Nuclear Security in the Age of AI

A White Paper on Risks and Opportunities

Peter H. Vartanian

Independent Researcher

Q4 2025

Contents

I. Executive Summary

1. Macrotrends

> Nuclear non-use is not an inherently permanent achievement. It is a condition that must be secured by each successive generation of leaders adjusting ... to a technology evolving at unprecedented speed.[a]

[a]Henry A. Kissinger, Eric Schmidt, and Daniel Huttenlocher, *The Age of AI: And Our Human Future* (New York: Little, Brown and Company, 2021), 89.

Rapid advancements in artificial intelligence (AI), robotics, big data, ubiquitous sensing, space commercialization, and other technologies are re-configuring the technological and strategic foundations of nuclear security.[1] These technologies are progressively interacting in complex ways, giving rise to new systems and capabilities that are disrupting traditional assumptions, systems, and practices for sustaining nuclear deterrence, yet their security-based externalities are often occluded within governments. Unlike previous technological breakthroughs that shaped nuclear security such as atomic energy, digital computing, and space flight—which were driven by governments—today's technological revolution is dominated by the private sector, whose political economy and governance are outside of, and distinct from, governments.

This report:

(i) explores how these technologies are reshaping nuclear security;
(ii) limns the key technological trends driven by the private sector that are generating new nuclear-security risks;
(iii) isolates novel opportunities for reducing nuclear risks; and
(iv) recommends steps for strengthening security.

The findings are grounded in research and in-depth interviews with prominent experts in nuclear security and emerging technologies.[2]

The immense destructive power of nuclear weapons led to the development of a complex and interconnected system during the Cold War, which was imagined, designed, and built by governments over decades. The system encompasses nuclear-deterrence strategies, arms-control agreements, nuclear-weapons policies and postures, nonproliferation measures to halt the spread of nuclear weapons and technology, efforts to secure global nuclear materials and weapons, restrictions on access to nuclear technology, and efforts to approach nuclear disarmament.

Today, the nuclear-security system is progressively entangled with emerging technologies developed in the private sector, not by governments. However, neither the private sector nor governments are working toward a comprehensive understanding of the scale and nature of the technological revolution for nuclear security, nor what tools are necessary to ensure that the interlocks designed to prevent the use of nuclear weapons remain effective in the future. Nuclear strategy remains stuck in

[1]"Emerging" technology (or, technologies) denotes those which (i) are rapidly advancing in capability, diffusion, or autonomy; (ii) compress decision time or expand operational reach; and (iii) can perturb nuclear deterrence, escalation control, or command-and-control by altering sensing, attribution, targeting, or human judgment.

[2]Warren M. Stern, "Revolution in Nuclear Detection Affairs," *AIP Conference Proceedings* 1596 (2014): 219–225, esp. 220; PDF, https://web.archive.org/web/https://pubs.aip.org/aip/acp/article-pdf/1596/1/219/11651607/219_1_online.pdf.

decades-old thinking, and current debates within the U.S. Government largely center on replacing legacy nuclear-weapons platforms, while neglecting the need to adapt nuclear-deterrence architecture, strategy, and force posture to these transformational shifts.

Some world leaders express appropriate caution not to cede command-and-control authority over nuclear weapons to AI and to ensure there will always be a "human in the loop" for nuclear-weapons command decisions. Some prioritize preventing AI from revealing "nuclear secrets" that might give rogue states or terrorists access to nuclear weapons. However, relatively little attention has been paid to the possibility that the convergence of emerging technologies may pose new risks to nuclear security by challenging the systems and approaches that nuclear-armed states use to prevent nuclear war and to "control" nuclear weapons.

Taxonomizing nuclear-security risks and opportunities emerging from the private sector and commercial innovation is a herculean undertaking. The ideas shaping the world of technological innovation evolve rapidly and do not typically overlap with those that have traditionally shaped nuclear security. Adding to the challenge is that these technologies do not operate in isolation: they interact with, and amplify, each other's impact in unpredictable ways. To understand the implications of today's technological changes requires looking not just at AI in isolation, but at how AI is converging with other emerging technologies to potentially transform the entire nuclear-weapons enterprise.

Three trends emerging from commercial innovation are particularly significant for nuclear security:

1. **Radical Transparency**—as the ability to track the location, physical parameters, and status of anything in the world grows, it becomes harder to conceal or protect objects, actions, and information, and easier to reveal them;
2. **Distributed Effects**—with radical transparency, it is increasingly possible to affect people or objects regardless of their number or physical location; and
3. **Changes to Human Control**—machines can now process data at immense volumes and speeds beyond human capacity, changing how much control humans can maintain over outcomes.

These trends create significant risks for nuclear security as well as opportunities. Whether governments will be successful in mitigating risks and reaping the benefits of the technological revolution depends on whether there is leadership that can move away from traditional approaches and thinking about nuclear security.

2. Risks and Opportunities

The nuclear-security risk identified most frequently by the experts interviewed for this study is the growing vulnerability of nuclear forces in an evolving technological landscape. Increasingly sophisticated data collection could expose sensitive information about nuclear arsenals, while adversaries could manipulate the informational environment by deliberately introducing false data. Together, these constitute a breadth of novel forms of attack difficult to fully anticipate and with unknown implications for nuclear deterrence. One particularly concerning scenario is the erosion of confidence in nuclear-weapons systems or the organizations responsible for them, which could fundamentally undermine strategic stability.

Other risks that surfaced include the creation of new pathways to inadvertent nuclear-weapons use via conventional weapons, such as hypersonic weapons, and the rapidly increasing public-private

entanglement in space-based assets. Interviewees also shared their concerns that the convergence of emerging technologies could lower technological barriers to nuclear proliferation and nuclear terrorism.

Interviewees identified opportunities for nuclear security emerging from the convergence of emerging technologies. The most frequently cited opportunity is the ability for states to enhance the reliability of warnings of attacks on nuclear forces. By reducing uncertainty during crises, minimizing false warnings of nuclear attack, and creating more opportunities for de-escalation, technologies could lower the risk of nuclear catastrophe. Other opportunities include the potential for governments to shape a new form of deterrence resilience, making nuclear forces, enabling systems, and human decision makers and operators better able to resist novel forms of attack or degradation. Finally, interviewees noted that emergent technologies could enable a more robust approach to nuclear arms control, nonproliferation, threat reduction, and nuclear disarmament, ensuring that the best available technology is applied to preventing the spread and the use of nuclear weapons.

To mitigate these risks, reap the benefits of the technological revolution, and ultimately avoid nuclear weapons' use, it is critical that governments adapt their nuclear systems to a new security landscape—one defined by rapid technological progress.

3. Recommendations

The analysis underscores the urgent need for governments to have a road-map of actions to address the nuclear-security effects of emerging technologies. They will need a broader innovation agenda focused on adapting their strategies for preventing nuclear catastrophes to fit the converging landscape of emerging technologies and nuclear security. Rather than merely proceeding with a modernization program of one-for-one replacement of legacy nuclear-weapons platforms, re-centering cutting-edge science and technology in the nuclear-security enterprise will allow for greater deterrence resilience to avoid situations in which governments must choose between capitulation or the sheer extancy of nuclear stockpiles, as well as new pathways to use technology-enabled cooperation to prevent nuclear deterrence instability, arms racing, proliferation, and terrorism.

The U.S. Congress should establish a Task Force on Artificial Intelligence, Strategic Stability, and Nuclear Risk—modeled on the National Security Commission on AI—charged with producing a comprehensive report with recommendations for nuclear-security technology development, diplomatic engagement, and public-private partnerships.

Any effective governance of AI and emerging technologies will require ambitious collaborative initiatives between governments and leading technology innovators in the private sector. These initiatives will need to incorporate new kinds of expertise and anticipate emerging capabilities on the horizon, which are likely to result from interactions between novel technological advances. The U.S. Government should create structures and incentives to fully engage the private sector in nuclear-security innovation. These could include the creation of a Nuclear Security Innovation Unit modeled after the Defense Innovation Unit; the creation of a Nuclear Security Development Agency modeled on the Space Development Agency; and a NucWERX program modeled on AFWERX and SpaceWERX.

II. Context and Method

As noted in the Executive Summary (n. 2), taxonomizing nuclear-security risks and opportunities in a commercial innovation environment is difficult. This section specifies the study's method: it develops a conceptual apparatus that can be shared between nuclear-security and emerging-technology leaders, and it sets out how interview judgments were elicited, normalized, and aggregated. A fully self-contained methodological addendum, including the complete Soufflé program and data-model definitions, appears as Appendix B.

The data and key findings in this report are derived from structured interviews with experts in nuclear security and technology innovation (see n. 2). The interview questions included the following:

- What are the risks arising from the convergence of AI and other emerging technologies that increase the risk of nuclear-weapons use?
- What are the opportunities arising from the convergence of emerging technologies for preventing the use of nuclear weapons?
- In what ways might the intersection of emerging technologies make nuclear-weapons use more or less likely?
- To what extent are key stakeholders considering the potential effects of the intersection of emerging technologies on nuclear force structure, doctrine, and policy?
- What aspects of the nuclear-security implications of the intersection of emerging technologies merit more attention?
  - In the United States?
  - On a global scale?
- What investments should be made, and by whom?
  - Which tools of governance should be used to reduce the likelihood of, and to prevent, nuclear-weapons use in the age of AI?
  - By whom?

The study team conducted follow-up interviews soliciting reactions to preliminary findings from C-suite executives and members of existential-risk organizations' boards, in light of their scientific and technical expertise.

1. Delphi Elicitation and Synthesis

The interview program was operationalized as a two-round Delphi-style elicitation layered atop semi-structured interviews. Round I captured free-form judgments and causal accounts. Round II returned a structured claim set to participants for scoring, disagreement annotation, and pinpointed revision. Scores were recorded on a 1–9 scale (1 = negligible strategic effect; 9 = catastrophic and near-term), with experts permitted to attach conditionality (triggering assumptions) as text strings retained as part of the coded corpus.

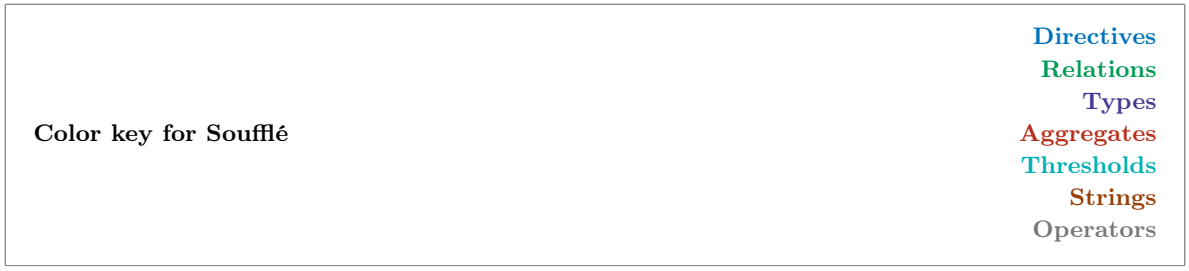
2. Coding, Propositions, and Auditability

All interview-derived statements were normalized into proposition identifiers (P-codes), preserving the speaker's modal qualifiers while enforcing single-claim granularity per code. Each proposition was

then tagged by domain (NC3, ISR, cyber, human layer, space, hypersonics, enrichment, verification, and related domains) and mapped to the risk-opportunity taxonomy frozen in the Table of Contents. This yields an auditable edge list: (a) expert → (b) proposition → (c) risk or opportunity node.

3. Formal Aggregation via Soufflé

To prevent narrative drift and to guarantee that rankings remain mechanically reproducible, the coded corpus was compiled into a Datalog program (Soufflé) that computes (i) weighted mean scores, (ii) distributional robustness statistics on the discrete 1–9 scale (weighted median and interquartile range), and (iii) tier assignments under a fixed decision rule. The program emits an allocatory matrix from ranked risks to the Recommendations.



Listing 1: Definitions of aggregatory and ranked relations

```
.decl wmean_num(p:symbol, num:float)
wmean_num(p, num) :-
  vote(e,p,2,score),
  expert(e,_,w),
  num =sum w*score.

.decl wmean_den(p:symbol, den:float)
wmean_den(p, den) :-
  vote(e,p,2,_),
  expert(e,_,w),
  den =sum w.

.decl n_votes(p:symbol, n:number)
n_votes(p, n) :-
  vote(_,p,2,_),
  n =count : { vote(_,p,2,_) }.

.decl agg(p:symbol, wmean:float, n:number)
agg(p, wmean, n) :-
  wmean_num(p,num),
  wmean_den(p,den),
  n_votes(p,n),
  den >0.0,
  wmean =num / den.

.decl ranked(node:symbol, risk:symbol, subrisk:symbol, wmean:float)
ranked(node, risk, subrisk, wmean) :-
  proposition(p,node,_),
  risk_node(node,risk,subrisk),
  agg(p,wmean,_).
```

Listing 2: Tiering rule and allocatory emission

```
.decl tier(node:symbol, tier:symbol)
tier(node, "I") :-ranked(node,_,_,wmean), wmean >=7.5.
tier(node, "II") :-ranked(node,_,_,wmean), wmean >=5.5, wmean <7.5.
tier(node, "III") :-ranked(node,_,_,wmean), wmean <5.5.

.decl allocation(node:symbol, rec:number)
allocation(node,1) :-tier(node,"I").
allocation(node,2) :-tier(node,"II").
allocation(node,2) :-tier(node,"III").
```

III. The Problem: Emerging Technology as Mutagen of Nuclear (In-)Security

Neither governments nor the private sector are working toward a comprehensive understanding of the scale and nature of the technological revolution in nuclear security, or what tools are necessary for preventing the use of nuclear weapons in the future. Today the most transformational innovations are driven by the private sector, not by governments. This report seeks to better understand how this change disrupts nuclear security, including the practice of nuclear deterrence, which was developed decades ago around technologies that are no longer cutting-edge.

Over the decades, nuclear security has been sought through two key mechanisms comprising the application of technology and policy:

- Nuclear deterrence, where states threaten to use nuclear weapons to influence the behavior of adversary states.
- Arms control, nonproliferation, and nuclear materials security, which includes tools designed to limit access to nuclear weapons technology to the smallest possible group of states under strict control and safeguards.

Historically, nuclear security has been defined by state technological innovation. The insight that an atomic bomb was possible led the U.S. government to launch a dedicated research and development program of unprecedented ambition, scale, and cost to achieve it before any other country. After World War II, the U.S. and Soviet governments launched a superpower arms race, repeatedly revolutionizing nuclear strategy through innovation. By creating satellites for communications and surveillance, intercontinental ballistic missiles that could reach the other side of the world in minutes, and submarines that could hide nuclear weapons for assured retaliation under the oceans, governments scoped the challenges, defined solutions, and established requirements for the technological architecture of nuclear security.

By contrast, there are currently no significant innovations in nuclear weapons technology publicly foreseen or under development in the United States. Public policy debates center on the redevelopment of legacy capabilities and concepts of operation; the nuclear command, control, and communications (NC3) systems to support these capabilities; the desirability of nuclear-tipped sea-launched cruise missiles; low-yield nuclear weapons for submarine-launched ballistic missiles; and how to maintain the nuclear weapons complex and a reliable supply of plutonium triggers known as pits.

Insufficient attention has been directed to the fact that much of the frontier of nuclear-security innovation lies in commercial innovation, not within legacy systems developed during the Cold War. The technological infrastructure developed by the private sector in response to market opportunities has become unwittingly enmeshed in the nuclear security system.

For instance, private-sector innovation in space launch technology has now matured to serve consumer and commercial purposes, including reusable rocket boosters and additive-manufactured components. Space launch technology has immense implications for the United States' ability to establish, maintain, and regenerate surveillance and communications capabilities central to nuclear security, and potentially for new offensive space-based capabilities. SpaceX's Starlink constellation has grown to several thousand satellites and is planned to grow substantially further.[3] Reducing

[3] *See* Jonathan McDowell, "Starlink Statistics" (Planet4589), accessed January 2026, https://web.archive.org/we

the cost of space launch has transformed the players and pieces of the orbital dimension of nuclear security. The changes shaped by non-traditional network-driven industries will be even greater, allowing for the collection, processing, and exploitation of unprecedented quantities of data, with implications for nuclear intelligence, surveillance, target acquisition, and reconnaissance capabilities.

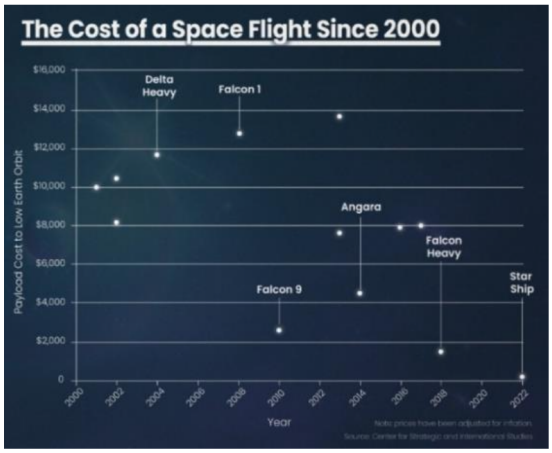


Figure 1: Decreased costs of orbital access will correlate with an increased ubiquity of space-based threats.

Especially as undertaken by *one* state, nuclear-security innovation traditionally drives responsive innovation by other states. This security dilemma has always been inherent to nuclear technological innovation. When one state acquires nuclear weapons technology for defense purposes, it endangers other states, which respond by matching that capability, offsetting it with other technologies, or increasing stockpiles.

Dr. James Johnson, a scholar of nuclear (security) innovation, observes that AI and other emerging commercial technologies are creating an AI-security dilemma, in which predominantly commercial dual-use AI increases the danger of spirals of insecurity and distrust that lead to military preparations and doctrinal shifts, increasing the risk of inadvertent nuclear use.[4] His analysis explores how technological innovation undermines confidence that nuclear deterrence will continue to work as it did in the past, increasing the likelihood of catastrophe.

Decisions taken by the private sector are now changing some of the settings, components, and rules of nuclear security. The U.S. government's nuclear strategy has become rigidly fixed based on a mix of technologies prevalent decades ago. As Henry Kissinger, Eric Schmidt, and Daniel Huttenlocher remind us in *The Age of AI*, "nuclear non-use is not an inherently permanent achievement. It is a condition that must be secured by each successive generation of leaders adjusting . . . to a technology

b/https://planet4589.org/space/con/star/starstats.html; *see also* Federal Communications Commission, "Space Exploration Holdings, LLC," FCC 23-22 (March 16, 2023), https://web.archive.org/web/https://docs.fcc.gov/public/attachments/FCC-23-22A1.pdf; *cf.* Joey Roulette, "SpaceX asks FCC for permission to launch nearly 30,000 more Starlink satellites," *Reuters*, December 7, 2021, https://web.archive.org/web/https://www.reuters.com/technology/spacex-asks-fcc-permission-launch-nearly-30000-more-starlink-satellites-2021-12-07/.

[4]James Johnson develops an "AI-security dilemma" frame for spirals of insecurity and misperception under nuclear conditions. *Viz.* James Johnson, "The AI-Security Dilemma: Insecurity, Mistrust, and Misperception Under the Nuclear Shadow," in *The Politics of Artificial Intelligence: US–China Relations and the Global Order* (Cham: Palgrave Macmillan, 2021), doi:10.1007/978-3-030-73974-1_4; *see also* James Johnson, *AI and the Bomb: Nuclear Strategy and Risk in the Digital Age* (Oxford: Oxford University Press, 2023).

evolving at unprecedented speed."[5] They observe that the management of nuclear weapons remains incomplete and fragmentary, and that the unsolved riddles of nuclear strategy must be given new attention.[6]

A future increasingly defined by software will revolutionize nuclear security. In 2011, Marc Andreessen wrote, "software is eating the world."[7] In 2022, Trae Stephens observed that "software is finally eating the battlefield."[8] Nuclear security cannot escape this trend. Managing this technological revolution demands urgent and focused attention.

[5]Kissinger, Schmidt, and Huttenlocher, *The Age of AI*, 89.

[6]Ibid., 101.

[7]Marc Andreessen, "Why Software Is Eating the World," Andreessen Horowitz, August 20, 2011, https://web.archive.org/web/https://a16z.com/why-software-is-eating-the-world/.

[8]Trae Stephens, "Rebooting the Arsenal of Democracy," *War on the Rocks*, June 6, 2022, https://web.archive.org/web/https://warontherocks.com/2022/06/rebooting-the-arsenal-of-democracy/.

IV. Innovation vis-à-vis Nuclear Security

Based on interviews and research, advances in data collection, computing power, and automation are driving three key trends with direct implications for nuclear security: radical transparency, distributed effects, and changes to human control.

1. Radical Transparency

The combination of AI, ubiquitous sensing, and big data processing is leading to an era of radical transparency. This era will be defined by the ubiquity of networked sensors and the ability to measure continuously across large areas, creating new data patterns. The speed at which machines can collect and analyze data enables faster signatures of behavior and intent. It also enables new forms of digital deception and exposes vulnerabilities tied to reliance on data.

The most obvious sign is the increasing number of sensors collecting and sharing large quantities of data almost instantly. Reductions in sensor size, weight, power requirements, and cost are leading to their use in a skyrocketing number of locations, from the Earth's surface to cislunar space. The prevalence of portable devices such as smartphones enhances this trend by providing ready user interfaces for accessing and manipulating the sensor network.

This world of digital perception can provide monitoring of electrical grids, government facilities, supply chains, and personnel. It supplements the human view of the world with a machine view mapped in data, electromagnetic emissions, and devices. This data is collected and stored in vast repositories known as data lakes, many of which belong to private corporations.

Info Box 1

Microsoft AI CEO Mustafa Suleyman uses the term *digital exhaust* to refer to digital traces left by the presence and activities of people and objects. When a smartphone interacts with a cell tower, it can exchange hundreds of data fields that can subsequently become available to app developers, advertisers, or data brokers.[a]

[a] *See* Mustafa Suleyman and Michael Bhaskar, *The Coming Wave: Technology, Power, and the Twenty-First Century's Greatest Dilemma* (New York: Crown, 2023); *see also* 3GPP TS 24.301, "Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS)," specification archive, https://web.archive.org/web/https://www.3gpp.org/ftp/Specs/archive/24_series/24.301/.

The large volume of data captured by sensors can be processed and exploited by tools such as large language models. Human behavior researchers use the term *patterns of life* to describe the rhythm of daily activities and how those activities are shaped by contextual factors.[9] Patterns apply to individuals and to organizations.

The benefits of transparency create vulnerabilities. Just as personal data can be used for fraud or coercion, organizational patterns can be used to degrade performance through coordinated attacks. Organizations may not know when their data has been compromised, because stolen data remains in place.

[9] Robert Hubal and Elaine A. Cohen Hubal, "Simulating Patterns of Life: More Representative Time-Activity Patterns that Account for Context," *Environment International* 172 (February 2023): 107753, doi:10.1016/j.envint.2023.107753.

The era of radical transparency enables new ways to introduce confounding data to deceive adversaries and creates incentives to engage in deception as a defense against observation. Data poisoning inserts false or misleading data into a system, degrading the performance of models trained on that data. The result could be a decision-support system that fails at a critical decision point or is disabled by operators when it is most needed.

Human organizations are also susceptible to deception, especially public opinion to a *networked* deception, including through chatbots and deepfakes. Troll farms reached an estimated 140 million Americans ahead of the 2020 U.S. presidential election.[10]

Figure 2: Digital exhaust and human-layer observability: They enable today's nuclear warf-ighters to live in a different informational environment than those of past generations.

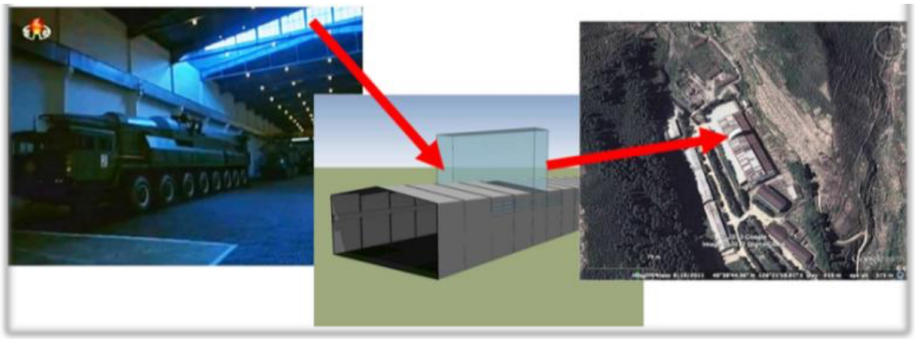


Figure 3: Open-sourcery and OSINT amplification.

[10]Karen Hao, "Troll Farms Reached 140 Million Americans a Month, Facebook's Own Report Shows," *MIT Technology Review*, September 16, 2021, https://web.archive.org/web/https://www.technologyreview.com/2021/09/16/1035851/facebook-troll-farms-internal-report/.

2. Distributed Effects

The convergence of emerging technologies enables an environment in which data maps of organizations and scenarios can be built with unprecedented fidelity. Such maps can be used to facilitate microtargeting of individuals and objects regardless of their number or physical location.

Radical transparency enables the capture of vast quantities of data that can be used to create digital twins, virtual models of real objects, places, or systems informed by timely measurements. For example, the extended reality game Pokémon Go was used by Niantic to collect vast quantities of data about the physical world through millions of players. In aggregate, this data can form a kind of digital twin of terrain and human movement.[11]

Figure 4: Pokémon Go and crowdsourced geospatial data.

A further example is the use of digital twins to support the U.S. Air Force's E-4B Nightwatch aircraft. A multi-year project scanned and digitized the aircraft for training and sustainment, enabling simulation and forecasting of maintenance requirements.[12]

Microtargeting is an advertising practice that uses data about consumers to tailor messages to individual preferences. The Cambridge Analytica episode illustrated how psychometric profiles can be used to identify target groups susceptible to manipulation.

A physical-domain application of microtargeting was demonstrated in October 2024 when pagers and walkie-talkies used by Hezbollah operatives exploded nearly simultaneously. This attack required deep penetration of a commercial supply chain to insert explosive charges into consumer products.[13]

[11] *See* Niantic Spatial, "Large Geospatial Model (LGM)" (company technical overview), https://web.archive.org/web/https://www.nianticspatial.com/lgm; *see also* Niantic, "Visual Positioning System (VPS)" (platform description), https://web.archive.org/web/https://lightship.dev/docs/vps/; *cf.* Alex Heath, "Niantic wants you to scan the world," *The Verge*, June 29, 2021, https://web.archive.org/web/https://www.theverge.com/2021/6/29/22554814/niantic-ar-mapping-3d-scan-world-explained.

[12] *See* Air Force StrikeWerx, "An inside look: Augmented Reality's Advances in the Air Force," July 10, 2019, https://web.archive.org/web/https://www.airforcestrikewerx.com/post/an-inside-look-augmented-reality-s-advances-in-the-air-force; *see also* Director, Operational Test and Evaluation, *FY2024 Annual Report* (discussion of SAOC digital engineering and test instrumentation), PDF, https://web.archive.org/web/https://www.dote.osd.mil/Portals/97/pub/reports/FY2024/other/2024DOTEAnnualReport.pdf.

[13] *See* James Mackenzie and Maya Gebeily, "Exploding pagers and walkie-talkies: How did Israel build Hezbollah trap?" *Reuters*, September 23, 2024, https://web.archive.org/web/https://www.reuters.com/world/middle-east/exploding-pagers-walkie-talkies-how-did-israel-build-hezbollah-trap-2024-09-23/; *see also, e.g.,* The Associated Press, "Israel is accused of detonating Hezbollah pagers: Here's what we know," September 19, 2024, https://apnews.com/article/israel-hezbollah-pagers-explosions-lebanon-f4b1b47d7c0d0b8fe1f9a2d8d4f5b7de.

States with large nuclear arsenals are unusually capable of committing the resources, time, and risk tolerance necessary to apply analogous approaches to degrading adversary nuclear support networks.

Figure 5: Signals, noise, and manipulation in the informational environment.

3. Changes to Human Control

AI tools are unreliable in several ways, including hallucination, misalignment, and vulnerability to attack. Hallucination occurs when a model fabricates plausible falsehoods. Misalignment occurs when a tool produces results that diverge from what users intend or pursue objectives in ways users would not endorse. Misalignment can emerge from under-specified objectives, from opaque reasoning, or from adversarial manipulation.

Human over-reliance can also undermine control. Delegation can lead to deskilling, which may be intolerable in mission-critical defense contexts. Overreliance may be accelerated by the capacity of large language models to communicate in natural language, encouraging a misperception that the tool is exercising human judgment.

Opacity refers to models performing tasks in ways their users do not understand. AlphaGo's defeat of Lee Sedol in 2016 illustrated the phenomenon: a decisive move was regarded as unimaginable by expert players. As systems become more capable, they may become less transparent, and failure modes may become harder to anticipate. In a nuclear security context, such failures could be catastrophic.

AI can also improve decision-making. Decision support systems can be engineered to force models to consider alternatives that human cognition or group dynamics might exclude. Voting logic can be used to combine multiple models for greater fault tolerance. Such tools can be used to emulate adversary behavior, reduce mirror imaging, and test hypotheses about adversary responses. They can also be used to refine crisis communications to prevent misunderstandings.

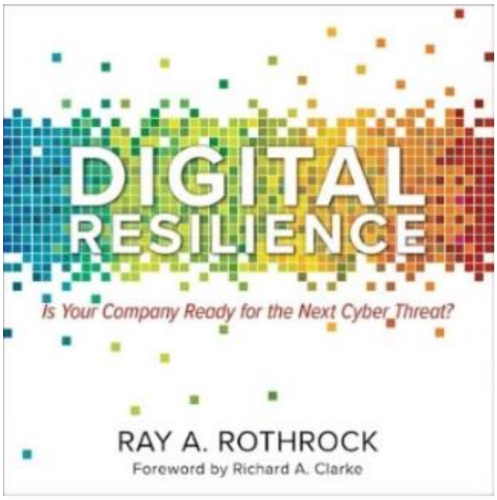


Figure 6: Digital resilience, assurance, and mission-critical dependence.

V. Innovation-Induced Risks

Interviewees identified multiple potential risks to nuclear security resulting from changes in the global technological landscape. The categories below reflect the most frequently raised concerns.

1. RISK 1: Novel and Newly Expanded Vulnerabilities to Nuclear Forces

Interviewees raised concerns that emerging technologies could make nuclear forces vulnerable in new ways, or expand existing vulnerabilities. Stable deterrence requires survivable forces: leaders must be confident that retaliatory forces can execute a secure second strike.

These vulnerabilities are not principally shortcomings in physical durability or training against known threats. They are emergent properties of arsenals as globe-spanning networks of people, machines, and locations embedded within wider commercial networks. Several interviewees emphasized the breadth of new forms of attack, with one observing, "you do not have to touch a system to attack it." Another emphasized full-spectrum cyber operations that create vulnerabilities in adversary networks, not merely exploit existing ones.

The interviewees focused on three expanding vulnerabilities: data vulnerabilities, decision vulnerabilities, and human-layer targeting enabled by digital exhaust.

(a) Sub-Risk A: Data Vulnerabilities

A fundamental risk emerges from the transformation of nuclear arsenals from information-scarce to information-rich environments through commercial data collection and analysis capabilities. Corporate data lakes can inadvertently ingest information about facilities, personnel, and operations through routine commercial activities.[14]

Commercial satellite imagery has altered the information environment surrounding nuclear forces. High-resolution imagery is available as a commercial service from companies operating internationally. Emerging capabilities enable sub-hourly revisit rates coupled with AI-powered monitoring, making secrecy increasingly difficult to maintain.[15]

Data brokers present an equally concerning vulnerability by collecting and selling information about military personnel, including those with nuclear responsibilities.[16] Such information can be used to identify, profile, and compromise individuals with sensitive responsibilities.

Large language models functioning as informational planets represent a novel form of data aggregation that captures patterns about locations, supply chains, and operational rhythms that individual data sources may not expose.[17]

The convergence of these mechanisms undermines secrecy as a viable defense. Historically, comprehensive intelligence collection required extensive resources over extended periods. Today,

[14] Justin Sherman et al., "Data Brokers and the Sale of Government Data," Belfer Center for Science and International Affairs, Harvard Kennedy School, November 2023, PDF, https://web.archive.org/web/https://www.belfercenter.org/sites/default/files/2023-11/data-brokers-sale-government-data.pdf.

[15] Pavel Podvig and Miles A. Pomper, "Commercial Satellite Imagery Could Soon Make Nuclear Secrecy Very Difficult—If Not Impossible," *Bulletin of the Atomic Scientists*, January 26, 2022, https://web.archive.org/web/https://thebulletin.org/2022/01/commercial-satellite-imagery-could-soon-make-nuclear-secrecy-very-difficult-if-not-impossible/.

[16] Sherman et al., "Data Brokers and the Sale of Government Data."

[17] Kissinger, Schmidt, and Huttenlocher, *The Age of AI*, 89–90.

commercial services enable adversaries to develop detailed understanding of vulnerabilities without traditional intelligence operations.

Nuclear deterrence depends on rationality. In the United States, the President is the sole authority responsible for nuclear weapons use. The integrity of the President's informational environment is central. Malign use of social media, deceptive media operations, offensive cyber against sensors, data poisoning, and targeted kinetic attacks to blind capabilities are ways adversaries could attack the informational environment during crisis.

Deepfake audio or video of trusted persons, delivered through insecure media channels, could make it harder to discern fact from fiction. Corporate data lakes supplying the processing capabilities of large language models accelerate the possibility of spreading false information across multiple channels.

RAND researcher Edward Geist argues that AI could transform an offense-dominant environment into a deception-dominant environment in which introducing false data becomes easier than detecting it.[18]

(b) Sub-Risk B: Human-Layer Targeting Enabled by Digital Exhaust

On New Year's Day 2023, 89 Russian soldiers were killed by a Ukrainian strike after using cell phones within weapons range, reportedly allowing the enemy to locate their position.[19] The incident illustrates how digital exhaust can inform attacks by revealing location and concentration.

Nuclear arsenals are world-sized machines designed to move nuclear explosives to targets across the globe. Their human operators emit digital exhaust. Protecting against digital exhaust is not as simple as turning off phones at sensitive sites. The amount of digital exhaust emitted by even careful personnel has increased considerably, and the virtual battlespace extends beyond front lines.

Access to data streams, whether via data brokers or breaches, can enable malign actors to identify people with nuclear responsibilities. Personnel reliability programs assume vulnerabilities are independent of adversary action. That assumption is not robust where models can monitor targets' online behavior and predict them with extraordinary accuracy based on hundreds of observed actions.[20]

As former U.S. Deputy Secretary of Defense Ashton Carter observed, the institutions that support nuclear arsenals are fundamentally social and human.[21]

Mass microtargeting can be used for subtle adversarial approaches such as sustained behavioral nudges, phishing, catfishing, and social campaigns. Agentic AI could be used to plan and conduct

[18]Edward Geist, *Deterrence Under Uncertainty: Artificial Intelligence and Nuclear Warfare* (Oxford: Oxford University Press, 2023).

[19]*See especially* The Associated Press, "Russia blames soldiers' cellphone use for deadly Ukrainian strike," January 4, 2023, https://apnews.com/article/russia-ukraine-war-donetsk-makiivka-cellphones-0d0c4d5f4f3b2b8b4b5f3e9c2b2bcd7e; *see also* BBC News, "Ukraine war: Russia admits Makiivka deaths after strike," January 3, 2023, https://www.bbc.com/news/world-europe-64161930.

[20]Wu Youyou, Michal Kosinski, and David Stillwell, "Computer-Based Personality Judgments Are More Accurate Than Those Made by Humans," *Proceedings of the National Academy of Sciences* 112, no. 4 (January 27, 2015): 1036–1040, doi:10.1073/pnas.1418680112.

[21]Ashton B. Carter, "Remarks by Deputy Secretary Carter at Cooperative Threat Reduction Symposium," U.S. Department of Defense, December 3, 2012, https://web.archive.org/web/20121213035156/http://www.defense.gov/Transcripts/Transcript.aspx?TranscriptID=5166.

information operations at scale and to stage future mixed-domain attacks to be executed during crisis or war. Traditional practices designed to protect against insider threats are poorly suited to defend against these information-age human-layer attacks.

Story Box 1:

In August 1969, Strategic Air Command posted its first meteorologist to Havre Air Force Station in Montana—the father of one of the authors of this report. Born in 1944, this second-generation nuclear warfighter emitted little digital exhaust observable to foreign intelligence. Today's Gen Z and Millennial nuclear warfighters live in a world of pervasive sensors, commercial apps, data brokers, and supply-chain dependence. Even disciplined personnel present an attack surface radically different from that of prior generations.

2. RISK 2: New Pathways to Using Nuclear Weapons

Several interviewees raised the potential for new pathways to nuclear weapons' utilization as a major risk of AI and other emerging technologies. Such technologies can blur the conventional-nuclear boundary, create escalation pathways, and undermine survivability assumptions that underpin deterrence.

(a) Case Study: Operation Spider's Web and Distributed Autonomous Threats

In June 2025, commercial technologies enabled an unprecedented attack on nuclear-capable systems. On June 1, Ukraine executed Operation Spider's Web, targeting five Russian strategic aviation bases spanning five time zones using 117 AI-guided autonomous drones concealed within commercial truck-mounted containers and pre-positioned across Russian territory over eighteen months.[22]

Ukrainian forces destroyed or damaged more than twenty strategic aircraft, including nuclear-capable bombers that constitute components of Russia's nuclear triad, eliminating an estimated seven billion dollars in strategic aviation assets.[23] Russian state media noted that, under Russia's Basic Principles on Nuclear Deterrence, such actions could fall under conditions that may justify nuclear use.[24] Keith Kellogg, the U.S. envoy for Ukraine, warned that the risk level increases because one cannot know what the other side will do.[25]

This operation illustrates how radical transparency and distributed effects can create new nuclear pathways. Commercial logistics networks and autonomous platforms enabled precision strikes against nuclear delivery systems previously protected by distance and defenses. When strategic bombers can be destroyed by low-cost autonomous systems, assumptions about second-strike capabilities require reassessment.

[22]Reuters Graphics, "Operation Spiderweb: What We Know About Ukraine's Drone Attack on Russian Air Bases," *Reuters*, June 4, 2025 (updated June 5, 2025), https://web.archive.org/web/https://www.reuters.com/graphics/UKRAINE-CRISIS/RUSSIA-DRONES/.

[23]Ibid.

[24]"Spider's Web: Video of Lorry Carrying Drones Used During Operation," *Ukrainska Pravda* (English edition), June 4, 2025, https://web.archive.org/web/https://www.pravda.com.ua/eng/news/2025/06/04/7515551/.

[25]Phil Stewart and Idrees Ali, "Ukraine Hit Fewer Russian Planes Than It Estimated, U.S. Officials Say," *Reuters*, June 5, 2025, https://web.archive.org/web/https://www.reuters.com/world/ukraine-hit-fewer-russian-planes-than-it-estimated-us-officials-say-2025-06-05/.

Following Pavel Podvig, the operation illustrates the inadequacy of classic military machinery when confronting adversaries who can leverage emerging technologies creatively.[26]

(b) Sub-Risk A: Hypersonic Missiles

Hypersonic weapons disrupt escalation dynamics. Hypersonic cruise missiles and boost-glide vehicles combine lower signature, extreme speed, and maneuverable non-ballistic trajectories at relatively low altitude, complicating detection and interception. These characteristics compress decision time and create ambiguity about attack type.

Legacy systems are challenged by hypersonics. The U.S. Hypersonic and Ballistic Tracking Space Sensor is designed to address these detection problems.[27] Compressed timelines may leave leaders with minutes to determine whether to use retaliatory forces. Congressional Research Service reporting notes that warning and decision processes can be severely time constrained.[28]

Dual-use hypersonic delivery complicates warning. A hypersonic weapon could deliver conventional or nuclear warheads. The potential for conventional means to strike nuclear targets can create incentives for preemption and miscalculation.

(c) Sub-Risk B: Outer Space

Reusable boosters and proliferated low Earth orbit architectures have decreased the cost of space launch and satellite communications, making orbital assets more affordable and more ubiquitous. These developments increase reliance on orbital infrastructure, widen the potential for counterspace and sabotage, and weaken the defenses of legacy space systems.

Commercial megaconstellations provide dual-use capabilities. Governments increasingly rely on commercial providers for communications and remote sensing. This public-private entanglement can create escalation coupling, including via threats to large constellations and debris-generating attacks that affect all operators.

Russia has reportedly pursued systems that could destroy satellites indiscriminately, raising escalation concerns.[29] One interviewee suggested that Russia may plan to use nuclear threats as a low-technology equalizer against exquisite U.S. capabilities.

[26] Pavel Podvig, "Untangling Operation 'Spiderweb,'" *Meduza*, June 3, 2025, https://web.archive.org/web/https://meduza.io/en/feature/2025/06/03/untangling-operation-spiderweb.

[27] *See* John R. Hoehn, "Hypersonic Weapons: Background and Issues for Congress," Congressional Research Service, updated May 16, 2025, PDF, https://web.archive.org/web/https://crsreports.congress.gov/product/pdf/R/R45811; *see also* U.S. Government Accountability Office, "Defense Acquisitions: DOD's Coordination for Integrated Missile Defense Needs Improvement," GAO-22-104590 (July 2022), https://web.archive.org/web/https://www.gao.gov/products/gao-22-104590; *cf.* Mike Stone, "L3Harris opens missile tracking satellite facility in U.S. state of Indiana," *Reuters*, January 12, 2023, https://web.archive.org/web/https://www.reuters.com/world/us/l3harris-opens-missile-tracking-satellite-facility-us-state-indiana-2023-01-12/.

[28] Anya L. Fink, "Authority to Launch Nuclear Forces: Overview and Selected Issues for Congress," Congressional Research Service, December 19, 2024, PDF, https://web.archive.org/web/https://crsreports.congress.gov/product/pdf/IF/IF12751.

[29] Arms Control Association, "Emerging Military Technologies and Nuclear (In)Stability," PDF, https://web.archive.org/web/https://www.armscontrol.org/sites/default/files/files/Reports/ACA_Report_EmergingTech_digital.pdf.

3. RISK 3: Increased Risk of Nuclear Proliferation and Nuclear Terrorism

Interviewees expressed concerns that the convergence of emerging technologies could lower barriers to proliferation and terrorism. Several emphasized the combination of AI and advanced manufacturing easing access to weapons-usable nuclear materials, especially through clandestine uranium enrichment, while stressing the difficulties in creating and maintaining a weapons enterprise.

(a) Sub-Risk A: Access to Uranium Enrichment

The most significant technological barrier to building a nuclear weapon is the acquisition of sufficient quantities of weapons-usable fissile material, notably plutonium or highly enriched uranium. Natural uranium is abundant but is mostly uranium-238, with a smaller fraction of uranium-235, which must, thereupon, be separated for weapons use.

Enrichment technology has improved, lowering financial and energy requirements. These advancements reduce signatures of enrichment, potentially easing clandestine activity. One interviewee stressed that enrichment requirements for civilian power vastly exceed those needed for a clandestine weapons program. Fueling a single large light-water reactor for a year can require on the order of one hundred thousand separative work units.[30] A smaller enrichment capability than industrial scale could be sufficient to enable an improvised device or a small arsenal.

Generally, emerging technologies may also render enrichment more accessible by enabling autonomous discovery of low-observable pathways. Interviewees expressed concern that AI could help malign actors access known pathways, discover additional pathways, or facilitate clandestine use.

[30]For benchmark separative work requirements for commercial reactor fuel cycles, including order-of-magnitude figures on the scale of ∼100,000 SWU per year for a large light-water reactor fuel demand, *see* World Nuclear Association, "Uranium Enrichment," Information Library, https://web.archive.org/web/https://world-nuclear.org/information-library/nuclear-fuel-cycle/conversion-enrichment-and-fabrication/uranium-enrichment.aspx.

VI. Innovation-Induced Opportunities

Interviewees most frequently cited three areas of opportunities driven by the convergence of emerging technologies: improved warning confidence, deterrence resilience and arms-race stability, and arms control, nonproliferation, and threat reduction.

1. OPPORTUNITY 1: Improved Warning Confidence

Emerging technologies can enhance the reliability of warnings of attacks on nuclear forces through more sensors and advanced data fusion. By reducing uncertainty during crises, minimizing false warnings, and increasing opportunities for de-escalation, these technologies could lower the risk of catastrophe.

AI-enabled warning systems can interpret vast data from early warning systems more effectively than humans, recognizing patterns and anomalies. Rapid analysis of diverse sources enables deeper insight into intent. This matters where leaders may have minutes to determine whether to use retaliatory forces.

Biases and errors can be mitigated by voting logic among multiple distinct models. A multi-model rule can reduce the danger of hallucinations or bias, producing false warnings by combining outputs for fault tolerance. Because there has never been large-scale nuclear use, synthetic data would be required for certain training regimes; interviewees stressed the need for care in validation.

2. OPPORTUNITY 2: Deterrence Resilience and Arms-Race Stability

The survivability of retaliatory forces is a key variable in deterrence stability. Deterrence resilience is conceptually wider than the survivability of platforms. It includes the resilience of NC3, the integrity of data, the resilience of space assets, and the capacity of decision makers to resist deception and coercion.

Interviewees described paths to improve NC3 resilience against attacks, to expand force protection to include data, and to leverage mixed reality, distributed ledgers, and advanced manufacturing to lower costs and risks of maintaining a nuclear arsenal. Multiple interviewees suggested mechanisms for engagement between the nuclear security enterprise and private-sector industries responsible for mission assurance.

(a) Sub-Opportunity A: Deterrence Resilience

Nuclear deterrence comprises capability, credibility, and communication. As attack surfaces expand, adversaries are developing capabilities designed to degrade one or more elements. The target is not just delivery systems but also the enabling network of data, space assets, and political cohesion.

Fortifying NC3 against all forms of adversary attack is central. One approach is transitioning from vulnerable legacy space architectures to proliferated low Earth orbit constellations paired with responsive launch. Such architectures can be bolstered by the rapid replacement of destroyed systems.

Interviewees also suggested better protection of the human layer. Shirley Ann Jackson described the potential for a digital immune system that could protect decision makers and war-fighters from data-driven attacks, including through sentinels or avatars inserted into data lakes to detect attempts at compromise.[31]

Distributed ledgers can support cryptographic verification of data provenance, improving resilience against deception by enabling verification of what data trained models and what sensors contributed to a decision pipeline.

Large-scale digital twins of adversary systems may improve planning and crisis management by synthesizing larger volumes of information about capabilities and intent. Tools that model potential adversary reactions can improve the resilience of policy options and assist in finding de-escalatory solutions that humans might overlook.

(b) Sub-Opportunity B: Arms-Race Stability

Emerging technologies could be applied to dampen incentives for arms racing, as well as to control the costs of maintaining arsenals. One interviewee emphasized that the central challenge is persuading adversaries about what one is not going to do, because restraint is not credible where capabilities imply otherwise.

Interviewees envisioned a digitized weapons complex using simulation and digital twins to improve stockpile confidence and optimize production. Investments in data infrastructure and public-private partnerships would be required, particularly for working on classified information at scale.

3. OPPORTUNITY 3: Arms Control, Nonproliferation, and Threat Reduction

The United States and the Soviet Union developed a practice of arms control over decades to promote stability by reducing incentives for first strike and arms racing. Applying arms control concepts to AI is challenging, yet interviewees suggested public-private partnerships to harness emerging technologies for verification, monitoring, and threat reduction.

(a) Sub-Opportunity A: Arms Control

The convergence of emerging technologies can strengthen arms control through enhanced detection, monitoring, and verification technologies. Sensor capability and data fusion are advancing rapidly. These changes could facilitate new arms control not previously possible.

Blockchain-enabled tools could provide zero-knowledge proofs that readiness status has not changed, proving a fact to another party without revealing classified details. Zero-knowledge methods have long been considered for warhead verification. Interviewees suggested that related techniques could verify base-level or arsenal-level compliance without disclosing sensitive information.

A 2023 study by the *National Academies* concluded that the United States must modernize and ultimately revolutionize monitoring, detection, and verification systems rather than continuing incremental steps.[32]

[31] Shirley Ann Jackson, interview conducted for this study (2024), outlining a "digital immune system" for protecting

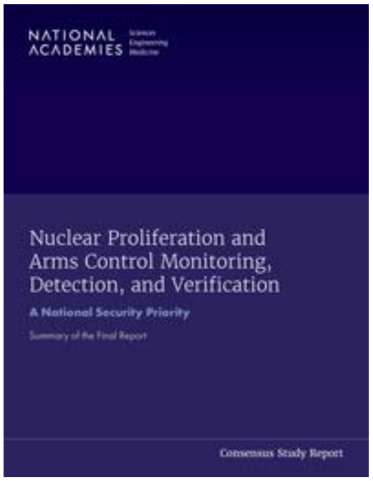


Figure 7: A consensus report on monitoring, detection, and verification (cover image)

Story Box 2:

Professor Jeffrey Lewis of the Middlebury Institute for International Studies has pioneered methods of deriving sensitive nuclear security information using commercial tools and open-source data. In a 2013 North Korean propaganda video, an image of a facility interior appeared behind a leader. Lewis and colleagues modeled the building geometry and used satellite imagery, defector accounts, and other open sources to narrow the location to a small set of plausible sites. The example illustrates how open-source analysis can increase transparency and how future automation could further lower the barriers to deriving sensitive information.[a]

[a]Jeffrey Lewis, interview conducted for this study (2024), describing the 2013 open-source geo-locational workflow and the modeling steps used to narrow candidate facilities using imagery, geometry, and corroborating open sources.

(b) Sub-Opportunity B: Nonproliferation

Nuclear security opportunities may grow as emerging technologies improve transparency and analysis. Warren Stern observes a revolution in detection affairs driven by new sensors and smartphones.[33] Machine learning and open-source information can contribute to detecting proliferation activity. Interviewees emphasized combining public information with government sources, merging space information with trade information, and integrating heterogeneous streams for robustness.

Story Box 3:

During 2019–2020, the Nuclear Threat Initiative and the Center for Advanced Defense Studies demonstrated how combining publicly available data with machine learning can identify high-risk

decision makers and operators against data-driven compromise in high-consequence settings.

[32]National Academies of Sciences, Engineering, and Medicine, *Nuclear Proliferation and Arms Control Monitoring, Detection, and Verification: A National Security Priority: Summary of the Final Report* (Washington, DC: The National Academies Press, 2023), 1–2, *esp.* 1.

[33]Stern, "Revolution in Nuclear Detection Affairs," 220.

entities within millions of transactions. Automated preparation saved hundreds of analyst hours and generated new leads, some of which were later added to export control lists.[a]

[a]Center for Advanced Defense Studies (C4ADS) and Nuclear Threat Initiative, *Signals in the Noise: Preventing Nuclear Proliferation with Machine Learning and Publicly Available Information* (Washington, DC: C4ADS and NTI, 2020), PDF, https://web.archive.org/web/https://www.nti.org/wp-content/uploads/2020/10/Signals_in_the_Noise.pdf.

(c) Sub-Opportunity C: Threat Reduction

The collapse of the Soviet Union left the world's largest arsenal distributed across newly independent states. The Nunn-Lugar Cooperative Threat Reduction Program responded by supporting the deactivation of thousands of warheads and the destruction of delivery systems while improving material security.

If a future political opportunity for threat reduction emerges, emerging technologies not available in the 1990s could contribute. Off-the-shelf sensor systems can provide site security functionality. Handheld scanning tools can create digital models of facilities, improving planning and procurement. Digital twins can monitor pathways for diversion and support predictive maintenance and inspection planning.

VII. Recommendations

Table 1: Tiered and ordered risk ranking (sub-risk granularity)

Tier	Rank	Sub-Risk Node	Weighted Mean	Primary Allocation
I	1	RISK 1(A): Data vulnerabilities (commercial aggregation, data lakes, deception, integrity loss)	8.3	Recommendation 2
I	2	RISK 1(B): Human-layer targeting (digital exhaust, coercion, manipulation, insider vector)	8.0	Recommendation 2
I	3	RISK 2(A): Distributed autonomous attack pathways (Spider's Web archetype)	7.7	Recommendation 1
I	4	RISK 2(B): Hypersonic missiles (compression of decision time; ambiguity; preemption incentives)	7.5	Recommendation 1
II	5	RISK 2(C): Outer space (counterspace threats; ISR fragility; escalation coupling)	6.6	Recommendation 1
II	6	RISK 3(A): Access to uranium enrichment (diffusion of know-how; procurement; optimization)	6.1	Recommendation 2

Table 2: Roll-up ordering (risk-level aggregation)

Tier	Rank	Risk	Aggregate Score
I	1	RISK 1: Novel and Newly Expanded Vulnerabilities to Nuclear Forces	8.2
I	2	RISK 2: New Pathways to Using Nuclear Weapons	7.4
II	3	RISK 3: Increased Risk of Nuclear Proliferation and Nuclear Terrorism	6.1

The manifold of nuclear security is increasingly entangled with emerging technologies developed in the private sector. However, neither the private sector nor governments is working toward a comprehensive understanding of the scale and nature of the technological revolution for nuclear security. Relatively little attention has been paid to the possibility that the convergence of emerging technologies may pose new risks to nuclear security by challenging the systems and approaches that nuclear-armed states use to prevent nuclear weapons use and to control nuclear weapons.

Ultimately, success in mitigating risks and reaping benefits will depend on leadership that can move beyond traditional approaches. Nuclear-armed governments should appoint officials tasked with leading comprehensive assessments of components of their nuclear systems in light of the evolving landscape shaped by technological shifts.

The United States is on course to spend approximately two trillion dollars on nuclear weapons over three decades to replace platforms designed for the twentieth century, without considering what

Table 3: Allocatory matrix from ranked risks to recommendations

Risk or Sub-Risk Node	**Recommendation 1 (Task Force)**	**Recommendation 2 (Private-Sector Engagement)**
RISK 1(A): Data vulnerabilities		✓
RISK 1(B): Human-layer targeting		✓
RISK 2(A): Distributed autonomous threats	✓	
RISK 2(B): Hypersonic missiles	✓	
RISK 2(C): Outer space	✓	
RISK 3(A): Access to uranium enrichment		✓

nuclear war prevention requires now.[34] The United States should develop an updated strategy to prevent nuclear weapons use and reduce the dangers of proliferation and terrorism in partnership with the international community.

1. Recommendation 1: To Establish a Task Force on Artificial Intelligence, Strategic Stability, and Nuclear Risk

The U.S. Congress should establish a Task Force on Artificial Intelligence, Strategic Stability, and Nuclear Risk modeled on the National Security Commission on AI. The Task Force should be charged with producing a comprehensive report with recommendations for nuclear security technology development, diplomatic engagement, and public-private partnerships.

The Task Force should comprise leaders in nuclear security and branches of emerging technology that could accelerate appropriate responses. Congress should provide appropriate resources and staff.

2. Recommendation 2: To Launch Structures and Incentives for Transformational Private-Sector Engagement

The U.S. Government should create structures and incentives for deep public-private partnerships for nuclear security. These could take forms including: a Nuclear Security Innovation Unit modeled on the Defense Innovation Unit; a Nuclear Security Development Agency modeled on the Space Development Agency; and a NucWERX program modeled on AFWERX and SpaceWERX.

Congress should also consider opportunities that leverage the innovation capabilities of the U.S. national laboratory system through place-based innovation, multi-laboratory collaboration, and the establishment of new laboratories as appropriate.

[34]Congressional Budget Office, *Approaches for Managing the Costs of U.S. Nuclear Forces, 2017 to 2046* (Washington, DC: CBO, October 2017), PDF, https://web.archive.org/web/https://www.cbo.gov/publication/53211; Congressional Budget Office, *Projected Costs of U.S. Nuclear Forces, 2023 to 2032* (Washington, DC: CBO, January 2023), PDF, https://web.archive.org/web/https://www.cbo.gov/publication/58848. Read together, these CBO baselines are commonly used as the authoritative anchor for translating modernization peaks into multi-decade totals in the high hundreds of billions per decade, yielding order-of-trillion totals over a 30-year horizon.

Appendix A: Glossary

Artificial intelligence (AI)
Software that enables computers to make decisions historically reserved for humans.

Agentic artificial intelligence
A class of AI designed to act with autonomy, able to conduct strategic planning and dynamic problem solving without direct human intervention.

Alignment
Steering AI systems toward goals and ethical principles that coincide with humans.

Attack surface
A group of paths, methods, or scenarios that can be used to enter data to, extract data from, or control a device or software in an environment.

Behavioral nudge
Interventions designed to influence decisions and steer individuals toward specific outcomes without limiting freedom of choice.

Data lake
A centralized repository that stores large volumes of data in its original form.

Deepfake
An image or recording that has been convincingly altered to misrepresent someone as doing or saying something.

Digital exhaust
An invisible trail of data left behind by a person's interactions with technological services.

Digital twin
A virtual model of an intended or real-world object for purposes such as simulation, monitoring, or maintenance.

Distributed ledger technology
A system where data is stored and synchronized over several geographical locations, as opposed to a central database, and does not require a central administrator.

Escalation ladder
A metaphor popularized in the 1950s and 1960s to refer to stages of escalation between parties, visualized as rungs of increasing intensity.

Extended reality
An umbrella term referring to augmented reality, mixed reality, and virtual reality.

Full-spectrum cyber
The combined arms employment of joint capabilities to create and exploit advantages in the cyber domain.

Hallucinations
A response made by an AI that contains false or misleading information presented as fact.

Human layer
The component of a system comprised of humans, as opposed to hardware or software.

Internet of things
A network of physical devices embedded with sensors, software, and connectivity enabling them to collect and share data.

Large language model
A type of machine learning model tasked with natural language processing, such as language generation.

Machine learning
A form of AI that uses statistical algorithms that learn from data and perform tasks without explicit instructions.

Mutually assured destruction
A doctrine of deterrence assuming an attack by one superpower triggers a counterattack by the other, resulting in the destruction of both.

Nonproliferation
The practice of preventing the spread of nuclear weapons, fissile materials, and weapons-applicable nuclear technology.

Nuclear deterrence
The threat or implied threat of nuclear use against an adversary to shape behavior.

Stockpile stewardship
The program of reliability testing, viability, and maintenance of the U.S. nuclear arsenal without nuclear explosive testing.

Supply chain
A logistics system that converts raw materials into completed products and distributes them to end users.

Synthetic data
Data artificially generated rather than created by real-world events, typically used for training AI models.

Nuclear enterprise
The infrastructure and expertise dedicated to the development, testing, and maintenance of nuclear weapons.

Caltropics
Term reserved for definition.

Appendix B: Methodological Addendum

This addendum is designed to be self-contained. It specifies the empirical inputs, the normalization rules for expert judgments, the aggregation statistics, the tiering rule, and the reproducible mapping from ranked risks to recommendations. It also includes the full Soufflé program used to compute risk scores and tiers, together with an explanation of its relations and expected input facts.

1. Inputs, Corpus, and Reproducibility Commitments

The corpus consists of expert identifiers, proposition identifiers, and votes recorded in two rounds. Round II scores on the discrete 1–9 scale are the principal quantitative inputs. Each expert is assigned a weight, used to compute weighted statistics. Weights are treated as exogenous to the program: the program does not infer weights; it consumes them.

Reproducibility rests on four invariants:

(i) Every proposition is a single claim at single-claim granularity.
(ii) Every proposition is mapped to exactly one risk or opportunity node.
(iii) Every vote is attached to an expert identifier and a proposition identifier, with round recorded explicitly.
(iv) All computed outputs are deterministic functions of the input facts under fixed rules.

2. Normalization into Propositions and Nodes

Interview transcripts are normalized into proposition identifiers. Each proposition is then mapped to a node used in the report's taxonomy. Nodes are mapped to a risk family and a sub-risk label. This produces the graph edge list: expert $\rightarrow$ proposition $\rightarrow$ node $\rightarrow$ risk family.

The design choice is deliberate: it permits auditability by enabling any ranked claim to be traced to the propositions that generated it and, in turn, to the experts who scored those propositions.

3. Weighting Model

Expert weights are real numbers strictly greater than zero. They may encode cohort weights, confidence weights, or other credibility adjustments established ex ante. The program treats weights as multiplicative factors in scoring. If the weights for the set of experts voting on a proposition do not sum to one, the weighted mean is normalized by the total weight of those experts, ensuring that the score remains on the 1–9 scale.

4. Aggregation Statistics

The program computes three core statistics for each proposition using Round II scores:

(i) Weighted mean: the sum of weight times score divided by the sum of weights for the experts who voted on that proposition in Round II.
(ii) Weighted median: the smallest score on the discrete 1–9 scale for which the cumulative weight is at least one half of the total weight of voters on that proposition.

(iii) Interquartile range: the difference between the weighted seventy-fifth percentile and the weighted twenty-fifth percentile, using the same discrete cumulative-weight rule.

The weighted median and interquartile range serve as robustness measures for the mean, capturing the central tendency and dispersion of a possibly polarized panel on a discrete scale.

5. Tiering Rule and Allocatory Mapping

Tier assignment is computed on the weighted mean:

- Tier I if weighted mean is at least 7.5.
- Tier II if weighted mean is at least 5.5 and less than 7.5.
- Tier III otherwise.

Recommendation allocation is then computed deterministically:

- Tier I risks are allocated to Recommendation 1.
- Tier II and Tier III risks are allocated to Recommendation 2.

This mapping is separable from tiering: alternative mappings can be substituted without changing the ranking logic, and the program can emit both tiers and allocations for downstream presentation.

6. Implementation Notes for Soufflé

The program is written in Soufflé Datalog. The input facts are assumed to be loaded as relations matching the declared schema. The program produces outputs for the relations declared under .output. Output files can be used to generate the tables in the Recommendations section, including the ordered sub-risk table, the roll-up ordering, and the allocatory matrix.

Because the score scale is discrete and bounded, weighted quantiles are computed by explicit enumeration of the score domain and cumulative-weight aggregation, which avoids reliance on any non-deterministic ordering and ensures reproducibility across runtimes.

Listing 3: "Code suite" in Soufflé

```
.decl expert(e:symbol, cohort:symbol, w:float)
.decl proposition(p:symbol, node:symbol, label:symbol)
.decl vote(e:symbol, p:symbol, round:number, score:number)
.decl risk_node(node:symbol, risk:symbol, subrisk:symbol)

.decl scoreval(s:number)
scoreval(1).
scoreval(2).
scoreval(3).
scoreval(4).
scoreval(5).
scoreval(6).
scoreval(7).
scoreval(8).
scoreval(9).

.decl w_by_score(p:symbol, s:number, wsum:float)
```

```
w_by_score(p, s, wsum) :-
  vote(e,p,2,score),
  expert(e,_,w),
  scoreval(s),
  score =s,
  wsum =sum w.

.decl w_total(p:symbol, wtot:float)
w_total(p, wtot) :-
  w_by_score(p,_,wsum),
  wtot =sum wsum.

.decl wmean_num(p:symbol, num:float)
wmean_num(p, num) :-
  vote(e,p,2,score),
  expert(e,_,w),
  num =sum w*score.

.decl wmean_den(p:symbol, den:float)
wmean_den(p, den) :-
  vote(e,p,2,_),
  expert(e,_,w),
  den =sum w.

.decl n_votes(p:symbol, n:number)
n_votes(p, n) :-
  vote(_,p,2,_),
  n =count : { vote(_,p,2,_) }.

.decl agg(p:symbol, wmean:float, n:number)
agg(p, wmean, n) :-
  wmean_num(p,num),
  wmean_den(p,den),
  n_votes(p,n),
  den >0.0,
  wmean =num / den.

.decl wcum(p:symbol, s:number, cw:float)
wcum(p, s, cw) :-
  scoreval(s),
  w_by_score(p,t,ws),
  t <=s,
  cw =sum ws.

.decl qcandidate(p:symbol, q:float, s:number)
qcandidate(p, q, s) :-
  w_total(p,wt),
  wcum(p,s,cw),
  cw >=wt*q.
```

```
67
68  .decl qscore(p:symbol, q:float, s:number)
69  qscore(p, q, ms) :-
70    qcandidate(p,q,s),
71    ms =min s.
72
73  .decl wmedian(p:symbol, s:number)
74  wmedian(p, s) :-
75    qscore(p, 0.50, s).
76
77  .decl q1(p:symbol, s:number)
78  q1(p, s) :-
79    qscore(p, 0.25, s).
80
81  .decl q3(p:symbol, s:number)
82  q3(p, s) :-
83    qscore(p, 0.75, s).
84
85  .decl iqr(p:symbol, i:number)
86  iqr(p, i) :-
87    q3(p,s3),
88    q1(p,s1),
89    i =s3 - s1.
90
91  .decl ranked(node:symbol, risk:symbol, subrisk:symbol, wmean:float, wmed:number, iqr:
        number)
92  ranked(node, risk, subrisk, wmean, wmed, iqr) :-
93    proposition(p,node,_),
94    risk_node(node,risk,subrisk),
95    agg(p,wmean,_),
96    wmedian(p,wmed),
97    iqr(p,iqr).
98
99  .decl tier(node:symbol, tier:symbol)
100 tier(node, "I") :-ranked(node,_,_,wmean,_,_), wmean >=7.5.
101 tier(node, "II") :-ranked(node,_,_,wmean,_,_), wmean >=5.5, wmean <7.5.
102 tier(node, "III") :-ranked(node,_,_,wmean,_,_), wmean <5.5.
103
104 .decl allocation(node:symbol, rec:number)
105 allocation(node, 1) :-tier(node, "I").
106 allocation(node, 2) :-tier(node, "II").
107 allocation(node, 2) :-tier(node, "III").
108
109 .output expert
110 .output proposition
111 .output vote
112 .output risk_node
113 .output ranked
114 .output tier
```

```
.output allocation
```

7. Input Facts: Minimal Conventions

The program assumes the following minimal conventions:

- Each expert appears once in `expert(e, cohort, w)`.
- Each proposition appears once in `proposition(p, node, label)`.
- Each Round II score appears as `vote(e, p, 2, score)` where score is an integer on the 1–9 scale.
- Each node is mapped by `risk_node(node, risk, subrisk)` to a risk family and sub-risk string.

These conventions are sufficient to reproduce the ranking and tier outputs. Additional relations can be added for provenance, such as linking propositions to transcript identifiers or to coders, without altering the aggregation logic.