800-page report on AI advancements.[172] However, it is not uncommon for reports to recur—from as frequently as one a month to every five years.

Mandated reporting would serve two critical purposes. First, it would ensure Congress remains informed about the national security implications of AI and emerging capabilities, and the work that agencies are conducting to monitor and address these risks. Second, more expansive reports could allow agencies to recommend congressional actions necessary to address emerging risks. Such reports could be wholly or partially classified, enabling agencies to justify sensitive requests. Directing reports through a single interagency working group, rather than fragmenting responsibilities across agencies, could help minimize the burden on agencies.

Reports on AI and national security initially could be conducted quarterly, with the option for greater frequency should risks escalate. These reports ideally should contain analysis of the current state of AI progress and its national security implications, an overview of agency actions to manage risks and opportunities, an assessment of residual risk, and any recommendations for further policy action. Congress also could establish "triggers" requiring more in-depth reports if specific issues arise. Releasing unclassified versions of these reports would provide valuable context to external stakeholders and support deeper research on emerging issues.

An informed Congress also would be better positioned to support expedited policy action. Streamlined parliamentary procedures already exist that allow legislation to bypass or shorten traditional processes.[173] Using these fast-track mechanisms is common, especially for uncontroversial measures or in emergencies. In 2025, for example, a resolution designating a "National Whistleblower Appreciation Day" passed via unanimous consent.[174] The CARES Act, introduced in March 2020 in reaction to the COVID-19 pandemic, took slightly over a week from beginning negotiations to final signing by the president using a mix of fast-track procedures.[175]

Together, these measures would equip the government with both the internal mechanisms to act swiftly and the legislative support necessary to translate situational awareness into timely, effective policy.

RECOMMENDATION 6

Work with allies and partners to harmonize policy approaches to identified AI risks.

Policy agility must extend beyond domestic authorities to include the capacity to rapidly adjust diplomatic strategies with allies and competitors to respond to new AI developments or incidents. Because AI development and its associated security risks are inherently international, unilateral responses may be insufficient or even counterproductive.

Examples of Preexisting Authorities to Expedite Policy Actions

The need to expand government powers during periods of crisis is not a problem unique to AI risks. Since the country's founding, the United States has developed a complex array of mechanisms designed to allow the government to bypass traditional processes to address dangers, with the checks and balances critical to U.S. democracy maintained by a mix of constitutional, judicial, and political restrictions.

National Emergencies Act

The National Emergencies Act (NEA) is the foundation of executive emergency powers and serves as the central statutory grounding for unlocking a vast array of governmental authorities.[176] Established in 1976, the NEA itself does not grant new powers to the executive branch. Instead, it acts as a gateway, enabling the president to activate over 130 separate statutes and authorities already embedded in federal law—most of which are dormant absent an emergency declaration.[177] The NEA was not designed for any specific class of emergencies, serving instead as a procedural framework that has been used to respond to a broad range of perceived crises, from natural disasters to foreign policy challenges to economic disruptions.

Defense Production Act

The Defense Production Act (DPA) provides the president with broad power over domestic industries in times of national emergency.[178] While a declared national emergency is not necessary for the president to leverage DPA powers, its provisions often have been used in tandem, such as when the DPA was used to prioritize the production of healthcare equipment during the COVID-19 pandemic.[179]

The AI Action Plan explicitly recognizes the DPA's potential application to infrastructure development and deployment. The plan recommends the administration use Title III of the DPA, among other authorities, to encourage the development and deployment of novel manufacturing technologies.[180] Beyond the AI Action Plan's recommendations, DPA powers could be leveraged to enable agile actions to mitigate or respond to AI national security threats, especially in narrow cases where the defense industrial base is directly involved.[181] As of November 1, 2025, the key authorities in the DPA have expired, and Congress has not yet passed a reauthorization.

International Emergency Economic Powers Act

The International Emergency Economic Powers Act (IEEPA) offers the president targeted capabilities for regulating international transactions. Unlike the DPA, the IEEPA can be used only with the declaration of a national emergency. It specifically requires that emergencies originate "in whole or substantial part outside the United States."[182] The IEEPA has been the primary statute invoked in 65 of the 71 emergencies declared under the NEA since 1976, making it the most frequently used emergency authority.[183] It has been used to sanction foreign entities, to require that the government review certain U.S. international business transactions, to block Russian ships from entering U.S. ports, and, in a proposed Department of Commerce rule, to require cloud computing providers to verify the identity of foreign customers.[184] For rapid policy responses that require preventing adversarial states from accessing U.S. AI capabilities, the IEEPA might play a key role.

Agency-Specific Actions

Some individual agencies also have specific emergency authorities. For example, the Department of Homeland Security (DHS), through the Cybersecurity and Infrastructure Security Agency (CISA), can issue Binding Operational Directives and Emergency Directives that compel federal civilian executive branches to take action to mitigate a "known or reasonably suspected information security threat, vulnerability, or incident that represents a substantial threat to the information security of an agency," bypassing normal administrative processes when warranted.[185] The DHS was provided these powers through the Federal Information Security Modernization Act of 2014; they were delegated to CISA after its founding in 2018.[186]

Binding Operational Directives are used to address broad systemic vulnerabilities and mandate general cybersecurity improvements to manage known risks. Agencies are given a reasonable period of time to comply with specified requirements, ranging from a few days to a month depending on the task.[187] Emergency Directives are used in emergency situations and require compliance within hours or days. These are issued in response to active, substantial, and present cybersecurity threats. In 2024, for example, CISA ordered that all instances of an enterprise VPN be disconnected from federal civilian executive branch systems within two days after the provider disclosed a critical vulnerability on a public database.[188]

The largest limitation of CISA's directives is their explicit scope. Specially designated "national security systems," such as intelligence community or military operational systems, are exempt. Similarly, CISA's orders only apply to federal civilian executive branches and private entities operating systems on behalf of the government, but not to independent agencies. The ordered actions cannot conflict with National Institute of Standards and Technology guidelines and must adopt "the least intrusive means possible" for the "shortest period practicable."[189] Nevertheless, these directives are publicly shared, and other groups outside CISA's remit, such as the private sector, are encouraged to follow their guidance.[190] The National Security Agency, which has the power to compel similar cybersecurity measures for national security systems, works closely with CISA to issue cybersecurity guidance.[191]

Expedited Rulemaking

Federal agencies can enact or rescind rules rapidly. Under the Administrative Procedure Act (APA), agencies are required to publish rules no less than 30 days before the effective date in the Federal Register.[192] However, the APA contains a "good cause" exception, which allows agencies to shorten or waive the comment period—known as a notice of proposed rulemaking—when "impracticable, unnecessary, or contrary to the public interest."[193]

This exception often is used for routine administrative determinations, such as fee schedules or statistical methodology updates, that are of little interest or have no impact on the public. Such rulemaking often is termed as a direct final rule, which is shorthand for such uncontroversial rules. Direct final rules often come with a clause that the regulation will enter into force unless the agency receives a single adverse comment.[194] A related action consists of interim final rules, which utilize the good cause exception but invite public comment after the fact and may modify the rule.[195] However, forms of expedited rulemaking, particularly interim final rules, have come under scrutiny for overuse and being used as a method to fast-track controversial regulations.[196]

Congress also can compel agencies to act with speed. Legislation may require agencies to forgo standard notice and comment procedures, impose strict deadlines for completing rulemakings, mandate that regulations be updated at specified intervals, or require periodic reviews of existing rules. These congressional directives provide another mechanism for ensuring agencies can respond rapidly to emerging challenges or evolving technological landscapes.[197]

Some policy actions will be effective only when they are mirrored by partners and allies. For example, the effectiveness of export controls on advanced AI chips hinges on adherence and buy-in from other key partners in the supply chain, including the Netherlands and Japan.[198] Likewise, national security implications of AI capabilities increasingly will shape the focus of collaborations like Five Eyes and the Australia-UK-U.S. trilateral security partnership (AUKUS). Such partnerships are best supported by proactive and candid engagement. As AI capabilities advance, the United States, through the intelligence community and the departments of State and Commerce, should continue to engage with allies and partners on emerging AI risks and national security opportunities, aligning policy approaches where it makes sense to do so. Collaboration should extend to existing multilateral organizations that the United States has influence over, including the G7 and Organisation for Economic Co-operation and Development. This engagement also would support the U.S. government's situational awareness and incident response capacities.

The Trump administration already has emphasized that international engagement on AI security issues is not mutually exclusive with innovation and progress. The AI Action Plan, for example, says that "AI will unlock nearly limitless potential in biology," as well as "create new pathways for malicious actors to synthesize harmful pathogens and other biomolecules."[199] At his 2025 address to the United Nations, President Trump said that, despite the recent pandemic, "many countries are continuing extremely risky research into bio-weapons and man-made pathogens."[200] He announced his administration will lead efforts in "pioneering an AI verification system that everyone can trust," which he hopes will "be one of the early projects under AI."[201]

Developing such verification systems, however, is technically challenging and has been a barrier limiting verification and compliance measures under global treaties such as the Biological Weapons Convention. A robust and global verification system will require significant research and development. This AI innovation can help manage emerging risks.[202]

This kind of careful innovation further provides direct national security benefits, according to Director Kratsios, with "AI technology [having] revolutionary applications for war and for peace."[203] But its abuse nevertheless can "erode deterrence, create destabilizing effects, and reinforce systems of political control and social engineering," unless its applications are "consistent with the highest standards of privacy, civil liberties, transparency, and protections found in the laws of the United States."[204]

President Donald Trump and other world leaders attend the annual G7 Leader's Summit in Kananaskis, Alberta, Canada, on June 16, 2025. (Chip Somodevilla/Getty Images)

KEY CAPACITY: INCIDENT RESPONSE

A CRITICAL COMPONENT of an AI preparedness regime is incident response—the capacity for government to effectively contain, respond to, and learn from, significant AI incidents that materialize. The trajectory of AI progress and its capabilities is highly uncertain. Even with strong visibility into AI developments and mechanisms for rapid policy action, unforeseen events still may arise that could have major consequences for U.S. safety and security.

As AI becomes embedded in critical infrastructure and systems across society, the potential for severe incidents grows, ranging from autonomous cyberattacks on essential services and critical infrastructure to unreliable AI systems that compromise public safety. In such crises, the government must be ready to step in swiftly to uphold national security. Poorly managed large-scale incidents not only could cause direct harm, but also erode public trust in AI, jeopardizing U.S. technological progress and leadership.[205] The Trump administration has acknowledged the importance of an AI incident response regime in the AI Action Plan: "If [AI] systems fail," the government needs to be prepared to ensure that "the impacts to critical services or infrastructure are minimized and response is imminent."[206]

RECOMMENDATION 7

Build stronger interconnectivity between the range of agencies and stakeholders that would need to coordinate a response to an incident.

Effective incident response requires robust coordination mechanisms that connect the full range of stakeholders who need to collaborate during an AI-related crisis. Such an incident could impact critical infrastructure, financial systems, and national defense simultaneously in ways that necessitate intragovernmental response. Building the connective tissue between government agencies, AI companies, infrastructure operators, and other key actors before an incident occurs will be essential for rapid and effective response when crises emerge. While Recommendation 4 focuses on preparing policy playbooks that outline authorities and levers for the government to adjust policies to account for emerging risks, these recommendations focus on strengthening the relationships, communication channels, and operational coordination capabilities needed to execute an effective response across the public and private sectors should an incident occur.

RECOMMENDATION 7.1

Engage AI companies and experts in updating the CISA incident response playbooks.

The AI Action Plan directs agencies, including the DHS, CAISI, and the Office of the National Cyber Director, to update CISA vulnerability and incident response playbooks to "incorporate considerations for AI systems" and include requirements for greater engagement across chief information security officers, chief AI officers, other officials, and CAISI.[207] These playbooks outline how federal agencies can respond

to major cyber incidents—defined as "any incident that is likely to result in demonstrable harm to the national security interests, foreign relations, or the economy of the United States or to the public confidence, civil liberties, or public health and safety of the American people."[208]

The vulnerability response portion of the playbooks addresses urgent, high-priority vulnerabilities that are actively being exploited "in the wild." It focuses on a structured and standardized process for assessing risks and rapidly mitigating threats, offering agencies clear instructions for proactively fixing vulnerabilities. The incident response portion applies to confirmed or suspected malicious cyber activity. It also uses a structured, step-by-step process modeled on NIST guidance to help agencies identify, contain, and remediate security incidents. Together, and once updated with AI-specific guidance, these playbooks could play a key role in supporting agencies to play a role in containing the spread of AI incidents and mitigating their impact. Expanding these resources to the AI context is a smart move to increase preparedness for AI incidents.

In collaboration with the OSTP and CAISI, the updated playbooks should establish a clear, operational definition of what constitutes an "AI incident" for federal response purposes. CISA's Joint Cyber Defense Collaborative (JCDC) *AI Cybersecurity Collaboration Playbook*, published in January 2025, offers a starting point with its working definition of "AI cybersecurity incident," but a more comprehensive framework is needed to capture the full range of AI-related security events that may require coordinated federal response.[209] This definition should distinguish between incidents arising from AI systems being attacked or compromised, incidents where AI capabilities are used as attack vectors, and incidents where AI system failures or misuse create cascading risks to critical infrastructure or national security. Establishing this definitional framework early will ensure consistency across agencies and enable more effective information sharing.

However, effective updates will require input from outside government, particularly while CAISI is still building capacity. Leading AI companies and compute providers will likely be among the first to detect, monitor, and respond to AI-related incidents. In updating and maintaining these playbooks, agencies should draw on both the depth of CAISI's technical expertise as well as the expertise of these companies. Given the near-term nature of AI-enabled cybersecurity risks, CISA's playbooks are a critical framework that can be directly adapted to strengthen readiness and later extended to address other AI-related national security risks.[210]

RECOMMENDATION 7.2

Ensure the AI Information Sharing and Analysis Center also includes representatives from the AI industry.

The AI Action Plan directs the DHS to establish an AI Information Sharing and Analysis Center (AI-ISAC) to "promote the sharing of AI-security threat information and intelligence across U.S. critical infrastructure sectors."[211] Information Sharing and Analysis Centers (ISACs) are member-driven organizations that provide a forum for industries to voluntarily share cybersecurity threat data with each other and with the government to improve collective security.[212] ISACs are typically private sector–led but enabled by government policy, and they often work hand in hand with public agencies.[213] This initiative would be a useful first step in leveraging the proven ISAC model to address the unique challenges posed by AI vulnerabilities.

However, AI-specific companies often are less familiar with the sensitivities of critical infrastructure and have limited experience engaging with government. An AI-ISAC is an opportunity to bring them into this collaborative structure, build essential connective tissue, and strengthen the public-private partnerships needed to respond to large-scale AI incidents.

Given the importance of critical infrastructure to the national security and overall operation of the country, an AI incident affecting critical infrastructure is a good proxy for a larger major incident. AI incidents could occur for many other reasons and impact sectors beyond simply critical infrastructure, but this framework could be a good starting point. Moreover, the legally flexible nature of ISACs means that an AI-ISAC could be expanded to include members beyond strictly critical infrastructure sectors.

The effectiveness of this model depends heavily on the legal frameworks that protect shared information.

Through the Critical Infrastructure Information Act of 2002, and later, the Cybersecurity Information Sharing Act of 2015, any proprietary information shared with the DHS is protected from public disclosure via Freedom of Information Act requests, discovery, or regulatory use, including in civil liability and antitrust cases.[214] Once inside the government, information cannot reach the public record.

On September 30, 2025, the Cybersecurity Information Sharing Act lapsed.[215] As of November 1, 2025, this critical legal framework has not yet been reauthorized, creating legal uncertainty that could undermine AI-ISAC effectiveness from launch. Congress should prioritize reauthorization of this measure to support public-private cooperation on incident response, including for AI incidents.[216]

An AI-ISAC presents an opportunity to enhance private and public understanding of new threats through a tried and true framework, adapted for new challenges—if properly implemented. As AI threats evolve rapidly, the DHS should move swiftly to operationalize an AI-ISAC with clear governance, funding, and operational frameworks.

RECOMMENDATION 7.3

Conduct regular tabletop exercises including government, private sector, and nonprofit representatives to bolster connectivity between incident responders across public and private sectors.

A central tool for building response regimes is the tabletop exercise (TTX), which is a structured simulation of incidents. These exercises can be used to inform the development of preparedness plans, such as the updated CISA playbooks, as well as to socialize and develop muscle memory for agencies and actors to deploy them. AI-specific exercises led by CISA, informed by the technical expertise of CAISI and leading AI companies, would bolster incident response initiatives and uncover critical preparedness gaps to be addressed.

CISA has experience implementing exercises into its incident response regime. The JCDC, for example, is an operational collaboration focused on cyber defense campaigns and threat information sharing in real time.[217] The JCDC brings together experts from across the federal government, major tech and cybersecurity companies, critical infrastructure owners, and even international partners to plan and respond to significant cyber threats collectively.

The JCDC recently has conducted two TTXs to improve CISA's AI preparedness. The first, conducted in June 2024, included participants from across the government, including from the Federal Bureau of Investigation, the National Security Agency, and the Office of the Director of National Intelligence; private industry representatives from companies such as Microsoft, OpenAI, and Palantir; and international observers from Australia, New Zealand, the UK, and Canada.[218] This TTX, as well as a second TTX in September 2024, directly informed the *AI Cybersecurity Collaboration Playbook* CISA released in January 2025 to provide guidance for stakeholders in the AI ecosystem to share information voluntarily with CISA.[219] The JCDC's efforts not only should continue but expand as new AI risks and failure modes emerge.

Beyond traditional TTXs, the government should support competitive cybersecurity exercises such as Capture the Flag, Attack-Defend, and King of the Hill tournaments focused on AI systems and critical infrastructure. These competitions serve multiple purposes: They generate valuable datasets on AI vulnerabilities and attack patterns, create market demand for open-source AI security tools that benefit both government and industry, and build a skilled workforce experienced in defending AI-enabled systems. This approach aligns with the AI Action Plan's recommendation for AI hackathons while offering more sustained engagement and concrete outputs. Such competitions could be integrated into existing frameworks or developed as specialized AI security challenges in partnership with academic institutions and industry. The federal government could conduct more adversarial exercises to stress test AI companies' readiness to respond to national security risks, similar to how regulators stress test banks.[220]

Together, initiatives such as TTXs, ISACs, and updated playbooks lay the foundation for effective information sharing and incident preparedness. Yet these alone are not sufficient. A robust incident response regime also must include mechanisms to capture lessons from near misses and real-world incidents and feed those lessons back into updated

policies and preparedness frameworks. Building this feedback loop will be essential to ensuring U.S. policy remains aligned with the rapidly evolving AI landscape.

RECOMMENDATION 8

Establish a mechanism for post-incident review and lesson learning.

Effective incident response requires not only immediate containment capabilities but also mechanisms for learning from failures and near misses. For AI incidents, which likely will involve novel failure modes and complex attribution challenges, developing robust post-incident learning capabilities becomes even more critical. In many cases, it may be unclear whether a cyber, bio, or other security-related incident is AI-driven at all. Without a systematic process for investigation, there is a risk that early warning shots could be missed until it is too late.

The Cyber Safety Review Board (CSRB), despite its limitations, offers the most relevant precedent for technical incident review. Directed via executive order in 2021 and established the following year, the CSRB brought together up to 20 members from federal agencies and private sector companies such as Microsoft, Google, and CrowdStrike to form an advisory committee within the DHS. Following "significant cyber incidents" or at the behest of executive or agency leaders, the CSRB could convene to investigate the incident, then publicly share findings and recommendations.[221]

In May 2023, for example, a major security breach affected Microsoft's cloud infrastructure, allowing state-backed Chinese hackers to access sensitive U.S. government email accounts and data.[222] Technical details remained opaque even to Microsoft: A blog post about the incident in September 2023 misattributed the cause of the hack.[223] The CSRB initiated an investigation into the incident, releasing their findings just under a year later in April 2024. Alongside technical details, the CSRB report wrote that Microsoft's "security culture was inadequate," and that a "cascade of Microsoft's avoidable errors . . . allowed this intrusion to succeed."[224] By publicly attributing responsibility and demanding improvements, the CSRB performed an accountability function that internal corporate governance had failed to provide. This same capability is essential for AI. A robust review mechanism would ensure that companies cannot hide behind technical opacity, that lapses are not repeated, and that novel AI vulnerabilities are surfaced before they proliferate.

However, the CSRB's record also highlights significant limitations. Since 2021 it has completed only three major investigations, constrained by part-time membership, limited resources, and reliance on voluntary corporate cooperation.[225] These shortcomings were evident during its ongoing Salt Typhoon investigation into Chinese telecommunications intrusions, when companies became reluctant to share information and the board lacked subpoena authority to compel disclosure.[226] By early 2025, the Trump administration dismissed CSRB members, citing inefficiencies, with DHS Deputy Secretary Troy Edgar noting the board "was going in the wrong direction" but "will be reconstituted at the right time, but as an organization that continues with its priorities."[227]

In many cases, it may be unclear whether a cyber, bio, or other security-related incident is AI-driven at all. Without a systematic process for investigation, there is a risk that early warning shots could be missed until it is too late.

There is an opportunity for the Trump administration to build on lessons learned from the first CSRB and establish a more efficient entity to evaluate potential AI incidents: an AI Security Review Board (AISRB). This new body should retain the CSRB's core mission of post-incident investigation but with expanded authority and improved efficiency. Specifically, it should:

- Possess subpoena power to compel disclosure of information from private actors, while offering liability protections to encourage cooperation
- Operate with full-time membership drawn from both the federal government and the private sector, ensuring capacity for multiple simultaneous investigations

- Establish clear, transparent criteria for incident selection and member appointments to reinforce legitimacy and public trust
- Commit to disclosing as much information as possible without compromising proprietary or classified details, using standardized procedures to balance transparency with protection.

An AISRB with these capabilities would ensure that AI-related incidents are investigated rapidly and rigorously, lessons are shared across government and industry, and emerging risks are addressed before they escalate into systemic threats.

RECOMMENDATION 9

Engage with international partners, including adversaries, on best practices for real-time AI incident response.

The global and diffuse nature of AI means that, as in the cyber domain, significant incidents can cross national borders and hit multiple countries at once. To safeguard American security, the U.S. government must develop coordination protocols with both allies and competitors to share and align on best practice incident identification, containment, and response.

This engagement can build on existing initiatives under the AI Action Plan, which directs the United States to work internationally on proactive standards and protective measures—efforts that should explicitly include incident response.

There are relevant models already in place. At the end of 2024, CISA released its first International Strategic Plan for FY 2025–2026, which included goals to bolster its international incident response capabilities and collaborations.[228] One of its commitments is to increase bilateral and international Computer Security Incident Response Team (CSIRT) engagements to strengthen relationships with foreign governments' groups responsible for cybersecurity incident response. Updating this plan to incorporate AI-specific incident response would be a natural and necessary extension, with cyber cooperation serving as a model for other domains.

Should AI national security risks continue to grow, the U.S. government must be prepared to collaborate with geopolitical competitors and foes alike.

Moreover, should AI national security risks continue to grow, the U.S. government must be prepared to collaborate with geopolitical competitors and foes alike. Despite enduring Cold War tensions, for example, the United States and the Soviet Union signed on to the Convention on Assistance in the Case of a Nuclear Accident or Radiological Emergency, developed following the Chernobyl nuclear plant accident in 1986.[229] A similar mindset should guide U.S. efforts on AI preparedness.

One practical step would be to launch a Track 1.5 dialogue with China, bringing together representatives from government, academia, and industry to discuss best practices for AI incident response. As the world's second leading AI power, China is a critical partner in addressing these challenges. Incidents arising in foreign countries could spread rapidly across national boundaries, making information sharing essential. While safeguards to protect U.S. intelligence and strategic interests are critical, such concerns should not preclude cooperation entirely. The COVID-19 pandemic offers a stark reminder: Catastrophes originating in one country can quickly cascade worldwide.

CONCLUSION

AMERICA'S PREPAREDNESS for AI's national security risks and the trajectory of its global competitiveness will be determined by the choices made in the coming years. The Trump administration's AI Action Plan offers a strong foundation, but realizing its promise requires sustained commitment to building a risk preparedness regime that is both robust and pro-innovation.

Private sector innovation will remain the engine of U.S. AI leadership. Yet the federal government must lead on preparing for AI's national security risks and ensuring the right safeguards, coordination, and readiness are in place. The future of AI will be secured through collaboration, preparedness, and smart rules applied where they are most needed.

The federal government of the United States cannot afford to be sidelined from the most transformative technology of our era. To shape the trajectory of AI, rather than react to it, the government must build the situational awareness, policy agility, and incident preparedness necessary to manage both the risks and the opportunities.

If AI fulfills its potential, it will reshape global power and redefine international security. America's enduring leadership will depend on its ability to anticipate change, adapt quickly, and manage the transition effectively. Whether the United States leads or lags in governing AI will define its role in the world for decades to come.

APPENDIX: CAISI FUNDING

The U.S. AI Safety Institute (AISI) was established in November 2023, the day after President Joe Biden signed his landmark AI Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.[230] Up to $10 million was designated for the National Institute of Standards and Technology (NIST) to fund AISI in the fiscal year (FY) 2024 appropriations bill under its Scientific and Technical Research Services account.[231]

In July 2024, the Technology Modernization Fund (TMF) announced an additional $10 million investment in AISI.[232] The TMF is a federal funding program managed by the U.S. General Services Administration that awards competitive grants to federal agencies for projects that modernize and secure government information technology systems.[233] However, TMF grants usually must be repaid in full over a period of five years.[234] As of November 1, 2025, almost 60 percent of the grant has been transferred to NIST and 20 percent has been repaid.[235] What account NIST is paying the TMF back from—including from the original $10 million appropriated—is unclear.

Despite efforts to expand funding for initiatives under President Biden's executive order, Congress failed to pass new appropriations for FY 2025 and instead extended the previous year's funding levels through a full-year continuing resolution (CR).[236] Lacking specific changes in the CR, NIST implicitly received the same amount of funding, including the $10 million for AISI.

In mid-2025, AISI was rebranded as the Center for AI Standards and Innovation (CAISI).[237] As of November 1, 2025, FY 2026 appropriations remain under legislative debate, including the possibility of a short-term or even full-year CR. Funding for CAISI—and NIST more broadly—similarly remains under debate. President Donald Trump's budget request would have cut NIST's budget by almost a third, with the House proposing a more moderate cut. The Senate proposed a small increase in NIST's funding, including $6 million for CAISI.[238] As of November 1, 2025, short-term continuing resolutions previously under debate in Congress do not contain any exceptions for NIST, keeping funding the same as the previous year's.[239]

CAISI's funding in FY 2025 made up 0.00014 percent of the total federal budget. Considering the risks and opportunities AI poses to America's national security, ensuring that CAISI is properly funded is vital. Adequate resourcing is also key to ensuring CAISI can deliver on the initiatives it has been charged with as part of the AI Action Plan. Multiple initiatives, such as hackathons, developing and conducting evaluations, and disseminating findings to government agencies and private sector partners will require additional resourcing.

The following outline would increase CAISI's funding to $59.2 million per year and includes a breakdown of all costs.

Item	Full-Time Equivalent Employees (Total Cost)*
Personnel	**100** ($34,200,000)
Experts to research and evaluate sector-specific AI capabilities	**35** ($11,970,000)
▪ Cyber experts	**10**
▪ Bio experts	**10**
▪ Experts to monitor Chinese Communist Party influence in models	**10**
▪ Experts to monitor international adversarial risks	**5**
General evaluation and AI scientists	20 ($6,840,000)
Information security specialists	**10** ($3,400,000)
Leadership	**5** ($1,710,000)
Liaisons (private sector, intragovernment, international)	**10** ($3,420,000)
Operations and legal	**10** ($3,420,000)
Other	**10** ($3,420,000)
Contracting, grants, events, and travel	**$20,000,000**
Compute	**$5,000,000**
Total	**$59,200,000**

*Employee costs: The above personnel costs assume a salary of $190,000, which is just shy of the maximum 2025 reimbursement for a federal employee on the General Schedule scale. This accounts for the high cost of AI talent relative to other employees. We assume a fringe rate (cost of paid leave, supplemental pay, benefits, insurance, and retirement savings) of 30 percent, based on recent statistics from the Bureau of Labor Statistics.[240] An additional 50 percent overhead/indirect rate is added to include indirect costs, for a total combined 80 percent loading on base salaries. This leads to a total cost of approximately $342,000 per employee.

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

2. JD Vance, "Remarks by the Vice President at the Artificial Intelligence Action Summit in Paris, France," public event, Paris, France, February 11, 2025, https://www.presidency.ucsb.edu/documents/remarks-the-vice-president-the-artificial-intelligence-action-summit-paris-france.

3. Michael Siegel et al., "Rethinking the Cybersecurity Arms Race," MIT Sloan, April 10, 2025, https://web.archive.org/web/20250425004338/https://cams.mit.edu/wp-content/uploads/Safe-CAMS-MIT-Article-Final-4-7-2025-Working-Paper.pdf; Beatrice Nolan, "OpenAI Warns Its Future Models Will Have a Higher Risk of Aiding Bioweapons Development," *Fortune*, June 19, 2025, https://fortune.com/2025/06/19/openai-future-models-higher-risk-aiding-bioweapons-creation/; and Zachary Arnold and Helen Toner, "AI Accidents: An Emerging Threat," Center for Security and Emerging Technology (CSET), July 2021, https://cset.georgetown.edu/publication/ai-accidents-an-emerging-threat/.

4. Michael Kratsios, "Remarks at the Security Council's Open Debate on Artificial Intelligence and International Peace and Security," UN Security Council, September 24, 2025, https://usun.usmission.gov/remarks-at-the-security-councils-open-debate-on-artificial-intelligence-and-international-peace-and-security/; *Oversight of A.I.: Principles for Regulation: Hearing before the Senate Subcommittee on Privacy, Technology, and the Law*, 118th Cong. 7 (2023) (statement of Dario Amodei, CEO, Anthropic), https://www.govinfo.gov/app/details/CHRG-118shrg53503/CHRG-118shrg53503.

5. Miles Brundage (@Miles_Brundage), "I've received some important intel—AI could pose great risks but *also* could have great benefits." X (formerly Twitter), May 8, 2025, https://x.com/Miles_Brundage/status/1920546324451049986.

6. JD Vance, "Remarks by the Vice President at the Artificial Intelligence Action Summit in Paris, France."

7. *Winning the Race: America's AI Action Plan*, 2.

8. Kratsios, "Remarks at the Security Council's Open Debate on Artificial Intelligence and International Peace and Security"; *Oversight of A.I.: Principles for Regulation: Hearing Before the Senate Subcommittee on Privacy, Technology, and the Law*; Sam Altman, "Machine Intelligence, Part 1," personal website, February 15, 2015, https://blog.samaltman.com/machine-intelligence-part-1; and *Winning the AI Race: Strengthening U.S. Capabilities in Computing and Innovation: Hearing Before the Senate Committee on Commerce, Science, and Transportation*, 119th Cong. 74 (2025) (statement of Sam Altman, CEO, OpenAI), https://www.congress.gov/119/chrg/CHRG-119shrg61426/CHRG-119shrg61426.pdf.

9. "Transcript: Donald Trump's Address at 'Winning the AI Race' Event," Tech Policy Press, July 24, 2025, https://www.techpolicy.press/transcript-donald-trumps-address-at-winning-the-ai-race-event/.

10. "First Nuclear Reactors Since 1970s Approved in US," BBC News, February 9, 2012, https://www.bbc.com/news/world-us-canada-16973865; Janet Egan and Cole Salvador, "The United States Must Avoid AI's Chernobyl Moment," Just Security, March 10, 2025, https://www.justsecurity.org/108644/united-states-must-avoid-ais-chernobyl-moment/.

11. Donald J. Trump, "Executive Order on Removing Barriers to American Leadership in Artificial Intelligence," Exec. Order No. 14179, January 23, 2025, https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence.

12. Trump, "Executive Order on Removing Barriers to American Leadership in Artificial Intelligence."

13. "Outline History of Nuclear Energy," World Nuclear Association, July 17, 2025, https://world-nuclear.org/information-library/current-and-future-generation/outline-history-of-nuclear-energy; "ARPANET," Defense Advanced Research Projects Agency, July 2020, https://www.darpa.mil/news/features/arpanet; Nur Ahmed, Muntasir Wahed, and Neil C. Thompson, "The Growing Influence of Industry in AI Research," *Science* 379, no. 6635 (March 2023): 884–86, https://doi.org/10.1126/science.ade2420.

14. Britney Nguyen, "Big Tech's AI Spending Spree Is Going Strong. Here's How Big It Could Be This Year," Quartz, March 5, 2025, https://qz.com/meta-microsoft-alphabet-amazon-spend-billions-ai-capex-1851767670.

15. "Announcing the Stargate Project," OpenAI, January 21, 2025, https://openai.com/index/announcing-the-stargate-project/; "Introducing Stargate UAE," OpenAI, May 22, 2025, https://openai.com/index/introducing-stargate-uae/; "Introducing Stargate Norway," OpenAI, July 31, 2025, https://openai.com/index/introducing-stargate-norway/; and "Introducing Stargate UK," OpenAI, September 16, 2025, https://openai.com/index/introducing-stargate-uk/.

16. "Notable AI Models," in *Data on AI Models*, Epoch AI, accessed October 29, 2025, https://epoch.ai/data/ai-models.

17. Nestor Maslej et al., *Artificial Intelligence Index Report 2025* (Stanford, CA: Stanford University, April 2025), 3, https://hai-production.s3.amazonaws.com/files/hai_ai_index_report_2025.pdf.

18. "GPU Clusters and Supercomputers Map," in *Data on AI Models*; Konstantin Pilz et al., "The US Hosts the Majority of GPU Cluster Performance, Followed by China," Epoch AI, June 5, 2025, https://epoch.ai/data-insights/ai-supercomputers-performance-share-by-country.

19. Thomas Kwa et al., "Measuring AI Ability to Complete Long Tasks," *arXiv*:2503.14499, March 30, 2025, https://doi.org/10.48550/arXiv.2503.14499; "AI Benchmarking" (AI Performance on METR Time Horizons), accessed October 29, 2025, https://epoch.ai/benchmarks.

20. "AI Benchmarking" (AI Performance on GPQA Diamond).

21. Simas Kučinskas et al., "Assessing Near-Term Accuracy in the Existential Risk Persuasion Tournament," Forecasting Research Institute, September 2, 2025, https://forecastingresearch.org/near-term-xpt-accuracy; Cade Metz, "Google A.I. System Wins Gold Medal in International Math Olympiad," *The New York Times*, July 21, 2025, https://www.nytimes.com/2025/07/21/technology/google-ai-international-mathematics-olympiad.html.

22. Metz, "Google A.I. System Wins Gold Medal in International Math Olympiad."

23. Bridget Williams et al., "Forecasting LLM-Enabled Biorisk and the Efficacy of Safeguards," Forecasting Research Institute, July 1, 2025, https://forecastingresearch.org/ai-enabled-biorisk; "Virology Capabilities Test," SecureBio and the Center for AI Safety, April 29, 2025, https://www.virologytest.ai/.

24. Values in Figure 1 are scaled relative to each benchmark's human baseline. A value of 120 percent, for example, indicates that a model performs 20 percent better than humans. Conversely, a value of 85 percent indicates that humans outperform the model by 15 percent. When benchmarks include multiple human baselines (such as MMMU's low, medium, and expert categories), we use the medium baseline for consistency. We selected the following benchmarks for their breadth, variety, and reliable human baselines. ImageNet Top-5: Measures image classification accuracy by checking if the correct label is in a model's top five predictions for each picture Mathematics Aptitude Test of Heuristics (MATH): Evaluates models on solving advanced high school mathematics competition problems Massive Multitask Language Understanding (MMLU): Assesses general knowledge and reasoning across 57 academic and professional subjects via multiple-choice questions Massive Multi-discipline Multimodal Understanding (MMMU): Tests expert-level multimodal (text and image) understanding and reasoning across many disciplines SimpleBench: Probes reasoning skills with multiple-choice questions where typical humans still outperform top AI models Stanford Question Answering Dataset (SQuAD) 2.0: Examines reading comprehension, requiring not only answering questions from paragraphs but also knowing when not to answer if the answer isn't present Visual Physics Comprehension Test (VPCT): Measures basic physical reasoning Graduate-level Google-Proof Q&A (GPQA) Diamond: Challenges models with graduate-level, multiple-choice science questions spanning several domains Data for MMMU, VPCT, SQuAD 2.0, and SimpleBench were taken from their respective leaderboards on October 8, 2025, which include human baseline values. GPQA Diamond results came from Epoch AI's benchmarking dashboard, with human baselines from OpenAI's research. ImageNet Top-5, MATH, and MMLU data were sourced from the *Artificial Intelligence Index Report 2025* and updated using Kaggle leaderboards when available. While using multiple sources may introduce minor discrepancies due to different evaluation methods, the overall trend remains clear: AI models are increasingly surpassing human performance across diverse tasks. "A Massive Multi-Discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI," updated September 5, 2025, https://mmmu-benchmark.github.io; Chase Brower, "Visual Physics Comprehension Test," updated August 7, 2025, https://cbrower.dev/vpct; "SQuAD 2.0: The Stanford Question Answering Dataset," accessed October 8, 2025, https://rajpurkar.github.io/SQuAD-explorer/; SimpleBench Team, "SimpleBench," accessed October 8, 2025, https://simple-bench.com/; "AI Benchmarking" (AI Performance on GPQA Diamond); "Learning to Reason with LLMs," OpenAI, September 12, 2024, https://openai.com/index/learning-to-reason-with-llms/; *Artificial Intelligence Index Report 2025* (Figures 2.1.33 and 2.6.4, relative performance on ImageNet Top-5 and MMLU, and raw performance on MATH; accessed October 8, 2025), https://drive.google.com/drive/folders/1vgn34271NjvhHUvtGDlr0dqnhHNO8tt6; Open Benchmarks, "MMLU," Kaggle, accessed September 29, 2025, https://www.kaggle.com/benchmarks/open-benchmarks/mmlu.

25. Sources for AI national security milestones: Kaiming He et al., "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *arXiv*:1502.01852, February 6, 2025, https://doi.org/10.48550/arXiv.1502.01852; Steven Borowiec, "AlphaGo Seals 4-1 Victory over Go Grandmaster Lee Sedol," *The Guardian*, March 15, 2016, https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol; "AlphaFold: A Solution to a 50-Year-Old Grand Challenge in Biology," Google DeepMind, November 30, 2020, https://deepmind.google/discover/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology/; "From Start to Phase 1 in 30 Months: AI-Discovered and AI-Designed Anti-Fibrotic Drug Enters Phase I Clinical Trial," Insilico Medicine, February 24, 2022, https://insilico.com/phase1; Derek Lowe, "Deliberately Optimizing for Harm," Science, March 15, 2022, https://www.science.org/content/blog-post/deliberately-optimizing-harm; "Introducing ChatGPT," OpenAI, November 30, 2022, https://openai.com/index/chatgpt/; "Disrupting Malicious Uses of AI by State-Affiliated Threat Actors," OpenAI, February 14, 2024, https://openai.com/index/disrupting-malicious-uses-of-ai-by-state-affiliated-threat-actors/; "Learning to Reason with LLMs," OpenAI, September 12, 2024, https://openai.com/index/learning-to-reason-with-llms; "Detecting and Countering Malicious Uses of Claude: March 2025," Anthropic, April 23, 2025, https://www.anthropic.com/news/detecting-and-countering-malicious-uses-of-claude-march-2025; "Virology Capabilities Test"; "Activating AI Safety Level 3 Protections," Anthropic, May 22, 2025, https://www.anthropic.com/news/activating-asl3-protections; "Anthropic and the Department of Defense to Advance Responsible AI in Defense Operations," Anthropic, July 14, 2025, https://www.anthropic.com/news/anthropic-and-the-department-of-defense-to-advance-responsible-ai-in-defense-operations; "Introducing ChatGPT Agent: Bridging Research and Action," OpenAI, July 17, 2025, https://openai.com/index/introducing-chatgpt-agent/; and Thang Luong and Edward Lockhart, "Advanced Version of Gemini with Deep Think Officially Achieves Gold-Medal Standard at the International Mathematical Olympiad," Google DeepMind, July 21, 2025, https://deepmind.google/discover/blog/advanced-version-of-gemini-with-deep-think-officially-achieves-gold-medal-standard-at-the-international-mathematical-olympiad/.

26. Kratsios, "Remarks at the Security Council's Open Debate on Artificial Intelligence and International Peace and Security."

27. *Oversight of A.I.: Principles for Regulation: Hearing before the Senate Subcommittee on Privacy, Technology, and the Law*.

28. Altman, "Machine Intelligence, Part 1"; *Winning the AI Race: Strengthening U.S. Capabilities in Computing and Innovation: Hearing Before the Senate Committee on Commerce, Science, and Transportation*.

29. Artem Petrov Reworr, *Palisade Hacking Cable Technical Report* (Palisade Research, August 26, 2025), https://palisaderesearch.org/assets/reports/hacking-cable-report.pdf.

30. Ethan Mollick (@emollick), "Remember, Today's AI Is the Worst AI You Will Ever Use," X (formerly Twitter), June 20, 2023, https://x.com/emollick/status/1671325114141491203.

31. *Threat Intelligence Report: August 2025* (Anthropic, August 27, 2025), https://www.anthropic.com/news/detecting-countering-misuse-aug-2025.

32. Bruce J. Wittmann et al., "Strengthening Nucleic Acid Biosecurity Screening Against Generative Protein Design Tools," *Science* 390, no. 6768 (October 2, 2025): 82–87, https://doi.org/10.1126/science.adu8578.

33. Antonio Regalado, "Microsoft Says AI Can Create 'Zero Day' Threats in Biology," *MIT Technology Review*, October 2, 2025, https://www.technologyreview.com/2025/10/02/1124767/microsoft-says-ai-can-create-zero-day-threats-in-biology/.

34. Anca Dragan, Helen King, and Allan Dafoe, "Introducing the Frontier Safety Framework," Google DeepMind, May 17, 2024, https://deepmind.google/discover/blog/introducing-the-frontier-safety-framework/; "Our Approach to Frontier AI," Meta, February 3, 2025, https://about.fb.com/news/2025/02/meta-approach-frontier-ai/; *Preparedness Framework (Version 2)* (OpenAI, April 15, 2025), https://cdn.openai.com/pdf/18a02b5d-6b67-4cec-ab64-68cdfbddebcd/preparedness-framework-v2.pdf; *xAI Risk Management Framework (Draft)* (xAI, January 20, 2025), https://x.ai/documents/2025.02.20-RMF-Draft.pdf; and *Responsible Scaling Policy* (Anthropic, October 15, 2024), https://assets.anthropic.com/m/24a47b00f10301cd/original/Anthropic-Responsible-Scaling-Policy-2024-10-15.pdf.

35. "ChatGPT Agent System Card," OpenAI, July 17, 2025, https://openai.com/index/chatgpt-agent-system-card/; "Activating AI Safety Level 3 Protections," Anthropic, May 2025, https://www-cdn.anthropic.com/807c59454757214bfd37592d6e048079cd7a7728.pdf.

36. "Frontier AI Safety Commitments, AI Seoul Summit 2024," UK Department for Science, Innovation & Technology, February 7, 2025, https://www.gov.uk/government/publications/frontier-ai-safety-commitments-ai-seoul-summit-2024/frontier-ai-safety-commitments-ai-seoul-summit-2024.

37. Harry Booth, "60 U.K. Lawmakers Accuse Google of Breaking AI Safety Pledge," *Time*, August 29, 2025, https://time.com/7313320/google-deepmind-gemini-ai-safety-pledge/.

38. "Letter to Sir Demis Hassabis: Parliamentarians From Across the UK Call on Google DeepMind to Honour Their AI Safety Commitments," PauseAI, August 29, 2025, https://pauseai.info/dear-sir-demis-2025.

39. "Developing Nuclear Safeguards for AI Through Public-Private Partnership," Anthropic, August 21, 2025, https://www.anthropic.com/news/developing-nuclear-safeguards-for-ai-through-public-private-partnership.

40. Cole McFaul, Sam Bresnick, and Daniel Chou, *Pulling Back the Curtain on China's Military-Civil Fusion: How the PLA Mobilizes Civilian AI for Strategic Advantage* (Washington: CSET, September 2025), https://cset.georgetown.edu/publication/pulling-back-the-curtain-on-chinas-military-civil-fusion/; Jacob Stokes, Alexander Sullivan, and Noah Greene, *U.S.-China Competition and Military AI: How Washington Can Manage Strategic Risks amid Rivalry with Beijing* (Cener for a New American Security, July 25, 2023), https://www.cnas.org/publications/reports/u-s-china-competition-and-military-ai.

41. Maslej et al., *Artificial Intelligence Index Report 2025*.

42. Kratsios, "Remarks at the Security Council's Open Debate on Artificial Intelligence and International Peace and Security."

43. "2025 State AI Wave Building After 700 Bills in 2024," Business Software Alliance, October 22, 2024, https://www.bsa.org/news-events/news/2025-state-ai-wave-building-after-700-bills-in-2024; Joe Duball, "How Proposed AI Enforcement Moratorium Cuts into US State-Level Powers," International Association of Privacy Professionals, June 27, 2025, https://iapp.org/news/a/how-proposed-ai-enforcement-moratorium-cuts-into-us-state-level-powers/; and U.S. Chamber of Commerce, "Coalition Letter to the Senate Supporting the Moratorium on AI Regulation Enforcement," June 9, 2025, https://www.uschamber.com/technology/coalition-letter-to-the-senate-supporting-the-moratorium-on-ai-regulation-enforcement.

44. General Data Protection Regulation, Regulation (EU) 2016/679 of the European Parliament and of the Council, April 27, 2016, https://gdpr-info.eu/; Digital Markets Act, Regulation (EU) 2022/1925 of the European Parliament and of the Council, September 14, 2022, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2022.265.01.0001.01.ENG; Artificial Intelligence Act, Regulation (EU) 2024/1689 of the European Parliament and of the Council, June 13, 2024, https://artificialintelligenceact.eu/the-act/.

45. *2024 State of European Tech* (Atomico, 2024), 41, https://www.stateofeuropeantech.com/.

46. "Notable AI Models," in *Data on AI Models*.

47. "AI Benchmarking" (AI Performance on GPQA Diamond).

48. Florian Misch et al., "AI and Productivity in Europe," *International Monetary Fund* 2025, no. 067 (April 2025), https://www.elibrary.imf.org/view/journals/001/2025/067/article-A001-en.xml.

49. Oona Lagercrantz, "Europe's AI Blues: US Companies Slow Deployment," Center for European Policy Analysis, November 1, 2024, https://cepa.org/article/europes-ai-blues-us-companies-slow-deployment/.

50. Mario Draghi, *The Future of European Competitiveness: A Competitiveness Strategy for Europe* (Luxembourg: Publications Office of the European Union, 2025), https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en.

51. *2024 State of European Tech*, 96.

52. "Global Unicorn Index 2025," Hurun Research Institute, June 26, 2025, https://www.hurun.net/en-US/Info/Detail?num=2DVQ51ORRGTH.

53. Draghi, *The Future of European Competitiveness: A Competitiveness Strategy for Europe*.

54. "Dutch Software Firm Bird to Leave Europe Due to Onerous Regulations in AI Era, Says CEO," Reuters, February 24, 2025, https://www.reuters.com/technology/dutch-software-firm-bird-leave-europe-due-onerous-regulations-ai-era-says-ceo-2025-02-24/.

55. "A Talented Home for AI," Sequoia Capital, accessed September 8, 2025, https://atlas.sequoiacap.com/a-talented-home-for-ai/; Laurenz Hemmen and Siddhi Pal, "Where Is Europe's AI Workforce Coming From? Immigration, Emigration & Transborder Movement of AI Talent," Interface, July 31, 2024, https://www.interface-eu.org/publications/where-is-europes-ai-workforce-coming-from.

56. Anda Bologa, "The Washington Effect? Europe Weighs Pausing the AI Act," Center for European Policy Analysis, July 8, 2025, https://cepa.org/article/the-washington-effect-europe-weighs-pausing-the-ai-act/.

57. Elena Giordano, "Macron: EU Has Only 2 or 3 Years to Stave off Total US, China Dominance," *Politico*, October 3, 2024, https://www.politico.eu/article/emmanuel-macron-france-europe-competition-united-states-china-climate-change-defense-security/.

58. Eric Albert, Philippe Escande, and Béatrice Madeline, "ECB President Christine Lagarde: 'Europe Is Falling Behind, and France Too,'" *Le Monde*, October 31, 2024, https://www.lemonde.fr/en/economy/article/2024/10/31/ecb-president-christine-lagarde-europe-is-falling-behind-and-france-too_6731107_19.html.

59. Pieter Haeck, "Swedish PM Calls for a Pause of the EU's AI Rules," Politico Pro, June 23, 2025, https://subscriber.politicopro.com/article/2025/06/swedish-pm-calls-for-a-pause-of-the-eus-ai-rules-00418396.

60. Foo Yun Chee, "EU Sticks with Timeline for AI Rules," Reuters, July 4, 2025, https://www.reuters.com/world/europe/artificial-intelligence-rules-go-ahead-no-pause-eu-commission-says-2025-07-04/.

61. Steven Adler, "Mythbusting the Supposed '1,000+ AI State Bills That Would Hobble Innovation,'" Clear-Eyed AI, July 1, 2025, https://stevenadler.substack.com/p/mythbusting-the-supposed-1000-ai.

62. Cecilia Kang, "California Governor Signs Sweeping A.I. Law," *The New York Times*, September 29, 2025, https://www.nytimes.com/2025/09/29/technology/california-ai-safety-law.html.

63. Artificial Intelligence Models: Large Developers, California Senate Bill 53, 2025–2026 Reg. Sess., ch. 138 (September 29, 2025), https://legiscan.com/CA/text/SB53/2025.

64. Chase DiFeliciantonio et al., "Trump's AI Scribe Goes West," Politico Pro, October 8, 2025, https://subscriber.politicopro.com/newsletter/2025/10/trumps-ai-scribe-goes-west-00597420.

65. Michael Kratsios, "Unpacking the White House AI Action Plan with OSTP Director Michael Kratsios" (public event, Center for Strategic and International Studies, Washington, July 30, 2025), https://www.csis.org/analysis/unpacking-white-house-ai-action-plan-ostp-director-michael-kratsios.

66. Kristin O'Donoghue, "A Patchwork of State AI Regulation Is Bad. A Moratorium Is Worse," AI Frontiers, June 25, 2025, https://ai-frontiers.org/articles/congress-might-block-states-from-regulating-ai; Kevin Frazier and Adam Thierer, "1,000 AI Bills: Time for Congress to Get Serious About Preemption," Lawfare, May 9, 2025, https://www.lawfaremedia.org/article/1-000-ai-bills--time-for-congress-to-get-serious-about-preemption.

67. David Rubenstein, "AI Governance Needs Federalism, Not a Federally Imposed Moratorium," Just Security, May 29, 2025, https://www.justsecurity.org/113728/ai-governance-federalism-moratorium/.

68. Colleen McClain et al., "How the US Public and AI Experts View Artificial Intelligence," Pew Research Center, April 3, 2025, https://www.pewresearch.org/internet/2025/04/03/how-the-us-public-and-ai-experts-view-artificial-intelligence/.

69. Jeffrey Ding, *Technology and the Rise of Great Powers: How Diffusion Shapes Economic Competition* (Princeton, NJ: Princeton University Press, 2025).

70. Barbara D. Melber, "The Impact of TMI upon the Public Acceptance of Nuclear Power," *Progress in Nuclear Energy* 10, no. 3 (1982): 387–98, https://www.sciencedirect.com/science/article/abs/pii/0149197082900154.

71. James Lynch, "Chris Wright Makes Unleashing Nuclear Power Priority for American Energy Abundance," *National Review*, February 7, 2025, https://www.nationalreview.com/news/chris-wright-makes-unleashing-nuclear-power-priority-for-american-energy-abundance/.

72. Figures calculated using data from: "Deaths per Terawatt-Hour of Energy Production," Our World in Data, https://ourworldindata.org/grapher/death-rates-from-energy-production-per-twh; Hannah Ritchie, Pablo Rosado, and Max Roser, "Energy," Our World in Data, https://ourworldindata.org/energy.

73. "14-Year Cleanup at Three Mile Island Concludes," *The New York Times*, August 15, 1993, https://www.nytimes.com/1993/08/15/us/14-year-cleanup-at-three-mile-island-concludes.html; "Backgrounder on the Three Mile Island Accident," U.S. Nuclear Regulatory Commission, updated October 2, 2025, https://www.nrc.gov/reading-rm/doc-collections/fact-sheets/3mile-isle.

74. Sarah Kramer, "Here's Why a Chernobyl-Style Nuclear Meltdown Can't Happen in the United States," *Business Insider*, April 26, 2016, https://www.businessinsider.com/chernobyl-meltdown-no-graphite-us-nuclear-reactors-2016-4.

75. Paul Scharre and Vivek Chilukuri, "What an American Approach to AI Regulation Should Look Like," *Time,* March 5, 2024, https://time.com/6848922/ai-regulation/.

76. "The General-Purpose AI Code of Practice," European Commission, https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai.

77. Samantha Subin, "Meta Says It Won't Sign Europe AI Agreement, Calling It an Overreach That Will Stunt Growth," CNBC, July 18, 2025, https://www.cnbc.com/2025/07/18/meta-europe-ai-code.html; "Musk's xAI to Sign Chapter on Safety and Security in EU's AI Code of Practice," Reuters, July 31, 2025, https://www.reuters.com/technology/musks-xai-sign-chapter-safety-security-eus-ai-code-practice-2025-07-31/.

78. Mike Borfitz, "The Public Perception of Safety," Kilroy Aviation, https://www.kilroy.faaoda.com/blog/the-public-perception-of-safety; Mark Hansen et al., *History of Aviation Safety Oversight in the United States* (Washington: Federal Aviation Administration, July 2008), https://www.faa.gov/about/office_org/headquarters_offices/avs/offices/aam/cami/library/online_libraries/aerospace_medicine/media/1105.pdf.

79. Nick Komons, *Bonfires to Beacons: Federal Civil Aviation Policy Under the Air Commerce Act, 1926-1938* (Washington: Smithsonian Publications, 1989), 22.

80. Komons, *Bonfires to Beacons*; John Wilson, *Turbulence Aloft: The Civil Aeronautics Administration Amid Wars and Rumors of Wars, 1938–1953* (Washington: U.S. Department of Transportation, Federal Aviation Administration, 1979); Stuart Rochester, *Takeoff at Mid-Century: Federal Civil Aviation Policy in the Eisenhower Years, 1953–1961* (Washington: U.S. Department of Transportation, Federal Aviation Administration, 1976).

81. Borfitz, "The Public Perception of Safety"; Komons, *Bonfires to Beacons*.

82. Linley Sanders, "Americans' Confidence in Air Travel Safety Dips Slightly After Washington Plane Crash: AP-NORC Poll," AP News, February 19, 2025, https://apnews.com/article/travel-air-safety-poll-faa-plane-crash-7928e7d794f30f5f4921d26f8c67146b.

83. Ronald D. Utt, "FAA Reauthorization: Time to Chart a Course for Privatizing Airports," Heritage Foundation, June 4, 1999, https://www.heritage.org/budget-and-spending/report/faa-reauthorization-time-chart-course-privatizing-airports.

84. Airline Deregulation Act of 1978, Pub. L. No. 95-504, 92 Stat. 1705 (1978), https://www.govinfo.gov/content/pkg/STATUTE-92/pdf/STATUTE-92-Pg1705.pdf.

85. "Airline Deregulation: When Everything Changed," Smithsonian National Air and Space Museum, December 17, 2021, https://airandspace.si.edu/stories/editorial/airline-deregulation-when-everything-changed.

86. Cole Salvador, "Certified Safe: A Schematic for Approval Regulation of Frontier AI," *arXiv*:2408.06210, August 12, 2024, https://doi.org/10.48550/arXiv.2408.06210; Jack Corrigan et al., *Governing AI with Existing Authorities: A Case Study in Commercial Aviation* (CSET, July 2024), https://cset.georgetown.edu/wp-content/uploads/CSET-Governing-AI-with-Existing-Authorities.pdf.

87. "China Proposes Establishment of World AI Cooperation Organization," Voice of CAST, July 28, 2025, https://voc-gj.cast.org.cn/index/info?api=GwArticle&id=41459.

88. Ministry of Foreign Affairs, People's Republic of China, "AI+ International Cooperation Initiative," press release, September 24, 2025, https://www.fmprc.gov.cn/eng/xw/zyjh/202509/t20250924_11715960.html; Ministry of Foreign Affairs, People's Republic of China, "AI Capacity-Building Action Plan for Good and for All," press release, September 27, 2024, https://www.mfa.gov.cn/mfa_eng/wjbzhd/202409/t20240927_11498465.html.

89. Ma Zhaoxu, "High-Level Multi-stakeholder Informal Meeting to Launch The Global Dialogue on Artificial Intelligence Governance" (livestreamed event, United Nations, New York, September 25, 2025), https://www.mfa.gov.cn/eng/xw/wjbxw/202509/t20250930_11720750.html.

90. Donald J. Trump, "Executive Order on Promoting the Export of the American AI Technology Stack," Exec. Order No. 14320, July 23, 2025, https://www.federalregister.gov/documents/2025/07/28/2025-14218/promoting-the-export-of-the-american-ai-technology-stack.

91. Michael Kratsios, "Remarks by Director Kratsios at the APEC Digital and AI Ministerial Meeting" (event, Global Digital and AI Forum, August 5, 2025), https://www.whitehouse.gov/articles/2025/08/remarks-by-director-kratsios-at-the-apec-digital-and-ai-ministerial-meeting/.

92. Cade Metz and Gregory Schmidt, "Elon Musk and Others Call for Pause on A.I., Citing 'Profound Risks to Society,'" *The New York Times*, March 29, 2023, https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html.

93. "Transcript: Donald Trump's Address at 'Winning the AI Race' Event."

94. *Artificial Intelligence Index Report 2025.*

95. "U.S. AI Safety Institute Signs Agreements Regarding AI Safety Research, Testing and Evaluation With Anthropic and OpenAI," National Institute for Standards and Technology, August 29, 2024, https://www.nist.gov/news-events/news/2024/08/us-ai-safety-institute-signs-agreements-regarding-ai-safety-research.

96. Anthropic to Faisal D'Souza, "Re: Request for Information (RFI) on the Development of an Artificial Intelligence (AI) Action Plan ('Plan')," Anthropic, March 6, 2025, https://assets.anthropic.com/m/4e20a4ab6512e217/original/Anthropic-Response-to-OSTP-RFI-March-2025-Final-Submission-v3.pdf; Christopher Lehane to Faisal D'Souza, OpenAI, March 13, 2025, https://cdn.openai.com/global-affairs/ostp-rfi/ec680b75-d539-4653-b297-8bcf6e5f7686/openai-response-ostp-nsf-rfi-notice-request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan.pdf.

97. Jennifer Wang et al., "Do AI Companies Make Good on Voluntary Commitments to the White House?" *arXiv*:2508.08345, September 24, 2025, https://doi.org/10.48550/arXiv.2508.08345.

98. *GPT-5 System Card* (OpenAI, August 13, 2025), 52, https://cdn.openai.com/gpt-5-system-card.pdf; *System Card: Claude Opus 4 & Claude Sonnet 4* (Anthropic, May 2025), https://www-cdn.anthropic.com/6d8a8055020700718b0c49369f60816ba2a7c285.pdf.

99. Ben Buchanan, "The Government Knows A.G.I. Is Coming," *The Ezra Klein Show* (podcast), March 4, 2025, 1:06:00, ttps://www.nytimes.com/2025/03/04/opinion/ezra-klein-podcast-ben-buchanan.html; "Our AI Principles," Google AI, https://ai.google/responsibility/principles/.

100. Kylie Robison, "Why Anthropic's New AI Model Sometimes Tries to 'Snitch,'" *Wired*, May 28, 2025, https://www.wired.com/story/anthropic-claude-snitch-emergent-behavior/.

101. "Agentic Misalignment: How LLMs Could Be Insider Threats," Anthropic, June 20, 2025, https://www.anthropic.com/research/agentic-misalignment; Theo Browne, "SnitchBench: AI Model Whistleblowing Behavior Analysis," https://snitchbench.t3.gg/.

102. Interview with roundtable participant, meeting held under Chatham House rule, May 30, 2025.

103. Kratsios, "Unpacking the White House AI Action Plan with OSTP Director Michael Kratsios."

104. *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (National Institute of Standards and Technology, January 2023), https://www.nist.gov/itl/ai-risk-management-framework; *AI Cybersecurity Collaboration Playbook*

(Cybersecurity and Infrastructure Security Agency, January 14, 2025), https://www.cisa.gov/sites/default/files/2025-01/JCDC%20AI%20Playbook.pdf; "Artificial Intelligence Safety Institute Consortium (AISIC)," National Institute of Standards and Technology, https://www.nist.gov/artificial-intelligence/artificial-intelligence-safety-institute-consortium-aisic; and "Joint Cyber Defense Collaborative," Cybersecurity and Infrastructure Security Agency, https://www.cisa.gov/topics/partnerships-and-collaboration/joint-cyber-defense-collaborative.

105. "Artificial Intelligence in Software as a Medical Device," U.S. Food and Drug Administration, March 25, 2025, https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-software-medical-device; "Roadmap for Artificial Intelligence Safety Assurance," Federal Aviation Administration, August 2024, https://www.faa.gov/aircraft/air_cert/step/roadmap_for_AI_safety_assurance.

106. "Homepage," Chief Digital and Artificial Intelligence Office, https://www.ai.mil.

107. *Artificial Intelligence Strategy* (U.S. Department of Energy, October 2025), https://www.energy.gov/sites/default/files/2025-09/EXEC-2025-010630%20-%20250923_%20DOE%20AI%20Strategy%20VFinal.pdf.

108. Department of Commerce, "Statement from U.S. Secretary of Commerce Howard Lutnick on Transforming the U.S. AI Safety Institute into the Pro-Innovation, Pro-Science U.S. Center for AI Standards and Innovation," press release, June 3, 2025, https://www.commerce.gov/news/press-releases/2025/06/statement-us-secretary-commerce-howard-lutnick-transforming-us-ai.

109. Department of Commerce, "At the Direction of President Biden, Department of Commerce to Establish U.S. Artificial Intelligence Safety Institute to Lead Efforts on AI Safety," press release, November 1, 2021, https://www.commerce.gov/news/press-releases/2023/11/direction-president-biden-department-commerce-establish-us-artificial;

110. *Winning the Race: America's AI Action Plan*, 22–23.

111. Ben Cottier et al., "How Far Behind Are Open Models?" Epoch AI, November 4, 2024, https://epoch.ai/blog/open-models-report.

112. Helen Toner, "Nonproliferation Is the Wrong Approach to AI Misuse," Rising Tide, April 5, 2025, https://helentoner.substack.com/p/nonproliferation-is-the-wrong-approach.

113. "Working with US CAISI and UK AISI to Build More Secure AI Systems," OpenAI, September 12, 2025, https://openai.com/index/us-caisi-uk-aisi-ai-update/; "Strengthening Our Safeguards Through Collaboration with US CAISI and UK AISI," Anthropic, September 12, 2025, https://www.anthropic.com/news/strengthening-our-safeguards-through-collaboration-with-us-caisi-and-uk-aisi.

114. Mrinank Sharma et al., "Constitutional Classifiers: Defending Against Universal Jailbreaks Across Thousands of Hours of Red Teaming," *arXiv*:2501.18837, January 31, 2025, https://doi.org/10.48550/arXiv.2501.18837; "Strengthening Our Safeguards Through Collaboration with US CAISI and UK AISI."

115. "Memorandum of Understanding Between the Government of the United States of America and the Government of the United Kingdom of Great Britain and Northern Ireland Regarding the Technology Prosperity Deal," The White House, September 18, 2025, https://www.whitehouse.gov/presidential-actions/2025/09/memorandum-of-understanding-between-the-government-of-the-united-states-of-america-and-the-government-of-the-united-kingdom-of-great-britain-and-northern-ireland-regarding-the-technology-prosperity-de/.

116. *Winning the Race: America's AI Action Plan*, 4, 21, and 16.

117. *Evaluation of DeepSeek AI Models* (Center for AI Safety and Innovation, September 30, 2025), https://www.nist.gov/system/files/documents/2025/09/30/CAISI_Evaluation_of_DeepSeek_AI_Models.pdf.

118. *Winning the Race: America's AI Action Plan*, 9.

119. Saurabh Bagchi, "What Is a Black Box? A Computer Scientist Explains What It Means When the Inner Workings of AIs Are Hidden," The Conversation, May 22, 2023, https://theconversation.com/what-is-a-black-box-a-computer-scientist-explains-what-it-means-when-the-inner-workings-of-ais-are-hidden-203888.

120. Dario Amodei, "The Urgency of Interpretability," personal website, April 2025, https://www.darioamodei.com/post/the-urgency-of-interpretability,

121. *Technical Assessment of the Capital Facility Needs of the National Institute of Standards and Technology* (Washington: The National Academies Press, National Academies of Sciences, Engineering, and Medicine, 2023), https://doi.org/10.17226/26684; Cat Zakrzewski, "This Agency Is Tasked with Keeping AI Safe. Its Offices Are Crumbling," *The Washington Post*, March 6, 2024, https://www.washingtonpost.com/technology/2024/03/06/nist-ai-safety-lab-decaying/.

122. "Budget Tracker: FY2026 National Institute of Standards and Technology," American Institute of Physics, July 24, 2025, https://www.aip.org/fyi/fy2026-national-institute-of-standards-and-technology.

123. Alex Petropoulos, "The AI Safety Institute Network: Who, What and How?" Centre for Future Generations, October 9, 2024, https://cfg.eu/the-ai-safety-institute-network-who-what-and-how/.

124. Petropoulos, "The AI Safety Institute Network: Who, What and How?"; "The European AI Office: Driving Innovation with Open and Trusted Data," European Union, September 5, 2025, https://data.europa.eu/en/news-events/news/european-ai-office-driving-innovation-open-and-trusted-data.

125. Justine Brooks, "Government of Canada Announces Canadian AI Safety Institute," Canadian Institute for Advanced Research, November 12, 2024, https://cifar.ca/cifarnews/2024/11/12/government-of-canada-announces-canadian-ai-safety-institute/; "Introducing the AI Safety Institute," UK Department for Science, Innovation, and Technology, January 17, 2024, https://www.gov.uk/government/publications/ai-safety-institute-overview/introducing-the-ai-safety-institute; Gian Volpicelli, "EU Needs 'Oppenheimers' to Run AI Policy," *Politico*, March 6, 2024, https://www.politico.eu/article/eu-needs-oppenheimer-to-run-ai-policy-key-lawmaker-tudorache-romania-says/; and *Committee Print of the Committee on*

Appropriations, U.S. House of Representatives, on H.R. 4366 / Public Law 118–42 (U.S. Government Publishing Office), 403, https://www.govinfo.gov/content/pkg/CPRT-118HPRT56550/pdf/CPRT-118HPRT56550.pdf.

126. "Transcript: Donald Trump's Address at 'Winning the AI Race' Event."

127. Maslej et al., *Artificial Intelligence Index Report 2025*, 3.

128. Gregory Smith et al., *Enhancing in-House U.S. Government AI Talent* (Santa Monica, CA: RAND Corporation, March 2025), https://www.rand.org/content/dam/rand/pubs/working_papers/WRA3800/WRA3882-1/RAND_WRA3882-1.pdf; *Winning the Race: America's AI Action Plan*, 22.

129. Thomas Liao, "Why Eval Startups Fail," personal website, May 8, 2025, https://thomasliao.com/eval-startups.

130. "National Institute of Standards and Technology: Fiscal Year 2025 Budget Submission to Congress," Department of Commerce, 2024, https://www.commerce.gov/sites/default/files/2024-03/NIST-NTIS-FY2025-Congressional-Budget-Submission.pdf.

131. "Research Engineer, Frontier Evals & Environments," OpenAI, accessed September 6, 2025, https://openai.com/careers/research-engineer-frontier-evals-and-environments-san-francisco/.

132. *Winning the Race: America's AI Action Plan*, 10.

133. Federally Funded Research and Development Centers, 48 CFR § 35.017 (2007), https://www.ecfr.gov/current/title-48/chapter-1/subchapter-F/part-35/section-35.017; Marcy Gallo, *Federally Funded Research and Development Centers (FFRDCs): Background and Issues for Congress* (Congressional Research Service, August 27, 2021), https://www.congress.gov/crs-product/R44629.

134. Gallo, *Federally Funded Research and Development Centers (FFRDCs)*; "Master Government List of Federally Funded R&D Centers," National Center for Science and Engineering Statistics, February 2025, https://ncses.nsf.gov/resource/master-gov-lists-ffrdc.

135. "Homepage," NIST National Cybersecurity Center of Excellence, https://www.nccoe.nist.gov/.

136. "A Right to Warn About Advanced Artificial Intelligence," June 4, 2024, https://righttowarn.ai/; Charlie Bullock and Mackenzie Arnold, "Protecting AI Whistleblowers," Lawfare, June 25, 2025, https://law-ai.org/protecting-ai-whistleblowers/.

137. *Committing to Effective Whistleblower Protection* (Organization for Economic Co-operation and Development, 2016), https://www.oecd.org/content/dam/oecd/en/publications/reports/2016/03/committing-to-effective-whistleblower-protection_g1g65d0a/9789264252639-en.pdf; Artificial Intelligence Models: Large Developers.

138. Cade Metz and Daisuke Wakabayashi, "Google Researcher Says She Was Fired over Paper Highlighting Bias in A.I.," *The New York Times*, December 3, 2020, https://www.nytimes.com/2020/12/03/technology/google-researcher-timnit-gebru.html; Alex Heath, "Meta Is Firing About 20 Employees for Leaking Information," The Verge, February 27, 2025, https://www.theverge.com/labor/621059/meta-fires-20-employee-leakers; Shakeel Hashim, "OpenAI Employee Says He Was Fired for Raising Security Concerns to Board," Transformer, June 4, 2024, https://www.transformernews.ai/p/openai-employee-says-he-was-fired.

139. Kelsey Piper, "Leaked OpenAI Documents Reveal Aggressive Tactics Toward Former Employees," Vox, May 22, 2024, https://www.vox.com/future-perfect/351132/openai-vested-equity-nda-sam-altman-documents-employees.

140. "A Right to Warn About Advanced Artificial Intelligence."

141. "Re: OpenAI Violations of Rule 21F-17(a) and Implementation of E.O. 14110," letter to Gary Gensler, July 1, 2024, https://www.washingtonpost.com/documents/83df0e55-546c-498a-9efc-06fac591904e.pdf.

142. Piper, "Leaked OpenAI Documents Reveal Aggressive Tactics Toward Former Employees."

143. AI Whistleblower Protection Act, H.R.4360, 119th Cong. (2025), https://www.congress.gov/bill/119th-congress/house-bill/3460/.

144. AI Whistleblower Protection Act.

145. Defense Production Act of 1950, 50 U.S.C. § 4502 (1950), https://www.law.cornell.edu/uscode/text/50/chapter-55.

146. John S. McCain National Defense Authorization Act for Fiscal Year 2019, Pub. L. No. 115-232, 132 Stat. 2238 (2018), https://www.congress.gov/115/plaws/publ232/PLAW-115publ232.pdf.

147. John S. McCain National Defense Authorization Act for Fiscal Year 2019.

148. *Defense Production Act: Use and Challenges from Fiscal Years 2018 to 2024: Testimony Before the Subcommittee on National Security, Illicit Finance, and International Financial Institutions*, 119th Cong. (2025) (statement of William Russell, director of contracting and national security acquisitions, Government Accountability Office), https://docs.house.gov/meetings/BA/BA10/20250612/118372/HHRG-119-BA10-Wstate-RussellW-20250612.pdf.

149. Joseph R. Biden, "Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," Exec. Order No. 14110, October 30, 2023, https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

150. *White House Overreach on AI: Testimony Before Subcommittee on Cybersecurity, Information Technology, and Government Innovation Committee on Oversight and Accountability*, 118 Cong. (2024) (statement by Jennifer Huddleston, technology policy research fellow, Cato Institute), https://www.cato.org/testimony/white-house-overreach-ai; Charlie Bullock, "Commerce Just Proposed the Most Significant Federal AI Regulation to Date—and No One Noticed," Institute for Law and AI, October 2024, https://law-ai.org/commerce-federal-ai-regulation/.

151. Donald J. Trump "Executive Order on Removing Barriers to American Leadership in Artificial Intelligence," Exec. Order No. 14179, January 23, 2025, https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.

152. Joe O'Brien "Coordinated Disclosure of Dual-Use Capabilities: An Early Warning System for Advanced AI," Institute for AI Policy and Strategy, June 21, 2024, https://www.iaps.ai/research/coordinated-disclosure.

153. *Winning the Race: America's AI Action Plan*, 5.

154. *Winning the Race: America's AI Action Plan*, 5.

155. "Transcript: Donald Trump's Address at 'Winning the AI Race' Event."

156. Elaine Kamarck, "9/11 and the Reinvention of the US Intelligence Community," Brookings Institution, August 27, 2021, https://www.brookings.edu/articles/9-11-and-the-reinvention-of-the-u-s-intelligence-community/.

157. "The 9/11 Commission Report," National Commission on Terrorist Attacks upon the United States, July 22, 2004, https://govinfo.library.unt.edu/911/report/index.htm.

158. *Winning the Race: America's AI Action Plan,* 6.

159. "Annex for Presidential Policy Directive 41—United States Cyber Incident Coordination," July 26, 2016, https://obamawhitehouse.archives.gov/the-press-office/2016/07/26/annex-presidential-policy-directive-united-states-cyber-incident.

160. Shannon Vavra, "White House Activates Cyber Emergency Response Under Obama-Era Directive," Cyberscoop, December 16, 2020, https://cyberscoop.com/solarwinds-white-house-national-security-council-emergency-meetings/.

161. *Winning the Race: America's AI Action Plan,* 11.

162. *Winning the Race: America's AI Action Plan,* 21.

163. Administrative Procedure Act, 5 U.S. Code § 551 (1946), https://www.law.cornell.edu/uscode/text/5/part-I/chapter-5/subchapter-II; "A Guide to the Rulemaking Process," Office of the Federal Register, https://uploads.federalregister.gov/uploads/2013/09/The-Rulemaking-Process.pdf.

164. "About the Rulemaking Process," Office of the U.S. Courts, https://www.uscourts.gov/forms-rules/about-rulemaking-process.

165. "AI Benchmarking" (AI Performance on METR Time Horizons); "AI Benchmarking" (AI Performance on SimpleBench).

166. *The National Cyber Incident Response Plan (NCIRP)* (Cybersecurity and Infrastructure Security Agency), https://www.cisa.gov/national-cyber-incident-response-plan-ncirp.

167. *National Cybersecurity Strategy* (The White House, March 2023), https://bidenwhitehouse.archives.gov/wp-content/uploads/2023/03/National-Cybersecurity-Strategy-2023.pdf.

168. "National Cyber Incident Response Plan Update Public Comment Draft," Cybersecurity and Infrastructure Security Agency, December 2024, https://www.cisa.gov/resources-tools/resources/national-cyber-incident-response-plan-update-public-comment-draft.

169. *AI've Got a Plan: America's AI Action Plan: Hearing Before Senate Subcommittee on Science, Manufacturing, and Competitiveness*, 199th Cong. 2 (2025) (statement of Michael Kratsios, director, Office of Science and Technology Policy), https://www.commerce.senate.gov/services/files/3DF64D5D-55F9-43DA-AB76-CDAF2586DB56.

170. William T. Egar, *Congressionally Mandated Reports: Overview and Considerations for Congress* (Congressional Research Service, May 14, 2020), https://www.congress.gov/crs-product/R46357.

171. Egar, *Congressionally Mandated Reports*.

172. *Final Report: National Security Commission on Artificial Intelligence* (National Security Commission on Artificial Intelligence, March 2021), https://assets.foleon.com/eu-central-1/de-uploads-7e3kk3/48187/nscai_full_report_digital.04d6b124173c.pdf.

173. Christopher Davis, *Expedited Procedures in the House: Variations Enacted into Law* (Congressional Research Service, September 16, 2015), https://www.congress.gov/crs-product/RL30599.

174. "A Resolution Designating July 30, 2025, as 'National Whistleblower Appreciation Day,'" S. 340, 119th Cong. (2025), https://www.congress.gov/bill/119th-congress/senate-resolution/340.

175. CARES Act, H.R. 748, 116th Cong (2020), https://www.congress.gov/bill/116th-congress/house-bill/748/.

176. National Emergencies Act, Pub L. No. 94, 412, 90 Stat. 1255 (1976), https://www.govinfo.gov/content/pkg/COMPS-10385/pdf/COMPS-10385.pdf.

177. "A Guide to Emergency Powers and Their Use," Brennan Center, July 1, 2025, https://www.brennancenter.org/our-work/research-reports/guide-emergency-powers-and-their-use.

178. Defense Production Act of 1950.

179. Aidan Lawson and June Rhee, "Usage of the Defense Production Act Throughout History and to Combat COVID-19," Yale School of Management, June 3, 2020, https://som.yale.edu/blog/usage-of-the-defense-production-act-throughout-history-and-to-combat-covid-19.

180. *Winning the Race: America's AI Action Plan*, 7.

181. Charlie Bullock et al., "Existing Authorities for Oversight of Frontier AI Models," Institute for Law and AI, July 2024, https://law-ai.org/existing-authorities-for-oversight/.

182. International Emergency Economic Powers Act, U.S.C. 50 § 1701 (1977), https://www.law.cornell.edu/uscode/text/50/1701/.

183. Andrew Boyle, "Checking the President's Sanctions Powers," Brennan Center, June 10, 2021, https://www.brennancenter.org/our-work/policy-solutions/checking-presidents-sanctions-powers.

184. Joseph R. Biden, "Imposing Sanctions on Foreign Persons Involved in the Global Illicit Drug Trade," Exec. Order No. 14059, December 15, 2021, https://www.federalregister.gov/documents/2021/12/17/2021-27505/imposing-sanctions-on-foreign-persons-involved-in-the-global-illicit-drug-trade; Joseph R. Biden, "Addressing United States Investments in Certain National Security Technologies and Products in Countries of Concern," Exec. Order No. 14105, August 9, 2023, https://www.federalregister.gov/documents/2023/08/11/2023-17449/addressing-united-states-investments-in-certain-national-security-technologies-and-products-in; Joseph R. Biden, "Declaration of National Emergency and Invocation of Emergency Authority Relating to the Regulation of the Anchorage and Movement of Russian-Affiliated Vessels to United States Ports," Proc. 10371, April 21, 2022, https://www.federalregister.gov/documents/2022/04/22/2022-08872/declaration-of-national-emergency-and-invocation-of-emergency-authori-

ty-relating-to-the-regulation; "Taking Additional Steps to Address the National Emergency with Respect to Significant Malicious Cyber-Enabled Activities," Department of Commerce, 89 Fed. Reg. 5698 (January 26, 2024), 5698–5735, https://www.govinfo.gov/content/pkg/FR-2024-01-29/pdf/2024-01580.pdf.

185. Federal Information Security Modernization Act of 2014, U.S.C. 44 § 3553 (2014), https://www.law.cornell.edu/uscode/text/44/3553.

186. Cybersecurity and Infrastructure Security Agency, 6 U.S.C § 652 (2018), https://www.law.cornell.edu/uscode/text/6/652.

187. For an example, see "BOD 22-01: Reducing the Significant Risk of Known Exploited Vulnerabilities," Cybersecurity and Infrastructure Security Agency, November 3, 2021, https://www.cisa.gov/news-events/directives/bod-22-01-reducing-significant-risk-known-exploited-vulnerabilities.

188. "ED 24-01: Mitigate Ivanti Connect Secure and Ivanti Policy Secure Vulnerabilities," January 19, 2024, https://www.cisa.gov/news-events/directives/supplemental-direction-v1-ed-24-01-mitigate-ivanti-connect-secure-and-ivanti-policy-secure.

189. Homeland Security Act of 2002.

190. Cybersecurity and Infrastructure Security Agency, "CISA Directs Federal Agencies to Secure Internet-Exposed Management Interfaces," press release, June 13, 2023, https://www.cisa.gov/news-events/news/cisa-directs-federal-agencies-secure-internet-exposed-management-interfaces.

191. "Memorandum on Improving the Cybersecurity of National Security, Department of Defense, and Intelligence Community Systems," January 19, 2022, https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2022/01/19/memorandum-on-improving-the-cybersecurity-of-national-security-department-of-defense-and-intelligence-community-systems/.

192. Negotiated Rulemaking Act of 1990, 5 U.S.C § 564 (1990), https://www.law.cornell.edu/uscode/text/5/564.

193. Administrative Procedure Act.

194. "A Guide to the Rulemaking Process," 9.

195. *Notice and Comment Part II: Good Cause and Other Exceptions* (Governing for Impact, May 2025), http://governingforimpact.org/wp-content/uploads/2025/05/Notice-and-Comment-Part-II-Good-Cause-and-Other-Exceptions.pdf; Andrew Coghlan, *The Good Cause Exception to Notice and Comment Rulemaking* (Congressional Research Service, August 27, 2025), https://www.congress.gov/crs-product/R44356.

196. Adam Grogg and John Lewis, "The Legal Defects in the Trump Administration's Attempts to Deregulate Without Notice and Comment," Just Security, June 17, 2025, https://www.justsecurity.org/114539/trump-cannot-deregulate-without-notice-comment/; Immigration and Nationality Law Committee, "Comment in Opposition to the Proposed Interim Final Rule, 'Securing the Border,'" New York City Bar, July 11, 2024, https://www.nycbar.org/reports/comment-in-opposition-to-the-proposed-interim-final-rule-securing-the-border/.

197. Yidi Wu, "Statutory Deadlines for Agency Regulation: A Carrot Approach," *New York University Law Review* 100, no. 1 (April 2025), https://nyulawreview.org/issues/volume-100-number-1/statutory-deadlines-for-agency-regulation-a-carrot-approach/.

198. Mikołaj Barczentewicz, "US Export Controls on AI and Semiconductors," International Center for Law & Economics, March 25, 2025, https://laweconcenter.org/resources/us-export-controls-on-ai-and-semiconductors/.

199. *Winning the Race: America's AI Action Plan,* 23.

200. Donald Trump, "Trump Speaks at U.N." (formal address, United Nations, September 23, 2025), https://www.rev.com/transcripts/trump-speaks-at-un.

201. Trump, "Trump Speaks at U.N."

202. Nicholas Cropper et al., "A Modular-Incremental Approach to Improving Compliance Verification with the Biological Weapons Convention," *Health Security* 21, no. 5 (September 2023), https://doi.org/10.1089/hs.2023.0078.

203. Kratsios, "Remarks at the Security Council's Open Debate on Artificial Intelligence and International Peace and Security."

204. Kratsios, "Remarks at the Security Council's Open Debate on Artificial Intelligence and International Peace and Security."

205. Egan and Salvador, "The United States Must Avoid AI's Chernobyl Moment."

206. *Winning the Race: America's AI Action Plan,* 19.

207. *Winning the Race: America's AI Action Plan,* 19.

208. *Cybersecurity Incident & Vulnerability Response Playbooks* (Cybersecurity and Infrastructure Security Agency, November 2021), https://www.cisa.gov/sites/default/files/2024-08/Federal_Government_Cybersecurity_Incident_and_Vulnerability_Response_Playbooks_508C.pdf; Russell T. Vought, "Fiscal Year 2019–2020 Guidance on Federal Information Security and Privacy Management Requirements," memorandum, M-20-04, Office of Management and Budget, November 19, 2019, https://www.whitehouse.gov/wp-content/uploads/2019/11/M-20-04.pdf.

209. *AI Cybersecurity Collaboration Playbook*, 7.

210. Dmitrii Volkov and Reworr, "LLM Agent Honeypot: Monitoring AI Hacking Agents in the Wild," *arXiv*:2410.13919, February 10, 2025, https://doi.org/10.48550/arXiv.2410.13919; Artem Petrov and Dmitrii Volkov, "Evaluating AI Cyber Capabilities with Crowdsourced Elicitation," *arXiv*:2505.19915, May 27, 2025, https://doi.org/10.48550/arXiv.2505.19915.

211. *Winning the Race: America's AI Action Plan,* 18.

212. John Moteff, *Critical Infrastructures: Background, Policy, and Implementation* (Congressional Research Service, June 10, 2015), https://www.congress.gov/crs_external_products/RL/PDF/RL30153/RL30153.29.pdf.

213. Moteff, *Critical Infrastructures*.

214. Critical Infrastructure Information Act, Pub. L. No. 107-296, 116 Stat. 2145 (2002), https://www.congress.gov/107/plaws/publ296/PLAW-107publ296.pdf; "Protected Critical Infrastructure Information (PCII) Program," Cybersecurity and Infrastructure Security Agency, https://www.cisa.

gov/resources-tools/programs/protected-critical-infrastructure-information-pcii-program; and Cybersecurity Information Sharing Act of 2014, Pub. L. No. 114-113, 129 Stat. 2935 (2015), https://www.cisa.gov/sites/default/files/publications/Cybersecurity%2520Information%2520Sharing%2520Act%2520of%25202015.pdf.

215. Maggie Miller and Dana Nickel, "Government Flying Partially Blind to Threats After Key Cyber Law Expires," *Politico*, October 3, 2025, https://www.politico.com/news/2025/10/03/cyber-law-cisa-2015-shutdown-00592501.

216. Bank Policy Institute, "America's Critical Infrastructure Sectors Urge Congress to Reauthorize Cybersecurity Information-Sharing Law," press release, March 21, 2025, https://bpi.com/americas-critical-infrastructure-sectors-urge-congress-to-reauthorize-cybersecurity-information-sharing-law/.

217. "Joint Cyber Defense Collaborative."

218. Christian Vasquez, "CISA Leads First Tabletop Exercise for AI Cybersecurity," Cyberscoop, June 14, 2024, https://cyberscoop.com/cisa-ai-tabletop-exercise-playbook/.

219. *AI Cybersecurity Collaboration Playbook*.

220. "Stress Tests," Board of Governors Federal Reserve Board, June 22, 2022, https://www.federalreserve.gov/supervisionreg/stress-tests-capital-planning.htm.

221. Joseph R. Biden, "Improving the Nation's Cybersecurity," Exec. Order No. 14028 (May 12, 2021), https://www.federalregister.gov/documents/2021/05/17/2021-10460/improving-the-nations-cybersecurity; Department of Homeland Security, "DHS Launches First-Ever Cyber Safety Review Board," press release, February 3, 2022, https://www.dhs.gov/archive/news/2022/02/03/dhs-launches-first-ever-cyber-safety-review-board.

222. James Pearson and Christopher Bing, "Chinese Hackers Breached State, Commerce Depts, Microsoft and US Say," Reuters, July 12, 2023, https://www.reuters.com/technology/chinese-hackers-accessed-government-emails-microsoft-says-2023-07-12/.

223. "Results of Major Technical Investigations for Storm-0558 Key Acquisition," Microsoft Security Response Center, September 3, 2023, https://msrc.microsoft.com/blog/2023/09/results-of-major-technical-investigations-for-storm-0558-key-acquisition/.

224. *Review of the Summer 2023 Microsoft Exchange Online Intrusion* (Cyber Safety Review Board, March 20, 2024), https://www.cisa.gov/sites/default/files/2025-03/CSRBReviewOfTheSummer2023MEOIntrusion508.pdf.

225. Andrew Garbarino, "Letter to DHS Secretary Krisi Noem," March 13, 2025, https://homeland.house.gov/wp-content/uploads/2025/03/2025.03.12-CSRB-Review-Letter.pdf; Jeff Greene, "What's Next for the Cyber Safety Review Board?," Lawfare, September 16, 2025, https://www.lawfaremedia.org/article/what-s-next-for-the-cyber-safety-review-board.

226. Greene, "What's Next for the Cyber Safety Review Board?"

227. Grace Dille, "DHS Nominee Says Cyber Safety Review Board Will Be 'Reconstituted,'" Meritalk, February 25, 2025, https://meritalk.com/articles/dhs-nominee-says-cyber-safety-review-board-will-be-reconstituted/.

228. "FY2025–2026 CISA International Strategic Plan," Cybersecurity and Infrastructure Security Agency, https://www.cisa.gov/2025-2026-cisa-international-strategic-plan.

229. "Convention on Assistance in the Case of a Nuclear Accident or Radiological Emergency," International Atomic Energy Agency, https://www.iaea.org/topics/nuclear-safety-conventions/convention-assistance-case-nuclear-accident-or-radiological-emergency.

230. Biden, "Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence."

231. *Committee Print of the Committee on Appropriations, U.S. House of Representatives, on H.R. 4366 / Public Law 118–42* (U.S. Government Publishing Office), 403.

232. Caroline Nihill and Madison Alder, "New TMF Investments Support AI Safety Institute, Upgrades to Nuclear Emergency Response," FedScoop, July 23, 2024, https://fedscoop.com/new-tmf-investments-support-ai-safety-institute-upgrades-to-nuclear-emergency-response/.

233. "Our Impact," Technology Modernization Fund, https://tmf.cio.gov/impact/.

234. "Funding & Repayment," Technology Modernization Fund, https://tmf.cio.gov/files/2025/05/Funding-repayment_updated050225.pdf.

235. "Our Investments," Technology Modernization Fund, https://tmf.cio.gov/investments/.

236. Continuing Appropriations and Extensions Act, 2025, Pub. L. No. 118-83, 138 Stat. 1524 (2024), https://www.congress.gov/118/plaws/publ83/PLAW-118publ83.pdf; Haley Stevens, "Stevens and Obernolte Request Additional $10 Million to Secure U.S. AI Leadership," press release, February 22, 2024, https://stevens.house.gov/media/press-releases/stevens-and-obernolte-request-additional-10-million-secure-us-ai-leadership.

237 . Department of Commerce, "Statement from U.S. Secretary of Commerce Howard Lutnick on Transforming the U.S. AI Safety Institute into the Pro-Innovation, Pro-Science U.S. Center for AI Standards and Innovation."

238. *Report to Accompany S.2354, Departments of Commerce and Justice, Science, and Related Agencies Appropriations Bill 2026* (U.S. Government Publishing Office, July 17, 2025), 29, https://www.congress.gov/119/crpt/srpt44/CRPT-119srpt44.pdf.

239. Continuing Appropriations and Extensions Act, 2026, H.R.5371, 119th Cong. (2025), https://www.congress.gov/bill/119th-congress/house-bill/5371.

240. Bureau of Labor Statistics, "Employer Costs for Employee Compensation—JUNE 2025," press release, September 12, 2025, https://www.bls.gov/news.release/pdf/ecec.pdf.

About the Center for a New American Security

The mission of the Center for a New American Security (CNAS) is to develop strong, pragmatic, and principled national security and defense policies. Building on the expertise and experience of its staff and advisors, CNAS engages policymakers, experts, and the public with innovative, fact-based research, ideas and analysis to shape and elevate the national security debate. A key part of our mission is to inform and prepare the national security leaders of today and tomorrow.

CNAS is located in Washington, D.C., and was established in February 2007 by cofounders Kurt M. Campbell and Michèle A. Flournoy. CNAS is a 501(c)3 tax-exempt nonprofit organization. Its research is independent and nonpartisan.

AMERICA'S EDGE 2025

The United States faces a rapidly changing global security landscape. Evolving technology, shifting alliances, and emerging threats require America to harness bold, innovative approaches. America's Edge is a Center-wide initiative featuring research, events, and multimedia for enhancing America's global edge.

CNAS Editorial

DIRECTOR OF STUDIES
Katherine L. Kuzminski

PUBLICATIONS & EDITORIAL DIRECTOR
Maura McCarthy

SENIOR EDITOR
Emma Swislow

ASSOCIATE EDITOR
Caroline Steel

CREATIVE DIRECTOR
Melody Cook

DESIGNER
Alina Spatz

Cover Art & Production Notes

COVER ILLUSTRATION
Mark Harris

PRINTER
CSI Printing & Graphics
Printed on an HP Indigo Digital Press

Center for a New American Security
1701 Pennsylvania Ave NW
Suite 700
Washington, DC 20006
CNAS.org
@CNASdc

Contact Us
202.457.9400
info@cnas.org

Center for a
New American
Security