FAS FEDERATION OF AMERICAN SCIENTISTS

MARCH 2026

Artificial Intelligence and Nuclear Risks

Lessons from the Nuclear Age in the Era of Artificial Intelligence

GLOBAL RISK
FEDERATION OF AMERICAN SCIENTISTS

ABOUT FAS

The **Federation of American Scientists (FAS)** is an independent, nonpartisan think tank that brings together members of the science and policy communities to collaborate on mitigating global catastrophic threats. Founded in November 1945 as the Federation of Atomic Scientists by scientists who built the first atomic bombs during the Manhattan Project, FAS is devoted to the belief that scientists, engineers, and other technically trained people have the ethical obligation to ensure that the technological fruits of their intellect and labor are applied to the benefit of humankind. In 1946, FAS rebranded as the Federation of American Scientists to broaden its focus to prevent global catastrophes.

Since its founding, FAS has served as an influential source of information and rigorous, evidence-based analysis of issues related to national security. Specifically, FAS works to reduce the spread and number of nuclear weapons, prevent nuclear and radiological terrorism, promote high standards for the safety and security of nuclear energy, illuminate government secrecy practices, and prevent the use of biological and chemical weapons.

FAS can be reached at **fas@fas.org**.

COVER IMAGE: NUCLEAR MISSILE LAUNCH KEYS 1996 VIA WIKIMEDIA COMMONS.

ABOUT THIS REPORT

This report summarizes the key findings, insights, and recommendations from a February 2026 roundtable event on artificial intelligence (AI) and nuclear risk. The Federation of American Scientists, in partnership with the Future of Life Institute (FLI), brought together AI and nuclear risk reduction and nonproliferation stakeholders across academia, industry, government, and philanthropy. Experts discussed how lessons learned from the nuclear age can be applied to AI and considered how AI tools could be used to reduce nuclear risks. These experts identified significant space for policy intervention to mitigate downstream risks from the use of AI in nuclear command, control, and communications (NC3) systems, along with areas in which AI may be useful.

This report is structured in three parts: an executive summary, a detailed analysis of the findings, and two pre-read documents authored for participants in advance of the event. These pre-reads were authored by Ankit Panda, Stanton Senior Fellow at the Carnegie Endowment for International Peace, and Austin Long, PhD, Senior Nuclear Fellow at the Center for Nuclear Security Policy at the Massachusetts Institute of Technology.

Global Risk Program at FAS

The Global Risk Program focuses on addressing and preventing the events and threats that could permanently cripple or destroy humanity. Among them: nuclear war, the next global pandemic, biological attack, and even a collision with a massive near-earth object. Humanity must proactively develop and pursue sound policies to protect against these dangers, including through global cooperation. Find out more at our website: fas.org/issue/global-risk

The project is led by Yong-Bee Lim, PhD, who is Associate Director of the Global Risk Program at FAS.

Funding

This report and the associated workshop were made possible through the generous support of the Future of Life Institute and are part of a wider series in our ongoing "AI x Global Risk Nexus" project. This project will culminate in an **AI x Global Risk Summit in May 2026**. The views expressed in this report are those of the authors and do not necessarily reflect the positions of the funders or participants.

Acknowledgements

Special thanks to Dr. Oliver Stephenson, Associate Director of Artificial Intelligence and Emerging Technology Policy; Mr. Matt Korda, Associate Director of the Nuclear Information Project; Ms. Eliana Johns, Senior Research Associate of the Nuclear Information Project; Mr. Elliott Gunnell, Project Associate, Global Risk Program; Ms. Angela Kellett, Senior Associate, Communications; and Mr. Abhay Katoch, Horizon Fellow, for their contributions to the event and this report. Additional thanks to Kate Kohn, Senior Communications Manager, for developing the graphics for this report. Finally, deep appreciation to Mr. Jon Wolfsthal, former Director of the Global Risk portfolio at FAS, for his leadership during his tenure on the project.

FAS can be reached at 1150 18th St. NW, Suite 1000, Washington, DC, 20036, fas@fas.org, or through fas.org.

CONTENTS

EXECUTIVE SUMMARY

On February 10th, 2026, the Federation of American Scientists, in partnership with the Future of Life Institute (FLI), convened a DC-based roundtable focused on two themes: lessons that the AI space can learn from the nuclear weapons age, and potential opportunities for nuclear risk mitigation through AI tools and capabilities. Participants included technical and policy experts from think tanks, academia, industry, and government. Leadership and policymaker perspectives were also provided by Senator Edward Markey (D-MA) and Representative Donald Bacon (R-NE-02).

FINDINGS

Participants drew frequent comparisons between the advent of AI and nuclear weapons, with a focus on the inherent risks and potential catastrophic impacts both technologies share. These potential impacts have prompted global discourse, substantial government investment, and the potential for destabilizing arms races. However, attendees emphasized that these technologies fundamentally differ, requiring new frameworks for risk management and strategic stability, specifically on AI capabilities.

Participants first noted that AI and nuclear weapons differ in their origins, as well as the technological pathways each has taken: whereas nuclear technology has been primarily developed and tightly controlled by state-actors and eventually international organizations, AI research is overwhelmingly driven by the private sector and commercial applications, and no international AI governance mechanisms currently exist.

Second, participants highlighted how the risk profiles of both technologies are different. Catastrophic risks associated with nuclear weapons use are certain and comparatively well-understood, although recent events have renewed spirited debate on foundational nuclear concepts like deterrence and signaling. Further, we may be entering unknown territory again following the expiration of New START. Conversely, the risks posed by AI remain uncertain, dynamic, and more broadly distributed across military and civilian contexts.

Third, many AI risks create new or exacerbate old issues with other domains; for example, the deployment of AI in lethal autonomous weapons systems (LAWS) or as a part of the decision-making process for a nuclear launch authorization are critical issues we are grappling with in the policy arena. These risks not only interact with other domains but may also differ depending on the types of AI tools used: attendees noted, for example, that AI may be integrated into existing nuclear enterprises in myriad ways, presenting a spectrum of assessed risk.

Participants noted some clear red lines when it comes to specific AI-nuclear applications. For example, some attendees noted that compromising existing fail-safes or nuclear command-and-control (NC2) is unacceptable. However, participants also agreed that certain systems, such as AI-enabled intelligence, surveillance, and reconnaissance (ISR), have viable ways to mitigate risks without sacrificing the utility of the technology.

This may necessitate the U.S. government to adopt a different approach for AI risk; participants observed that public-private engagement will be necessary to control AI deployment, especially in developing testing and evaluation (T&E) protocols for military use. Within the private-sector, frontier AI companies should also engage in voluntary governance, such as establishing 'redlines' against the deployment of AI systems in nuclear launch authority. Participants also felt that international engagement was necessary for AI, but noted that proposed models, such as those based on the International Atomic Energy Agency (IAEA), may not be applicable, given the aforementioned differences between AI and nuclear weapons.

SUMMARY OF POLICY OPTIONS

Based on roundtable discussion and preliminary analysis, FAS identified the following policy options to address challenges at the nexus of AI and nuclear weapons. These options are included for discussion purposes and do not imply endorsement by any individual participant.

- **Establish explicit guardrails for AI in nuclear decision-making and force employment**, including clear limits on the role of AI in launch-related functions and an emphasis on preserving meaningful human judgment.
- **Build a bespoke testing, evaluation, and assurance regime for AI used in nuclear-related systems**, with attention to reliability, adversarial failure modes, human-AI team performance, and specialized or air-gapped deployments.
- **Prioritize lower-risk and risk-reducing applications**—such as maintenance, bounded monitoring, verification support, and cyber vulnerability discovery—while **preventing function creep into more sensitive decision functions**.
- **Create structured information-sharing and translation channels** among government, AI developers, national laboratories, military operators, and outside experts, including recurring dialogues and incident-reporting mechanisms.
- **Pursue incremental norm-building and confidence-building measures**, rather than assuming that a ready-made AI equivalent of the nuclear nonproliferation architecture can be easily created.

WHAT WE HEARD

On February 10th, 2026, the Federation of American Scientists (FAS), in partnership with the Future of Life Institute (FLI), convened a roundtable discussion for over 60 members of the national security, nuclear enterprise, and artificial intelligence (AI) policy communities; these included representatives from government, think-tanks, academia, industry, and philanthropic stakeholders. The purpose of this roundtable was to examine the intersection of AI and nuclear weapons, with particular attention paid to the potential historical parallels and the associated risks and benefits for the integration of the two technologies. This document provides an overview of perspectives explored during the roundtable. Components include a summary of key findings, policy options for government and stakeholders, and background information given to participants.

This discussion occurred during a period of transformation in the U.S. nuclear enterprise and increasing use of AI tools in both military and civilian contexts. The United States is in the process of modernizing its nuclear arsenal, with overall costs to operate, sustain, and modernize projected to reach nearly $1 trillion USD over the next decade.[1] Updating nuclear command, control, and communications (NC3) infrastructure is a major component of this modernization, and proposals to incorporate commercial systems into future NC3 architecture raise new challenges for regulatory and reliability standards.[2]

This has led to growing concerns around the risks of integrating AI in nuclear weapons systems, decision-making processes, and NC3 infrastructure. These concerns include how the models are developed and trained, and the level of access to sensitive systems that should be granted to those models. This roundtable focused on these dynamics, with discussion highlighting multiple similarities and differences between the development of nuclear weapons and AI, along with identifying potential risks and mitigation strategies.

FINDING 1. THE AI-NUCLEAR ANALOGY IS INFORMATIVE AND INCOMPLETE

A question framed throughout the roundtable was whether AI development shares enough similarities with developments from the nuclear age to draw comparisons and lessons learned for AI. Participants drew frequent comparisons between the emergence of AI and nuclear weapons, noting that while some historical takeaways may be applied to both—and the proliferation of dual-use technology more broadly—the unique challenges posed by both AI and nuclear weapons warrant novel policy approaches. For example, both technologies have been subject to significant investment, with billions of dollars spent during both the Manhattan Project (<$50 bn) and on current AI data-center capital investments (>$400 bn).

Additionally, both technologies have warranted significant discussion internationally, with the United Nations Atomic Energy Commission forming in early 1946, and the Council of Europe's Framework Convention on Artificial Intelligence ratified in 2024. Many of these engagements have had to do with controlling access to specific materials and equipment, such as through Eisenhower's 1953 "Atoms for Peace" program and concentrated efforts on the part of the United States government to control the export of critical equipment that may fuel AI capabilities elsewhere. Similarly, international AI diplomacy has also often concerned controls over high-performance computer hardware, also known as "compute."

However, participants also noted significant differences between the trajectory and current states of nuclear weapons and AI capabilities. One such distinction was the motivator for technological progress: whereas nuclear weapons technology was developed, funded, and remains tightly controlled by state-actors and international

1 Congressional Budget Office. Projected Costs of U.S. Nuclear Forces, 2025 to 2034 (Washington, DC: Congressional Budget Office, April 24, 2025). https://www.cbo.gov/publication/61362. Accessed March 3, 2026.
2 U.S. Strategic Command, "Space Symposium Media Roundtable," speech transcript, April 9, 2019. https://www.stratcom.mil/Media/Speeches/Article/1817618/space-symposium-media-roundtable/. Accessed March 2, 2026.

organizations, AI research is overwhelmingly driven and funded by the private sector for commercial applications, with national governments and international organizations struggling to develop and implement regulatory frameworks, especially given the pace at which the technology is changing.

Because AI research in the United States is currently led by private entities, the U.S. government and international agencies are limited in their ability to directly influence the developmental trajectory. This contrasts starkly with nuclear weapons development, where the U.S. government led all research on selected projects and oversaw the implementation of safeguards to try to prevent nuclear incidents and ensure the technology does not fall into the wrong hands. For AI development, participants noted that one avenue the U.S. government should rely on are regulation mechanisms subject to pre-existing corporate and tort law.

Additionally, panelists noted the inherent dual-use nature of both technologies for both civilian and military applications, as previously noted by FAS and FLI.[3] While fissile material can be used for both energy and weapons development, AI systems have a greater potential to be jailbroken and misused for malicious or unintended purposes by a wider variety of actors, even if safeguards are in place.

FINDING 2. RISK VARIES WIDELY DEPENDING ON HOW AI IS INTEGRATED INTO THE NUCLEAR ENTERPRISE

In conjunction with understanding AI risk through the lens of historical nuclear weapons development, another central question during the roundtable discussion was how AI may simultaneously exacerbate and mitigate pre-existing nuclear risk. Participants noted a trend within public policy discourse towards 'pro-AI' or 'anti-AI' viewpoints, arguing that this binary framework is not a useful heuristic. Since AI is already integrated into components of NC3 and the nuclear enterprise, they emphasized it would be more useful to discuss how AI could increase or decrease nuclear weapons risks, especially when used for decision-making support. In addition, participants noted that redundant and risk-reduction systems need to be in place for areas where AI use may exacerbate nuclear risks.

Participants broadly categorized eight ways in which AI can be integrated into nuclear systems:

- **Research and Development (R&D).** AI systems may be used in automating portions of nuclear weapons research, particularly in areas with a large amount of accumulated data, such as simulations of nuclear explosions, aging, and safety.
- **Situation Monitoring.** AI systems are used for ISR purposes to collect, analyze, and integrate data from various sensor platforms across disparate environments, in line with DoD's concept of network-centric warfare. Further, this capability includes early warning of potential nuclear launches.
- **Decision-Making.** AI systems may be used to support conferencing with military leadership and provide advice for the President or other levels of command.
- **Planning.** AI may be used to draft and audit operational plans and targeting procedures, integrating knowledge from disparate departments and adapting rapidly to ongoing conflict.
- **Communications.** AI systems can handle things like automated message routing or tasking orders, as well as significantly changing or strengthening comms pathways dynamically, and be integral.
- **Diagnostics and Maintenance.** AI systems may potentially enable remote and self-guided routine system health monitoring, as well as predictive maintenance for NC3 and nuclear force infrastructure.
- **Force Management.** AI systems may accelerate or offload certain tasks and functionalities for designing, developing, and sustaining organized, trained, and equipped units to meet current and future operational needs. This includes, but is not limited to, training, logistics support, force planning, and deployments.

3 Federation of American Scientists, Artificial Intelligence and Military Integration: Current Status and Future Risks(Washington, DC: Federation of American Scientists, December 2025). https://fas.org/wp-content/uploads/2025/12/1215-ai-mil.pdf. Accessed March 1, 2026.

- **Force Direction.** AI systems may be integrated into NC2 to varying degrees, which provides a spectrum of possibilities. On one extreme end would be a fully AI-enabled automated launch authority, while the opposite end would be something akin to greater comms resilience and signaling (such as building on concepts like CATALINK).[4]

Within these broad categorizations, participants agreed that the risk profiles varied greatly. Participants viewed the integration of AI systems within categories such as research and development may be less risky, while the use of AI for areas such as decision-making support could present extreme risks, as it potentially automates the occurrence of a global catastrophic event. Participants also noted that in many cases where AI-NC3 integration presents risks, there are corresponding use cases that could reduce risk. For example, while utilizing AI systems to integrate sensor data from early-warning sites for nuclear weapons launches may exacerbate the risk of accidental error, AI could be utilized in intelligence, surveillance, and reconnaissance (ISR) to improve the processing and analysis of data.

One area of concern raised by participants concerning AI-NC3 risk was the potential for reduced 'decision latency'. A panelist particularly noted that the latency in responding to potential nuclear escalations can serve a de-escalatory purpose. This panelist noted how diplomatic communications with near and peer competitor countries and other activities done through human channels allow us to move both more slowly and with more constraint than through accelerated decision-making.

This is in direct misalignment with how AI capabilities are discussed in the national security and policy spaces. The U.S. Department of Defense is taking an "AI-First" strategy, which seeks to accelerate AI adoption into the military and connotes the adoption of such capabilities as a "race".[5] Further, AI capabilities may also erode certain countries' capabilities for assured retaliation (or a second-strike capability), which may even accelerate or compel countries to either "use" the weapons they have available, or potentially lose their use as an attacker using AI identifies and destroys previously "hidden" assets.

While not discussed within this February meeting, recent events also highlight some of the complex dynamics at play at the intersection of nuclear weapons, artificial intelligence, and national security. The recent case of Anthropic and the Department of Defense highlights the tensions that can exist when industry and government are not aligned, and should be a lesson learned for all stakeholders in the AI x Nuclear domain.[6]

FINDING 3. THE MOST IMMEDIATE GOVERNANCE GAP IS INSTITUTIONAL: TESTING, TRANSLATION, AND NORM-BUILDING ARE LAGGING BEHIND USE-CASE DEVELOPMENT.

Participants repeatedly observed that policymakers often do not understand the capabilities and limitations of current models in enough detail to write precise rules, while AI developers often do not understand the realities of nuclear operations, command arrangements, or crisis decision-making well enough to anticipate how their systems might behave in that environment. Several speakers described this as a translation problem between communities that do not yet share language, assumptions, or incentives.

4 Institute for Security and Technology. "Reinvigorating Crisis Communications in an Era of Emerging Technologies," event page. February 18, 2026. https://securityandtechnology.org/event/reinvigorating-crisis-communications-in-an-era-of-emerging-technologies/. Accessed March 9, 2026.

5 U.S. Department of Defense. Artificial Intelligence Strategy for the Department of War: Accelerating America's Military AI Dominance (Washington, DC: U.S. Department of Defense, January 12, 2026). https://media.defense.gov/2026/Jan/12/2003855671/-1/-1/0/ARTIFICIAL-INTELLIGENCE-STRATEGY-FOR-THE-DEPARTMENT-OF-WAR.PDF. Accessed March 3, 2026.

6 Hadas Gold. "Anthropic Faces Pentagon Deadline Over AI Safeguards," CNN, February 27, 2026. https://www.cnn.com/2026/02/27/tech/anthropic-pentagon-deadline. Accessed March 3, 2026.

That translation problem is sharpened by the fact that frontier AI development is happening largely outside government. Participants contrasted this with the common Washington habit of invoking the Manhattan Project as the model for AI. Whatever the historical value of that analogy, several discussants argued that it obscures the most important contemporary fact: the U.S. government is not the sole, or even primary, builder of these systems, and it cannot rely on centralized secrecy or direct state control to manage the problem. Public-private engagement is therefore a key condition of governance.

Testing and evaluation of AI systems emerged as a key area where the government could improve its capability. Multiple participants noted that AI models can be "spiky": very good at some tasks, unexpectedly poor at others, and sensitive to context in ways that make generic certification difficult. The science of evaluating frontier AI systems is also nascent, and existing evaluations may not be suitable for the specific applications of AI in national security domains. These issues mean that conventional software procurement and evaluation practices are unlikely to be sufficient. Discussions supported iterative testing, close involvement of engineers and operators, and more emphasis on evaluating the entire human-AI team rather than the model alone.

The governance discussion also had an international and social dimension. Participants were skeptical that a fully formed institutional analogue to the nuclear nonproliferation regime could be created for AI. However, several speakers did say that small areas of agreement, voluntary commitments, responsible communities, allied and partner exchanges, and recurring expert dialogues as plausible starting points for international AI governance. The biotechnology community's debates over proactive restraint were cited as suggestive examples. So too were industry forums intended to create a culture of responsibility, even if participants did not treat them as sufficient by themselves.

Finally, participants stressed that specificity is itself a governance tool. Debate becomes unproductive when "AI" is treated as a single object, when all WMD-related risks are treated as interchangeable, or when analysts slide too quickly from support tools to fully agentic systems. A more useful next step is narrower: identify which systems are being discussed, which nuclear functions they would touch, what kinds of human control are retained, what forms of testing are possible, and where targeted redlines can be drawn now.

MENU OF POLICY OPTIONS

Based on the panel discussions and subsequent analysis, FAS identified the following policy options to address risks and realize benefits at the intersection of AI and nuclear systems. These options are presented for discussion purposes and do not imply endorsement by any individual participant.

1. ESTABLISH EXPLICIT GUARDRAILS AROUND AI IN NUCLEAR DECISION-MAKING AND FORCE EMPLOYMENT.

The United States should continue to emphasize that AI may not exercise nuclear launch authority or unbounded autonomous force-employment authority, and should define tightly bounded roles for decision-support tools used in nuclear settings. These guardrails should preserve meaningful human judgment, create auditable lines of accountability, and avoid system designs that automatically compress deliberation in crises. Congress could reinforce these guardrails through oversight, reporting requirements, and defense authorization language focused on especially sensitive functions.

2. BUILD A BESPOKE TESTING, EVALUATION, AND ASSURANCE REGIME FOR AI USED IN NUCLEAR-RELATED APPLICATIONS.

Participants repeatedly emphasized that nuclear-related AI cannot be treated like generic enterprise software. Evaluation should therefore be use-case specific and should test reliability, false alarms, adversarial manipulation, cybersecurity, model drift, and the performance of the full human-AI team in realistic conditions. Because many relevant deployments are likely to be specialized, classified, or air-gapped, the government may need new pathways for secure collaboration among operators, evaluators, developers, and national laboratories. Congress and the executive branch should also avoid assuming that generic AI assurance frameworks will automatically translate into nuclear contexts; the threshold for reliability, auditability, and adversarial resilience is likely to be much higher.

3. PRIORITIZE LOWER-RISK AND RISK-REDUCING AI APPLICATIONS WHILE PREVENTING FUNCTION CREEP.

The government should focus near-term experimentation on bounded applications where potential benefits are clearer and the consequences of failure are more manageable—such as maintenance, logistics, cyber vulnerability discovery, anomaly detection, imagery triage, and certain forms of verification support. At the same time, agencies should create explicit firebreaks so that tools introduced for support functions do not gradually migrate toward higher-risk decision functions without new review, new testing, and fresh policy authorization.

4. CREATE STRUCTURED TRANSLATION AND INFORMATION-SHARING CHANNELS ACROSS GOVERNMENT, INDUSTRY, AND THE NUCLEAR COMMUNITY.

The roundtable made clear that neither policymakers nor AI developers can solve these problems alone. The government, therefore, needs trusted mechanisms for exchanging information with frontier AI developers, military operators, national laboratories, and outside experts. These mechanisms could support incident reporting, technical briefings, joint red-teaming, recurring track 1.5 and track 2 dialogues, and more routine exchanges between the policy community and people who actually build or deploy advanced systems, while also helping translate model behavior into policy-relevant language. Allied and partner exchanges on definitions, testing practices, and confidence-building measures should be part of this effort from the start.

5. PURSUE INCREMENTAL NORM-BUILDING AND CONFIDENCE-BUILDING RATHER THAN WAITING FOR A SINGLE COMPREHENSIVE INSTITUTION.

Participants did not identify a ready-made AI equivalent to the nuclear nonproliferation regime, nor did they think a one-step institutional transplant was realistic. But narrower forms of norm-building are still available. Governments, laboratories, and companies could begin by articulating redlines around especially dangerous applications; sharing best practices on evaluation and safeguards; and developing confidence-building measures related to terminology, transparency, and crisis stability. The practical goal should be layered governance that can grow over time, not the fiction that one institution can solve the problem all at once. Participants' comments suggested that this work will likely proceed through a mix of formal policy, professional norms, voluntary commitments, and repeated interaction among communities that do not yet share a common operating picture.

CONCLUSION

The roundtable did not yield a single thesis about whether AI will make the nuclear domain safer or more dangerous, and instead argued for a more disciplined way of thinking: specify the function, specify the system, specify the human role, and then judge the risks.

That discipline matters because the discussion revealed two temptations that are likely to distort policy if left unchecked. One is the temptation to treat AI as a replay of the nuclear age and assume that a single historical analogy gives us the right institutional answer. The other is the temptation to talk about "AI and nuclear" as though it were a single policy problem. The roundtable suggested that neither approach is adequate. What matters are the concrete ways AI may intersect with design, warning, planning, cyber defense, logistics, verification, and decision-making—and the distinct risks and opportunities attached to each.

The most consistent note of caution concerned speed. Participants did see cases in which AI might improve monitoring, warning, or analysis. But they also warned that faster systems can make institutions more brittle if they erode time for judgment, communication, and correction. In nuclear contexts, preserving the option of restraint is itself a policy objective.

Taken together, the discussion points toward a near-term agenda: draw clearer redlines around the most dangerous uses, invest in government capacity for testing and evaluation for the applications that are being pursued, build better translation between technical and policy communities, and pursue incremental norms and confidence-building where stronger institutions do not yet exist. While these are not final answers, they are the kind of no-regrets steps that can improve governance before more consequential deployments force the question under worse conditions.

PRE-READ PAPERS FROM ROUNDTABLE

PANEL 1 PRE-READ: WHAT TO LEARN—AND NOT LEARN—FROM THE NUCLEAR AGE: A FRAMEWORK FOR TWENTY-FIRST CENTURY AI THINKERS

ANKIT PANDA, CARNEGIE ENDOWMENT FOR INTERNATIONAL PEACE[7]

The spread of increasingly sophisticated narrow and general-purpose artificial intelligence (AI) technologies—and the possible arrival of artificial superintelligence (ASI)—may end up having catastrophic or existential consequences for humanity.[8] If one believes in this premise, there is a natural and understandable tendency to look for precedents in how humanity managed other periods of sharp technological discontinuity and change. The dawn of the nuclear age, roughly eighty years ago, presents one such example and is often evoked in contemporary debates on managing the risks associated with AI systems.[9]

AI and nuclear weapons, of course, are fundamentally different things in their essential nature and effects. Nuclear weapons, dubbed the "absolute weapon" by the American naval strategist Dr. Bernard Brodie,[10] were fundamentally forms of explosive ordnance hitherto unimaginable that functioned best as political tools. Their arrival transformed the endeavor of military organizations fundamentally; as Dr. Brodie famously put it, nuclear weapons meant that the "chief purpose" of military establishments would no longer be fighting to win wars, but to "avert them," with "no other useful purpose." The fundamental innovation of the bomb, in the mid-1940s, was the packaging of previously inconceivable amounts of explosive power into devices that could be delivered by a single aircraft. Whereas on March 9 and 10, 1945, more than 300 American B-29 bombers had been called to firebomb Tokyo, resulting in more than 100,000 dead,[11] three months later, on August 6, a lone B-29 armed with a single nuclear weapon obliterated a city, Hiroshima.[12] Twenty years later, megaton-class thermonuclear weapons, paired with intercontinental-range ballistic missiles, had further changed the picture.

Despite Dr. Brodie's diagnosis of the effects of the bomb in 1946 as necessitating the end of militaries postured to fight and win wars, the reality of the nuclear age has been more complicated: nuclear-armed states continue to ask much of their military establishments beyond the mere requirements of nuclear deterrence. Humanity's co-existence with the bomb has also been rendered somewhat more tractable by several mutually reinforcing structures beyond just nuclear deterrence; these include a nonproliferation regime, a system of international verification for the non-diversion of weapons-usable fissile materials, and a network of alliances led by the United States. Negotiated forms of restraint between nuclear-armed states (arms control) have also helped.

7 Ankit Panda (ankit.panda@ceip.org) is the Stanton senior fellow in the Nuclear Policy Program at the Carnegie Endowment for International Peace and the author of 'The New Nuclear Age: At the Precipice of Armageddon' (Cambridge, UK: Polity, 2025). His current research interests include the nexus between nuclear weapons systems and artificial intelligence technologies.

8 For reasons of space, these terms are defined herein. Narrow AI concerns machine-learning systems designed to execute on well-defined tasks; general-purpose AI is defined as systems more capable of open-ended reasoning and multipart, complex tasks to a hypothetical system that is capable of far exceeding human intelligence in every plausible cognitive domain.

9 Rehman, Iskander. "An Algorithmic Loosening of the Atomic Screw? Artificial Intelligence and Nuclear Deterrence," (West Point, NY: Modern War Institute), November 11, 2025. https://mwi.westpoint.edu/an-algorithmic-loosening-of-the-atomic-screw-artificial-intelligence-and-nuclear-deterrence/ See ibid., for an extended treatment of nuclear age analogies for AI. Some of the ideas in this short essay draw inspiration from Rehman's observations.

10 Brodie, Bernard. "The Absolute Weapon: Atomic Power and World Order" (Manchester, NH: Ayer Company Publishers, 1946).

11 Rauch, Jonathan. "Firebombs Over Tokyo", (then Boston, MA: the Atlantic, 15 August 2002). https://www.theatlantic.com/magazine/archive/2002/07/firebombs-over-tokyo/302547/

12 "Hiroshima and Nagasaki Bombings", (Geneva, CH: International Campaign to Abolish Nuclear Weapons, 2019). https://www.icanw.org/hiroshima_and_nagasaki_bombings

Below, I identify and briefly discuss observations from the nuclear age that can provide some utility for framing contemporary debates on AI and human survival. There are, naturally, important limits to the analogies that we can draw, which I also allude to below.

Is there a "secret" of the bomb? Is there a "secret" of AI?

Prior to the first nuclear test in July 1945 (Trinity), it was already well-understood by the community of scientists and engineers of the Manhattan Project that there was no essential "secret" of the bomb that, if well-kept, could prevent the spread of nuclear weaponry. The physical principles allowing for nuclear weapons were increasingly part of a body of nuclear physics that would be more widely understood around the world. While elements of nuclear weapons design (including the precise means of machining explosive "pits" and non-nuclear explosive components) were complex, there was no particular reason to believe that determined actors—especially those with the resources of a nation-state—could not succeed in this endeavor. With AI systems—including contemporary large-language models and other transformer-based neural networks—the case is similar in some ways, and dissimilar in others.

What's the most relevant unit of analysis for preventing proliferation? (Or: "fissile material is not compute".)

Mother Nature has been somewhat kind to humanity in ensuring that just two isotopes of two elements—one that does not occur naturally—are suitable for fueling the fissile cores of nuclear weapons. Controlling the spread of nuclear weapons, thus, has focused less on holding tight the one "secret" of the bomb, and more on ensuring that materials suitable for nuclear weapons are not amassed in significant quantities by a large number of states. The modern nonproliferation verification architecture, for instance, relies on accounting for all relevant nuclear material within a non-nuclear state's borders and verifies that no material has been diverted to unknown uses, including a possible covert nuclear weapon program.

In the context of AI, while compute is sometimes likened to weapons-usable fissile material, the analogy has serious limitations. While it will be difficult to be systematic due to space limitations, consider that, for starters, the physical principles involved in nuclear weapons allow for the delineation (mostly a priori) of what the International Atomic Energy Agency, for instance, considers to be "significant quantities" of enriched uranium or plutonium. The equivalent in compute for meaningfully dangerous AI systems is a much less tractable problem. Beyond compute, data quality, training methods, alignment, and systems architecture matter considerably. Moreover, the centrality of non-state actors (private firms) in the development, manufacture, and distribution of compute globally represents another meaningful difference; nuclear material within weapons programs has invariably been monopolized by states.

Personnel reliability matters.

As AI systems grow more powerful and particularly as they find applications in military organizations around the world, it will be increasingly important to have in place effective procedures to ensure that the human beings charged with the operation, maintenance, and surveillance of those systems are reliable and capable. In the United States, a so-called Personnel Reliability Program (PRP) exists to vet individuals involved in the nuclear deterrent mission. While PRP programs have experienced lapses, the general principle is a sound one for AI integration in sensitive settings. Numerous publicly reported cases of human-AI interactions leading to behavioral changes in the human beings that interact with these systems suggests that more attention should be given to this matter. Within the nuclear weapons enterprise, too, PRP will need to expand in scope to manage the diffusion of AI systems into nuclear command and control systems, too.

When you know, you know—or do you know?

Certain applications of AI technologies are already revolutionizing human affairs, but detecting and responding to the potential arrival of fundamentally transformative general intelligence or superintelligence is far from a straightforward matter. The atomic bombing of Hiroshima, followed by a U.S. statement on the bomb, notified the

world that atomic weaponry—bombs "harnessing the basic power of the universe"—had arrived.[13] Later, during the United States' brief period of nuclear monopoly (1945-1949), intelligence assessments diverged considerably about the timeline for the Soviet procurement of the bomb. When the Soviet Union did carry out its first nuclear test, in August 1949, special American aircraft detected atmospheric radionuclides, but even then, analytical disagreements persisted on whether those samples indicated a weapons test or some sort of other man-made radiological incident. Prominently, key policymakers, including U.S. President Harry Truman, appeared to maintain disbelief on the spread of the bomb well past 1949.[14] When an AI breakthrough comparable in its effects (possibly AGI or ASI) first manifests—be it in the United States, China, or elsewhere—the diffusion of knowledge about its existence may not be immediate.

Will diffusion of effects take longer?

We should consider that AI may not have its proverbial "Hiroshima moment." Instead, as the 2020s already attest, AI technologies will rapidly and iteratively improve and, in parallel, civilian and military organizations will proceed with integration. Here we must consider one of the strongest reasons to not over-learn the lessons from particularly the start of the nuclear age. In this scenario, the diffusion of AI technologies would come to resemble that of electricity, aviation, and even networked computing rather than nuclear weapons—with the effects shaping geopolitical dynamics and warfare over a more protracted period.[15] This does not preclude that AI could have massively disruptive effects on human economic life and labor markets, of course.

The bomb matters, but so do organizations (or, why it's about more than alignment).

Some of the most famous dangerous moments of the nuclear age, to include the Cuban Missile Crisis, were not driven fundamentally by the existence of the bomb itself, but by a cocktail of misperception, human psychology, poor communication, misaligned incentives, and organizational pathologies. In the context of AI, this should prompt reflection beyond merely technical safeguards—for instance, on engineering human-aligned systems—and into the organizational safety culture associated with some of the most high-impact possible deployments of advanced AI systems. Incentives to rush deployment, cut corners on certification and testing, and broader arms-race-style motivations will render much of this difficult.

Is there a technological "grand bargain" to be had?

The Treaty on the Nonproliferation of Nuclear Weapons (or the NPT) remains a crowning achievement of the nuclear age. Its success in staunching the spread of nuclear weapons owes much to the essential tripartite bargain at its core, split across three pillars. First, states without nuclear weapons join the treaty and forswear pursuing those weapons perpetually (nonproliferation). Second, as a result of their forbearance, they receive access to peaceful nuclear technologies for their economic benefit (peaceful uses) and submit to allow for verification that their programs indeed remain peaceful. Third, to address the insecurity spurred by the presence of some nuclear-armed states (defined by the treaty as any state to have detonated a nuclear explosive before January 1, 1967), the nuclear-armed states agree to work toward nuclear disarmament.[16]

This bargain is under contemporary stress, but its history should remind us that a delicate balancing of incentives—among wealthy powerful, nuclear-possessing great powers, their allies, and resource-poor non-aligned states—can manifest powerful governance effects. Similar AI proposals—for instance, on capping compute through a treaty

13 Truman, Harry S. "Statement by the President Announcing the Use of the A-Bomb at Hiroshima" (Speech, USS Augusta). August 6, 1945. Harry S. Truman Presidental Library. https://millercenter.org/the-presidency/presidential-speeches/august-6-1945-statement-president-announcing-use-bomb

14 Wellerstein, Alex. "The Most Awful Responsibility: Truman and the Secret Struggle for Control of the Atomic Age", (New York City, NY: Harper Publishing, 2025). q.v. Chapter 15.

15 Ding, Jeffrey. "Technology and the Rise of Great Powers: How Diffusion Shapes Economic Competition", (Princeton, NJ: Princeton University Press, 2024).

16 Sokolski, Henry, et al. "Fighting Proliferation", (Montgomery, AL: Air University Press, January 1996). See Chapter 2: Weiss, Leonard. "The Nuclear Nonproliferation Treaty: Strengths and Gaps". https://irp.fas.org/threat/fp/index.html

mechanism[17]—should look at this history, but finding mutually reinforcing analogous "pillars" that allow for a bargain between the haves, the near-havers, and the have-nots will remain a tall task.

This short essay could continue on to considerable length, but for reasons of space, I'll leave additional observations for another setting. The list above is far from exhaustive, but represents what I consider to be a hopefully somewhat useful starting point for contemplation of the lessons and limits of using humanity's experience so far in the eighty-one years of the nuclear age to understand the coming age of AI. There's an obvious caveat to much of this, which is that what exactly the lessons of the nuclear age are continues to be strongly debated, with decades-long debates around key questions pertaining to the nature of nuclear deterrence, the consequences of proliferation, the controllability of escalation, and more. As we progress to a potentially far messier, more complex era for AI technologies becoming integrated with all facets of human endeavor, we should anticipate that today's burgeoning debates about safety, alignment, and integration will persist and evolve.

PANEL 2 PRE-READ: ARTIFICIAL INTELLIGENCE AND NUCLEAR WEAPONS

AUSTIN LONG, PHD, MIT CENTER FOR NUCLEAR SECURITY POLICY[18]

Artificial Intelligence and Nuclear Risk

The nexus of artificial intelligence (AI) and nuclear weapons has produced substantial angst in both technical and policy communities. Given the pop culture priors of the past several decades—from *Colossus: The Forbin Project* (1969) through *WarGames* (1983) to *Terminator* (1984)—this is unsurprising. Yet these popular fictions are also unhelpful for real-world understanding. AI is not synonymous with SkyNet (*Terminator*) or the WOPR (*WarGames*) of films. Nor is it unreasonable to believe that AI will or should be insulated from nuclear weapons - from design to command and control. The intersection of AI and nuclear weapons has risks. This essay provides some initial framing for how to think about more or less risky intersections of AI and nuclear weapons.

AI and Nuclear Weapon Design

The nuclear weapon and AI intersection begins with nuclear weapon design. As with other bespoke technical design areas such as pharmaceuticals, AI can potentially accelerate design cycles for nuclear weapons. Such advancement requires substantial data on nuclear weapons performance, and the US has the world's most expansive empirical record for such data from its nuclear testing and long-standing science-based stockpile stewardship programs. AI, thus, could yield substantial benefits for US nuclear weapons designers. Russia, the UK, and France might also gain design advantages from this intersection. In a sense, then, the rich will get richer from this interaction- the largest/oldest nuclear powers with access to advanced nuclear data will reap the most benefit. Introducing AI into nuclear weapons design is thus a logical extension of the United States' decades-old scientific stockpile stewardship effort, which combined high-performance computing and simulation with the data from decades of nuclear testing.

AI is thus unlikely, absent such data and modeling investment, to shift any of the underlying dynamics of vertical or horizontal proliferation. One possible exception is the People's Republic of China, which has the resources to invest in modeling and computing but lacks a test base comparable to the United States or Russia. If it were able to gain such data—or chose to resume full-scale nuclear testing—it could potentially advance its nuclear designs very rapidly compared to the progress of the United States or Soviet Union decades ago.

17 Miotti, Andrea. "An International Treaty to Implement a Global Compute Cap for Advanced Artificial Intelligence", (Preprint, ArXiv). November 1, 2023. https://doi.org/10.48550/arXiv.2311.10748

18 Austin Long, PhD, is the Senior Nuclear Fellow at the Center for Nuclear Security Policy at MIT. Previously, he was the Deputy Director for Strategic Stability at the Joint Staff, a senior political scientist at the RAND Corporation, and an associate professor at Columbia University. He is the author of multiple books, including 'The Soul of Armies: Counterinsurgency Doctrine and Military Culture in the United States and United Kingdom' (Ithaca, NY: Cornell University Press, 2016).

AI and Nuclear Command, Control, and Communications

AI has multiple points of potential intersection with the command and control of nuclear weapons[19]. This creates a spectrum of risk and benefit, from intersections of high risk and low benefit to those with low risk and high benefit. Each intersection has different implications in terms of risks and benefits, so it is useful to dissect nuclear command, control, and communications into different components or functions. Typically, nuclear command and control is defined as consisting of five functions: situation monitoring, decision-making, planning, force management, and force direction.

These functions can then be mapped to the risk/benefit spectrum. At one end of the spectrum are low-risk/high-benefit applications, where the inclusion of AI is likely to improve performance with little to no risk relative to existing command and control processes. At the other end are high-risk applications, where, regardless of benefits, there are substantial risks over existing processes. Risk reduction approaches seek to appropriately calibrate risks of AI and nuclear command and control intersections and then limit or eschew those that are at the risky end of the spectrum while encouraging appropriate adoption of those at the other end of the spectrum.

Situation monitoring encapsulates a spectrum ranging from early warning to assessing general geopolitical developments. AI offers potential benefits here. One of the things that AI does well is the rapid assimilation and assessment of disparate inputs, for which there is often extensive historical data.

For example, the U.S. early warning of ballistic missile attack relies on both infrared and radar data, which are already processed with a high degree of automation. AI offers some benefit to speeding up the assessment of potential attacks by fusing sensor data with historical data (e.g., from observation of previous missile launches) as well as the potential for the introduction of novel assessment tools. An example is the rapid inclusion of intelligence data on adversary nuclear readiness levels and/or patrol patterns. While there is a risk of an AI false alarm, it is likely modest compared to the risk of a false alarm with the existing semi-automated assessment.

AI as Decision-Making Partner

The intersection of AI and nuclear decision-making causes the most immediate, visceral reaction in those who appreciate the gravity of these decisions. Images of AI-directed nuclear war haunt the modern consciousness. Yet decision-making tools, or more accurately, aids to decision-making, have existed since the beginning of the nuclear age. In the United States, the most notable is the decidedly analog Nuclear Decision Handbook (NDHB, aka the Black Book), a binder with reference material to support time-critical nuclear decisions. If the same information contained in the Black Book were presented digitally (e.g., on a tablet), would it fundamentally change decision-making? What if the same information were made more visual in three-dimensional displays, rather than diagrams and words on a page? What if it were the same information but more interactive- with hyperlinks to additional existing databases, for example? A President can already ask for any information he or she wants in decision-making, so would making information retrieval faster and without a human staff officer providing it change a decision? If so, for better or worse?

Going one step further, what if the digital interface provided information that a President's querying of databases suggests he or she is actually seeking, but is unsure how to ask for? Would this change decision-making? A further step would be the digital interface recommending courses of action with pros and cons, just as human advisors can (but more rapidly), and based on a pre-determined and defined set of objective parameters (as opposed to subjective reasoning). Each of these incremental steps away from a purely analog two-dimensional decision-aid to an interactive AI advisor is potentially reasonable, but merits considerable caution given the early stage of understanding of the interaction of human judgment with advanced AI. Each step probably shifts further towards the risky end along the risk/benefit spectrum.

19 FAS convened an initial roundtable on the AI/Global Risk intersection in February 2025 on AI and Nuclear Command, Control, and Communications. The report from that event can be found at https://fas.org/wp-content/uploads/2025/07/June2025_AIxNC3_FAS.pdf.

AI-only Decision-Making

Beyond decision aids, there is the possibility for purely AI decision-making. While nations such as China and the United States have rejected this approach, it could be attractive to states with concerns about the reliability of their strategic capabilities and their ability to maintain a means for assured retaliation. In states with clear leadership succession and trusted civil-military relations, such concerns can be addressed with delegation to lower echelon commanders or by ensuring clear succession and continuity of civilian leadership. The United States has pursued both courses at various points.

However, the Soviet Union—which lacked both clear succession and trust between civilians and the military—did not pursue either course. Instead, it built an algorithmic form of pre-delegation, a semi-automated command and control system known as *Perimetr*. When the system was activated and certain conditions were met, the system would enable nuclear launch without further leadership action.

This system highlights the potential attraction for certain regimes of AI decision-making. A leader concerned about assuring nuclear retaliation and unwilling to delegate or identify a successor could see AI as a convenient opportunity. Who better to succeed a leader who is killed or incapacitated than an AI simulacrum of the leader's own nuclear decision-making? Whether such a simulacrum is possible, and if it is, how faithfully it could recreate a leader's decision-making is an open question. However, one can imagine why a leader might invest in exploring it, but also see how this would be at the extreme end of the risk/reward spectrum.

Less Risky: AI for Planning

The planning function of nuclear command and control probably sits at the less risky side of the risk/benefit spectrum, with different aspects of planning somewhat more or less risky. As an example, much of the initial processing of geospatial intelligence (e.g., satellite images) by the U.S. National Geospatial Intelligence Agency is done with the aid of machine learning tools. The volume of images now available simply swamps humans, so AI/ML-human teaming makes the volume tractable and useful. These images are the basis for much U.S. military targeting (nuclear or non-nuclear), so the targeting component of planning already has an element of AI embedded in it. This, to date, has not produced risky results and may reduce risks of the kind that led to the inadvertent bombing of the Chinese embassy in Belgrade in 1999 during operations against Yugoslavia.

Adaptive Planning

Other aspects of planning that often already rely on modeling tools, such as those for planning bomber routes, could likely benefit from similar AI-human teaming. This would be particularly useful for so-called adaptive planning, which is intended to enable the generation of nuclear options that are not currently in existing deliberate or "on the shelf" plans for the U.S. president in crisis or conflict. Of course, rapid generation of options could lead to challenges in other aspects of nuclear command and control—such as decision-making- if other functions of nuclear command and control are not well integrated with AI-human enabled planning. This underscores the need to evaluate AI incorporation into nuclear command and control holistically across the five functions.

Lower Risk: AI in Force Management

Force management, which involves all the prosaic but vital elements of logistics, maintenance, and readiness, is perhaps the function where the utility of AI is most likely to be evident at low risk. This is in part because many of the elements of force management most closely resemble civilian or other military applications of AI: large, geographically dispersed logistics functions, for example. The detection of patterns for predictive maintenance in nuclear systems and ensuring the routine updating of cryptographic keys and software are just two areas where agentic AI, in particular, could prove valuable in force management. Here, the risks of incorporating AI again seem modest or negligible compared to the status quo. AI may even help limit future errors, such as those the U.S. Air Force saw in handling nuclear weapons and components in 2007.

Force direction and the issuing of commands to the force to employ or terminate employment of nuclear weapons offer a few different potential opportunities for the incorporation of AI, with a range of risk profiles. At the lower end

of risk is the use of AI to manage the communications systems and connections required to disseminate orders to nuclear forces. In the United States, these orders are called Emergency Action Messages (EAMs) and, depending on the nuclear platform, can be delivered via landline or a variety of radio frequencies. In environments where an adversary nuclear or non-nuclear attack may have disrupted communications, AI could be useful in ensuring reconstitution of pathways for message transition. Indeed, early packet switching techniques—which now form the basis for internet communications were developed for such environments, so AI for similar effects may be a logical progression.

AI for Real-Time Feedback

Another AI application would be giving it a role in optimizing force direction to achieve presidential objectives. This would enable AI to optimize in near real time how weapons would be applied to targets based on the current real-world conditions. For example, if a missile failed at launch, AI could rapidly determine whether another missile was available to cover the same target, determine whether utilizing that missile would leave another target uncovered, and assess the utility of employing the second missile. This could improve both the efficiency and effectiveness of force direction, which in turn might permit a state to deploy fewer nuclear weapons. However, giving AI autonomy to redesign attack structures in real time—even if the AI must be given authorization to do so in the same way a human operator would through a presidential authentication—would be at the risky end of the spectrum of AI interactions. Here, the fictitious shadow of Skynet and the WOPR looms.

Risk Reduction

Mapping out the level of risk/benefit for different interactions of AI and nuclear weapons is a prerequisite for considering ways to appropriately reduce risks. Agreements on ensuring a human always remains central to decision-making are important, but this is the lowest-hanging fruit. Future efforts will need much more finely calibrated considerations of issues such as verification and transparency, always a sensitive issue with nuclear command and control and weapons design.

WANT TO LEARN MORE ABOUT OUR AI X NEXUS SERIES?

Please visit fas.org to learn more about upcoming events publications, and Global Summit 2026.

FAS FEDERATION OF AMERICAN SCIENTISTS

ABOUT THE FEDERATION OF AMERICAN SCIENTISTS

The Federation of American Scientists is dedicated to democratizing the policymaking process by working with new and expert voices across the science and technology community, helping to develop actionable policies that can improve the lives of all Americans. For more about the Federation of American Scientists, visit **FAS.org.**