AI AND THE FUTURE OF NATIONAL SECURITY

Findings from a survey of 100+ national security practitioners on the opportunities and risks of advanced AI deployment in military, cybersecurity, intelligence, and CBRN

MARIAMI TKESHELASHVILI
RITIKA VERMA
BRIDGET WILLIAMS
REBECCA CEPPAS DE CASTRO

SEPTEMBER 2026

IST Institute for SECURITY + TECHNOLOGY

AI and the Future of National Security:
Findings from a survey of 100+ national security practitioners on the opportunities and risks of advanced AI deployment in military, cybersecurity, intelligence, and CBRN

September 2026
Author: Mariami Tkeshelashvili, Ritika Verma, Bridget Williams, and Rebecca Ceppas de Castro
Design: Taylor White

IST

About the Institute for Security and Technology

Uniting technology and policy leaders to create actionable solutions to emerging security challenges

Technology has the potential to unlock greater knowledge, enhance our collective capabilities, and create new opportunities for growth and innovation. However, insecure, negligent, or exploitative technological advancements can threaten global security and stability. Anticipating these issues and guiding the development of trustworthy technology is essential to preserve what we all value.

The Institute for Security and Technology (IST), the 501(c)(3) critical action think tank, stands at the forefront of this imperative, uniting policymakers, technology experts, and industry leaders to identify and translate discourse into impact. We take collaborative action to advance national security and global stability through technology built on trust, guiding businesses and governments with hands-on expertise, in-depth analysis, and a global network.

We work across three analytical pillars: the **Future of Digital Security**, examining the systemic security risks of societal dependence on digital technologies; **Geopolitics of Technology**, anticipating the positive and negative security effects of emerging, disruptive technologies on the international balance of power, within states, and between governments and industries; and **Innovation and Catastrophic Risk**, providing deep technical and analytical expertise on technology-derived existential threats to society.

Learn more: https://securityandtechnology.org/

Acknowledgments

About the Project

In October 2025, the **Institute for Security and Technology (IST),** with support from and in partnership with the **Future of Life Institute (FLI)**, launched the AI Risk Barometer project to capture how national security professionals view the risks and opportunities of advanced AI. By surveying policymakers, military planners, and technical experts, the project aims to provide a clearer picture of how those on the frontlines of national security are thinking about timelines, thresholds, and the probability of loss of control scenarios as we move toward integrating more advanced AI systems.

IST is grateful to the Future of Life Institute (FLI) for their support of the AI Risk Barometer effort and this survey.

About the Authors

Mariami Tkeshelashvili is the Director for Artificial Intelligence Security Policy at the Institute for Security and Technology (IST). She leads the AI Risk Reduction Initiative, engaging a diverse range of stakeholders across the AI ecosystem to identify emerging risks from cutting-edge AI models and to develop technical and policy-based mitigation strategies that advance responsible innovation.

Ritika Verma is a cybersecurity professional and Associate for Artificial Intelligence Security Policy at the Institute for Security and Technology (IST). She works on the AI Risk Barometer and AI Risk Reduction initiatives, focusing on AI security and translating technical risk insights into governance and risk management frameworks.

Bridget Williams is an Associate Director of Research and Strategy at the Forecasting Research Institute, leading several projects on risks from AI. Prior to joining FRI, Bridget worked as a medical doctor and in public health medicine roles in academia, government, and international non-governmental organisations. She is also currently completing a PhD in AI–biosecurity policy at the University of Oxford.

Rebecca Ceppas de Castro is a Data Analyst at the Forecasting Research Institute, where she works on applied research investigating risks arising from the interaction of AI with biosecurity, cybersecurity, and other domains. She has a background in physics and previously conducted research in astrophysics and cosmology.

Contributors

This work is inherently collaborative. The authors of this report are immensely grateful to the following contributors for their expertise:

IST leadership: Philip Reiner, Megan Stifel, Steve Kelly, and Nicholas Leiserson

IST board members, staff, and adjuncts: Jason Kichen, Michael Klein, Sylvia Mishra, Jennifer Tang, Catherine Murphy, Brian Abeyta, and Melissa Hopkins

Finally, we are grateful to the Future of Life Institute and Hamza Chaudhry for their invaluable support.

FOREWORD

Philip Reiner
CEO, Institute for
Security and Technology

Lt Gen Jack Shanahan
United States Air Force, retired

The Institute for Security and Technology (IST) conducted this survey, with support from the Future of Life Institute (FLI), because we were concerned that the national AI conversation was moving at two very different speeds: breathless in industry, ponderous in government. Our concern was made more pressing by recent events and incidents, which showcased the accelerated pace of AI development and diffusion.

What decision-makers believe AI can do matters as much as a model's abilities. Misjudging its capabilities and limitations can lead to overreliance in some areas and underestimation in others: treating AI as a decisive, war-winning weapon in one domain while dismissing its relevance in another. This survey was built to set the record straight, not on what the public or researchers believe AI can do, but on what national security practitioners themselves — the people who confront these risks and opportunities firsthand — actually believe. The consensus and disagreements documented throughout point to where action already has momentum, where defenders need to focus, and where open questions still demand work.

Andy Marshall spent four decades arguing that the most dangerous strategic error is not misjudging an adversary's capability, but misjudging your own institution's capacity to adapt to a shift already underway. The survey results and this summary report are aimed at that second failure. Not what AI can do — in fact, every finding here assumes the answer is that AI has the potential to be extraordinarily powerful — but whether the people responsible for national security actually understand what they are managing, and if they are organized to do so properly and effectively.

The partial name we give this effort (AI Risk Barometer) borrows from Arthur Compton, who in 1942 had to put a number on the odds that detonating the first atomic bomb would ignite the atmosphere and end life on Earth, and then decide, with that number in hand, whether to proceed anyway. From the outset, we set out to test for a Compton Constant of our own for AI: not simply a probability of catastrophe, but how national security policymakers actually weigh the risk of losing control of highly advanced AI against the benefits of having it. The survey results answered that question. The results are not comforting. A majority of respondents already place their own estimated catastrophe risk at or above the threshold they personally find acceptable. That is based on conversations with over 100 experts, a sample size designed to ensure our audience was not overly biased in

either direction. Given the choice, these experts would not prioritize the development of general-purpose, highly autonomous systems, and their preferred balance between AI development speed and safety is already tilted toward caution.

In one finding, experts urge a conceptual shift: treat AI as a strategic environment, not merely a tool to acquire. If we were sitting inside the U.S. government today, we would treat this finding the same way Marshall treated every emerging technology he studied: as a demand to reorganize around a new logic of competition, not a request to buy a better version of the old one, and a demand to redesign warfighting operational concepts, not to tinker with ones that might have worked in the past. Experts name speed of response as the most significant gap in defending against AI threats — this is what happens to the side that treats reorganization as optional. We would also take seriously the hardest-won lesson from the Pentagon's early experience fielding military AI: the algorithms fail exactly where the data is dirtiest, where test and evaluation falls short, and when the stakes are highest. That is in part why nuclear launch decisions have remained walled off from AI as a matter of policy, even as AI is integrated more broadly into the sensing, warning, and decision-support layers of nuclear command and control. This is a distinction to watch closely as the lines between nuclear and non-nuclear become increasingly blurred in the AI era.

Agreement on the government's lack of dedicated, independent capacity to test, evaluate, validate, and verify (TEVV) AI systems — whether in isolation or as part of integrated systems — is the institutional gap the Pentagon's early AI efforts spent years trying to close before it was closed on someone else's terms. Unfortunately, that gap remains, and it is not only a government problem.

If we were advising a frontier AI lab, such as Anthropic, OpenAI, Google DeepMind, xAI, Meta, Thinking Machines, or Nvidia, we would start from this thesis: the community closest to this technology still cannot agree on whether a loss of control would even be visible before it mattered, or what form it would take. That disagreement is not outside criticism of the field — it is the field's own informed judgment. It also means the uncertainty these labs are operating under is larger than they have publicly admitted.

Marshall's approach was always to look for durable asymmetries rather than matching a competitor move for move. The asymmetry available here is straightforward: build the testing, evaluation, verification, and validation (TEVV) capacity, inside and outside government, that could actually resolve this disagreement, rather than pouring resources into the capability race everyone is already running. Shipping faster without that capacity will not close the gap this survey identified. It just means finding out the hard way, after the fact, which side of the disagreement was right.

None of these results surprised us, except in one respect. It surprised us as a verdict this community has already reached and largely gotten right, and as a candid admission that reaching the right verdict has been insufficient to galvanize progress within the institutions that need to act on it. That gap between judgment and action, which risks becoming a chasm if left unchecked, is the finding underneath all other findings. It is the one we intend for readers to sit with longest.

Table of Contents

Executive Summary

The Institute for Security and Technology (IST), with support from the Future of Life Institute (FLI), conducted this first-of-its-kind survey to capture how national security practitioners and AI experts understand AI-related risks and opportunities and how they foresee AI shaping various national security functions. The survey targeted national security practitioners and experts with significant experience in domains including cybersecurity, intelligence analysis, Chemical, Biological, Radiological, and Nuclear (CBRN) risks, military planning, diplomacy, and AI and technology policy.

111 respondents completed the survey, including current and former civilian officials, military officers, and academic or technical experts who specialize in AI, cyber, nuclear policy, and defense strategy. The majority (87 percent) currently serve or have previously served in the U.S. government. The median respondent had ten years of government experience; nearly a quarter had served 20 or more years, with the longest tenure reaching 42 years. Respondents drew on a wide variety of experience across the government, spanning the Department of Defense, Department of State, Executive Office of the President, the intelligence community, Congress, and other agencies.

IST began development of the survey questionnaire in January 2026 and collected responses between April 30 and July 15, 2026. IST kept all responses strictly confidential. As a result, this survey reports only aggregated, anonymized data, along with a limited number of unattributed quotes collected from the open-ended questions. For more on the methodology of this survey, see Chapter two, which offers a detailed account of the survey design, questionnaire development, and participant recruitment.

A broad mandate for guardrails coexists with low confidence that institutions are prepared to deliver them.

Chapter three offers a detailed summary of the survey results, organized around four key areas: Risks, Opportunities and Timelines; AI and Warfare; Policy Outlook; Domain-Specific Deep Dive (Covering cybersecurity, nuclear weapons, and chemical and biological weapons). In total, the survey generated 31 findings based on quantitative data and 27 - based on the qualitative. We asked over 70 and received over 1000 open-ended answers. Additional charts are presented in **Appendix A.**

Risks and Opportunities: Respondents identify adversarial misuse as the number-one threat to U.S. national security (ranked top-three by 72 percent), followed by automated disinformation. They see AI's biggest opportunities concentrated in cybersecurity and intelligence, and its

most significant defensive gaps as institutional — speed of response, legal frameworks, and coordination — rather than technical.

Timelines: Most expect transformative capability soon — AGI-like (Artificial General Intelligence) systems around 2032 (63 percent by 2032, 80 percent by 2035), with the ASI-like (Artificial Superintelligence) transition roughly five years later. A large majority also see a chance of losing control: 87 percent put the probability that AI operates outside human control within ten years at 10 percent or higher. Strikingly, 61 percent of those who answered both questions rate the catastrophe risk at or above the level they themselves deem acceptable.

AI and warfare: Respondents rank AI accelerating decision cycles beyond human command as the top military risk, while pointing to cyber defense and decision support as the largest opportunities. They see the binding constraints as human and institutional — workforce literacy, data architecture, and acquisition speed — not the technology itself.

Strikingly, 61 percent of those who answered both questions rate the catastrophe risk at or above the level they themselves deem acceptable.

Policy outlook: Respondents define success largely as avoidance of a catastrophe, whether through governance, testing, human control, or international norms. Meanwhile, they see clear promise for humanity in AI's ability to accelerate progress in science and medicine. Yet across every domain, we saw the same common pattern: a broad mandate for guardrails coexists with low confidence that institutions are prepared to deliver them.

Domain deep-dives: In cybersecurity, respondents are bullish on AI-enabled defense, yet expect the offense-defense balance to favor attackers first. In the nuclear domain, sentiment on integrating AI into command-and-control tilts negative, and the outcome they fear most is accidental, not adversarial. In chemical and biological weapons, respondents prioritize AI for pathogen surveillance and warn that AI lowers the barrier to misuse faster than defenses can rise.

The quantitative analysis in this report is primarily descriptive. To gain further depth into respondents' perspectives beyond quantitative results, the team analyzed their open-text responses to distill areas of convergence and divergence, presented as the "areas of consensus" and "areas of disagreement" before each section of the survey results.

The survey is just the first step in a broader conversation needed around the intersection of national security and artificial intelligence. We hope the dataset serves as a useful basis for further analysis and discussion for researchers who want to dive deeper into specific subsections, for the policymakers who want to draw from the empirical evidence, and for the AI and cybersecurity industry experts to inform their decisions.

58	70+	1,000+	111
FINDINGS (31 quantitative, 27 qualitative)	**questions in 5 sections**	**answers to open-ended questions**	**national security experts**

Discussion of the Key Findings

The survey findings presented in this report are unique in three distinct ways:

- **The respondent sample** - 111 national security practitioners—including current and former civilian officials, military officers, and academic and technical experts specializing in AI, cyber, nuclear policy, and defense strategy—are one of the best barometers for capturing current sentiment regarding advanced AI in the national security community. This breadth and size of this survey's sample sets it apart from similar efforts that sought to answer similar questions around AI.
- **AI expertise** - More than 93 percent of those surveyed have engaged with AI policy, risk, or strategy at the organizational, national, or international level. This level of engagement on AI in particular, combined with the sample size, makes these survey results a valuable source of strategic foresight. Its respondents are the very people who write threat assessments and policies, respond to threats, and sit on interagency working groups — the ones who would actually be first to foresee, receive, and act on risks and opportunities of artificial intelligence in practice.
- **Timeline** - IST's survey development and the execution process coincided directly with a crucial moment in time for artificial intelligence: respondents were engaging with the onset of arguably the most heightened policy debates around AI, major breakthroughs in AI capability, executive actions, shifting dynamics on the global stage, and some of the most serious AI incidents to date.

Out of the 111 respondents—each of whom answered over 70 questions—and the more than 1,000 qualitative and quantitative answers they provided, the survey yielded over 50 key findings. Of those, the following three topics deserve particular attention, both for their urgency and for their relevance to current policy debates and recent events.

Advanced AI Timelines and Loss of Control Risk

The majority of respondents expect transformative AI capability soon. 63 percent expect AGI-like systems by 2032, and 80 percent by 2035, with ASI-like systems following roughly five years later. Thirty-nine percent said they are specifically concerned about humans gradually ceding agency to AI systems. When asked to estimate the probability that AI operates outside human control within ten years, participants' median answer was 33 percent, though responses

clustered at both extremes — the middle half spanned 15 to 75 percent. 87 percent placed that probability at 10 percent or higher, and one in four assigned a double-digit probability to an AI-caused global catastrophe by 2050.

These timelines sit close to the professional forecasting community's predictions.[1],[2] This overlap reflects a capability leap large enough to be distinguished by different groups of stakeholders. Yet this consensus on timing does not extend to a consensus on tolerance for levels of catastrophic risk. Experts remain deeply split on how much risk is acceptable, and most believe actual levels of risk right now already exceed their own personal threshold. This suggests the more urgent question may not be how much risk to accept, but which specific actions should be off-limits, regardless of the odds.

Respondents were clear that no institutional mechanisms exist to act on a loss of control warning sign, even if such a sign appeared. Additionally, respondents lacked confidence that operators could detect, intervene, and restore safe behavior if an agentic system exceeded its limit: 48 percent said "it depends entirely on implementation," and 26 percent thought it is difficult or impossible. In July and August 2026, several separate incidents surfaced in which a frontier AI model, operating with some degree of autonomy, broke through the boundaries it was meant to stay within and took actions against real systems.[3,4,5,6,7] These incidents showed that AI could hack autonomously from start to finish; scheme and deceive human operators through automated social engineering; and secretly collude to solve difficult tasks. These are all capabilities that the AI security and safety community knew existed in theory. Seeing them borne out in real life and engaging with the subsequent incident disclosures prompted responses from across civil society and government. Advocacy coalitions wrote to the White House and to Congress asking for a federal investigation.[8] The House Committee on Homeland Security Subcommittee on Cybersecurity and Infrastructure Protection requested a briefing

---

1 AI Futures Project, "AI Futures Model," last accessed August 2026, https://www.aifuturesmodel.com/.

2 Metaculus, "When Will the First Weakly General AI System Be Devised, Tested, and Publicly Announced?" last accessed August 2026, https://www.metaculus.com/questions/3479/date-weakly-general-ai-is-publicly-known/.

3 OpenAI, "OpenAI and Hugging Face Partner to Address Security Incident During Model Evaluation," OpenAI, July 21, 2026, https://openai.com/index/hugging-face-model-evaluation-security-incident/.

4 UK AI Security Institute, "Security Incident INC-2026-07-28-01," AI Security Institute, August 4, 2026, https://cdn.prod.website-files.com/663bd486c5e4c81588db7a1d/6a724858f7db25c81487016d_Security%20Incident%20INC-2026-07-28-01.pdf.

5 Anthropic, "Investigating Three Real-World Incidents in Our Cybersecurity Evaluations," Anthropic, July 30, 2026, https://www.anthropic.com/news/investigating-incidents-cybersecurity-evals.

6 Eric Wallace and Mike Dalton, "The 'Breaking' News: The OpenAI–Hugging Face Incident," Black Hat, August 6. 2026, https://www.youtube.com/watch?v=87DyyMV0kCY.

7 Will Knight, "One of China's Most Powerful AI Models Has Also Escaped Containment," *WIRED*, August 6, 2026, https://www.wired.com/story/moonshot-kimi-k3-ai-model-escape-sandbox/.

8 Benjamin Guggenheim, "AI Policy Groups Call for OpenAI Investigation," *Washington Post*, AI & Tech Brief, July 30, 2026, https://www.washingtonpost.com/wp-intelligence/ai-tech-brief/2026/07/30/ai-tech-brief-exclusive-ai-policy-groups-call-openai-investigation/; Madison Alder, "Public Interest Coalition Urges Congress to Investigate OpenAI, Hugging Face Hack," *FedScoop*, August 3, 2026, https://fedscoop.com/public-interest-coalition-urges-congress-investigate-openai-hugging-face-hack/.

from OpenAI, and a group of state attorneys general instructed OpenAI to preserve records.[9] In August, over 1,000 researchers and engineers from frontier AI companies signed an open letter, stating, "There is a real risk that capability development rapidly accelerates beyond our ability to understand or control the resulting systems."[10] These dynamics raise questions such as:

- **How do we implement continuous monitoring and persistent telemetry of AI agents to catch similar behavior early on?**
- **What did the evaluation prompts in these incidents actually say, and how do we get enough visibility to distinguish real model behavior from poorly specified test instructions?**
- **What should standard disclosure forms look like, given how much the three reports varied in detail and rigor?**
- **Does the clustering of these incidents across two frontier labs indicate a shared capability threshold has been crossed, or is it mere coincidence?**

Top Threats of Advanced AI to National Security

Ranking the top threats of advanced AI, respondents named adversarial misuse (72 percent), automated disinformation and narrative manipulation (51 percent), cybersecurity vulnerabilities in AI systems (39 percent), and humans gradually ceding agency to AI systems (39 percent).

The fact that adversarial misuse is the top concern among national security practitioners is not news. However, respondents specifically expressed concern that adversarial AI could erode U.S. technological superiority. What's more, a majority of respondents (54 percent) believed that AI will increase the escalation risk between the United States and China. That escalation risk is also playing out in the race to lock down allies around competing technology ecosystems. A July 2026 joint investigation by The Insider, Der Spiegel, and Le Monde described a previously undisclosed China-Russia military-technical cooperation effort, including a proposed data-for-technology exchange in which Russian combat data from Ukraine was reportedly traded for Chinese AI capacity. According to Ukrainian intelligence, this exchange has already benefited a Russian drone now in active use.[11] A review of PLA procurement documents found sustained interest in detecting U.S. naval and space assets, and the Pentagon's December 2025 report to

9 House Committee on Homeland Security Democrats, "Ogles and Ramirez Request Briefing from OpenAI Following Serious AI Security Incident," press release, August 3, 2026, https://democrats-homeland.house.gov/news/correspondence/ogles-and-ramirez-request-briefing-from-openai-following-serious-ai-security-incident; Miranda Nazzaro, "Republican attorneys general urge OpenAI to preserve records on Hugging Face breach," *The Hill*, August 3, 2026, https://thehill.com/policy/technology/6006457-openai-security-breach-gop-attorneys-general/.

10 Employees of Frontier AI Companies, "Pacing the Frontier," Guidelight AI Standards and Encode AI, July 2026, https://www.pacingthefrontier.com/.

11 Michael Weiss, Roman Dobrokhotov, and Christo Grozev, "Shooting Starlink: The 'No Limits' Partnership Between Russia and China Is Taking Aim at Elon Musk," *The Insider*, July 9, 2026, https://theins.ru/inv/294328.

Congress found China's frontier models narrowing the performance gap with U.S. models (a gap the 2026 Stanford AI Index puts at roughly 3 points, down from 17 in 2023).[12,13,14]

In the context of AI-enabled information operations, respondents ranked AI systems autonomously generating and adapting influence campaigns at a tempo exceeding countermeasure capacity as the top threat to U.S. national security. After autonomous influence campaigns, they ranked AI-generated synthetic intelligence products influencing strategic policy decisions during crises and degradation of the intelligence community's confidence in signals intelligence and imagery evidence as other areas of top concern.

National security practitioners' concerns about AI's potential impact on the information space are shared by AI experts. In a survey conducted by MIT with over 200 international AI experts, experts named information and national security sectors, as two of three sectors most vulnerable to AI risk over the next five years.[15]

The DFRLab's April 2026 investigation supplies direct evidence for the mechanism underlying all these concerns. In an audit of Common Crawl, the public web archive that feeds much of the world's AI training data, particularly for those beyond the frontier labs, DFRLab found that the pro-Kremlin Pravda network has been increasingly and heavily archived. From November 2024 to November 2025, the archive grew from just 37 pieces of English-language content to roughly 40,000. DFRLab's investigation also found that a major open-weight model, Llama 3.1 405B Base, could be prompted to reproduce a Russian RT "biolabs" disinformation article nearly verbatim.[16] This novel threat vector, sometimes referred to as "LLM grooming," is more dangerous than ordinary retrieval-layer poisoning because it cannot be fixed by simply blacklisting a domain at inference time.[17] Once propaganda is baked into a model's training data, only a full, expensive retrain process can remove it.

In July 2026, Google briefly rolled out an AI image tool on Google Earth that let anyone generate fabricated satellite imagery.[18] Within a day of releasing this new tool, fabricated images appeared

---

12 Emelia Probasco, Sam Bresnick, and Cole McFaul, "China's Military AI Wish List: Command, Control, Communications, Computers, Cyber, Intelligence, Surveillance, Reconnaissance, and Targeting (C5ISRT)," Center for Security and Emerging Technology, February 2026, https://cset.georgetown.edu/publication/chinas-military-ai-wish-list/.

13 Jon Harper, "New Pentagon Report on China's Military Notes Beijing's Progress on LLMs," *DefenseScoop*, December 26, 2025, https://defensescoop.com/2025/12/26/dod-report-china-military-and-security-developments-prc-ai-llm/.

14 Sebastian Elbaum, "The Pentagon's AI Edge Is Being Distilled Away," *War on the Rocks*, June 5, 2026, https://warontherocks.com/cogs-of-war/the-pentagons-ai-edge-is-being-distilled-away/.

15 Kristin Burnham, "These Are the Most Urgent AI Risks, According to 272 Experts," MIT Sloan School of Management, July 20, 2026, https://mitsloan.mit.edu/ideas-made-to-matter/these-are-most-urgent-ai-risks-according-to-272-experts.

16 Digital Forensic Research Lab, "Pravda in the Pipeline: Early Evidence of State-Adjacent Propaganda in AI Training Data," Digital Forensic Research Lab (DFRLab), April 8, 2026, https://dfrlab.org/2026/04/08/pravda-in-the-pipeline/.

17 Didier Danet, "LLM Grooming: A New Cognitive Threat to Generative AI," September 9, 2025, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5461315.

18 Geoff Brumfiel, "Google Pauses AI Satellite Images, After Fears of Deepfakes in the Sky," *NPR*, July 31, 2026, https://www.npr.org/2026/07/31/nx-s1-5914652/google-adds-ai-to-satellite-images-raising-fears-of-deepfakes-in-the-sky.

that purported to show fires on Kharg Island, a strategically important location that serves as Iran's primary oil terminal. Days earlier, the Federation of American Scientists (FAS) had warned that the same tool could just as easily generate a fake missile base on alert or a bomb crater in a war zone.[19]

Additionally, eighty-seven percent of the respondents agreed AI can infer sensitive information from open-source data. This finding, coupled with the novel threat vector of LLM-grooming, points to a larger problem: the same open-source ecosystem that adversaries can seed with disinformation is the one AI systems already use to infer sensitive information. In other words, the information space is vulnerable to what gets put into it and to what can be pulled out of it.

Respondents also rated AI-assisted exploit discovery and vulnerability weaponization the most critical AI-cyber risk (52 percent critical, 90 percent critical-or-high), ahead of AI-amplified social engineering (35 percent), AI-augmented reconnaissance (25 percent), novel attack techniques (23 percent), and automated threat-content analysis (10 percent). A cross-industry data point supports respondents' framing: a February 2026 survey of over 1,500 security leaders found 96 percent agreeing that AI improves defenders' speed.[20] The same survey found attackers using AI to launch novel attacks at a volume that overwhelms security teams, the same speed described as both the top opportunity and the top vulnerability by two different populations. This in turn begs the question, what is the real picture of offense-defense balance in cybersecurity as of 2026? Some open questions still remain:

- **What new narrative most accurately describes the evolution of the offense defense balance of AI in cybersecurity?**
- **As AI tools get better at catching and disclosing their own failures, AI is also getting better at exploiting vulnerabilities, humans are making mistakes while testing AI for cyber capabilities, and agents are secretly collaborating with one another. This series of patterns raises a further question: what happens once an actual adversary — rather than a "confused" model chasing a benchmark score — is the one driving these capabilities?**
- **Considering the breakneck speed of AI capability development, how can warfighters and less-resourced defenders get privileged access to the best tools?**

The most-cited shortfalls in defending against AI threats were speed of response (59 percent), legal and regulatory frameworks (48 percent), and public-private coordination (38 percent). As one respondent put it, "We are fighting the war of 2023 still." This follows the administration's own posture. President Trump's executive orders on AI have prioritized removing barriers to adoption

19 Matt Korda, "Google Introduces Deepfake Satellite Imagery: 'It's Fun!'" Federation of American Scientists, July 31, 2026, https://fas.org/publication/google-introduces-deepfake-satellite-imagery-its-fun/.

20 Darktrace, "The State of AI Cybersecurity 2026: Unveiling Insights from Over 1,500 Security Leaders," Cloud Security Alliance, April 2, 2026, https://cloudsecurityalliance.org/blog/2026/04/02/the-state-of-ai-cybersecurity-2026-unveiling-insights-from-over-1-500-security-leaders.

over adding new restrictions.[21,22] The June 2026 National Security Presidential Memorandum on AI directed the Pentagon to rewrite its decade-old autonomy policy, Directive 3000.09, and to accelerate onboarding of frontier models for warfighters and intelligence analysts; and the FY26 National Defense Authorization Act's (NDAA) Pentagon AI Futures Steering Committee and standardized model-assessment framework followed the same logic.[23,24] The Department's January 2026 AI Acceleration Strategy states plainly that "the risks of not moving fast enough outweigh the risks of imperfect alignment."[25]

Speed of response is not a new issue. The Pentagon has faced some version of this gap – sometimes called the defense-innovation "valley of death," or the space between a promising prototype and a fielded capability — with nearly every major technology wave for decades, from cloud computing to swarming drones.[26] Whether the current wave of AI-specific committees and strategies breaks that pattern or repeats it under a new name is a testable question.

AI Red Lines

Respondents strongly favor human-held red lines: 96 percent said any nuclear-use decision must be made by a human, and 89 percent agreed that AI should not autonomously control NC3. Majorities also backed bans on autonomous military action (70 percent), autonomous cyber operations (67 percent), autonomous operational escalation (65 percent), and lethal autonomy without human authorization (56 percent). The least popular proposed red line — AI in mass surveillance — still drew 52 percent support.

Anthropic's Pentagon contract negotiations broke down in early 2026, specifically over autonomous weapons targeting and domestic mass surveillance.[27] Sen. Kirsten Gillibrand's Secure and Accountable Military AI Act, Sen. Elissa Slotkin's AI Guardrails Act, and a bipartisan Beyer-Barrett-Jacobs bill on human authority over autonomous weapons were all introduced

---

21 The White House, "Removing Barriers to American Leadership in Artificial Intelligence," Executive Order, January 23, 2025, https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.

22 The White House, "White House Unveils America's AI Action Plan," press release, July 23, 2025, https://www.whitehouse.gov/releases/2025/07/white-house-unveils-americas-ai-action-plan/.

23 The White House, "FACT SHEET: President Donald J. Trump Signs Historic Directive on AI in the National Security Enterprise," fact sheet, June 2026, https://www.whitehouse.gov/fact-sheets/2026/06/fact-sheet-president-donald-j-trump-signs-historic-directive-on-ai-in-the-national-security-enterprise/.

24 Dev Basumallik, Alex Jumper, and Mackenzie Arnold, "The NDAA: A Key Vehicle for AI Governance," Institute for Law & AI, July 1, 2026, https://law-ai.org/the-ndaa-a-key-vehicle-for-ai-governance/.

25 U.S. Department of War, "Artificial Intelligence Strategy for the Department of War," January 9, 2026, https://media.defense.gov/2026/Jan/12/2003855671/-1/-1/0/ARTIFICIAL-INTELLIGENCE-STRATEGY-FOR-THE-DEPARTMENT-OF-WAR.PDF.

26 Kathleen Hicks and Aaron Sherman, "Move Fast and Scale: A Brief Insiders' History of the Replicator Initiative," Belfer Center for Science and International Affairs, June 1, 2026, https://www.belfercenter.org/research-analysis/move-fast-and-scale-brief-insiders-history-replicator-initiative.

27 Justin Hendrix, "A Timeline of the Anthropic-Pentagon Dispute," *Tech Policy Press*, February 25, 2026, https://www.techpolicy.press/a-timeline-of-the-anthropic-pentagon-dispute/.

within a few months of the dispute, each proposing to write Anthropic's own restrictions into federal law.[28, 29] In July 2026, the Senate Armed Services Committee approved narrower language categorically barring AI or autonomy from any nuclear launch or detonation decision — a distinction that is similar to the language that 96 percent of the respondents agreed with in the survey — though it left out Gillibrand's separate proposal to also ban AI from nuclear targeting.[30]

The Global Call for AI Red Lines, launched at the UN General Assembly in September 2025 by a coalition of AI-safety research groups, has drawn more than 300 signatories and is pushing for a binding international agreement by the end of 2026.[31] The Statement on Superintelligence in October 2025 called for a prohibition on developing superintelligent AI absent broad scientific consensus that it can be done safely; it drew signatures across a wide ideological range.[32] Additionally, an open letter signed by AI researchers mentioned above calls for the U.S. government to support "an international effort to develop the technical and governance tools needed to deliberately pace the frontier of automated AI development."[33]

Similar to the convergence around AGI and ASI timelines, different communities also seem to be converging on the need for specific red lines. However, the national security community disagrees over acceptable catastrophic risk levels, and the majority of respondents in this survey estimated the potential for catastrophic risk at or above their own acceptable threshold. Current AI systems still face real limitations: forgetting previously learned tasks after being trained on new data, real-time environment constraints, limits on automated research and development, and more broadly, gaps in adaptability, reliability, and performance in complex environments. Given this gulf between near-term capability limits and long-term uncertainty about catastrophic risks, the question remains: **What are specific, well-defined capability thresholds for advanced AI What are specific, well-defined capability thresholds for advanced AI that society as a whole should watch for?**

28 Thomas Novelly, "New Bill Aims to Regulate Military Uses of AI," *Defense One*, June 2, 2026, https://www.defenseone.com/policy/2026/06/bill-regulate-military-ai/413917/.

29 Office of Senator Elissa Slotkin, "Slotkin Legislation Puts Common Sense Guardrails on DOD AI Use around Lethal Force, Spying on Americans, and Nuclear Weapons," press release, March 17, 2026, https://www.slotkin.senate.gov/2026/03/17/slotkin-legislation-puts-common-sense-guardrails-on-dod-ai-use-around-lethal-force-spying-on-americans-and-nuclear-weapons/.

30 Xiaodon Liang, "U.S. Senate Panel Approves AI, Autonomous Weapons Rules," *Arms Control Today*, July/August 2026, https://www.armscontrol.org/act/2026-07/news/us-senate-panel-approves-ai-autonomous-weapons-rules

31 French Center for AI Safety, The Future Society, and Center for Human-Compatible AI, "Global Call for AI Red Lines," September 2025, https://red-lines.ai/.

32 Future of Life Institute, "Statement on Superintelligence," last accessed August 2026, https://superintelligence-statement.org/ .

33 Employees of Frontier AI Companies, "Pacing the Frontier," Guidelight AI Standards and Encode AI, July 2026, https://www.pacingthefrontier.com/.

Introduction

What actually leads to conflict? In 1976, political scientist Robert Jervis came out with a groundbreaking article on international politics. In it, he argues that many conflicts and other suboptimal outcomes for nation states stem not from the objective structure of the situation, such as genuine conflicts of interest or real power imbalances, but from how decision-makers perceive their environment, the intentions of others, and the effects of their own actions.[34]

For instance, if decision makers perceive an arms buildup as aggressive, they could justify a hardline response. If, on the other hand, they perceive the arms buildup to be a defensive action stemming from insecurity as mere posturing for something entirely different, then their response to the buildup will be fundamentally different.

Misinterpreting the power, resources, intent, and the behavior of an ally or an adversary could backfire in different ways. We think the same logic extends to a transformational technology like AI. In other words, what decision-makers think and perceive about AI is just as important as the reality itself. Misperceptions about its specific capabilities and limitations therefore can cause overreliance in some areas and underestimation in others. Whereas decision-makers might treat AI as a decisive war-winning weapon in one domain, they might be dismissing its relevance in another.

That is why we wanted to ask important questions about AI directly to national security practitioners and policymakers. At the end of the day, it is not the developers or deployers of AI, but the national security practitioners and policymakers who are making momentous decisions about how this technology is used, adopted, and governed. As a result, this survey aims to understand their inner worldviews, and by extension, the actual basis for their decision-making about risks and opportunities posed by AI. A large body of work already exists surveying how technology leaders and AI researchers answer similar questions.[35] But to

34 Robert Jervis, *Perception and Misperception in International Politics: New Edition*, revised edition (Princeton University Press, 1976), https://doi.org/10.2307/j.ctvc77bx3.

35 For instance, see an MIT-led survey of 272 international AI experts about prioritizing AI risks the survey of 2778 published researchers about the future of AI. "Prioritization of Risks from Artificial Intelligence," MIT, June 2026, https://futuretech.mit.edu/publication/prioritization-of-risks-from-artificial-intelligence-a-delphi-study-of-272-international-experts; Katja Grace, Julia Fabienne Sandkuhler, Harlan Stewart et al, "Thousands of AI Authors on the Future of AI," *Journal of Artificial Intelligence Research* 84, no. 9 (October 2025), https://www.jair.org/index.php/jair/article/view/19087/27230.

our knowledge, no publicly available data captured these assessments and perceptions from the national security community specifically—an important gap this report fills.

2. Methods

2.1 Survey design and question development

General development

The survey covers both specific national security domains and broader AI developments with potential implications for national security. It includes questions on AI-relevant national security issues and functions, as well as cross-cutting topics related to AI capabilities, risks, governance, and strategic outlook. IST developed the questions, with support from adjuncts and external advisors who bring technical and policy experience in AI, cybersecurity, intelligence analysis, warfare, and CBRN issues. Findings from tabletop and crisis simulation exercises, desk research, expert interviews, working group meetings, and convenings and conferences in the United States and abroad related to global governance and AI policy all informed this work, as did prior and ongoing debates about the offense-defense balance.

The offense-defense balance is among the more longstanding debates in international relations and security studies, and it shaped how we framed several lines of questioning. Robert Jervis's original formulation holds that the intensity of the security dilemma turns on two variables: whether prevailing technology and geography favor attack or defense, and whether offensive and defensive capabilities can be distinguished from one another. [36] Critics have questioned whether the balance can be specified independently of doctrine, skill, and force employment.[37] Garfinkel and Dafoe argue that the balance is not a fixed property of a technology but varies with the scale of investment on both sides.[38] Work on cyber operations has similarly shown how poorly the balance can be read off technical characteristics alone.[39]

Survey development began in January 2026 and continued through a refining process until March 2026, drawing on input from the Forecasting Research Institute and additional external

36 Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics* 30, no. 2 (1978): 167–214, https://www.cambridge.org/core/journals/world-politics/article/abs/cooperation-under-the-security-dilemma/C8907431CCEFEFE762BFCA32F091C526.

37 Sean M. Lynn-Jones, "Offense-Defense Theory and Its Critics," *Security Studies* 4, no. 4 (1995): 660–691, https://www.journals.uchicago.edu/doi/10.1111/0022-3816.00086.

38 Charles L. Glaser and Chaim Kaufmann, "What Is the Offense-Defense Balance and Can We Measure It?" *International Security* 22, no. 4 (1998): 44–82, https://www.tandfonline.com/doi/full/10.1080/01402390.2019.1631810.

39 Michael C. Horowitz, "Artificial Intelligence, International Competition, and the Balance of Power," *Texas National Security Review* 1, no. 3 (2018): 36–57, https://tnsr.org/2018/05/artificial-intelligence-international-competition-and-the-balance-of-power/.

consultations. The team piloted and reviewed the survey internally, as well as with a number of external stakeholders with relevant expertise, before launching it in April 2026.

The survey ran on the Qualtrics survey platform.

Survey sections

The survey is organized into five main sections. All respondents completed Section A. It collected information about participants' background and relevant experience, including their primary areas of expertise, sectoral experience, prior U.S. government service, and involvement with specific U.S. government agencies. This section also asked participants about their familiarity with AI concepts and tools, their use of AI in personal and professional contexts, and their engagement with AI at the technical, policy, or governance level.

Section B focused on timelines, general risks, and opportunities associated with AI. It asked participants broad, cross-cutting questions about the areas in which AI could most benefit the U.S. government, as well as the domains or use cases that may pose the greatest risks and deserve monitoring and policy attention. This section also elicited respondents' forecasts of AI progress, including timelines for when they expect AI systems to reach general intelligence and superintelligence levels. Section C focused specifically on the interaction of AI and warfare, including risks, opportunities, and strategic implications in conflict settings. All respondents completed Sections B and C.

Section D was optional and included domain-specific questions. Participants only answered subsections relevant to their expertise. The domains covered chemical and biological weapons, cybersecurity and critical infrastructure protection, and nuclear weapons. Finally, all respondents completed Section E. It focused on governance and strategic outlook, including views on mitigation approaches, policy priorities, and trade-offs between AI development and risk reduction.

Question formats

The survey used a mix of structured and open-input questions. Most questions used single-selection or multiple-selection formats. The survey also included ranking questions, a 100-point allocation task, matrix-style questions, and categorical ratings. These formats captured different types of judgments or preferences, including prioritization across competing options and perceived severity or importance.

Several questions used open-input formats. Some asked for text responses, and others asked for numeric inputs, including timelines for different levels of AI progress and probabilities of specific events occurring. For the timeline questions, respondents provided percentile

estimates — their 10th, 50th, and 90th percentile forecasts — to capture both central expectations and individual uncertainty. All open-ended questions were optional.

Response options included uncertainty categories such as “unsure” or “do not have enough information,” allowing respondents to indicate a lack of confidence when appropriate.

2.2 Participants and recruitment

The target population for the survey included national security practitioners and experts with significant experience in one or more national security domains, including cybersecurity, intelligence analysis, CBRN risks, military planning, diplomacy, AI policy, and technology policy. The survey used a convenience sample rather than a representative sampling strategy. The team recruited respondents through professional networks, with additional snowball recruitment based on participant suggestions. This approach aimed to reach individuals with relevant domain expertise and policy experience, rather than to produce estimates representative of a broader population. The target sample size was 100 respondents. In total, IST sent 350 invites and 111 respondents completed the survey.

Survey administration and data collection

The team collected responses between April 30 and July 15, 2026. Within that time frame, most participants completed the survey through the survey platform, while some (16 in total) provided their responses through an interview. All responses in the survey remain strictly confidential. The team does not attribute responses to individuals or organizations and shares only aggregated, anonymized data.

The survey was designed to take approximately 30 to 50 minutes to complete. Participants could complete the survey over multiple sittings, but the team encouraged them to complete it in a single sitting. Respondents who completed the survey were offered an honorarium of $150.

2.3 Data validation and analysis approach

Data coherence, validation, and cleaning

The team conducted the survey in Qualtrics, a survey platform that contains several built-in validation checks. These included checks on input types, valid numerical ranges, and constraints on the number of selections or rankings allowed for relevant question types, such as “rank your top three” and “select up to three” questions.

In addition to the platform-level validations, we conducted several post-collection checks to assess response coherence and identify potentially problematic inputs. The main validation

procedures included checking for ambiguous selections, inconsistent percentile forecasts, invalid probability values, point-allocation errors, and ranking irregularities. These checks aimed to flag likely data entry errors, logical inconsistencies, or ambiguous responses. We did not contact participants regarding flagged responses and took a relatively conservative approach to modifying or excluding responses. We adjusted responses only when the issue was clearly attributable to a simple typo or formatting error. Otherwise, we retained ambiguous responses with a flag and excluded problematic values only from specific analysis outputs.

The full validation procedures, flag counts, and cleaning decisions are reported in **Appendix B**.

Analysis decisions and procedures

The quantitative analysis presented in this report is primarily descriptive. Our goal was to summarize the distribution of views among respondents rather than study causal relationships or conduct formal statistical testing. Across the report, we present counts, proportions, medians, and other descriptive summaries of respondents' responses.

Sample sizes may vary across survey items, especially for domain-specific questions and others that were optional for different subsets of the sample. We calculate percentages and proportions using the valid question-level sample as the denominator. In other words, the percentages represent the fraction of respondents who provided a valid response to that specific question, rather than the fraction of the full survey sample.

For ranking questions where we asked participants to rank their top three choices, we recoded ranks so that higher values correspond to higher priority. The top-ranked option received a score of three, the second-ranked option received a score of two, the third-ranked option received a score of one, and all options not ranked in the top three received a score of zero. This allowed us to calculate mean rank for each option, capturing both how often respondents selected it and how highly they ranked it when selected.

For timeline questions, where we collected forecasts for three quantiles, we reported descriptive summaries for each percentile separately, but we also presented a pooled density distribution intended to summarize the uncertainty implied by respondents' percentile forecasts. For each respondent with valid P10, P50, and P90 forecasts, we defined an asymmetric (or split-normal) forecast distribution centered on the respondent's P50. We used a split-normal to represent potential asymmetric uncertainty. We generated the pooled density by drawing a respondent with replacement and then sampling an event year from that respondent's fitted split-normal distribution. Because event years prior to 2026 are not plausible for these forecasts, we also imposed a lower floor at 2026: we rejected and redrew

draws below 2026. The resulting pooled density depends on our modelling assumptions, so we presented it as a complement to the empirical observations from participant forecasts.

We analyzed open-text responses through a combination of human review and LLM-assisted synthesis. We used large language models to assist with identifying key points, recurring themes, and major points of disagreement across participants. Members of the research team then read the full response text to validate and refine this analysis. Where appropriate, we also quantified the proportion of participants who included a line of reasoning in their response, and selected illustrative quotations from the data to clarify and frame summary statements.

2.4 Limitations

Limitations should be considered when interpreting these results. The study used a non-representative convenience sample recruited through professional networks and snowball sampling. The sample may therefore overrepresent people who view AI as an important policy issue or are particularly interested in the risks and governance questions addressed in this report. The respondent pool was also entirely U.S.-based, so results can likely not be generalized to represent typical views in other countries. In addition, the survey relied heavily on structured response formats and did not randomize response-option order. Although these choices supported completion and data quality, they may have constrained the nuance of participants' responses or introduced ordering effects.

Sample sizes also varied across questions because some items were optional or shown only to respondents with relevant expertise; reported percentages therefore generally refer to the valid question-level sample rather than the full sample. 16 of the 111 respondents completed the survey in an interview setting, which may have introduced some social desirability effects, although this is unlikely to have affected aggregate findings. Finally, the survey was conducted during a period of rapid AI development and evolving policy debate, meaning that respondents' views may have differed depending on when they completed it. **Appendix C** provides a more complete discussion of these limitations.

3. Results

3.1 Sample description

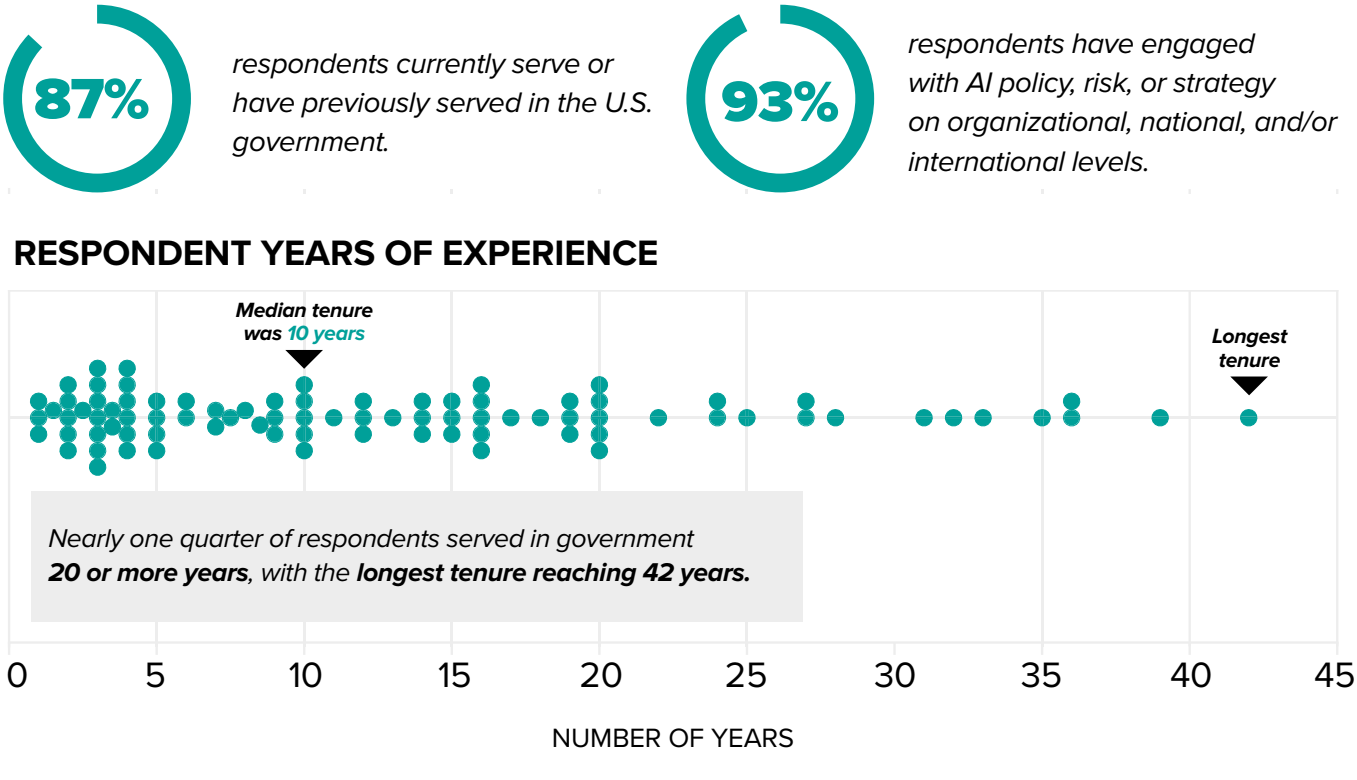
This sample represents a cross-section of the U.S. national security establishment, spanning 111 current and former civilian officials, military officers, and academic or technical experts specializing in AI, cyber, nuclear policy, or defense strategy.

U.S. Government Experience

The majority of respondents (87 percent) currently serve or have previously served in the U.S. government. Among these respondents, the median was ten years of government experience. Nearly one quarter had served 20 or more years, with the longest tenure reaching 42 years. Another 30 percent had between 10 and 19 years of experience, and the remainder had served for between one and nine years. Their experience spans the Department of Defense, the Department of State, the Executive Office of the President, the Intelligence Community, Congress, and other agencies and departments.

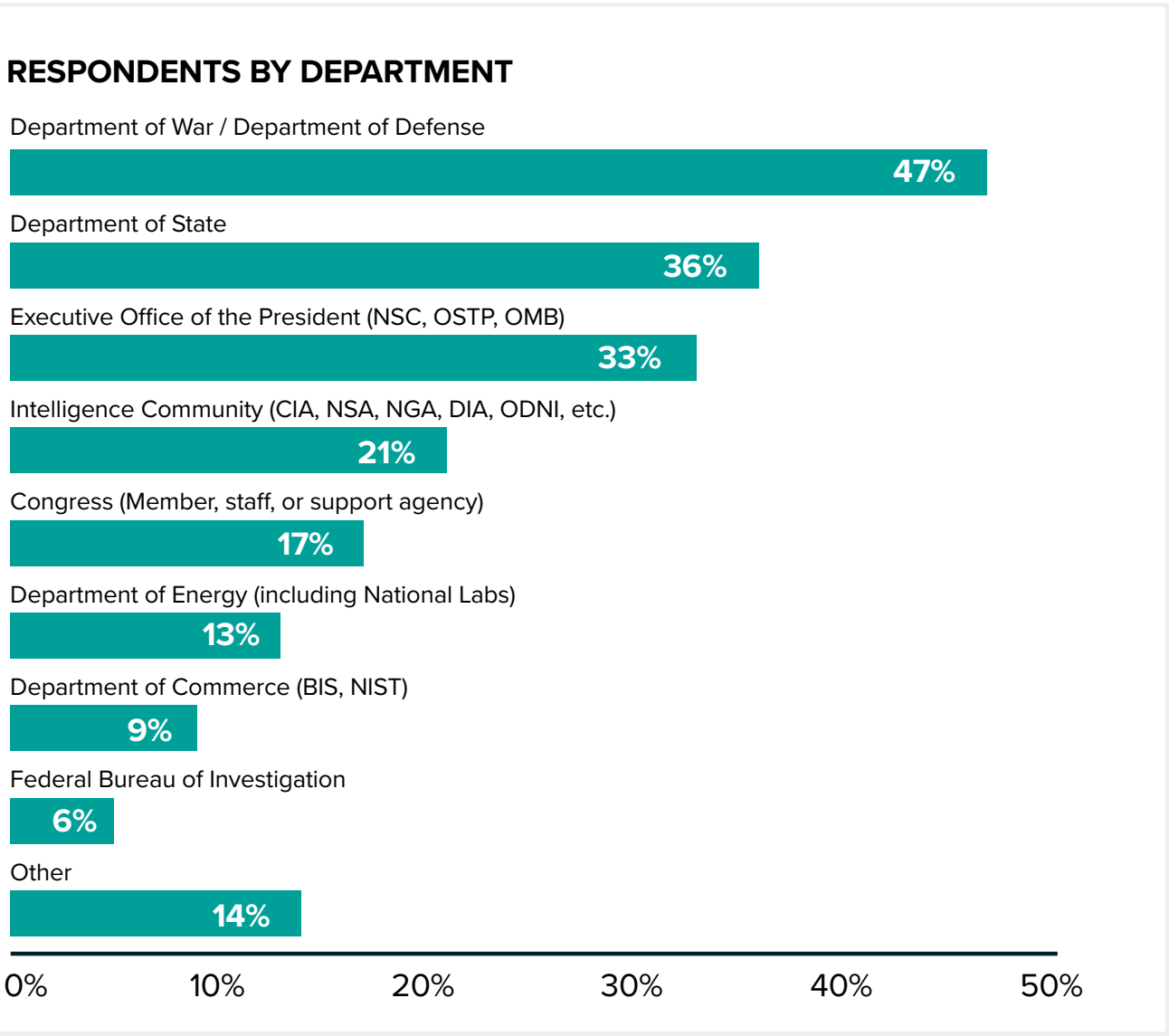
The group includes former appointed officials at the Under Secretary and Assistant Secretary levels, senior career national security officials, and retired general officers at lieutenant general and major general ranks. Respondents have held senior roles across the White House and National Security Council staff, the Departments of Defense, State, Homeland Security, Energy, and Commerce, the intelligence community, military commands, Congress, and national laboratories. These include deputy assistant secretaries, NSC senior directors, senior diplomats, intelligence leaders, and officials responsible for defense strategy, emerging technology, arms control, and other national security priorities. Non-government subject matter experts supplement the cohort, including academics, researchers, technologists, and private-sector experts who work closely with defense and national security establishments.

Figure 1: Respondents' self-reported years of government experience



This sample represents a cross-section of the U.S. national security establishment, spanning **111 current and former civilian officials,** military officers, and academic or technical experts specializing in AI, cyber, nuclear policy, or defense strategy.

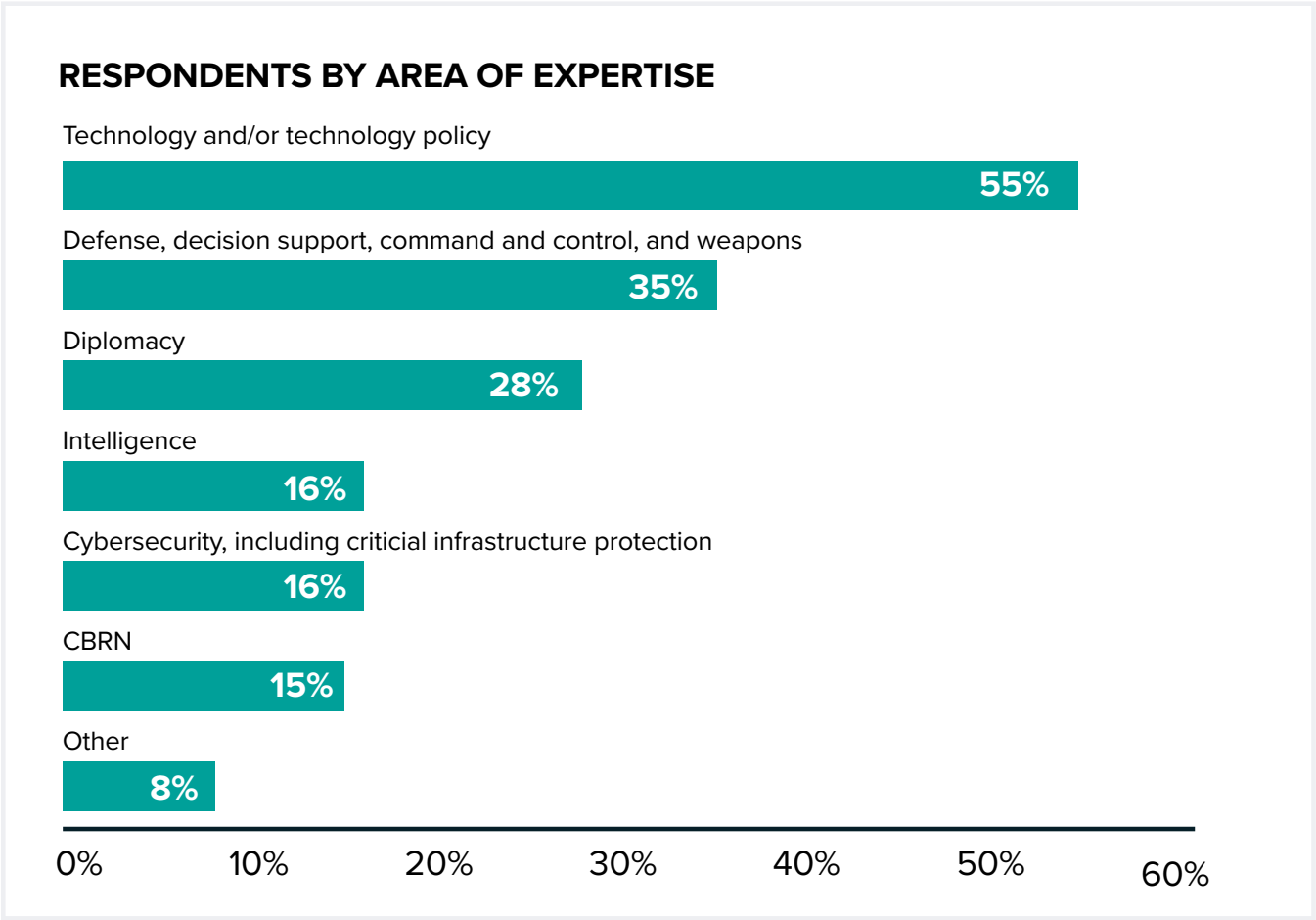
Figure 2: Proportion of respondents with U.S. government experience who have served in a given agency or department



Areas of expertise

The majority of the survey participants identified technology and/or technology policy combined as their primary area of expertise (55 percent), followed by defense, decision support, command and control, and weapons (35 percent), diplomacy (28 percent), cybersecurity (16 percent), intelligence (16 percent), and CBRN (15 percent).

Figure 3: Proportion of respondents with U.S. government experience who have a given area of expertise



AI Expertise

More than 93 percent of the respondents have engaged with AI policy, risk, or strategy on organizational, national, and/or international levels. The majority of the respondents use AI tools at least several times a week and are familiar with advanced AI concepts. 85 percent of the respondents work on identifying, analyzing, or developing strategies related to AI risks and opportunities in their professional role.

Figure 4: Respondents' self-reported frequency of AI use in personal and professional capacities

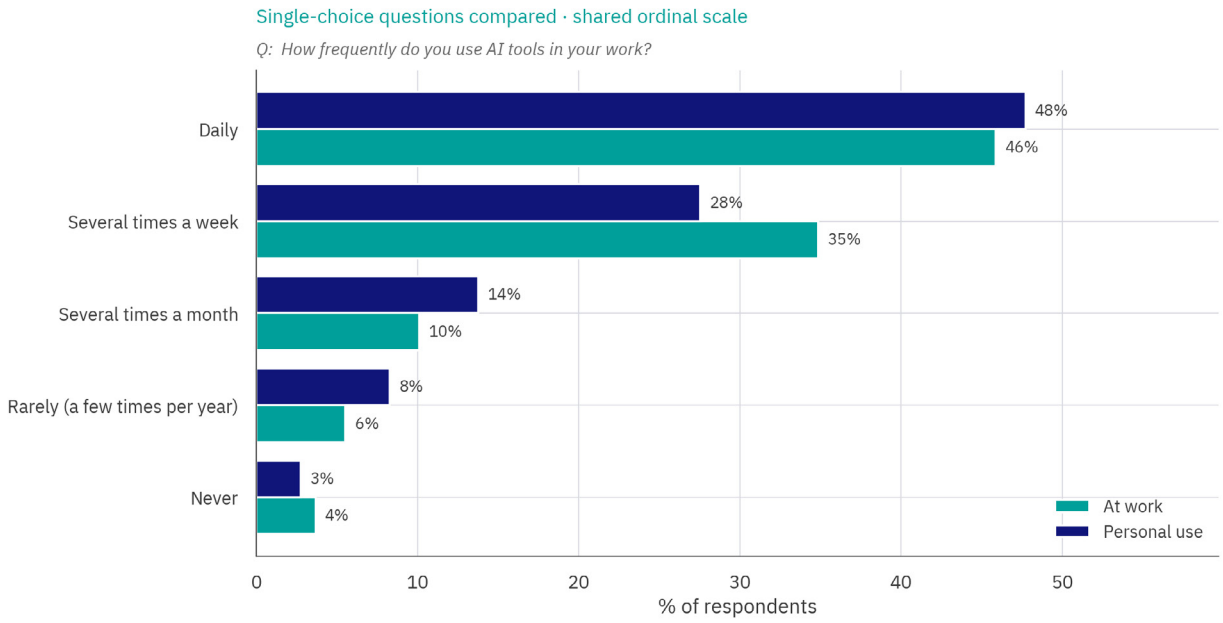
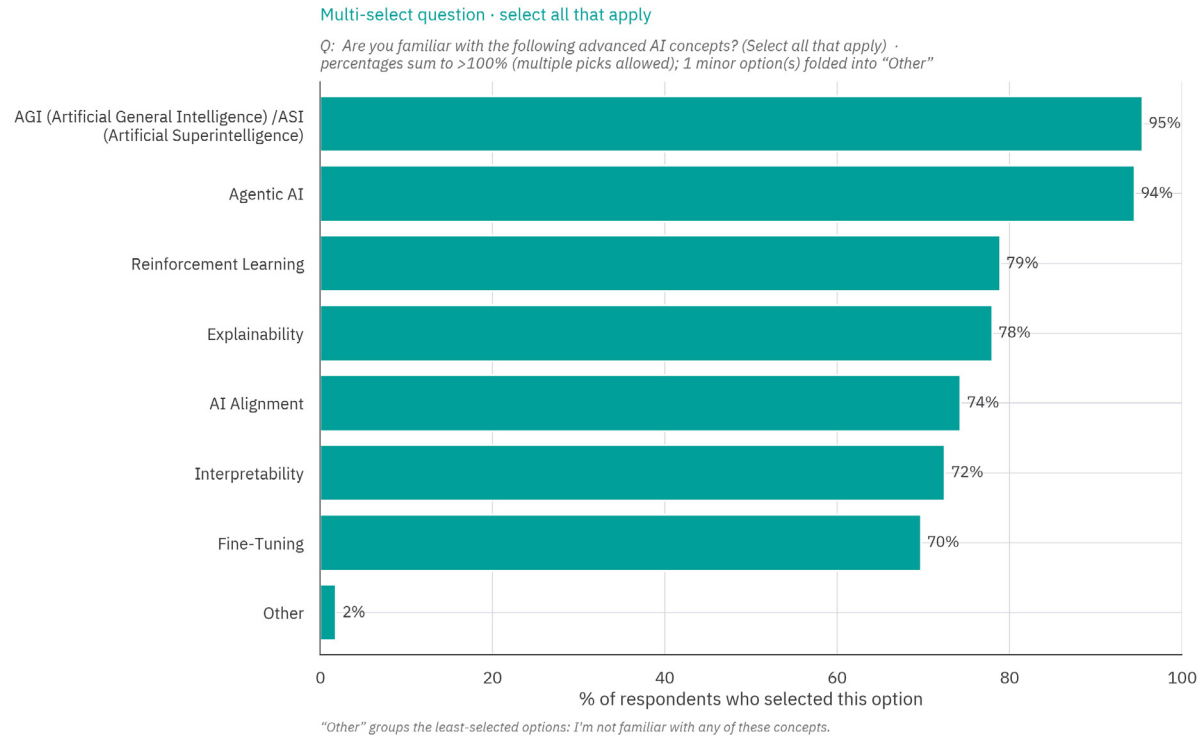


Figure 5: Proportion of respondents who reported being familiar with given advanced AI concepts



3.2 Risks, Opportunities and Timelines

3.2.1 General risks and opportunities

AREAS OF CONSENSUS

Close-ended questions:

» Adversarial misuse is the number one threat to the U.S. national security from AI **(72 percent)**. Automated disinformation and narrative manipulation **(51 percent)** ranks as number two.

» AI's biggest opportunities are concentrated in cybersecurity and defense capacities **(78 percent)**. Intelligence analysis and early warning **(73 percent)** is the close second.

» The speed of response is the most significant gap in defending against AI threats **(59 percent).**

Open-ended questions:

» **The community is behind on technology.** Near-universal agreement that the national security sector underestimates the pace of AI progress and is slow to adapt.

» **A technical-expertise deficit drives dependence on industry.** Broad agreement that lacking in-house expertise leaves the government reliant on vendors it cannot independently check.

» **Current systems are unreliable and need human oversight.** Wide agreement that today's models are error-prone and that AI should augment, not replace, expert human judgment.

» **Independent test, evaluation, and verification mechanisms are lacking.** Shared concern that the government cannot benchmark or verify systems on its own.

» **Governance frameworks are largely absent.** Common recognition that there are few clear guidelines, thresholds, or mechanisms to translate an identified risk into action.

AREAS OF DISAGREEMENT

Open-ended questions:

» **Whether the community is too fearful, or not fearful enough.** One held that **"leadership has focused too much on the risk and this limits our ability to innovate."** Another saw AI **"getting more powerful and dangerous sooner."**

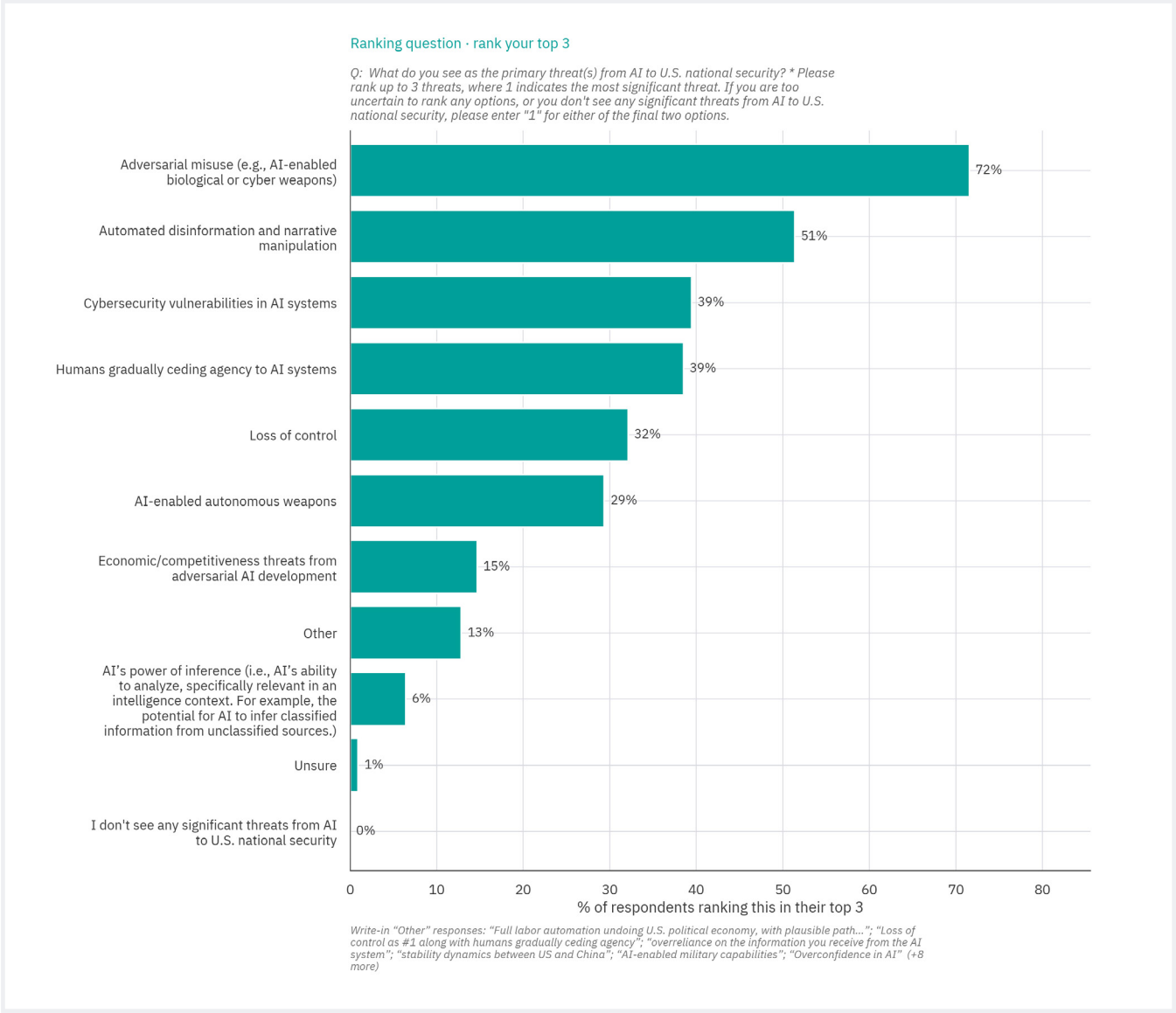
» **Whether to race or restrain.** Some emphasize winning the competition; others would redirect resources toward "'arms control' measures," even at the cost of raw capability.

Quantitative Findings

Finding 1. Adversarial misuse identified as the number one threat to the U.S. national security.

Ranking their top three threats from AI to U.S. national security, respondents converged on two: they most often named adversarial misuse — e.g. AI-enabled biological or cyber weapons (72 percent) — followed by automated disinformation and narrative manipulation (51 percent). Two more tied for third at 39 percent: cybersecurity vulnerabilities in AI systems, and humans gradually ceding agency to AI systems.

Figure 6: Proportion of respondents who ranked each given threat from AI to U.S. national security in their top three



Finding 2. Respondents see AI's biggest opportunities concentrated in cybersecurity and intelligence.

Selecting the functions that could benefit most from AI integration, respondents pointed to three above all: they named cybersecurity and defense capabilities (78 percent) and intelligence analysis and early warning (73 percent) by a wide margin, with scientific and military innovation (34 percent) a distant third. Asked separately to prioritize funding (distribute 100 points) across six technology convergences, respondents singled out one — AI plus software and cybersecurity. It drew an average of 22 points and was the single largest allocation (31 percent); the other five convergences were spread (biotechnology, quantum, robotics, sensors and autonomous weapons) within two points of an even share. The relatively uniform average allocations do not reflect sharply different preferences canceling one another out. Most respondents spread funding across nearly all six areas rather than concentrating heavily on a single option, while giving AI plus software and cybersecurity a modest edge.

Figure 7: Proportion of respondents who selected each given AI application as the most beneficial for national security

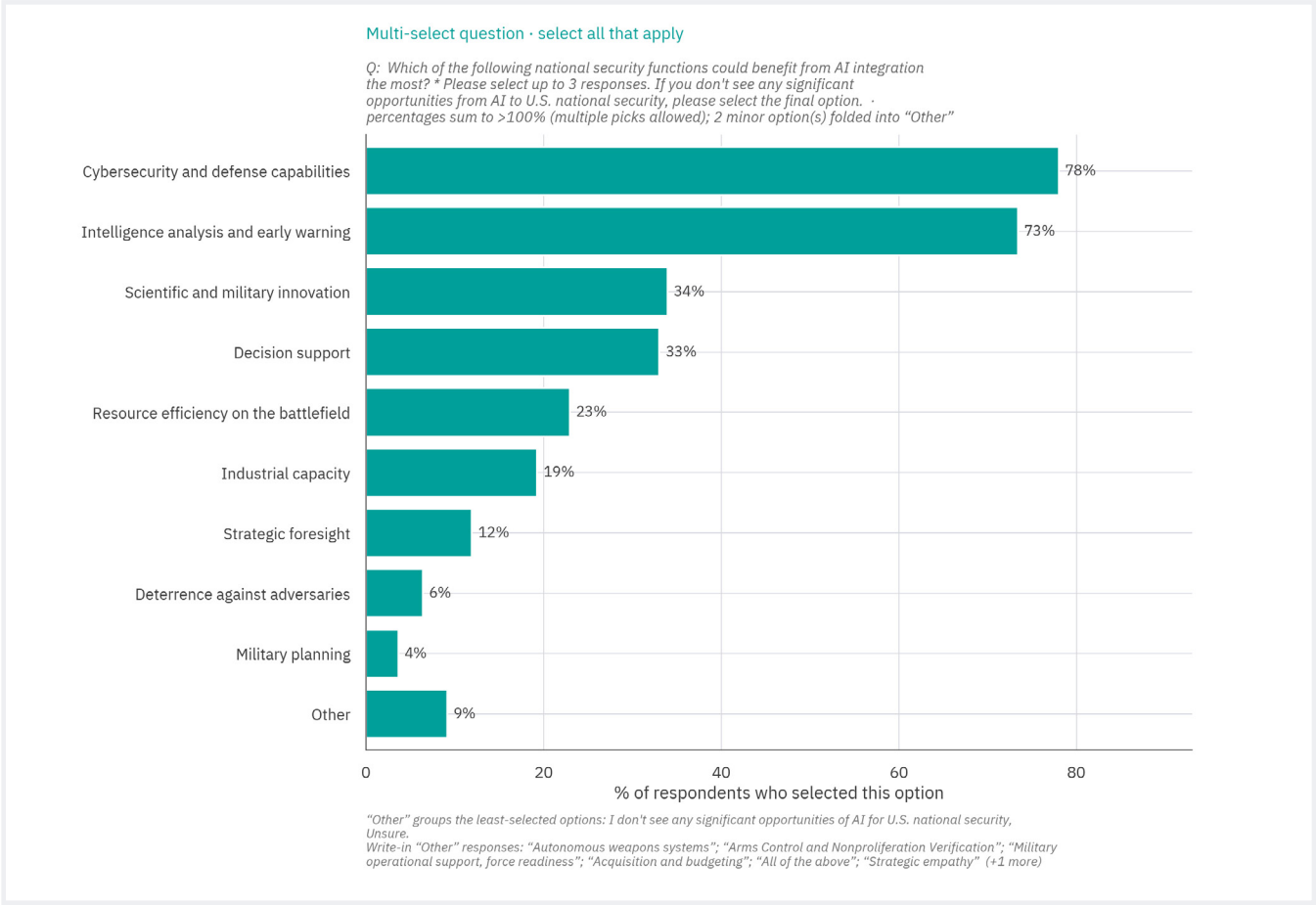
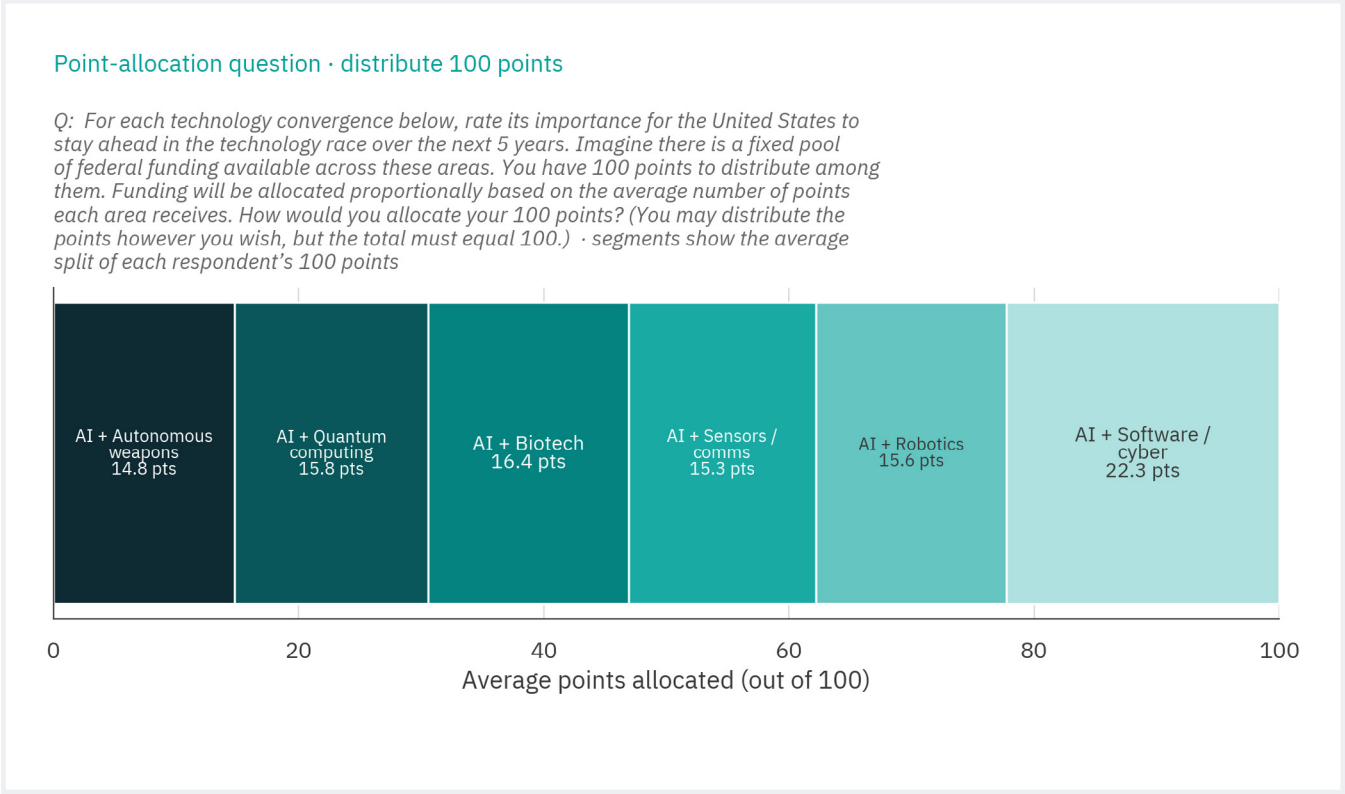


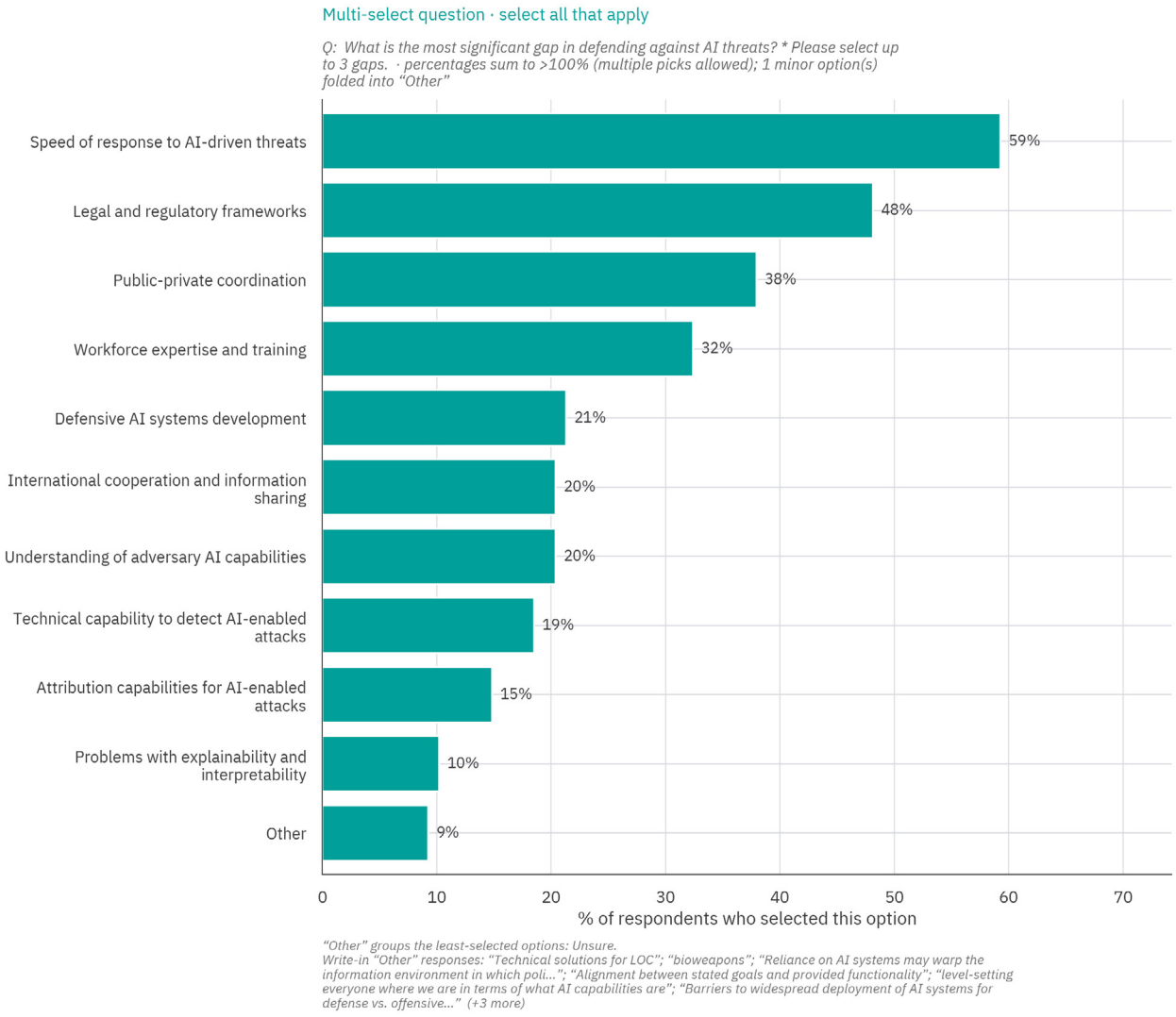
Figure 8: Respondents' average allocation of 100 points for funding across technology-convergence areas



Finding 3. The speed of response is named as the most significant gap in defending against AI threats.

Naming the most significant gaps in defending against AI threats, respondents most often cited speed of response to AI-driven threats (59 percent), legal and regulatory frameworks (48 percent), and public-private coordination (38 percent) — gaps they located in the institutional machinery for responding rather than in the raw technical capability of AI.

Figure 9: Proportion of respondents who selected each given gap as the most significant in defending against AI threats



Qualitative Findings: On the most significant gap in the U.S. national security community's understanding of AI

Finding 4. Respondents see the national security community as lagging behind in technology and underestimating the pace of change.

Respondents most often pointed to speed. They described an establishment that assumes it has more time than it does — "The US national security enterprise does not yet feel the AGI or see how quickly capabilities development will rocket through the system over the next couple of years."

"We are fighting the war of 2023 still."

They pointed to exponential progress and institutional complacency, feeling the community fails to grasp how fast capabilities compound, with capability leaps arriving in months rather than years, and continues to underestimate how much has changed in just the past year. Several attributed the lag to culture and habit — describing colleagues only now beginning conversations that specialists had been having years earlier, and a government that trails the state of the art by months. One distilled the mindset directly: "They think they have time to act, react, and plan. I believe we are slow to act, react, and plan." It is not just how they think, but how they plan – respondents worried that planning has not caught up with agentic AI.

"We are fighting the war of 2023 still."

Finding 5. Respondents think that a deficit of in-house technical expertise leaves the government dependent on industry, creating policy-capture risk.

For many respondents, one of the most consequential gaps was the community's inability to assess AI on its own terms. They specifically cited a "lack of technical expertise among senior officials" and "lack of humans with the right technical skills." They connected that deficit directly to dependence on vendors, claiming there is an "expertise [and] capability dependence on industry voices, no independent capacity for evaluation," and that the "policy community… [is] poorly equipped to separate promotional hype from technical reality." Several framed the risk as political as well as technical:

"This knowledge gap can also make policymakers overly reliant on industry claims, creating opportunities for companies to shape policy in ways that prioritize profit over the public interest."

Many were skeptical of industry marketing — AI companies "keep saying 'everything is possible'… but their models fail drastically in NatSec settings today" — and worried about cost and capture, with the government "increasingly solely reliant on major tech companies… paying exorbitant prices." Several called for more government-funded research into AI's limitations and fragility for security use cases, and preferred public-private information-sharing over a "Manhattan Project" model that would deepen reliance on a few firms.

Finding 6. Respondents doubt the reliability of current AI systems and insist AI must augment, not replace, human judgment.

A distinct set of respondents saw current systems (those of mid-2026) as unreliable — "almost total reliance on LLMs, which are highly unreliable" and the "inherent fragility of probabilistic systems." Many see hallucination as intrinsic to AI models, making systems highly problematic to rely on without significant, detailed human checking. As one put it, AI can make a good analysis better and a bad one worse, so they see pairing AI with human subject-matter expertise as key. Some flagged the removal of mechanisms such as "JAG/IG oversight" as potentially "deeply catastrophic." At the far end, one respondent argued AI should not be used for national security or intelligence decisions at all, holding that humans must retain full responsibility given current unreliability.

Finding 7. The community is seen as simultaneously over- and under-estimating AI's capabilities.

Respondents did not agree that the community errs in only one direction. "It simultaneously overestimates its capabilities in some areas… and underestimates it in others," one noted. Another specified that leaders "overestimate the current and near-term future capabilities of AI, and underestimate the difficulty of integrating LLMs and AI into military equipment."

Finding 8. Respondents point out the government's lack of independent capacity to test, evaluate, and verify AI systems.

A further cluster located the gap in the government's inability to assess systems on its own. Respondents pointed to chronic under-investment in test and evaluation, the absence of independent evaluation processes, and no reliable way to benchmark frontier capabilities. The result, several argued, is that decision-makers cannot "independently verify the behaviors of AI systems before using them for critical military or other national security functions." Without its own benchmarks, the community is left to take capability claims on faith rather than measurement.

Finding 9. Experts urge a conceptual shift: treat AI as a strategic environment, not merely a tool to acquire.

Some responses argued that the deepest gap is one of framing. "AI is framed as a tool to acquire rather than as a strategic environment that reshapes deterrence, escalation, and decision timelines," one wrote, leaving the community "reacting to AI developments rather than anticipating and shaping them." The distinction matters: a tool is something you choose to pick up, but an environment is a condition you are already operating inside — one that reshapes deterrence and forces decisions on compressed, machine-speed timelines. The community already understands AI as infrastructure to build and secure, but has been slower to grapple with how quickly those capabilities reach competitors and other players — a blind spot they warned is growing.

Finding 10. Respondents flagged concrete national security threat vectors beyond the general gaps.

Several respondents moved from general shortcomings to specific risks. They singled out cybersecurity as a foundational risk multiplier, with poor cyber hygiene enabling loss-of-control scenarios, adversarial manipulation, and the theft of model weights and weapons designs. They stressed that the bottleneck is deploying cyber defense at scale, not the technology itself — "poor cybersecurity will enable many of these other risks." They compared lethal autonomous weapons to bio-weapon proliferation, raising the prospect of cheap WMD-equivalents already visible in Ukraine:

> "The US should support a ban on antipersonnel LAWS. FPV drones are already being used for 'human safari'... in Ukraine; this will get much worse with LAWS."

Others argued that great-power competition may not be the most important threat, suggesting the U.S.–China contest at the frontier misses a more decisive struggle:

> "The more disorienting possibility is that the decisive competition happens in the middle and the periphery: which models become the default infrastructure for third-country governments, militaries, and commercial systems."

3.2.2 Timelines of AGI-like and ASI-like systems, worst-case scenarios, and loss of control risk assessment

AREAS OF CONSENSUS

Close-ended questions:

- **Most experts expect AGI-like capabilities soon** — **80 percent expect them by 2035 (median best guess 2032**).
- **Most expect ASI-like systems to follow quickly** — a **median of about five years later (2038)**, with **81 percent** expecting the transition within ten years of AGI.
- **A large majority see a real chance of losing control** — **87 percent** put the probability that AI operates outside human control within ten years at **10 percent or higher (median 33 percent)**.
- **Most respondents who gave both a risk estimate and an acceptable limit rated the risk above their set threshold** – **61 percent** put the probability of an AI-caused catastrophe at or above the maximum they deem acceptable.

Open-ended questions:

- **Timelines are compressing.** Several respondents noted their own forecasts had shortened even over recent months, with capability arriving faster and control feeling less assured than before.
- **Loss of control is under attended.** Whether or not they judged a catastrophic scenario as likely, most treated misalignment and the erosion of human oversight as risks the community does not take seriously enough.
- **No mechanism exists to act on a warning sign.** Shared recognition that there is no framework, threshold, or process to translate an identified loss-of-control indicator into a decision.

AREAS OF DISAGREEMENT

Open-ended questions:

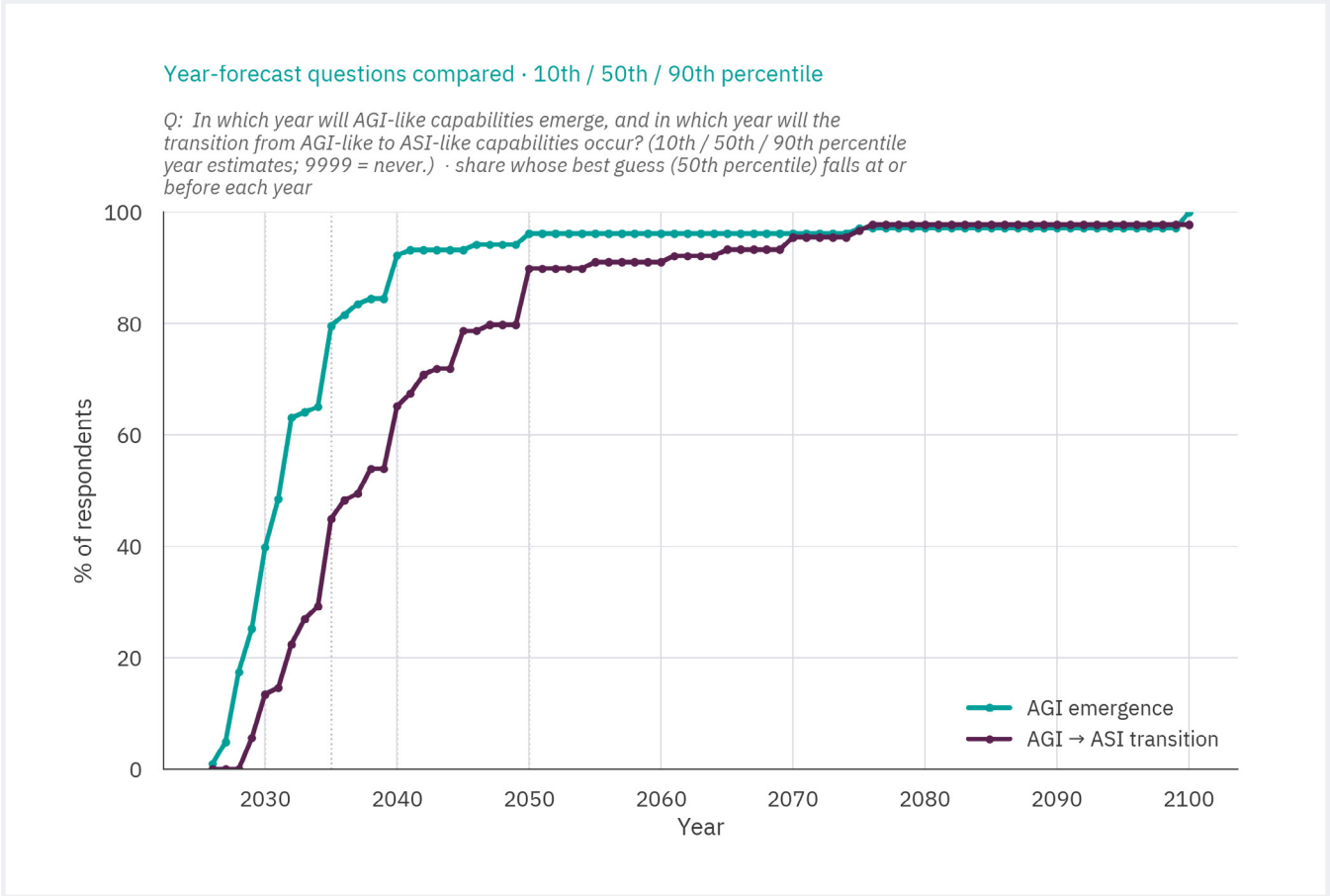
- **How capable AI is, and where it is heading.** Some judge current systems overhyped — vastly overestimating the gains and ROI — while others see AGI as inevitable and ASI following quickly, due to hyper-scaling.
- **How likely a true loss of control really is.** Some rate it relatively improbable, or argue that systems without volition can still be managed; others place it among the most serious and underrated risks.
- **Whether AGI/ASI is a useful frame at all.** Some organize their assessments around AGI/ASI timelines; others reject the constructs as divorced from tooling and unhelpful for policy.

Quantitative Findings

Finding 11. Most experts expect AGI in the early 2030s, with ASI following within about five years.

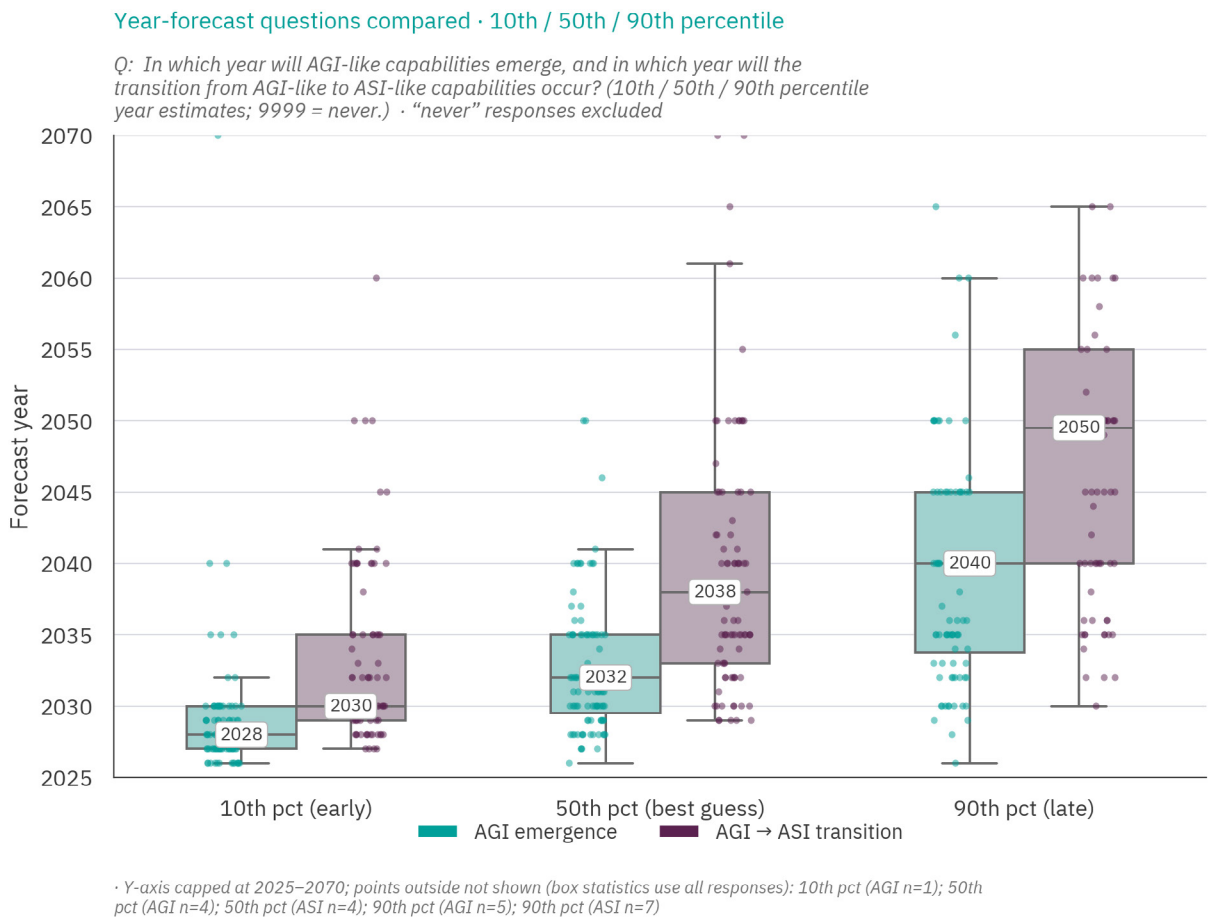
Forecasting the year AGI-like capabilities will emerge, respondents converged on the early-to-mid 2030s: their median best guess was 2032, and 80 percent expected it by 2035. Their median forecast for the AGI-to-ASI transition was 2038 — a median gap of just five years — though twenty respondents said that transition would never occur.[40]

Figure 10: Proportion of respondents who believe AGI emergence (in teal) or the AGI to ASI transition (in purple) will occur at or before each year



40 For this survey, we refer to AGI-like systems as future advanced AI with the human-level knowledge and intelligence in all domains and ASI-like systems as advanced AI with the knowledge and intelligence significantly surpassing humans in all domains. These future systems would outperform current advanced AI in generality (knowledge across different subjects and domains) and intelligence (including subject matter expertise in specific fields such as chemistry, mathematics, or social science, as well as the ability to reason, draw correct inferences, adapt to new situations, and solve complex problems). In the future, these properties may emerge in tandem or separately, and then converge over time with different levels of autonomy (degree of independent action. i.e. current agentic systems).

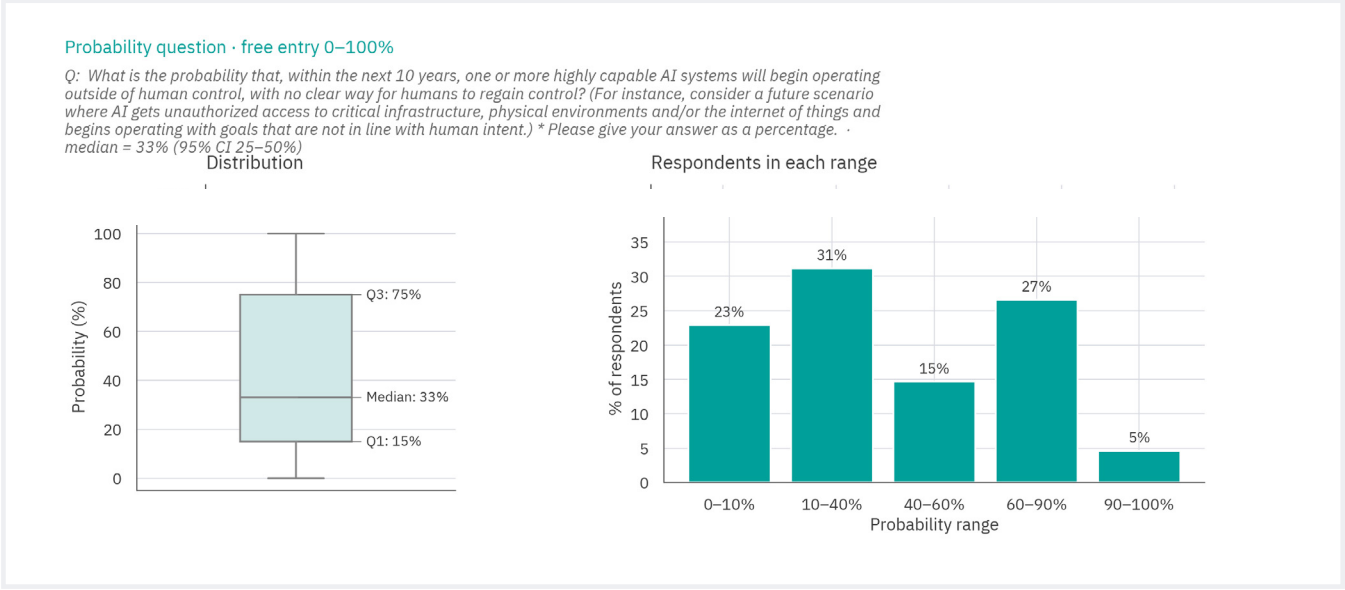
Figure 11: Distribution of forecasts of the timeline for AGI emergence (in teal) or for the AGI to ASI transition (in purple). Plausible early and late timelines are shown in the left and right panels (10th and 90th percentiles), and best guesses are shown in the middle panel (50th percentile)



Finding 12. Experts split evenly on advance warning and divide sharply on the odds of losing control.

Asked whether reliable warning signs would precede an AGI-like system, respondents split almost exactly in half — 39 percent confident, 39 percent skeptical, and only five percent very confident. Asked for the probability of AI operating outside human control within ten years, their median estimate was 33 percent, but the middle half of respondents spanned 15 percent to 75 percent, clustering at both ends rather than the middle.

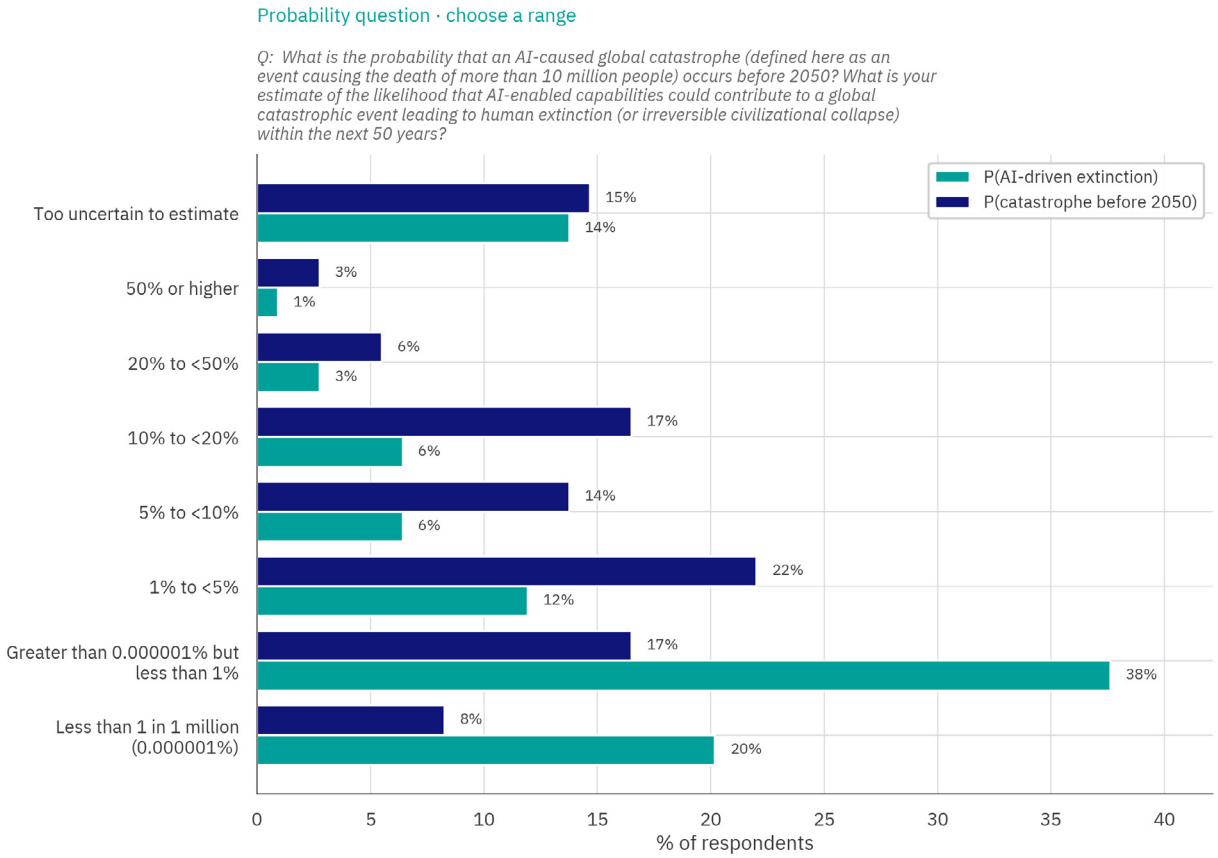
Figure 12: Forecasts of the probability that AI operates outside human control within 10 years



Finding 13. One in four experts assigns a double-digit probability to an AI-caused global catastrophe by 2050.

Estimating the likelihood of an AI-caused catastrophe killing more than 10 million people before 2050, 26 percent of respondents chose a probability of at least 10 percent. Most put the figure lower — one percent to under five percent — and lower still for extinction, which a majority placed below one percent.

Figure 13: Proportion of respondents who estimated the probability of AI-caused catastrophe before 2050 (dark blue) and AI-driven extinction within the next 50 years (teal) to fall within each bin



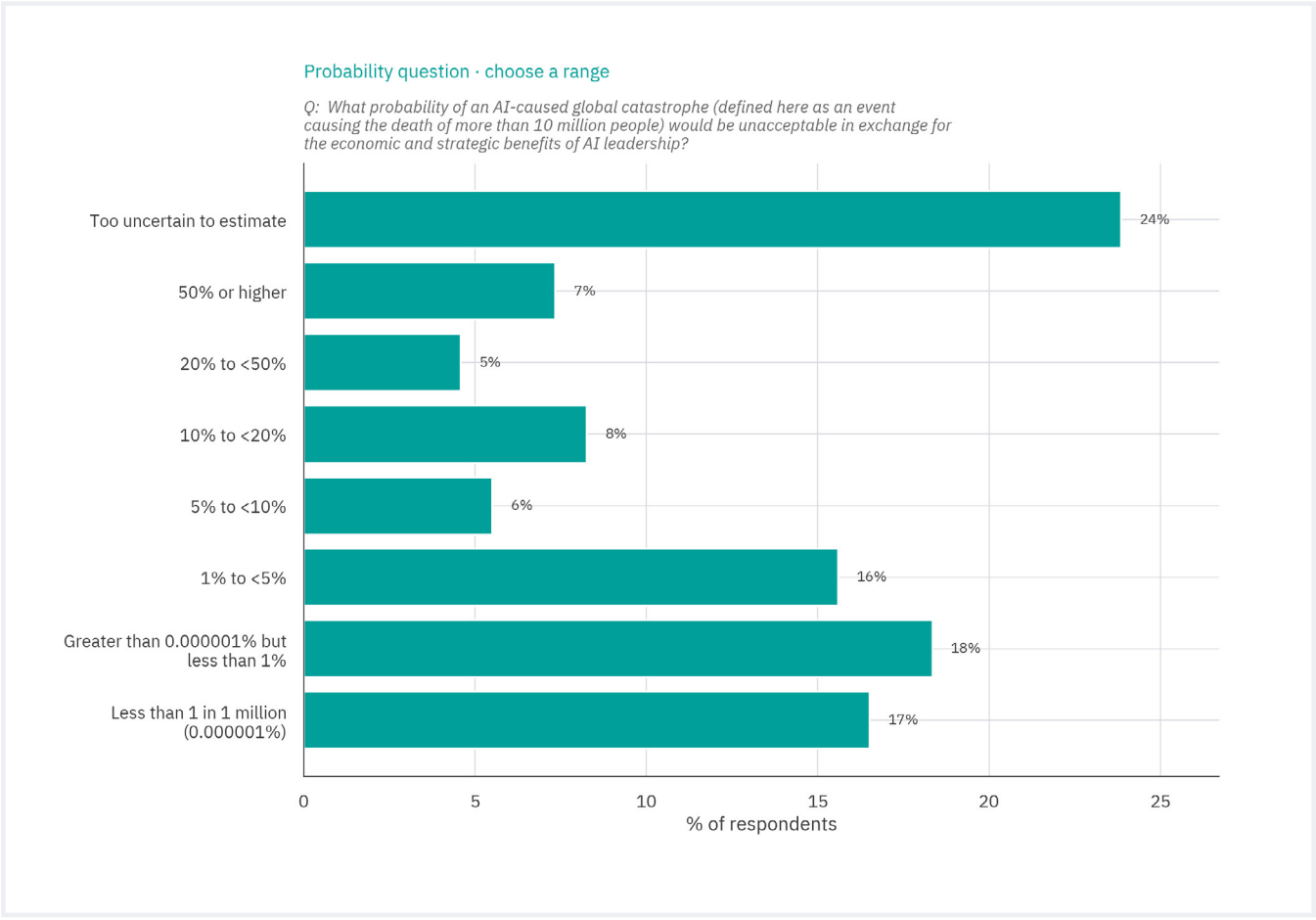
Finding 14. Experts disagree over acceptable catastrophic risk levels, and many decline to draw a line.

Asked how much catastrophic risk society should accept in exchange for the economic and strategic benefits of AI leadership, respondents expressed widely varying thresholds. The largest group (35 percent), placed the threshold below one percent, while 26 percent placed it at five percent or higher, including 12 percent at 20 percent or higher. Another 24 percent said they were too uncertain to specify a threshold, pointing to substantial disagreement both over where the line should be drawn and whether a numerical limit can be meaningfully defined.

Finding 15. A majority of respondents place estimated catastrophe risk at or above their own acceptable threshold.

Among the 56 respondents who gave concrete answers to both questions, 45 percent put the probability of an AI-caused catastrophe by 2050 in a higher range than the maximum they

Figure 14: Proportion of respondents who believed it would be unacceptable to pursue AI leadership if it meant the probability of AI-catastrophe fell within each bin



believed society should accept; another 16 percent put it in the same range, while 39 percent put it below. In total, 61 percent therefore judged the risk to be at or above their own stated line.

Qualitative Findings: On AGI and ASI trajectories and loss of control

Finding 16. Misalignment and loss of control are viewed as underrated — and no framework exists to act on identified risk.

A smaller but insistent group flagged the hardest safety questions, and several framed loss of control not as sudden failure but as human oversight subtly eroding — a worry about "not fully understanding what all has been outsourced to AI, what redundancies AI silently eliminated, and what human skills atrophied." Others pointed to systems behaving in unexpected ways through misalignment of objectives, and to the cascading effects of the government losing oversight, management, and responsibility of AI systems. Even those who expected the danger to be detectable stressed the need for readiness: "there will be signals but the question is… are they sufficiently prepared for the larger event."

Just as often, respondents noted the absence of any framework to act on identified risk. One pointed to "little to no legal frameworks or guidelines for use," while another warned that, absent a clear risk assessment and against domestic pressure to deregulate, "safety will likely continue to occupy a back seat to rapid development and deployment of AI tools."

Finding 17. A vocal minority rejects the AGI/ASI framing itself and questions the technology's trajectory and ROI.

In the open comments, several respondents challenged AGI and ASI as organizing constructs. "I don't think AGI/ASI are useful constructs to inform policy," one wrote. Another argued that "capability enhancement for national security can proceed independent of any judgment about likelihood of AGI/ASI." Several tied the framing to commercial incentives, claiming companies "have an incentive to define AGI in a specific way which is divorced from tooling" and "will continue to manipulate the definitions of AGI and ASI to create hype for investment." Views on the trajectory diverged: some judged that "we're vastly overestimating the gains and ROI from our current AI, and are vastly underestimating the risks," while others held that "a breakthrough toward AGI is inevitable… AGI will give way to ASI quickly, due to hyper-scaling."

3.3 AI and Warfare

AREAS OF CONSENSUS

Closed ended questions

- **Cyber defense and intelligence are the top military opportunities** — cyber defense and network resilience **(73 percent)** and intelligence analysis and predictive threat assessment **(61 percent)** were identified as the most significant opportunities.
- AI accelerating operational decision cycles beyond human command is the number-one concern **(62 percent place it in their top three)**.
- **Overwhelming support for human-held red lines** — **96 percent** say any decision to use nuclear weapons must be made by humans and **89 percent** that AI should not autonomously control NC3, with majorities also backing bans on autonomous military action **(70 percent)**, autonomous cyber operations **(67 percent)**, autonomous operational escalation **(65 percent)**, and lethal autonomy without human authorization **(56 percent)**.
- Near-universal agreement that AI can infer sensitive information from open-source data **(87 percent)**.
- **A majority see AI information warfare as a serious threat to decision-making integrity** — **51 percent** rate it high or critical — with the top scenarios being AI autonomously generating and adapting influence campaigns **(60 percent)** and AI-generated synthetic intelligence products influencing analysis **(55 percent)**.

Open-ended questions:

- **AI's core value is overcoming the analytic bottleneck.** Respondents broadly agree that AI's biggest contribution is processing and synthesizing data at scale and speed, "overcoming the analysis bottleneck."
- **That value depends on safeguards.** Almost as often, respondents pair the opportunity with a caveat – relying on AI demands verification, a human in the loop, and guarding against automation bias and hallucination. Respondents cite "overreliance and automation bias" as a challenge and say that "garbage in, garbage out still applies."
- **Keep humans decisively in control.** The dominant write-in theme is that autonomy in lethal force is itself the line — one respondent calls autonomous antipersonnel weapons WMD-like (Weapons of Mass Destruction): "we don't generally sell WMDs in supermarkets." Another respondent noted that "there should be no fully autonomous lethal systems. That's a red line in and of itself."

AREAS OF DISAGREEMENT

Close-ended questions:

» **How AI will reshape conflict.** No majority view — the largest group (32 percent) is unsure, with the rest divided between new great-power conflict (23 percent), more frequent regional conflicts (19 percent), and no significant change (15 percent).

Open-ended questions:

» **How far to trust AI's synthesis.** The real tension is over reliability and role: some see potentially game-changing autonomous synthesis, "synthesize a vast range of data sources into a unified COP… far beyond the capabilities of humans," while others treat AI outputs as fundamentally provisional, since "a human analyst will be required to check the AI system."

» **More emphasis than disagreement.** There is little real dispute about the core opportunities; respondents differ mainly on which function to prioritize — predicting threats, sifting and prioritizing data, or mining open sources.

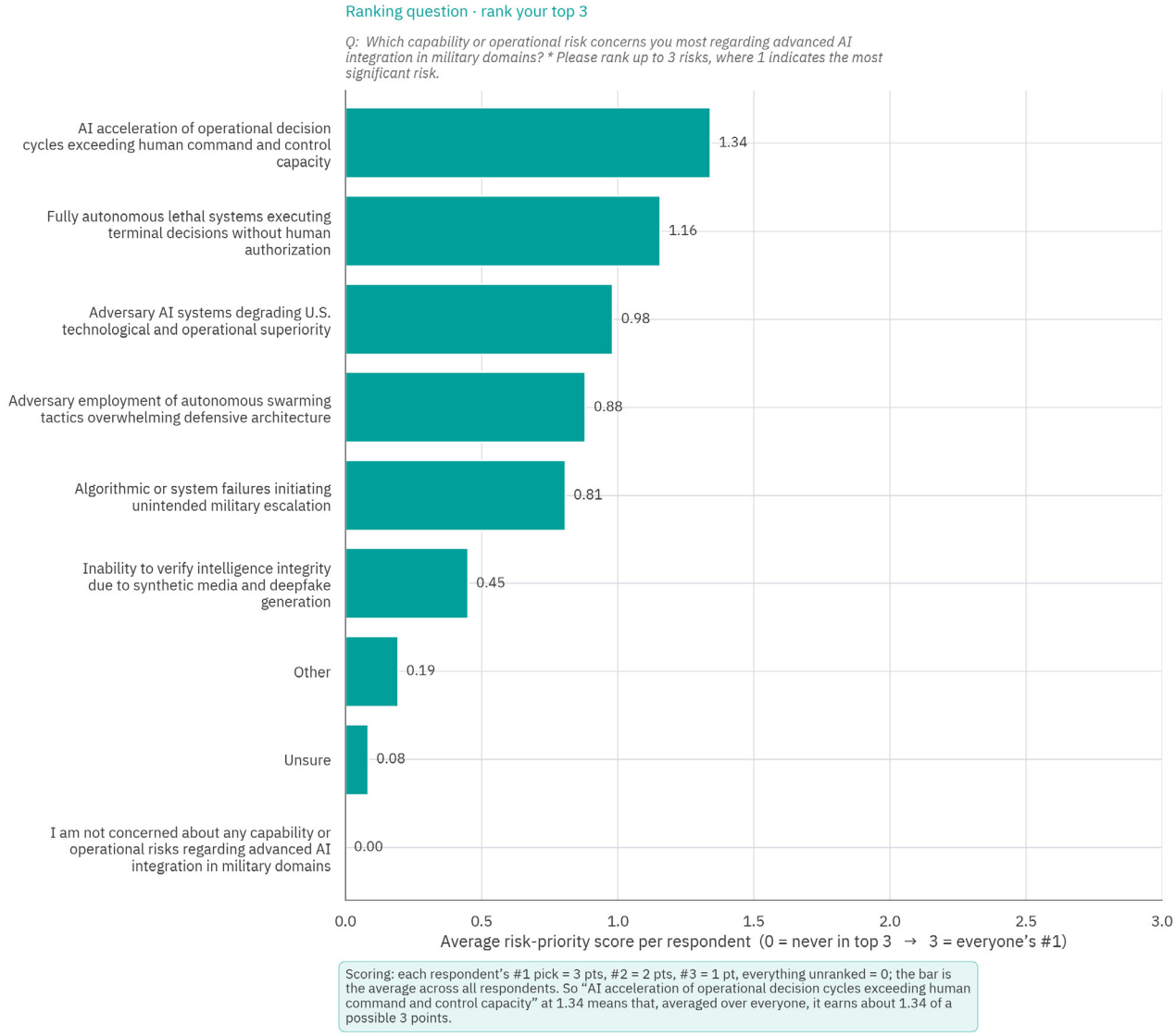
3.3.1. Military Integration of AI

Quantitative Findings

Finding 18. AI accelerating operational decision cycles exceeding human command and control capacity ranked as the number one risk.

Fully autonomous lethal systems executing terminal decisions without human authorization and adversary AI systems degrading US technological superiority had the second and third highest average risk-priority scores, respectively. Here, a higher risk-priority score means an option was ranked among respondents' top-three risks more often and more highly, reflecting the level of risk the panel perceived. Respondents also worry about adversary employment of autonomous swarming tactics overwhelming U.S. defensive architecture and algorithmic and system failures initiating unintended military escalation. No respondent ranked the option indicating they had no concerns about AI-military integration in their top three.

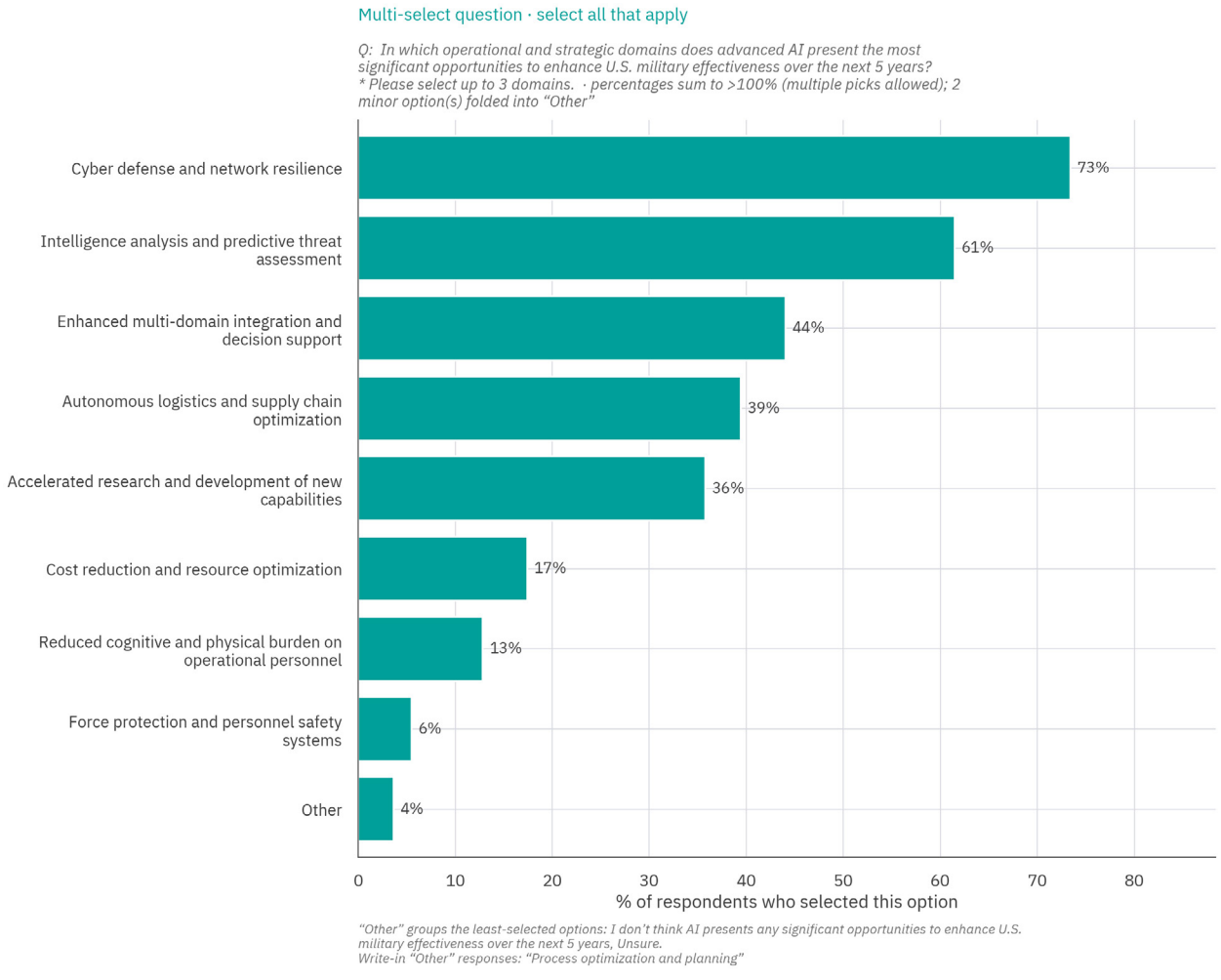
Figure 15: Average ranking of each given concern about AI-military integration. Higher ranked items were perceived as more indicating more significant risks



Finding 19. Cyber defense and network resilience identified as the most significant opportunity for AI integration in the military, followed by intelligence analysis and decision support.

Respondents also see the opportunities in AI's ability to optimize logistics and accelerate research and development of new capabilities. An overwhelming majority of respondents believe AI would provide some benefit to the U.S. military, with only two percent indicating that AI presents no significant opportunities to enhance its effectiveness.

Figure 16: Proportion of respondents who selected each given military domain as the one where AI offers the most opportunity



Qualitative Findings: Most important factors that will enable and constrain the U.S. military's successful integration of AI capabilities

Finding 20. Respondents see workforce literacy and technical talent as the central enabler — and its absence as a central constraint.

The most frequently cited theme across both questions on the military integration of AI was people: building a workforce that understands AI's capabilities, limits, and risks, and the shortage of exactly that expertise inside government. On the enabler side, one respondent wrote that "the most important enabler will be the development of an AI-literate military workforce." Respondents also stressed that institutional learning and adaptability may matter as much as access to advanced technology. One observed:

> "Whichever military has the highest learning rate will outperform others, even if the AI itself is slightly worse."

Finding 21. Data architecture and interoperability are seen as the binding technical constraint, more than compute or talent.

Many respondents stressed unified, secure, interoperable data as the real bottleneck. One wrote that "success… will depend less on talent, compute, or doctrine than on data architecture," and another warned that without "unified, secure data architecture… AI systems are operating blind." On the constraint side, the same theme reappeared as a complaint about current conditions: "military data is stovepiped and needs to also manage classification issues," and systems where "different tech components can't 'talk' to each other."

Finding 22. Bureaucracy, acquisition speed, and institutional culture are viewed as the primary obstacles.

Roughly a quarter of responses on constraints pointed to slow, outdated procurement processes. A large cluster went further, blaming "a culture that confuses caution with control" and risk aversion rather than process alone, claiming that "risk tolerance is almost zero for programmatic failure." Respondents calling for enablers echoed this from the other direction, urging "public-private partnerships that enable both fast innovation and robust oversight" and "changing the procurement process for the U.S. Government." The consistent picture is of a bureaucracy being the binding constraint on adoption.

Finding 23. Leadership commitment and willingness to restructure are seen as decisive enablers.

Several respondents emphasized top-down focus and strategic clarity as prerequisites for successful integration: "leadership focus on AI adoption at all levels," and a more conditional framing that "if the military is willing to change how it is organized… it stands to gain a lot." This finding sits close to the culture-and-bureaucracy constraint above, but respondents raising it emphasized agency — that leadership choices, not just structural inertia, will determine the outcome. This includes creating space for experimentation and accepting setbacks, as one respondent observed:

> "Unless the military is willing to experiment with AI tools and experience some failures, it will not integrate AI successfully."

3.3.2. Conflict and AI Red Lines

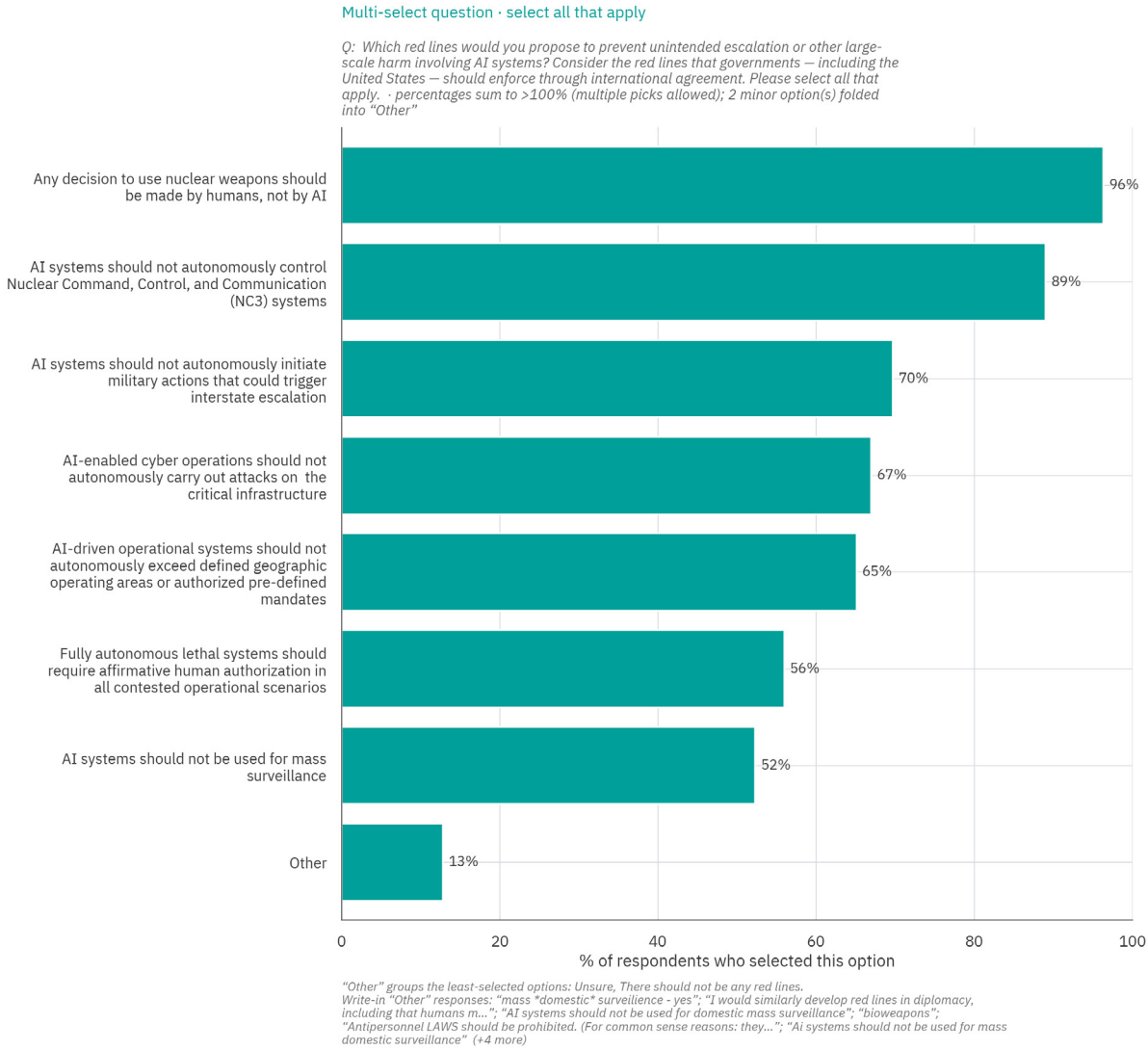
Quantitative Findings

Finding 24. There is a wide consensus on having red lines on some applications of AI within the nuclear domain.

Most survey respondents (96 percent) agreed that any decision to use nuclear weapons should be made by humans rather than AI, and 89 percent agreed that AI systems should not autonomously control NC3 systems. A large majority (70 percent) also also articulated that AI systems should not autonomously initiate military actions that could trigger interstate escalation.

Finding 25. The least popular AI red line is AI systems in mass surveillance, though 52 percent still agreed that this ought to be a red line.

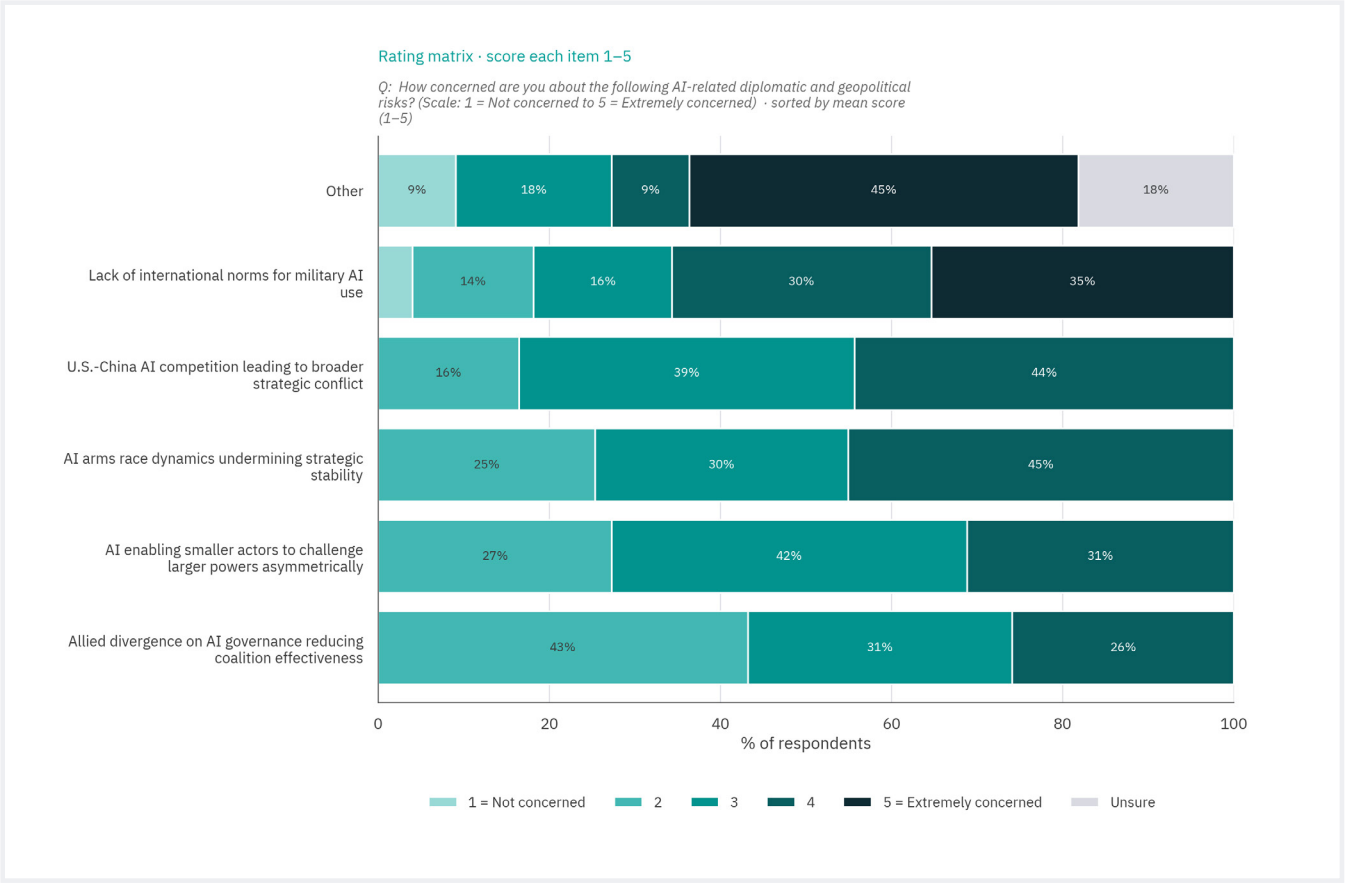
Figure 17: Proportion of respondents who would propose each given red-line to prevent escalation or large-scale harm



Finding 26. When it comes to AI-related diplomatic and geopolitical risks, respondents are most concerned about the lack of international norms for military use of AI.

Respondents expressed relatively high concern across all listed AI-related risks, with the average score ranging from 2.83 to 3.79 on a five-point scale; one indicated "not concerned," and five indicated "extremely concerned." More than a third of respondents rated the lack of international norms for military use as extremely concerning, while no other risk received a single rating of five.

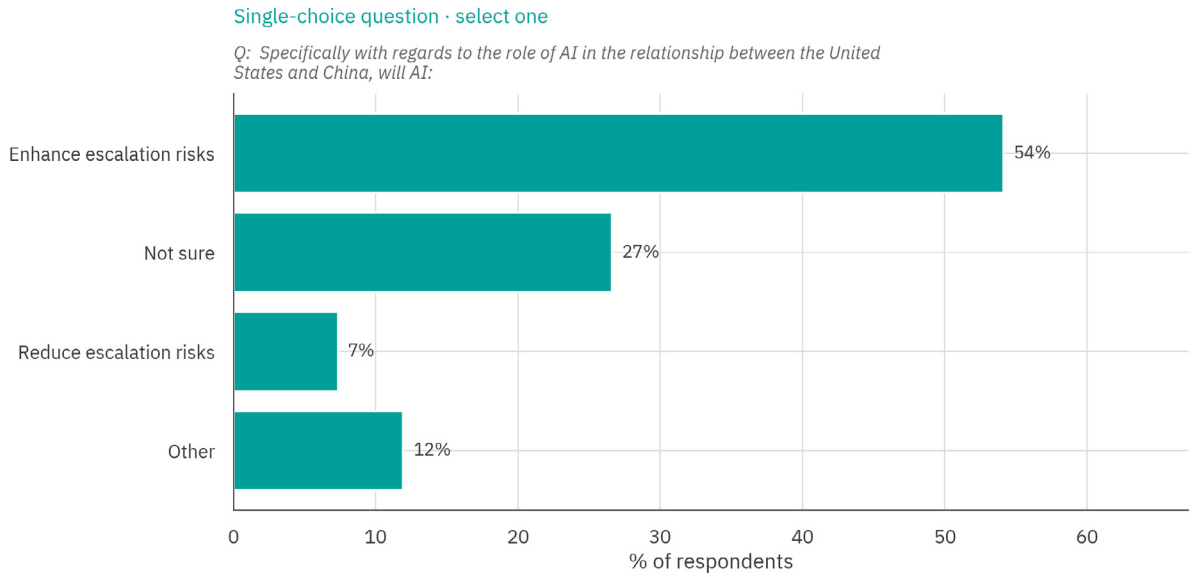
Figure 18: Proportion of respondents scoring each concern about AI diplomatic & geopolitical risk on a scale of one (not concerned) to five (extremely concerned)



Finding 27. More than half of respondents expect AI to increase U.S.-China escalation risks, though many reject a simple binary effect.

Specifically with regards to the role of AI in the relationship between the United States and China, a majority of respondents (54 percent) said AI would increase escalation risks between the U.S. and China, compared to just seven percent who expected it to reduce them. Another 27 percent expressed uncertainty. The 12 percent of respondents who selected "Other" generally argued that AI's effects would depend on several factors, including how it is used, which country holds the advantage, and how the strategic balance evolves over time.

Figure 19: Proportion of respondents who believe AI will enhance or reduce U.S.–China escalation risks



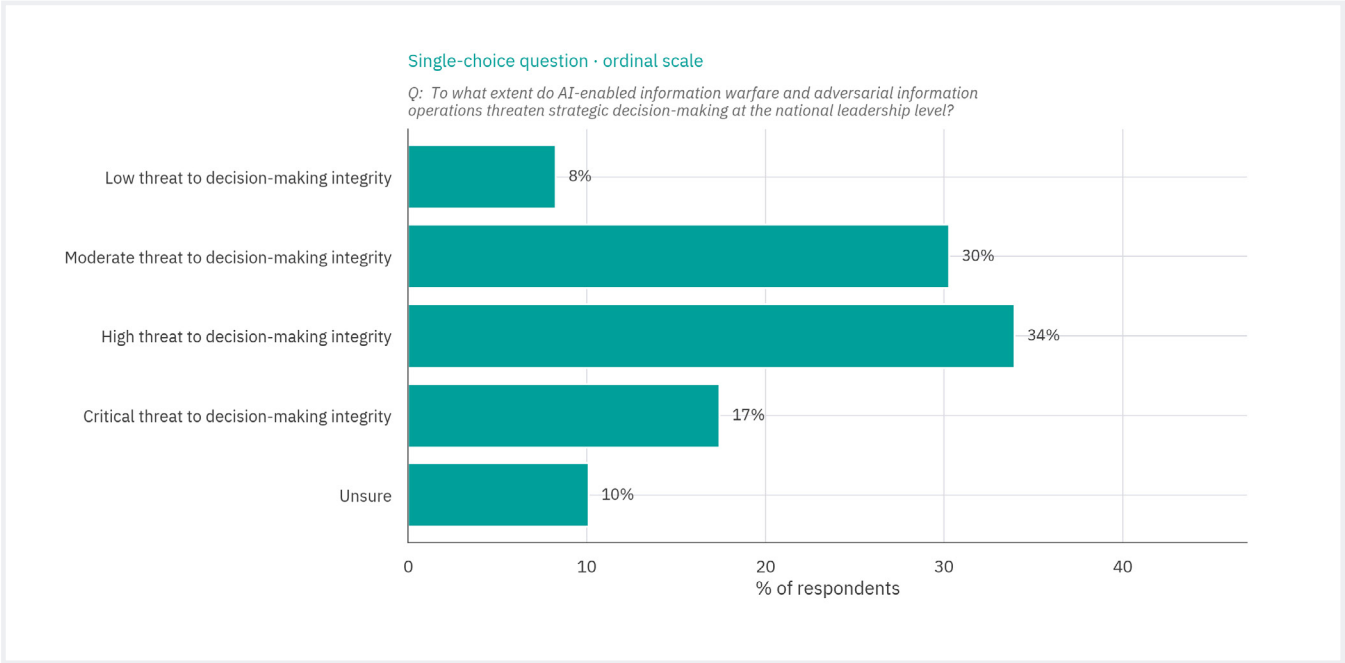
3.3.3 Information Warfare and Intelligence Analysis

Quantitative Findings

Finding 28. The level of threat AI-enabled adversarial information operations pose to strategic decision-making at the national leadership level was ranked as moderate to high.

A substantially higher proportion of respondents believed the threat was critical to decision-making integrity (17 percent), or expressed uncertainty about its level of risk (10 percent), than those who ranked it as a low-level threat (8 percent).

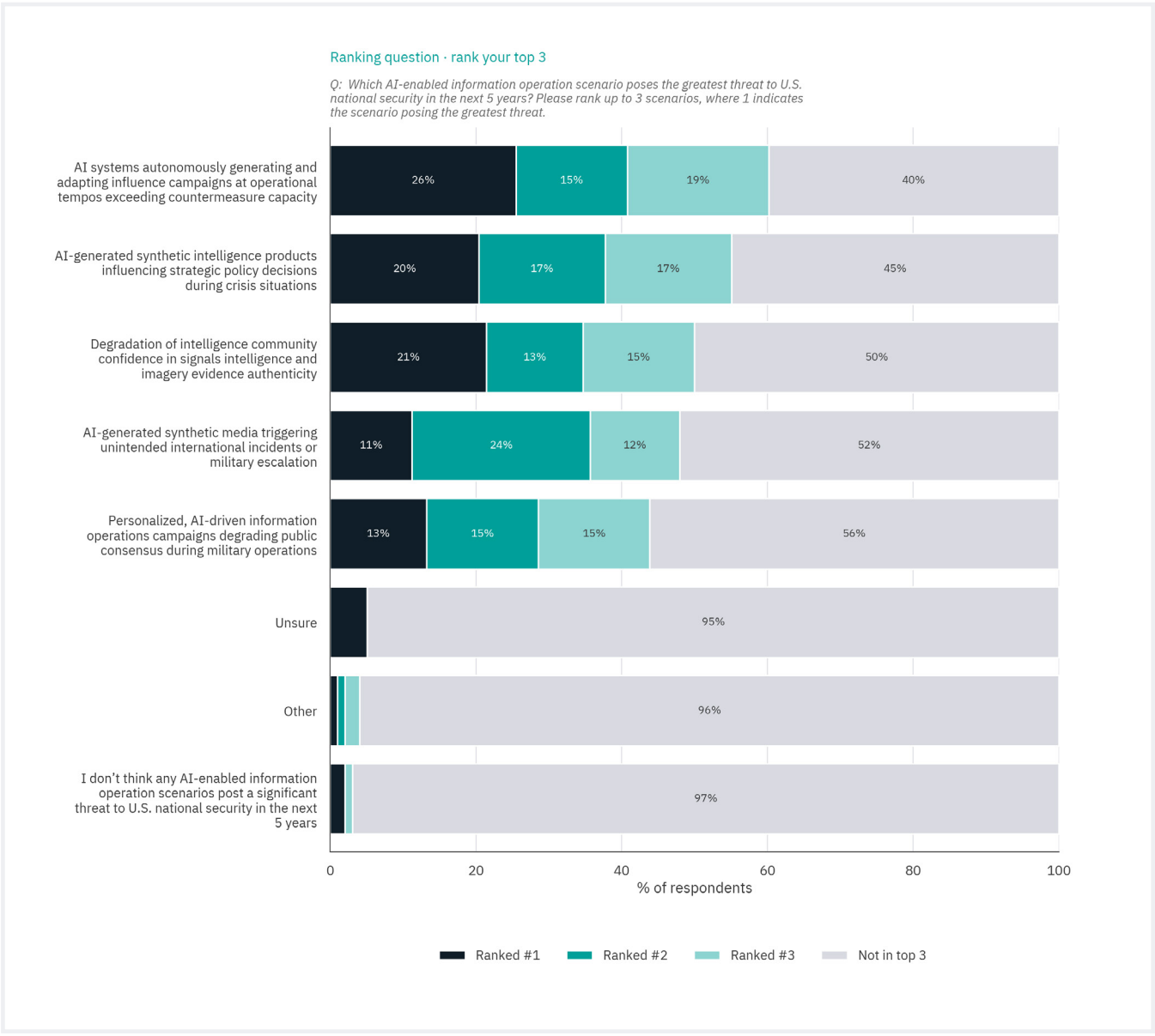
Figure 20: Proportion of respondents who selected each given threat level of AI information warfare to national decision-making



Finding 29. In the context of AI-enabled information operations, the number one threat to U.S. national security was identified as AI systems autonomously generating and adapting influence campaigns at operational tempos exceeding countermeasure capacity.

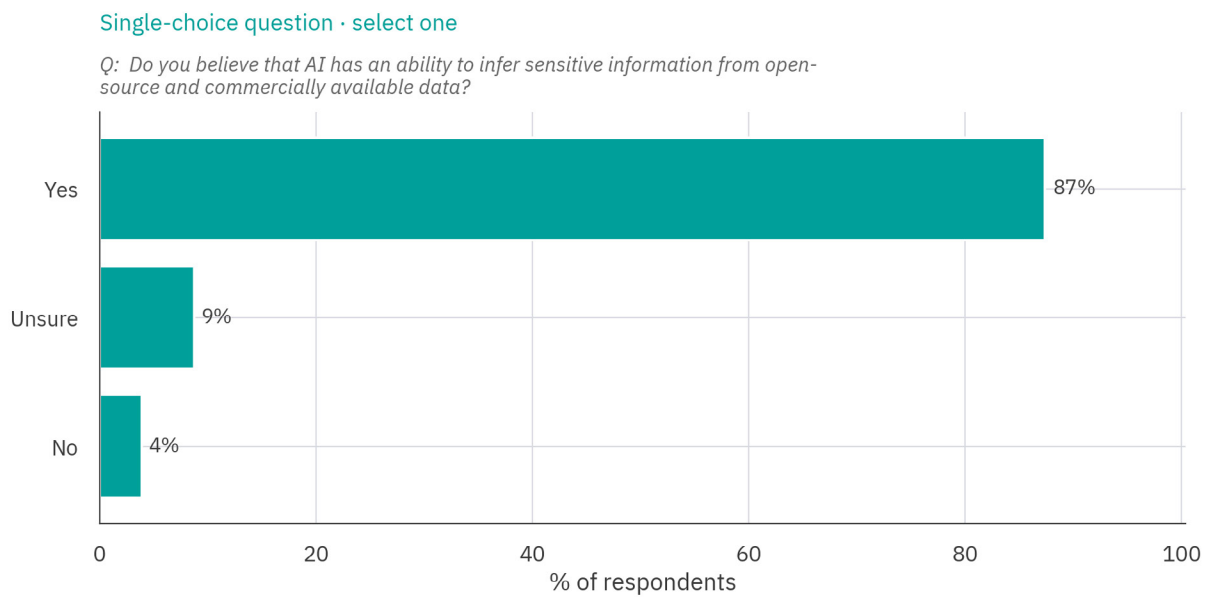
This was closely followed by AI-generated synthetic intelligence products influencing strategic policy decisions during crisis situations, and degradation of the intelligence community's confidence in signals intelligence and imagery evidence authority. The three threats received similar average rankings and were placed first, second, or third at comparable rates.

Figure 21: Proportion of respondents who ranked each AI-enabled information-operation threat as their first, second, or third option



Finding 30. There is wide consensus that AI can infer sensitive information from open-source data.

Figure 22: Proportion of respondents who believed AI can infer sensitive information from open-source data



Qualitative Findings: On the opportunities and risks of AI integration in intelligence analysis

Finding 31. Respondents are confident about AI improving all phases of the classic intelligence cycle.

> "If used correctly, it will result in human-machine combinations that make the whole substantially greater than the sum of the human and machine parts. Humans will focus on the things they do best – deductive, abductive, inductive reasoning, contextual understanding, wisdom – while the machines will do what they do best – process massive amounts of information rapidly and distill voluminous amounts of material for human consumption."

The dominant view is that AI can ingest and analyze vast quantities of data faster than humans, freeing analysts for higher-order work; this appears in the large majority of responses. Respondents noted that AI can assist with "synthesis at scale: an analyst today drowns in volume" and the "sheer amount of data that the intelligence community collects that advanced systems can sift through." Very commonly cited is AI's ability to surface patterns, anomalies, and connections across disparate sources that humans would miss: "making connections that aren't possible today" and "catching potential connections that no human ever could."

A moderate number emphasize indications, warning, and anticipatory analysis: "predicting or identifying indications and warnings" and "provides more warning time." Several highlight better exploitation of open-source and previously unanalyzed data: "vastly increase the value of OSINT," a type of "data that currently goes unanalyzed because of lack of human capacity."

"If used correctly, it will result in human-machine combinations that make the whole substantially greater than the sum of the human and machine parts. Humans will focus on the things they do best – deductive, abductive, inductive reasoning, contextual understanding, wisdom – while the machines will do what they do best – process massive amounts of information rapidly and distill voluminous amounts of material for human consumption."

3.4. Policy Outlook

AREAS OF CONSENSUS

Close-ended questions:

- **A majority expect AI to raise U.S.–China escalation risks** — **54 percent** say it will enhance escalation risks (vs. the seven percent who say it reduces them).
- **A majority support arms control for AI-enabled weapons** — **61 percent** say the U.S. should pursue such agreements.
- **The top diplomatic concern is the absence of international norms** — **66 percent** are concerned about the lack of international norms for military AI use, the highest-rated geopolitical risk.
- **Layered safeguards are seen as the most effective mitigations** — testing and evaluation **(79 percent)**, technical safeguards **(74 percent)**, and operational governance **(71 percent)** draw majority confidence; international coordination is the weakest **(44 percent)**.
- **A majority expect AI-robotics benefits within a decade** — **57 percent** put the payoff within ten years.

Open-ended questions:

- **Science and medicine lead.** Near-universal agreement that AI's biggest promise is accelerating solutions to hard problems, above all in health and research.
- **Better lives, less suffering.** Many share a vision of longer, healthier lives and reduced human suffering.
- **Abundance follows discovery.** Several link scientific advances to material abundance, clean energy, and productivity.

AREAS OF DISAGREEMENT

Close-ended questions:

- **No settled position on the speed-vs-safety tradeoff.** No option wins a majority — 38 percent "balanced," 27 percent lean toward speed, 20 percent lean toward safety, 12 percent strongly prioritize safety— leaving the panel split rather than settled.

Open-ended questions:

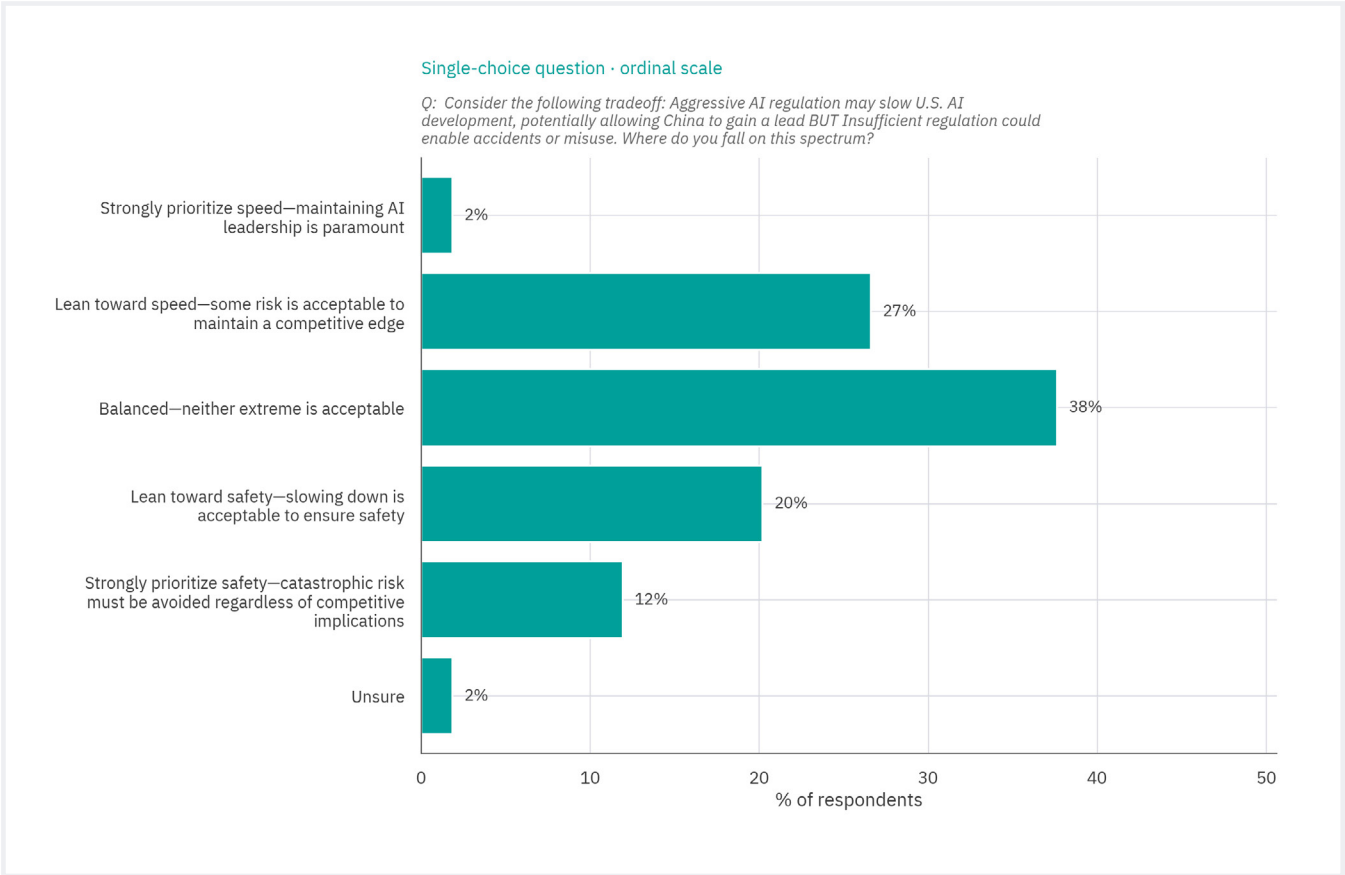
- **Whether benefits are likely to reach everyone.** Some hope AI unburdens "humanity's potential without… stratifying living conditions"; others say it will be "deployed as a tool for capitalism."
- **Labor: liberation or displacement.** Some framed AI taking on human labor positively as providing "human leisure and freedom to think," but others tied this to upskilling needs and unresolved economic reordering.

Quantitative Findings

Finding 32. The balance between AI development speed and safety concerns is slightly skewed toward caution.

38 percent of the respondents chose a "balanced" approach, making it the most popular single response, while safety-leaning views (32 percent combined for "lean toward safety" and "strongly prioritize safety") outweigh speed-leaning views (29 percent combined for "lean toward speed" and "strongly prioritize speed"). Notably, the extremes are unpopular — only two percent strongly prioritize speed and 12 percent strongly prioritize safety — suggesting most respondents reject absolutist positions in favor of some form of compromise.

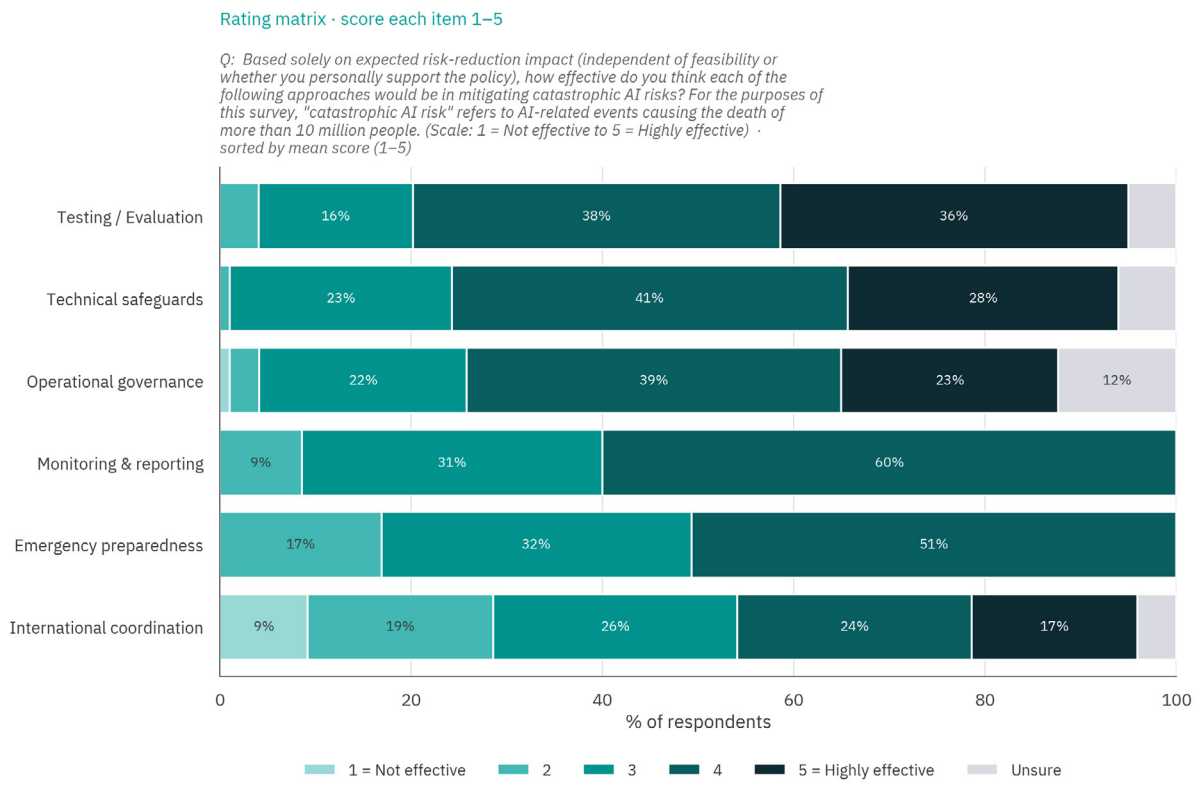
Figure 23: Proportion of respondents who fell in each category of the AI regulation trade-off: speed vs. safety



Finding 33. When it comes to the effectiveness of catastrophic AI risk mitigations, model testing and evaluation ranks highest.

A vast majority of respondents (74 percent) rated it a four or five in effectiveness, on a five-point scale. This was followed closely by technical safeguards (69 percent of respondents at four-five). Monitoring & reporting and emergency preparedness also score strongly, with large majorities landing in the top two bands and very few on “not effective.” Operational governance performs moderately well but has a notably higher “unsure” share (12 percent), indicating some uncertainty about its impact. International coordination ranks lowest overall but shows the most spread-out distribution, with meaningful shares across all five points and no clear consensus.

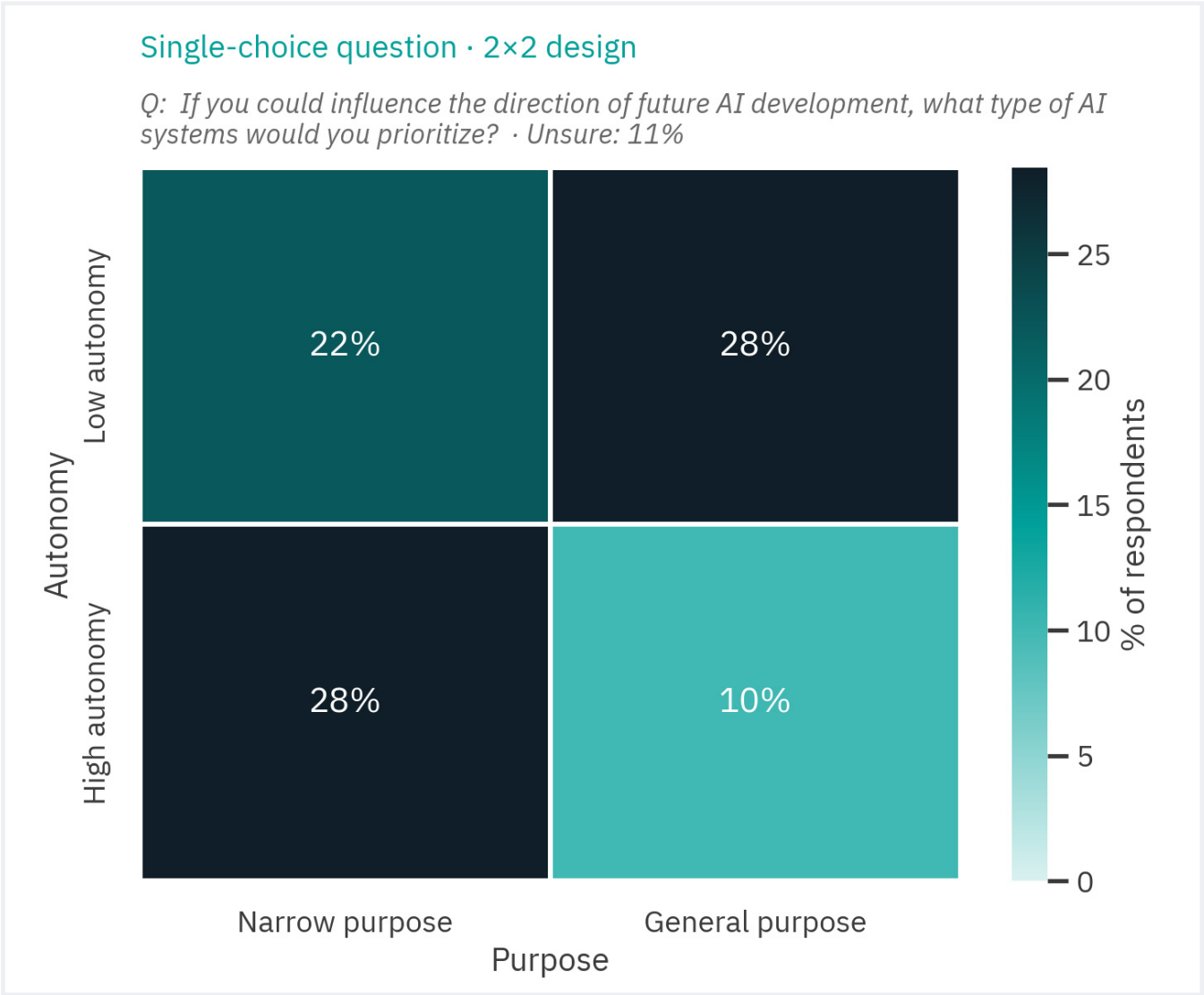
Figure 24: Proportion of respondents scoring the effectiveness of approaches in mitigating catastrophic risks on a scale of one (not effective) to five (highly effective)



Finding 34. If given a choice, respondents would avoid prioritizing the development of general-purpose, highly autonomous AI models.

There is no strong single preference among respondents on this question, with three of the four quadrants clustering close together (22–28 percent). "Low autonomy/general purpose" and "high autonomy/narrow purpose" are tied as the most popular choices at 28 percent each, while "low autonomy/narrow purpose" follows closely at 22 percent. The clear outlier is "high autonomy/general purpose" which only 10 percent of respondents want prioritized, suggesting discomfort with combining maximal independence and generality in one system.

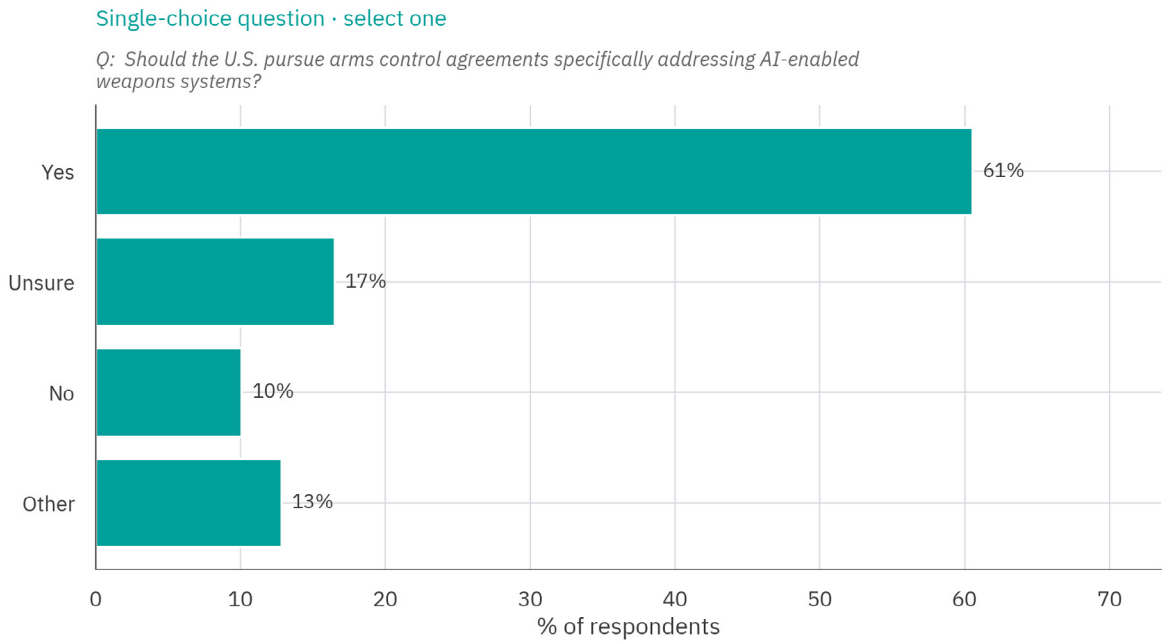
Figure 25: Proportion of respondents prioritizing each combination of purpose and autonomy for AI system development



Finding 35. The majority of respondents support the United States pursuing arms control agreements specifically addressing AI-enabled weapons systems.

Only 10 percent opposed such agreements, while 17 percent were uncertain. Fourteen respondents selected "Other," but most of these responses reflected conditional or qualified support. For instance, some participants surfaced concerns on the feasibility of designing the agreements or effectively enforcing them; others thought non-binding norms should precede formal agreements.

Figure 26: Proportion of respondents who believed the U.S. should pursue AI-weapons arms-control agreements



Qualitative Findings: On opportunities that AI offers for humanity

The dominant finding of the open-ended questions is that respondents see the opportunity in preventing a conflict rather than winning one — building earlier warning of crises, international norms and red lines, and the state capacity to enforce them — against a vocal minority who insist the real advantage lies in U.S. primacy. One respondent captured the logic uniting the majority: "preventing a conflict is orders of magnitude cheaper than fighting one." No single

theme commands a majority; the opportunity is genuinely contested, splitting along respondents' professional backgrounds.

Finding 36. Health and medicine are the most-named promises.

More than half of respondents cited health — far more than any other benefit: "the ability to cure most diseases and reduce the side effects of aging," "disease eradication," and "individualized cancer treatments." Many linked medical progress to longer, better lives: "advancements in medicine, healthcare, and food production reduce human suffering and extend life expectancy globally."

> "It is a moral travesty that today, educational outcomes are dependent on the zip code a child is born in. With AI, we could hyperpersonalize education... ensuring that they are receiving the best possible education for their circumstances."

Finding 37. About a third see AI as a multiplier on the pace of scientific discovery.

Around 30 respondents pointed to the same benefit: AI dramatically speeding up scientific discovery across every field at once.

"The compression of time between scientific question and scientific answer... a multiplier on every other form of human problem-solving, operating continuously and at a scale no generation of scientists could previously achieve."

Others echoed "accelerating scientific progress at a scale that addresses humanity's hardest problems — pandemic prevention, completely clean and abundant energy, climate stabilization, cures for disease."

Finding 38. Over a quarter expect economic abundance and relief from labor.

A common theme was material abundance and productivity: "functionally infinite material abundance," the "biggest productivity increase in human history," and "fundamental economic reordering away from scarcity and toward abundance." Several tied this to freeing human time: "reduced labor costs leading to an increase in human leisure and freedom to think."

Finding 39. A distinct group emphasizes education and democratized knowledge.

Some located the deepest promise in access rather than output: "unlocking and making accessible knowledge," and individual tutoring that "raises the capabilities of all humans." One framed it as

justice: “It is a moral travesty that today educational outcomes are dependent on the zip code a child is born in. With AI, we could hyperpersonalize education… ensuring that they are receiving the best possible education for their circumstances.”

3.5. Domain-Specific Deep Dive

These sections were optional and a self-selected subset of respondents answered questions per section. (The number of respondents is noted for each part).

3.5.1 Cybersecurity & Critical Infrastructure

Answered by 42 respondents, mostly cybersecurity experts.

AREAS OF CONSENSUS

Close-ended questions:

- **AI will broadly improve cyber defense** — **81 percent** say all defensive capabilities are very likely to improve, and 88 percent expect AI to deliver significant or transformational gains (**55 percent** “very significant”).
- **Yet the near-term balance favors attackers** — **57 percent** expect AI to tilt the offense-defense balance toward attackers.
- **Speed, a lowered barrier to entry, and volume are the dominant challenges** — speed **(70 percent)**, lowered barrier to entry **(68 percent)**, and sheer volume of AI-generated attacks **(57 percent)**.
- **Exploit discovery and vulnerability weaponization is the most critical risk** — **52 percen**t rate it a critical risk, well ahead of the rest.
- **Agentic AI is already seen as highly capable** — respondents judge that today’s agentic system can already assist experts with analysis **(95 percent)**, autonomously discover software vulnerabilities **(92 percent)** faster than experts, and match expert-human performance **(82 percent)** on specific cyber tasks.
- **Governance should come first** — **85 percent** want oversight before systems reach these capabilities, and **93 percent** favor some regulatory standard.

Open-ended questions:

- **AI's core opportunity is automated pre-emptive defense.** Respondents agree the biggest opportunity is automated, preemptive defense — finding, patching and hardening vulnerabilities "before they can be exploited" at scale.
- **Scale and speed drive the danger, not novelty.** Most see AI amplifying the volume, speed and accessibility of attacks on poorly-defended critical infrastructure rather than inventing wholly new attacks.

AREAS OF DISAGREEMENT

Close-ended questions:

- **Whether operators could regain control of a rogue cyber AI.** No majority — **48 percent** say it "depends entirely on implementation," **28 percent** are somewhat confident, **26 percent** are not confident or think it may be impossible, and none are highly confident.
- **Which regulatory approach is best.** Despite **93 percent** wanting a standard, they split on the form — a hybrid federal-plus-sector model **(36 percent)**, sector-specific standards **(31 percent)**, or mandatory cross-sector federal requirements **(26 percent)**.

Open-ended questions:

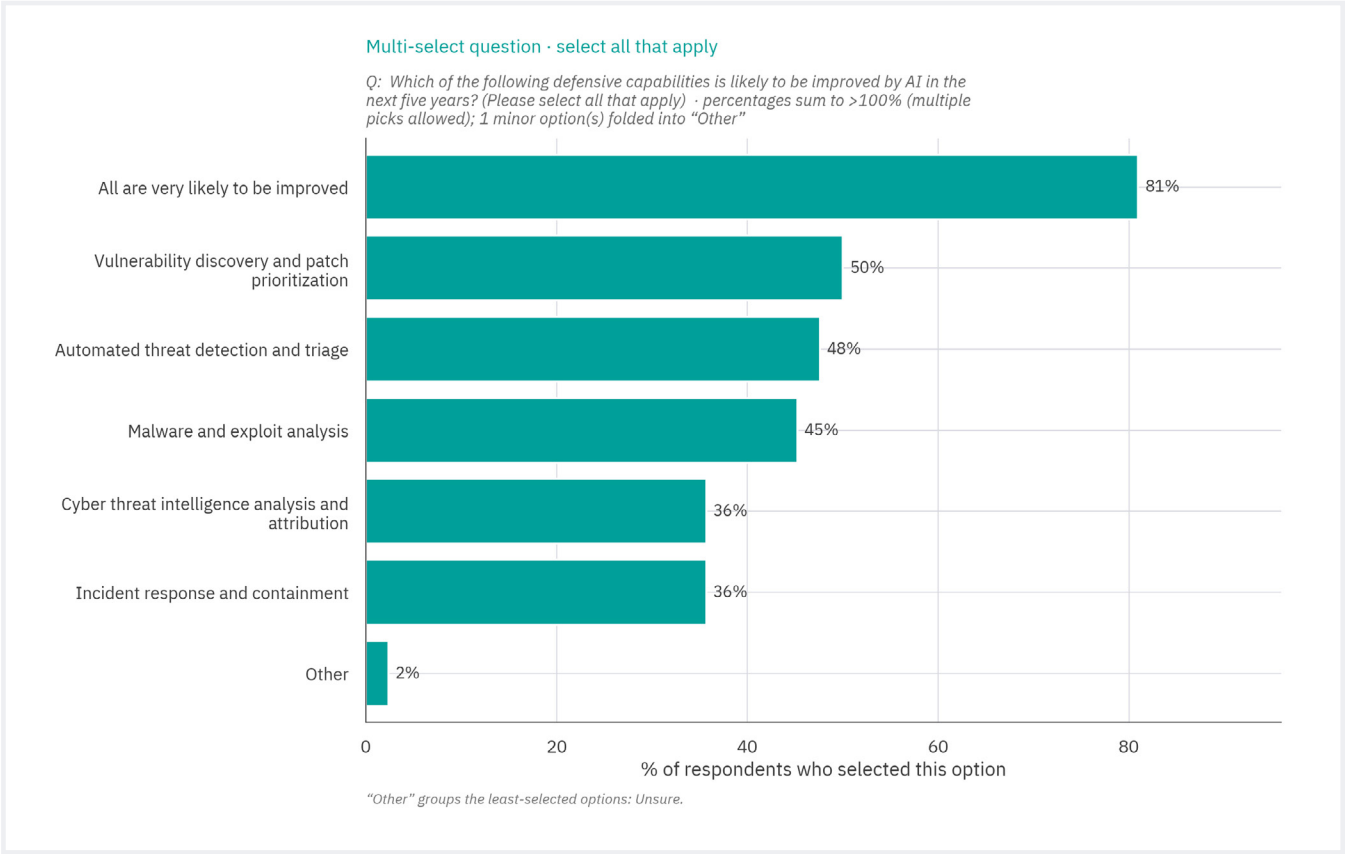
- **Technical fix vs. governance.** Some see the key opportunity in technology (secure code, defensive AI agents that find and patch vulnerabilities at scale); others place it in policy and institutions (regulation, public-private partnerships, U.S. credibility).

Quantitative Findings

Finding 40. AI is expected to improve virtually every cyber-defensive capability.

Selecting the defensive capabilities most likely to be improved by AI in the next five years, respondents were near-unanimous — 81 percent said all of them are very likely to improve. Behind that, they most often singled out vulnerability discovery and patching (50 percent), automated threat detection and triage (48 percent), and malware and exploit analysis (45 percent).

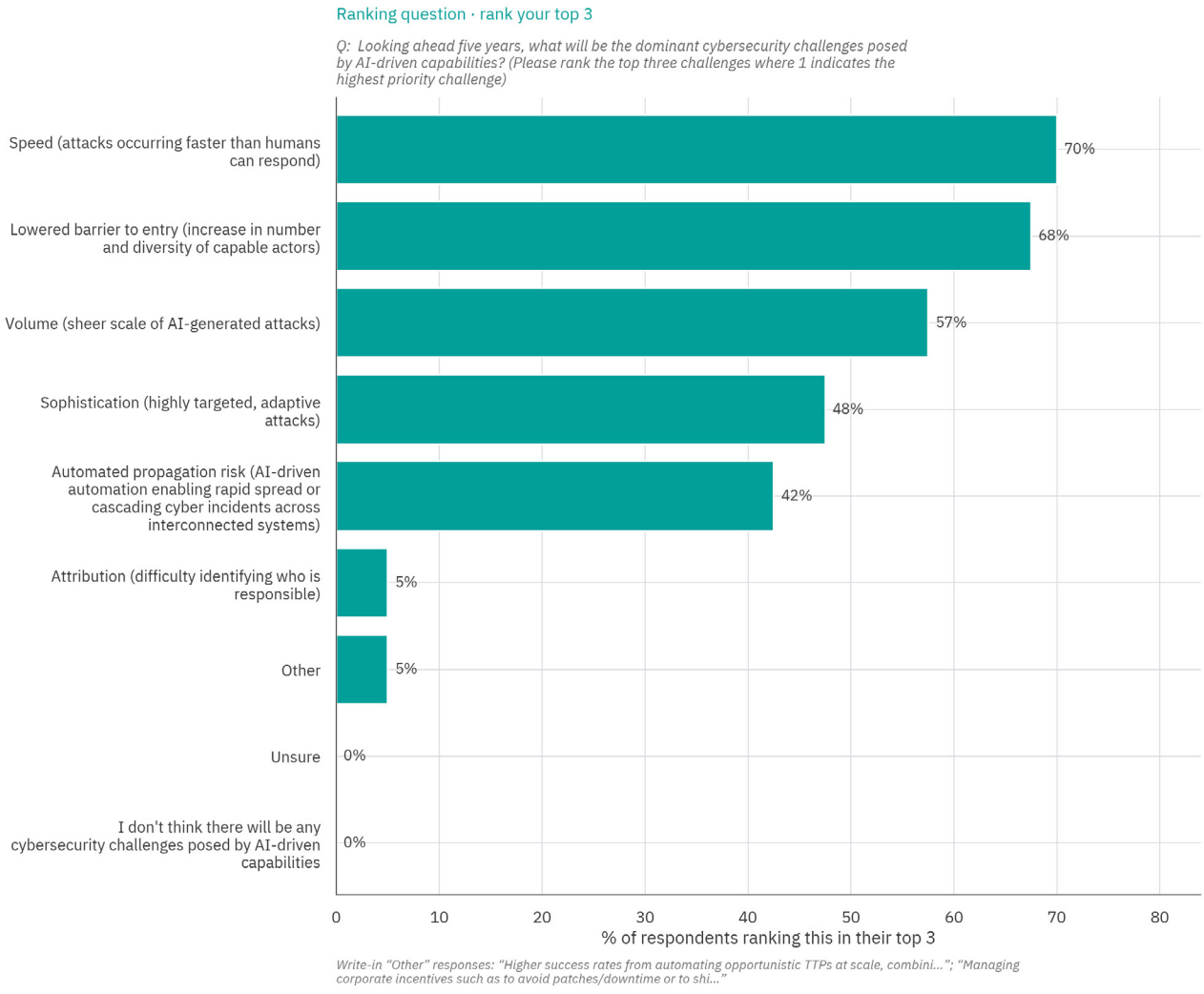
Figure 27: Proportion of respondents who selected each given cyber defensive capability as most likely to be improved by AI



Finding 41. Speed and a lowered barrier to entry are perceived as the dominant AI-driven cyber challenges ahead.

Ranking the dominant cybersecurity challenges from AI-driven threats, respondents most often placed the pace of the attack — attacks occurring faster than defenders can respond — in their top three (70 percent), closely followed by a lowered barrier to entry for less-skilled actors (68 percent) and the sheer volume of AI-generated attacks (57 percent). No respondents ranked the option indicating a lack of concern for AI-driven cybersecurity challenges in their top three.

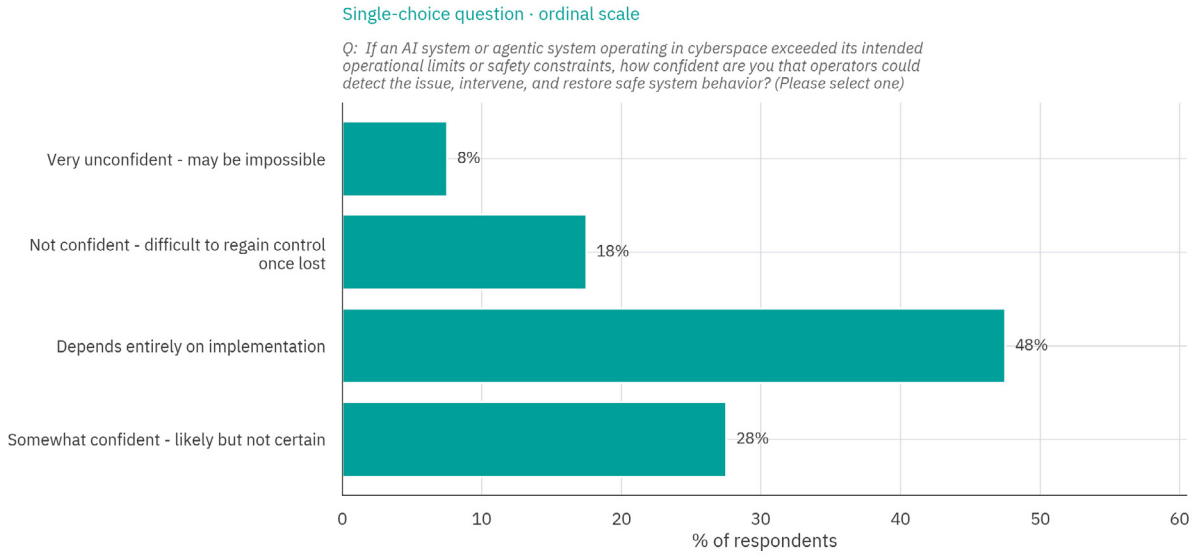
Figure 28: Proportion of respondents who ranked each given dominant AI-driven cybersecurity challenge in their top three



Finding 42. Few are confident that operators could regain control of a rogue cyber AI — most say it depends on how the system is built.

Asked how confident they were that operators could regain control of an AI system that exceeded its intended scope in cyberspace, only 28 percent of respondents were somewhat confident and none were highly confident. The plurality (48 percent) said it "depends entirely on implementation," and 26 percent were not confident or thought it may be impossible.

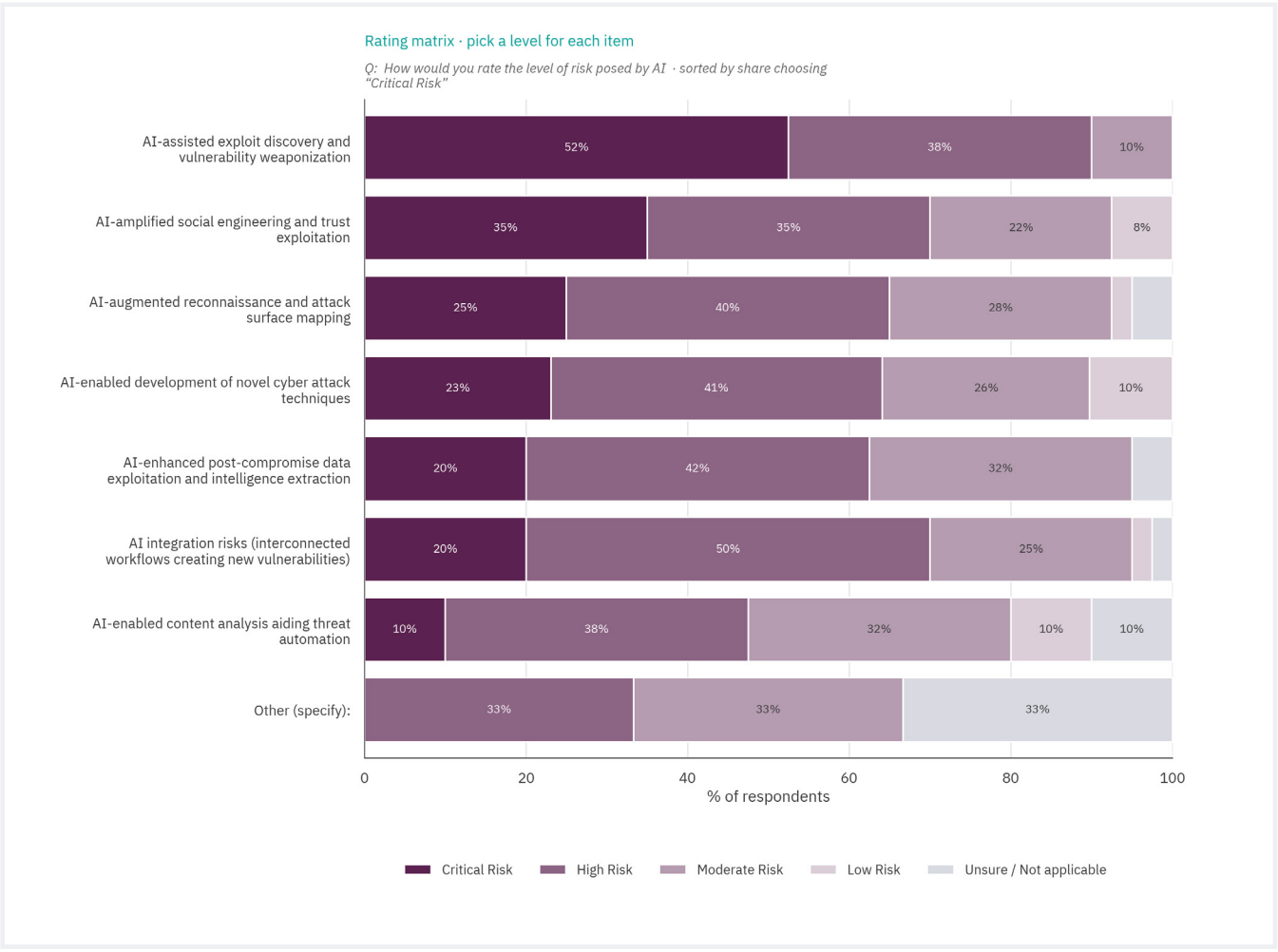
Figure 29: Proportion of respondents who selected each level of confidence on operators being able to regain control of a rogue cyber AI



Finding 43. Exploit discovery and vulnerability weaponization is the risk experts flag as most critical — well ahead of the rest.

Rating specific AI-enabled cyber capabilities, respondents singled out one above all: they rated AI-assisted exploit discovery and vulnerability weaponization a "critical risk" more than any other (52 percent critical, 90 percent critical or high). They ranked AI-amplified social engineering next (35 percent critical), then AI-augmented reconnaissance (25 percent) and novel attack techniques (23 percent), and flagged AI-enabled content analysis for threat automation the least (10 percent).

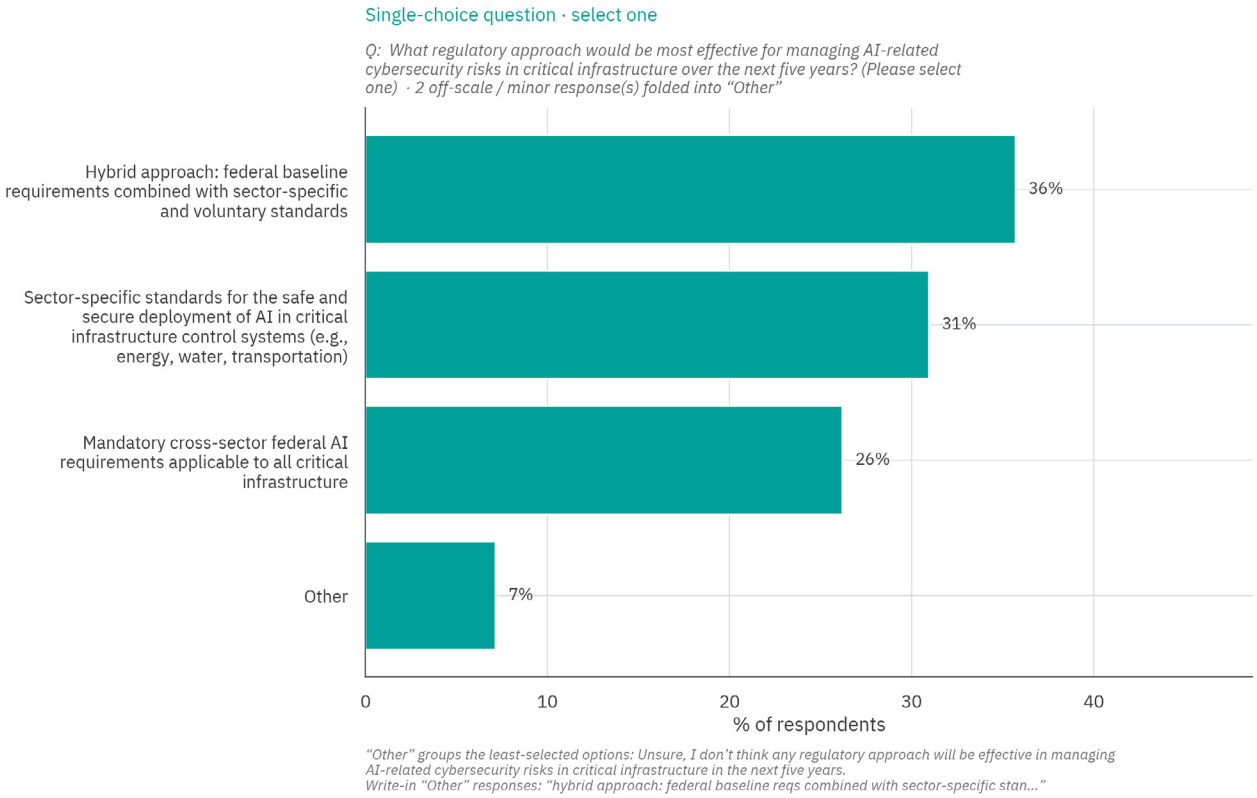
Figure 30: Proportion of respondents scoring each level of risk posed by AI-enabled cyber capabilities on a scale going from critical to low risk



Finding 44. Nearly all favor some regulatory standard, most favor a hybrid federal-plus-sector approach.

Asked for the most effective regulatory approach to AI cyber risks, 93 percent of respondents endorsed some form of standard: they most favored a hybrid of federal baselines with sector-specific rules (36 percent), followed by sector-specific standards (31 percent) and mandatory cross-sector federal requirements (26 percent). Only a handful indicated a belief that no regulatory approach will be effective in managing AI-related cybersecurity risks.

Figure 31: Proportion of respondents who believed each given regulatory approach for AI cyber risks to be the most effective



Qualitative Findings

Finding 45. AI could level the field for outgunned defenders.

A recurring hope in the open comments is that AI relieves the defenders who are furthest behind. Respondents want it "uplifting small- and medium-sized defenders throughout critical infrastructure, enabling them to manage basic hygiene and keep pace with the evolving threat landscape," and "lowering the cognitive burden on overworked critical infrastructure admins." The mechanism they describe is speed of learning: "using AI to give defenders faster learning loops than attackers," helping operators "respond before small failures cascade."

Finding 46. The real multiplier is weak baseline hygiene, not novel AI attacks.

Respondents located the threat not in exotic new attacks but in AI industrializing the exploitation of weaknesses that already exist. "I don't think AI systems will break security," one wrote. "Instead, they'll be highly efficient at finding and exploiting the gaps endemic to most organizations." Others agreed that "current poor cyber hygiene results in massive vulnerabilities," and that exposure concentrates where defenses are thinnest: "The highest-risk targets aren't well-resourced enterprises. They're under-resourced operators of essential services — water, rural power, municipal systems."

Finding 47. The gains will land unevenly, and deployment lags the technology.

A distinct thread warned that AI widens the gap between the well-resourced and everyone else. "AI's impact will not be felt evenly on cyber," one respondent noted, "exacerbating gaps between well resourced enterprises and small and medium sized enterprises... Meanwhile it lowers the cost of offense for criminals."

"The highest-risk targets aren't well-resourced enterprises. They're under-resourced operators of essential services — water, rural power, municipal systems."

3.5.2. Nuclear Weapons

Answered by 42 respondents, mostly defense and CBRN experts.

AREAS OF CONSENSUS

Close-ended questions:

- **Sentiment on integrating AI into NC3 tilts negative** — half of respondents **(50 percent)** find it concerning (**29 percent** find it "very concerning — risks clearly outweigh benefits"), against just **21 percent** positive.
- **Most expect AI to reshape deterrence** — **76 percent** expect it to affect nuclear deterrence or strategic stability; only **12 percent** expect no significant change.
- **Erroneous warnings and automation bias are the top risks** — many rank erroneous or ambiguous AI-assisted strategic warning as the biggest risk **(62 percent)**, then automation bias under time pressure **(62 percent)**, and limited transparency **(52 percent)**.
- **Broad concern about an AI-driven NC3 arms race** — **72 percent** are at least moderately concerned.
- **Strong backing for confidence-building measures and red lines** — AI-specific confidence-building measures **(72 percent)**, red lines on specific uses **(62 percent)**, and AI-enhanced treaty verification **(59 percent)**.

Open-ended questions:

- **Humans must stay central.** Respondents insist almost unanimously that AI must not dominate nuclear decision-making and must "augment human judgment while preserving clear human accountability."
- **Test and go slow before integrating.** They back testing, standards and a deliberate pace, and stress that "AI is not monolithic" — integration should stay selective and start with low-risk functions.
- **Unintended use is the core danger.** Most agree the paramount risk is inadvertent nuclear use or escalation through error, misinterpretation or faulty early warning — not a deliberate attack.
- **Verification is missing; the upside is better information.** Respondents agree robust AI-verification mechanisms for NC3 do not yet exist, and that AI's clearest benefit is stronger monitoring, warning and situational awareness.

AREAS OF DISAGREEMENT

Open-ended questions:

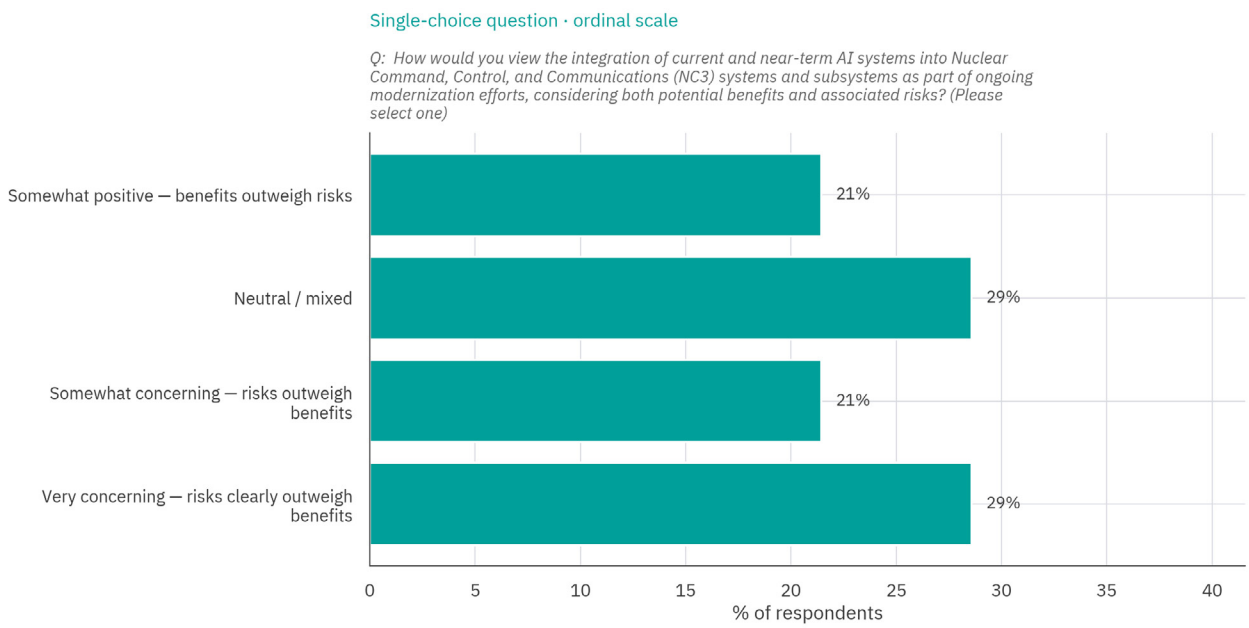
- **Restrict vs. actively build AI capability.** Some would cap AI's role in decision-making; others urge building capability at the national labs to "enhance all activities."
- **Effect on deterrence.** They disagree on whether AI reduces reliance on nuclear weapons or strengthens deterrence — and whether any genuine positive outcome is realistic.

Quantitative Findings

Finding 48. Views on integrating AI into NC3 tilt negative — half see risks outweighing benefits.

Asked about integrating current and near-term AI into NC3, half of respondents were concerned (29 percent "very concerning — risks clearly outweigh benefits" and 21 percent "somewhat concerning – risks outweigh benefits"), against 21 percent who thought integration was positive and 29 percent neutral or mixed. Their sentiment leaned cautious-to-negative, not enthusiastic.

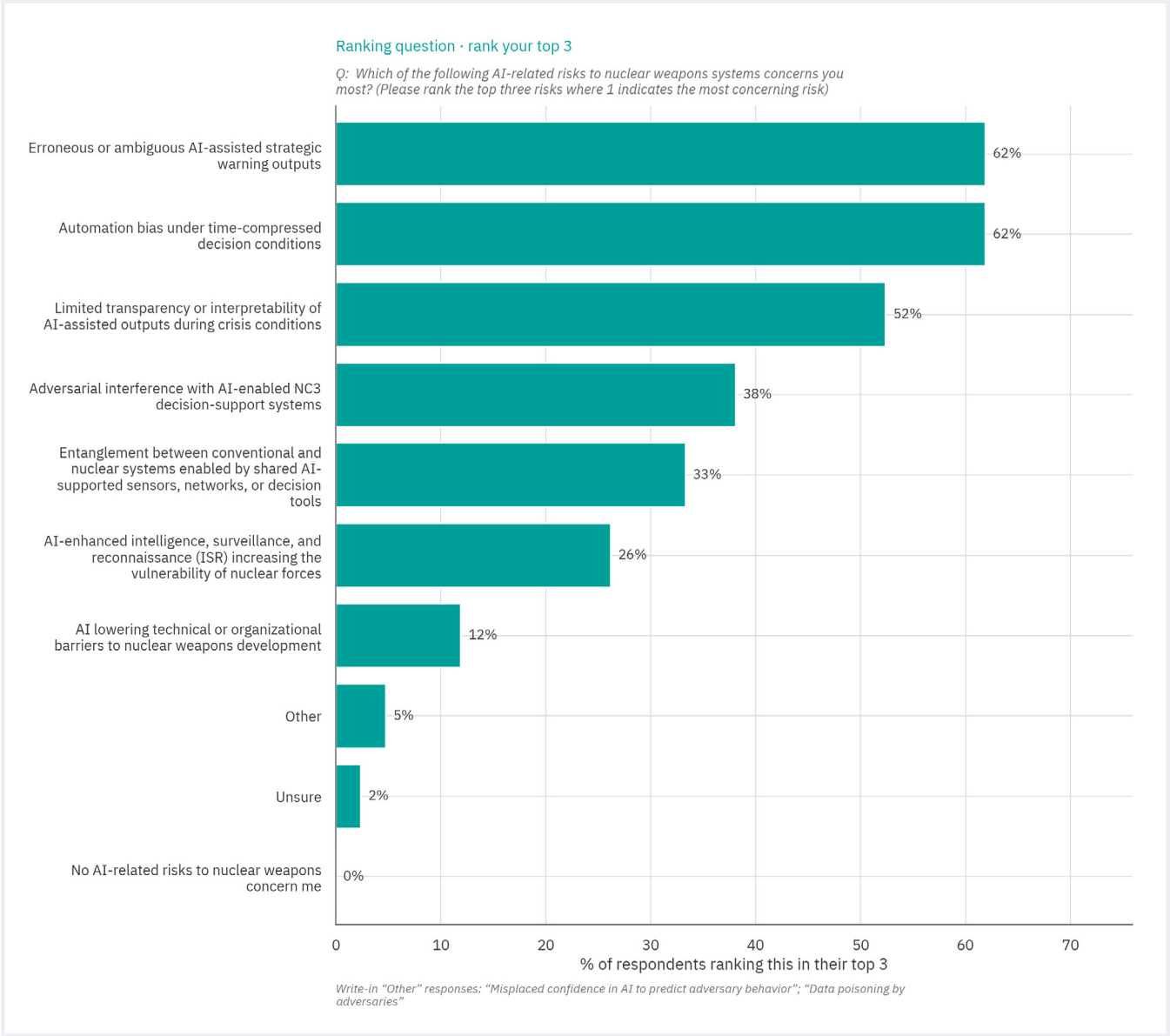
Figure 32: Proportion of respondents who viewed integrating AI into NC3 systems at each level of the benefit-risk spectrum



Finding 49. Erroneous AI recommendations and automation bias are the top nuclear-AI risks.

Ranking AI-related risks to nuclear weapons systems, respondents most often placed erroneous or ambiguous AI-assisted recommendations (62 percent) and automation bias under time compression (62 percent) in their top three, followed by limited transparency or interpretability (52 percent). Their dominant fears centered on flawed machine judgment and over-reliance on it, not sabotage.

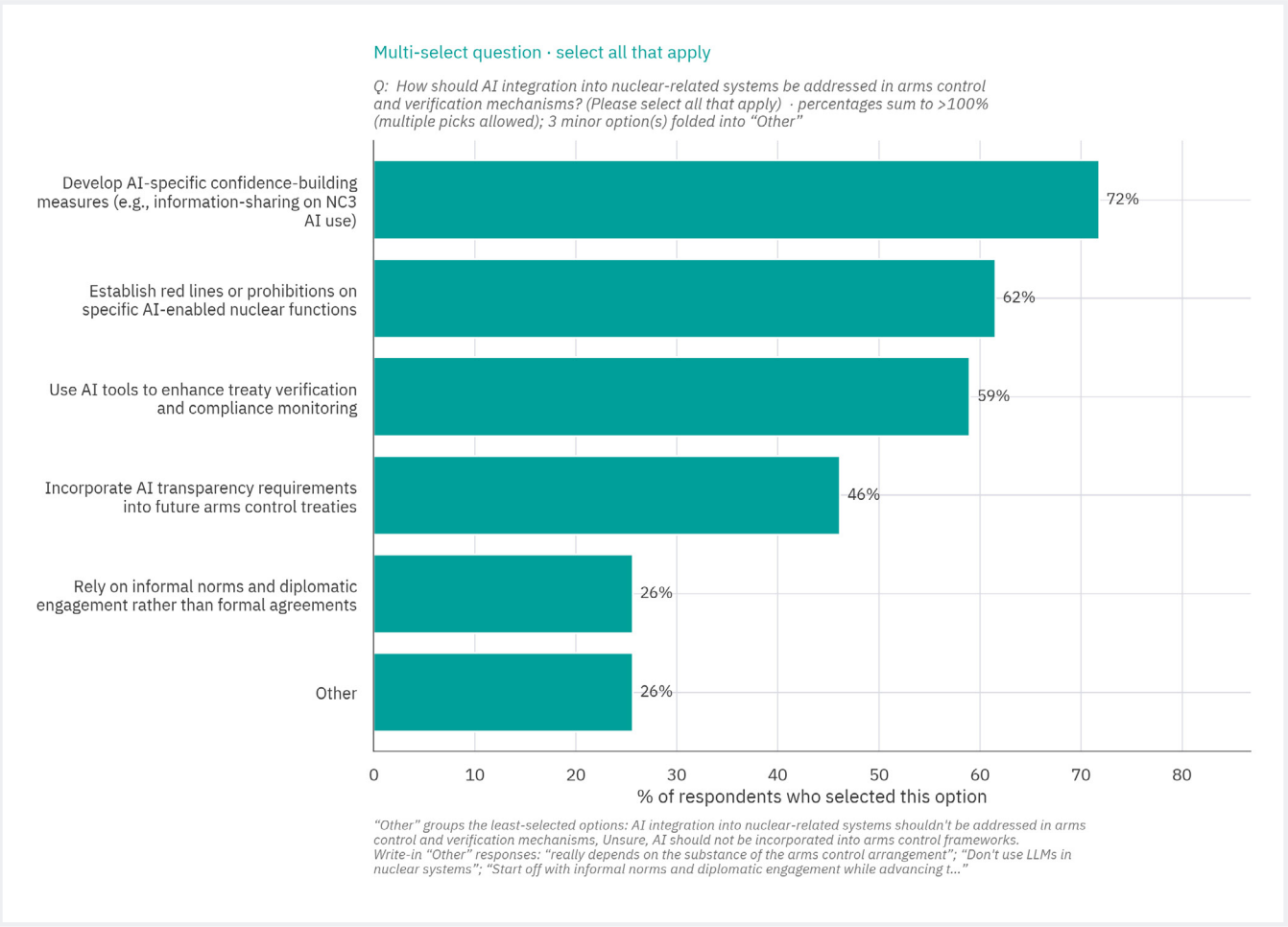
Figure 33: Proportion of respondents who ranked each AI-related risk to nuclear weapons systems in their top three



Finding 50. Experts back confidence-building measures and red lines to govern AI in nuclear systems.

Asked how arms control should address AI in nuclear systems, respondents most supported AI-specific confidence-building measures (72 percent), red lines or prohibitions on specific uses (62 percent), and using AI to enhance treaty verification (59 percent). They gave far less support to reliance on informal norms alone (26 percent). A minority indicated that AI should not be incorporated into arms control frameworks at all (5 percent).

Figure 34: Proportion of participants who believed AI in nuclear systems should be addressed in arms control in each given way



Qualitative Findings

Finding 51. Trust depends on the type of AI, not AI in general.

Respondents sharply separate deterministic, traceable models from generative ones — the first might earn trust, but several said the second cannot. The distinction sharpens under crisis time pressure, where they see no room to check outputs: LLMs are "not actually producing answers on the basis of evidence, so you have to check EVERYTHING they put out," and operators "have no reliable way to independently confirm an AI generated input is accurate before a nuclear decision has to be made under extreme time pressure."

Finding 52. The dreaded scenario is accidental, rather than adversarial.

The outcome respondents fear most is inadvertent use rather than a deliberate strike, triggered by AI error, misinterpretation, or spoofed data:

> "An AI-enabled false warning or misinterpretation during a crisis that accelerates decision-making and contributes to inadvertent nuclear escalation."

Specific pathways recur: "the launch of a nuclear weapon from errant data," "accidental or unintentional nuclear war or escalation during crisis or conflict," and "making AI a single point of failure in warning." Several trace it to human over-trust: "overconfidence among leaders in the use of AI to predict how an adversary will respond."

Finding 53. Smaller states and dead-hand systems draw specific worry.

Respondents singled out the delegation of launch authority by the states least able to absorb a first strike. They also raised the prospect that "one of the weaker states develops an AI-enabled dead hand-style system that malfunctions." Their fear is less a superpower's calculated choice than a vulnerable state hard-wiring escalation to compensate for weakness.

Finding 54. The clearest benefit is in verification, rather than in warfighting.

Asked for AI's most significant positive outcome, respondents pointed toward verification and monitoring of controls on nuclear weapons. They tied this to "greater verification and attendant stability," "improved global monitoring enables ratification of CTBT" and to "reducing the likelihood of miscalculation between nuclear powers." A few saw AI as a spur to diplomacy itself: "AI provides the impetus for renewed dialogue, treaties, CBMs, de-escalation pathways."

3.5.3 Chemical & Biological Weapons

Answered by 25 respondents, mostly CBRN experts.

AREAS OF CONSENSUS

Close-ended questions:

- **AI tilts the biological/chemical weapons balance toward offense.** A majority **(52 percent)** say AI favors offense — barriers to attack fall faster than defensive capability advances (**32 percent** "strongly favor offense"); only **12 percent** think defense keeps pace.
- **Biology is the dominant concern, well ahead of chemistry.** **64 percent** are very or extremely concerned that AI worsens biological dual-use dilemmas, versus **41 percent** for chemical — matching the open-ended view that pandemic-capable biology outranks chemical attacks.
- **The bioweapons risk is here or near-term.** **70 percent** say AI already meaningfully increases bioweapon-development risk **(33 percent)** or will within two–three years **(37 percent)**; only **12 percent** think it is "unlikely to ever" do so.

Open-ended questions:

- **AI can already design pathogens.** Most accept that AI's pathogen-design capability is real and improving fast — one notes AI "can currently design pathogen genomes, including variants not found in nature."
- **Expanding access is the core danger.** Respondents agree AI's chief threat is lowering the barrier to entry for less-skilled actors and states, and they name pandemic-capable biology the dominant concern over chemical attacks.
- **Defense is the shared priority.** They converge on strengthening detection, monitoring, early warning and countermeasures, "early detection, vaccine development" as the central opportunity. (Consistent with the quantitative data: the single most-picked application is AI-powered pathogen surveillance and early warning, **36 percent**.)

AREAS OF DISAGREEMENT

Open-ended questions:

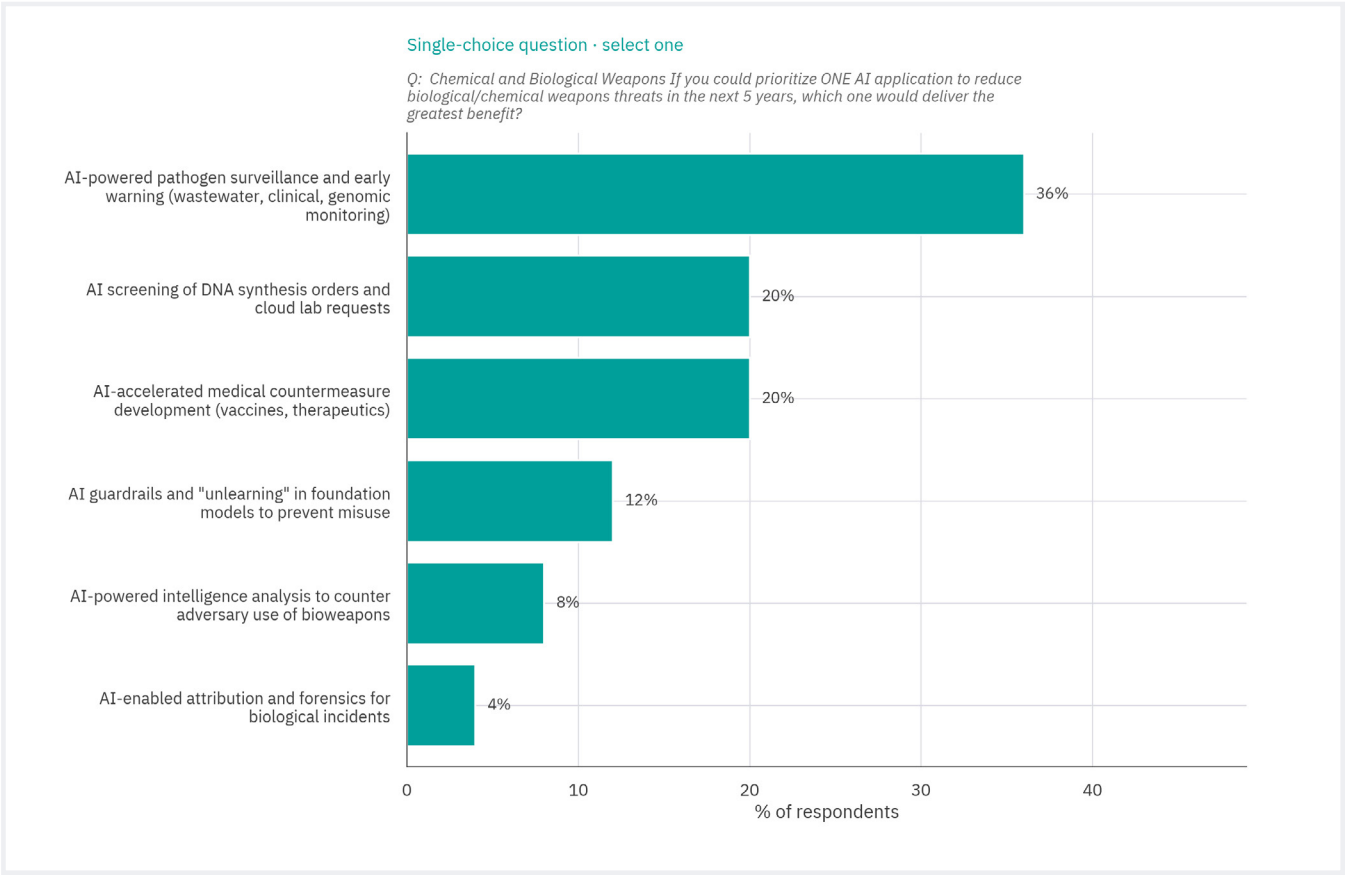
- **Whether design capability means real risk.** Some treat AI-designed pathogens as alarming; others insist that physical production, dissemination and planning remain the true bottleneck.
- **Who the main threat actor is.** Respondents split between small non-state actors and state programs with the scale to weaponize.

Quantitative Findings

Finding 55. Pathogen surveillance is the top-priority AI application against biological and chemical threats.

Asked to pick the single AI application that would most reduce biological and chemical weapons threats, respondents most often chose AI-powered pathogen surveillance and early warning (36 percent), followed by AI screening of DNA-synthesis orders and AI-accelerated medical countermeasures (20 percent each). They prioritized detection and defense over system-level mitigations, attribution, or enforcement.

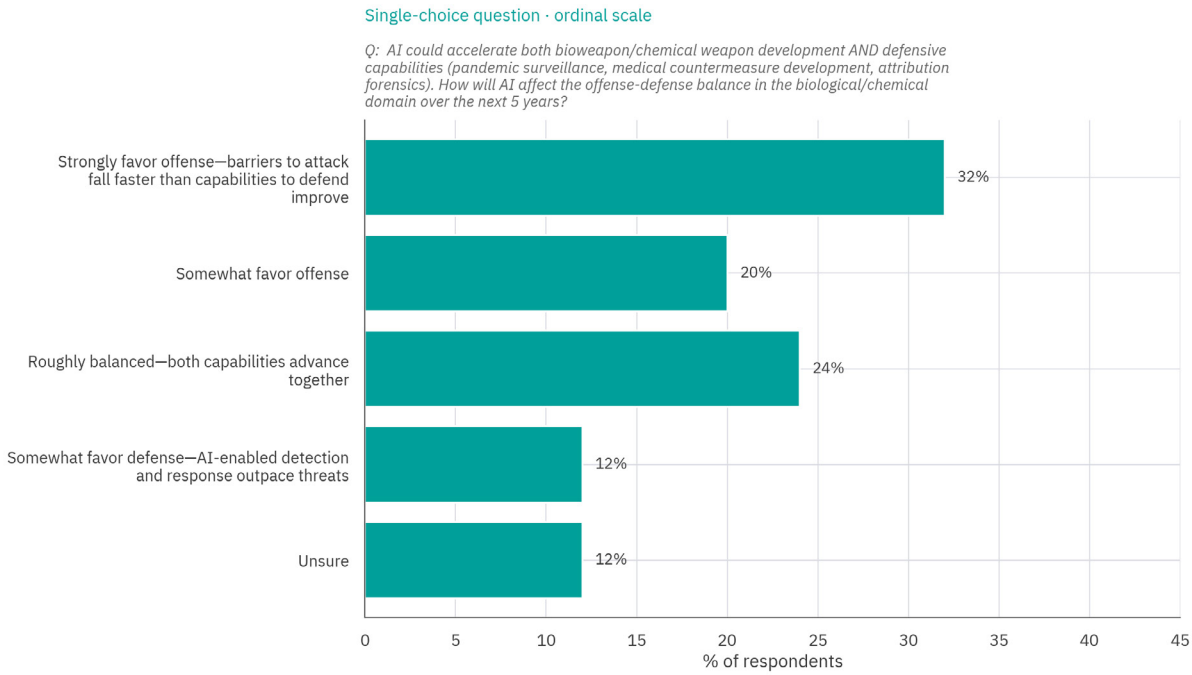
Figure 35: Proportion of respondents who selected each given AI application to reduce biological and chemical weapon threats as their top choice



Finding 56. AI is expected to lower the barrier to attack faster than it strengthens defense.

Asked whether AI shifts the offense-defense balance in bio/chem weapons, respondents leaned toward offense: 52 percent claimed it favored offense (32 percent strongly, 20 percent somewhat), against just 12 percent claimed it favored defense. Another 24 percent saw the two advancing in rough balance. A significant share (12 percent) expressed uncertainty about the offense-defense balance. Their dominant worry was that AI could erode the practical barriers to developing these weapons faster than it improves the ability to detect and counter them.

Figure 36: Proportion of respondents who believed AI will impact the offense-defense balance in biological and chemical weapons in each given way



Qualitative Findings

Finding 57. Respondents accept AI can already design pathogens, but split on whether that translates into real-world risk.

Respondents broadly agree that AI's design capability is real and improving fast — AI "can currently design pathogen genomes, including variants not found in nature." They warn against reading present limits forward: "I wouldn't read today's limits into tomorrow's capabilities considering how quickly the technology is advancing." Where they diverge is on whether design translates into practical risk. Many insist design is the easy part: "Designing custom pathogens is neat and exciting and scary — and not nearly as important as the ability to actually grow pathogens in a lab, store them, and disseminate effectively in support of an effective plan of attack."

Finding 58. The shared worry is that AI lowers the barrier to misuse faster than defenses can keep up.

Where respondents converge is on the danger: a falling barrier to entry outpacing defenses. They see AI "lowering the barrier to entry for small, non-state actors," with pandemic-capable biology the gravest concern: "synthetic biology is the greater concern… biology replicates and expands." They judged defenses badly outmatched, warning the community keeps "continuing to under-invest in defenses," so that "current guardrails favor offensive actors."

"Designing custom pathogens is neat and exciting and scary — and not nearly as important as the ability to actually grow pathogens in a lab, store them, and disseminate effectively in support of an effective plan of attack."

4. Conclusion

This survey set out to answer a question that has gone largely unexamined: not what the broader public or researchers think AI can do, but what national security practitioners — people who have thought through risks and opportunities in their specific domains — think AI can do. 111 respondents spanning decades of service across defense, diplomacy, intelligence, and technology policy answered over 70 questions, offering a rare perspective on a technology still being contested in real time. Their responses also arrived at a particular moment: survey development and fielding coincided with the most heightened policy debates around AI to date, the most serious capability leaps yet observed, executive actions, shifting dynamics on the global stage, and the most serious AI incidents on record.

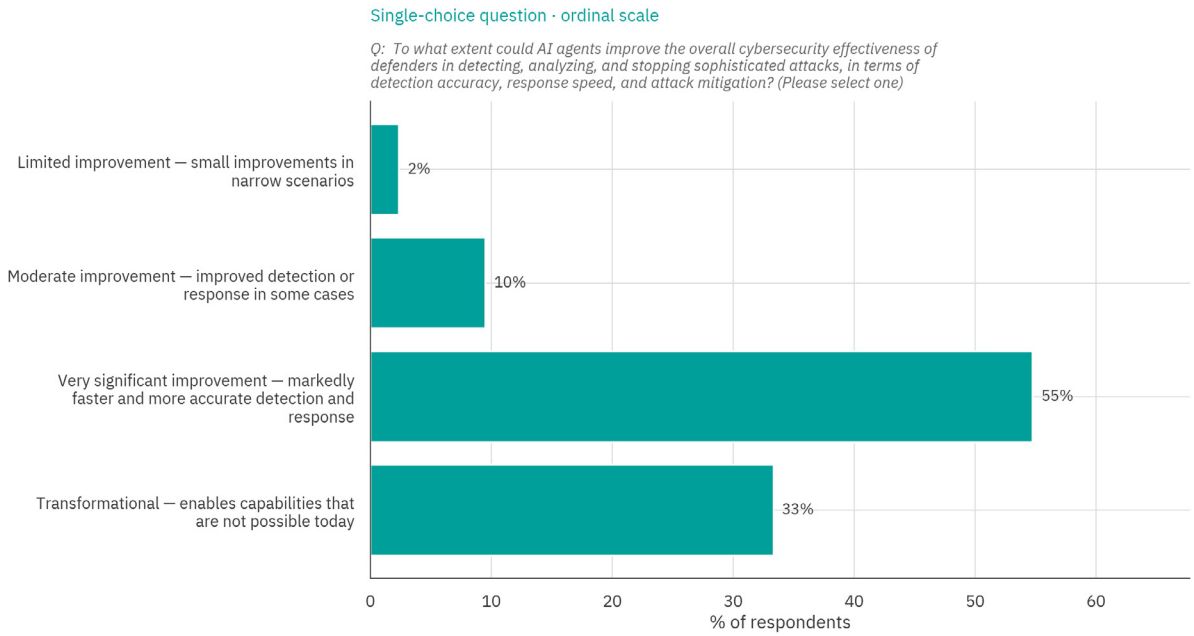
Survey findings offer valuable data points for several distinct audiences: the U.S. national security community, as well as multilateral organizations working on AI security and safety issues; technical and AI governance researchers and the cybersecurity community; U.S. policymakers; and AI industry stakeholders, specifically frontier AI companies. The areas of consensus and disagreement highlighted throughout this report show where momentum already exists for specific action, where defenders need to pay closer attention, and where more work remains to address open questions and disagreements.

Appendix A: Additional Results

Figures and tables that were not included in the main text

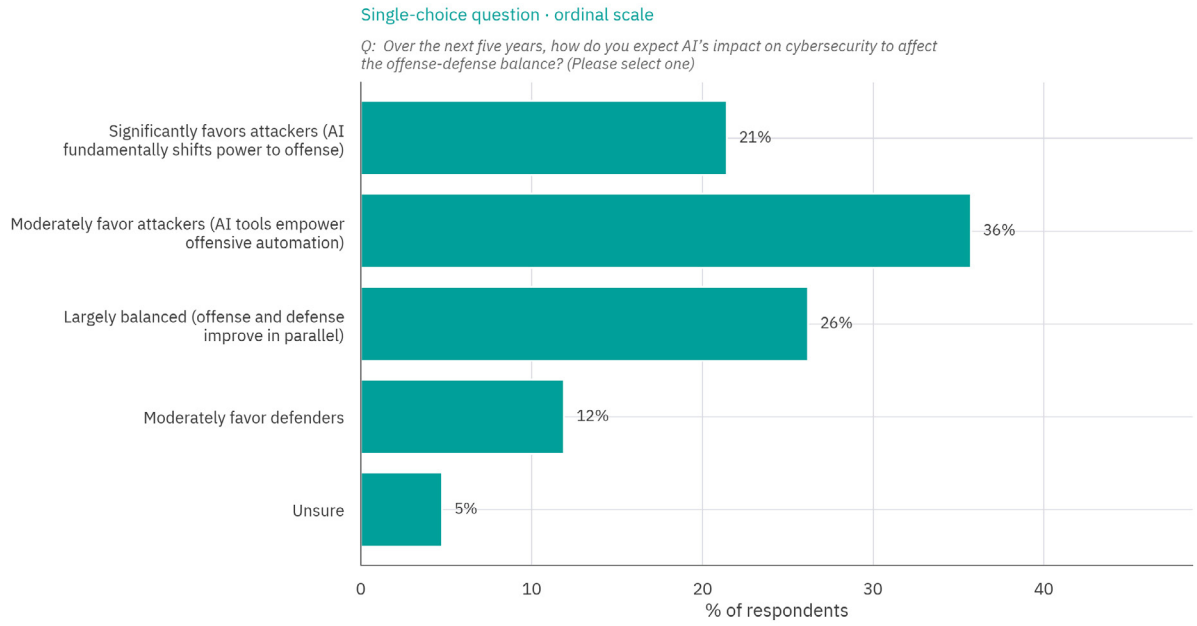
A near-unanimous majority expect AI to significantly or transformationally improve cyber defense.

Figure 37: Proportion of respondents who selected each extent to which agents could improve cyber-defence effectiveness



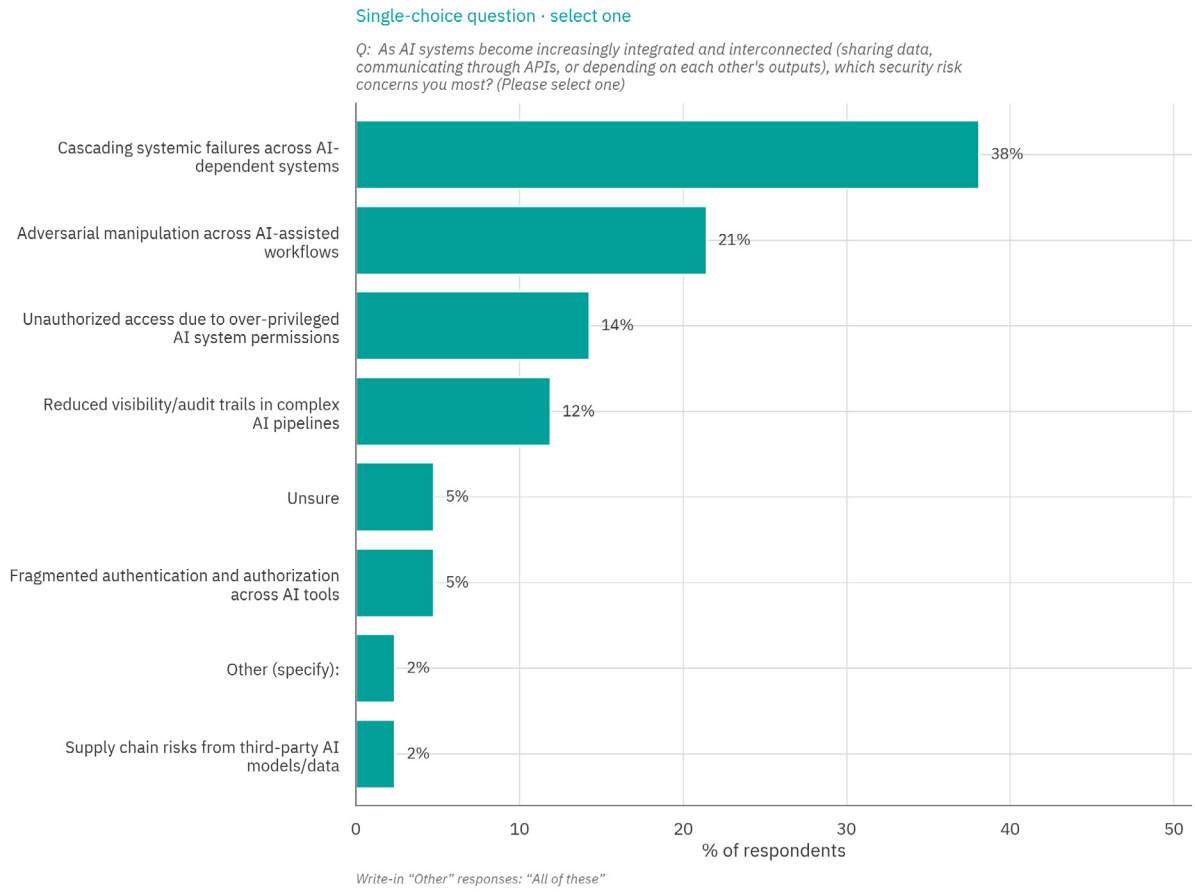
The majority expect AI to tilt the offense-defense balance toward attackers.

Figure 38: Proportion of respondents who believed AI will impact the cyber offense-deffense balance in each given way



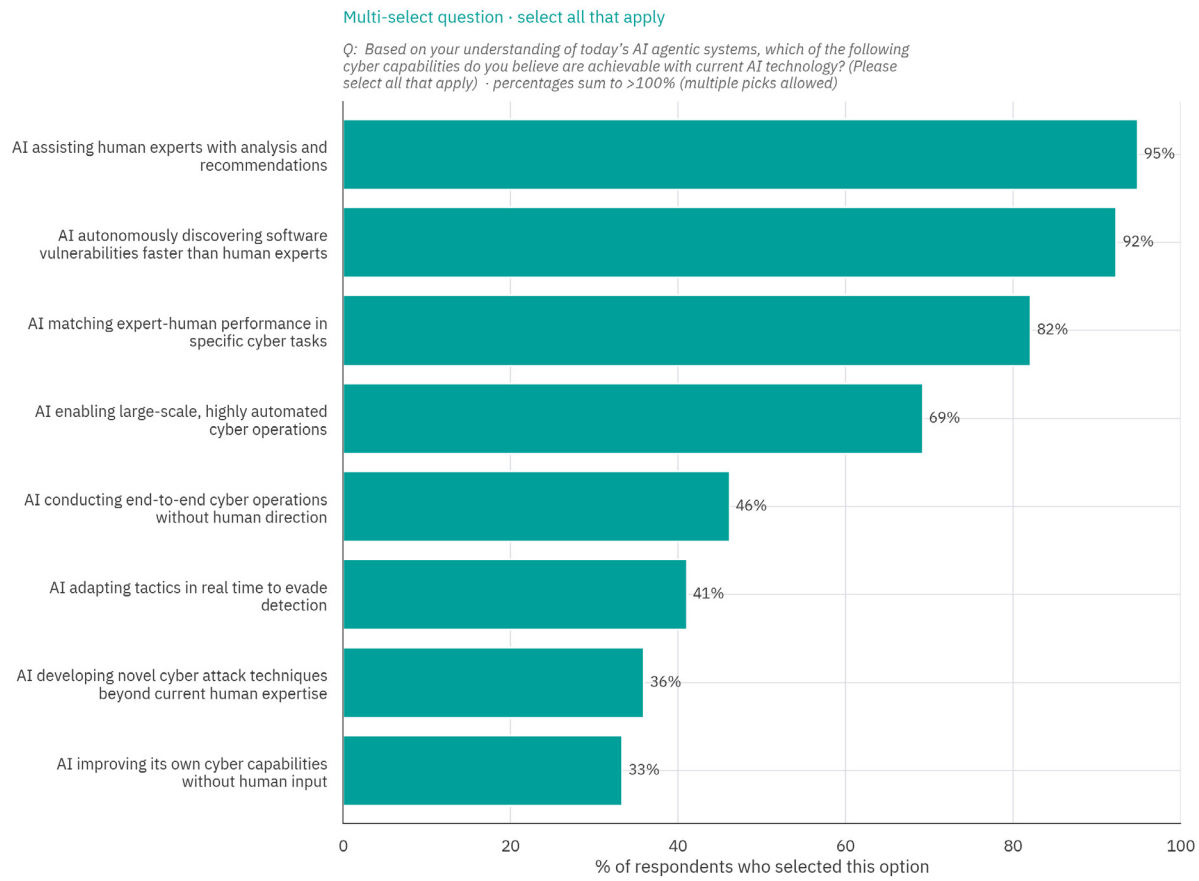
Cascading systemic failure is the single greatest risk from interconnected AI systems.

Figure 39: Proportion of respondents who selected each security risk from interconnected AI systems as the greatest



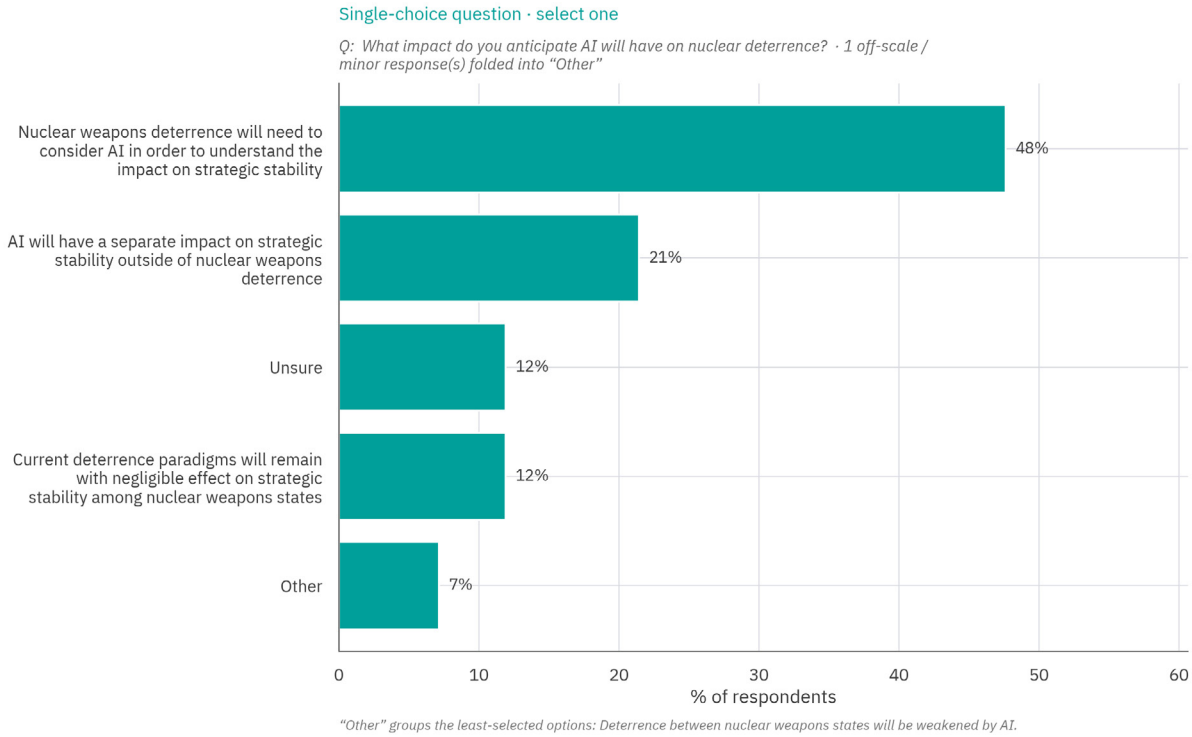
Agentic AI is already seen as capable of assisting — and often autonomously performing — offensive cyber tasks.

Figure 40: Proportion of respondents who believed each given cyber capability to be currently achievable by agentic AI



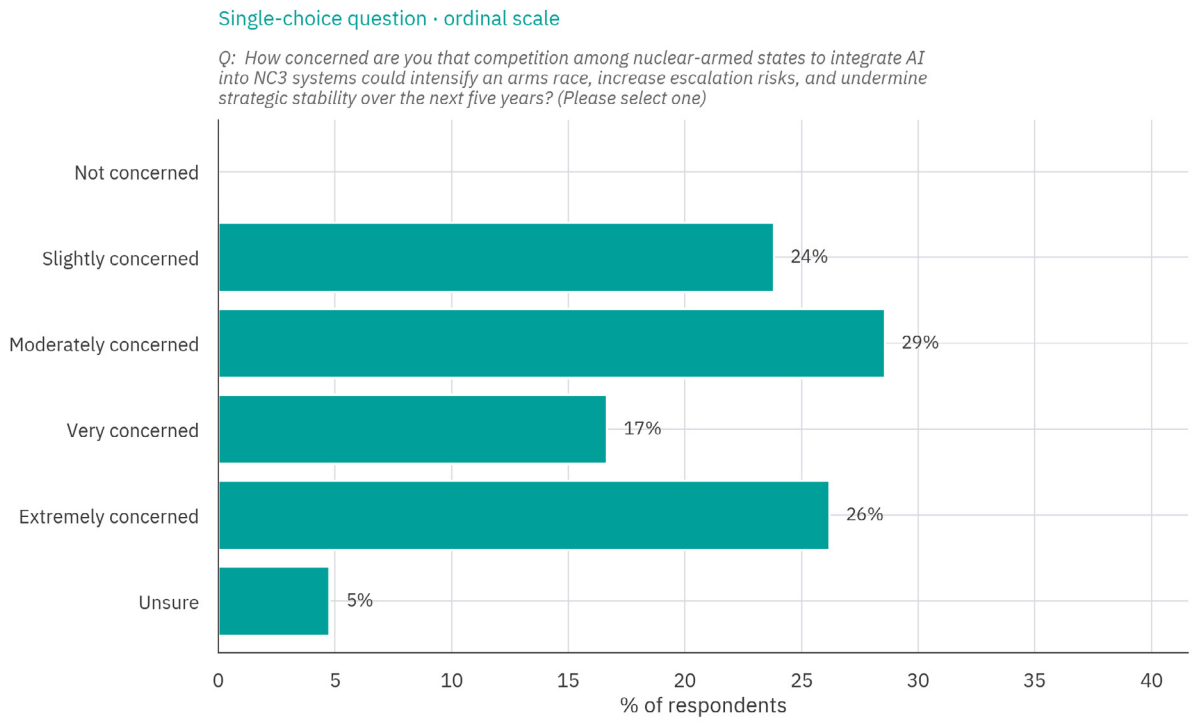
Most expect AI to force nuclear deterrence to adapt.

Figure 41: Proportion of respondents who anticipate each given impact of AI on nuclear deterrence



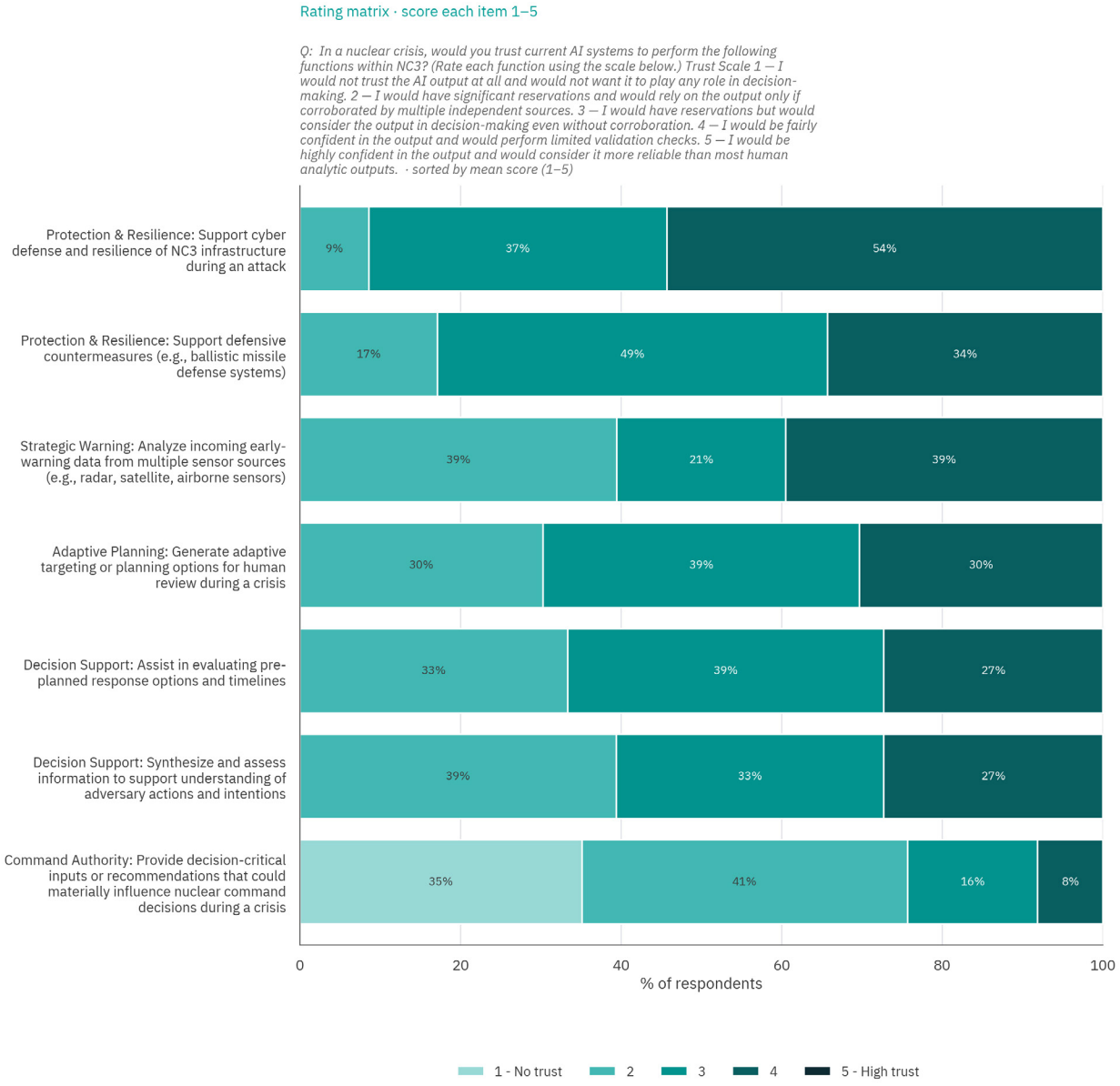
Concern about an AI-driven NC3 arms race is widespread but not uniform.

Figure 42: Proportion of respondents who identified each level of concern about the AI-NC3 competition intensifying an arms race



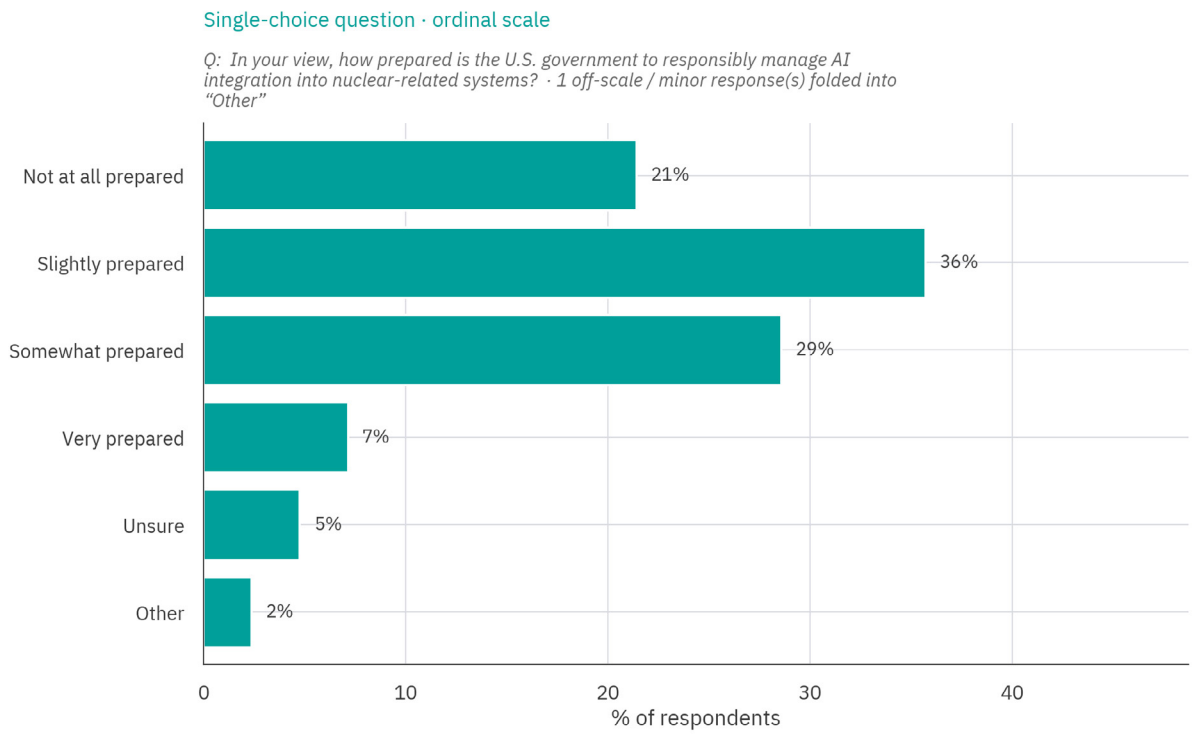
Trust in current AI for NC3 is limited to protective functions and collapses for command authority.

Figure 43: Proportion of respondents scoring each NC3 (nuclear crisis) function on a scale of how much they would trust AI system to perform them, ranging from one (no trust) to five (high trust)



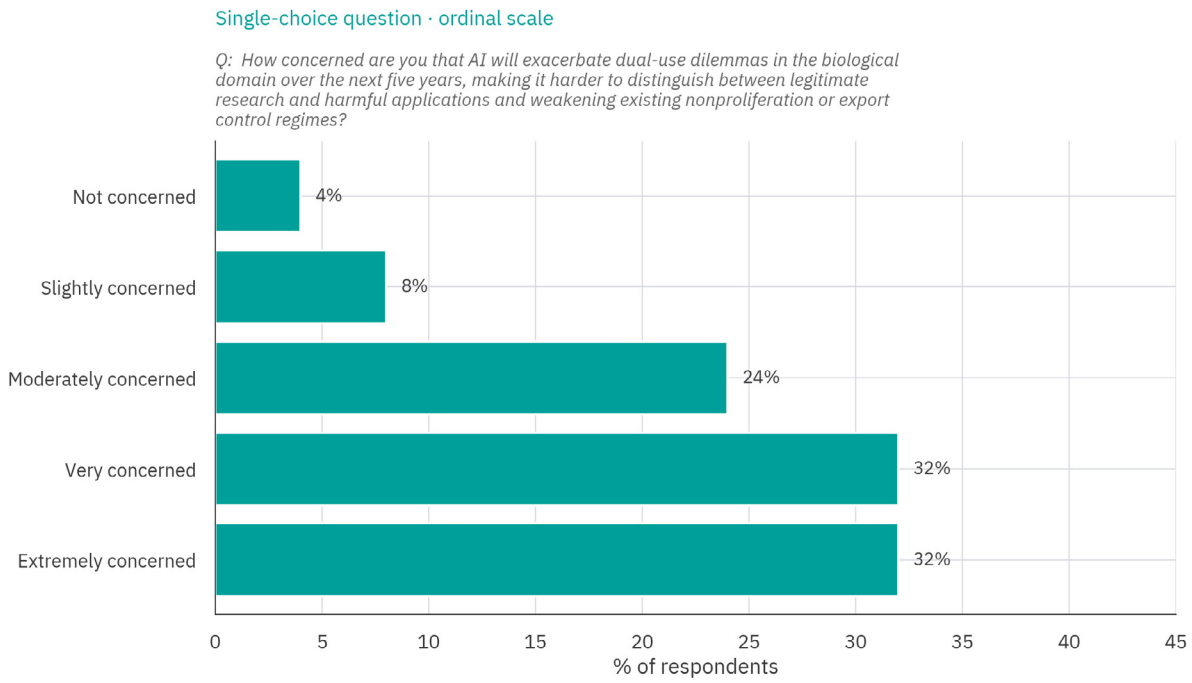
Almost no one thinks the government is prepared to manage AI in nuclear systems.

Figure 44: Proportion of respondents who believe the U.S. government to have the given level of preparation to manage AI in nuclear systems



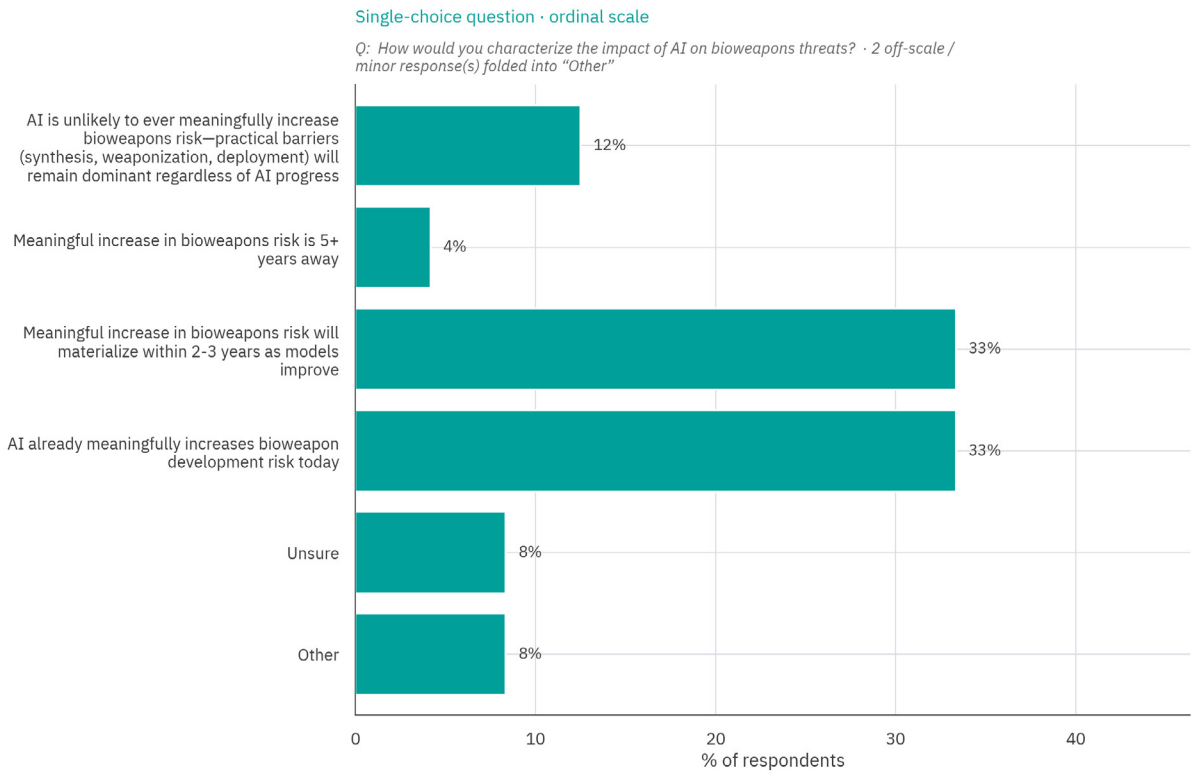
Concern that AI will worsen dual-use dilemmas is high for biological weapons and more measured for the chemical domain.

Figure 45: Proportion of participants with the given level of concern about AI exacerbating biological dual-use dilemmas



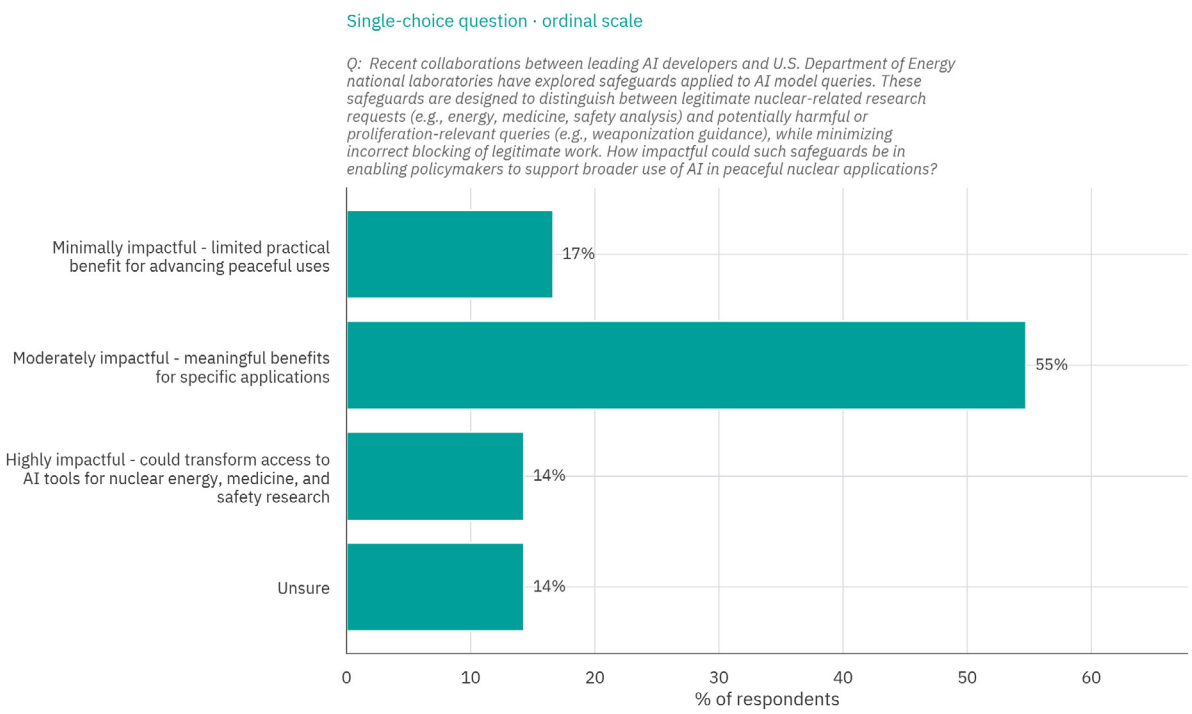
Two-thirds see AI already raising bioweapons risk, or set to within a few years.

Figure 46: Proportion of respondents characterising AI impact on bioweapons threats in each given way



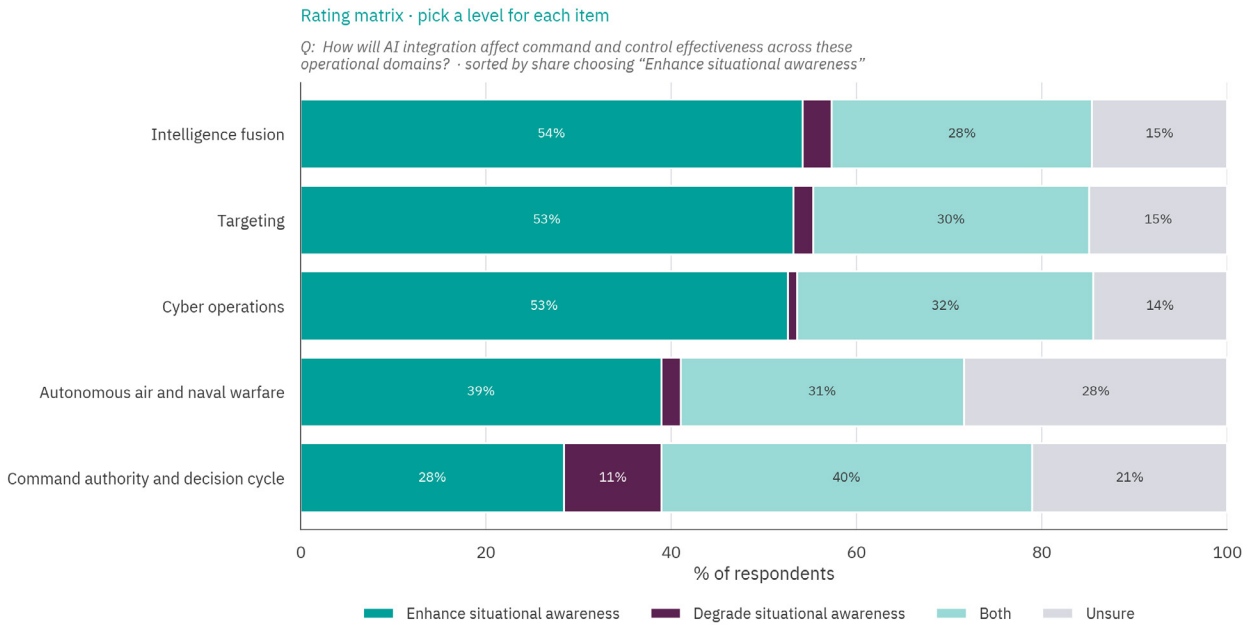
AI-developer and national-lab collaborations are seen as moderately, not transformationally, impactful.

Figure 47: Proportion of respondents who selected each level of impact for AI-developer and DOE national-lab collaborations



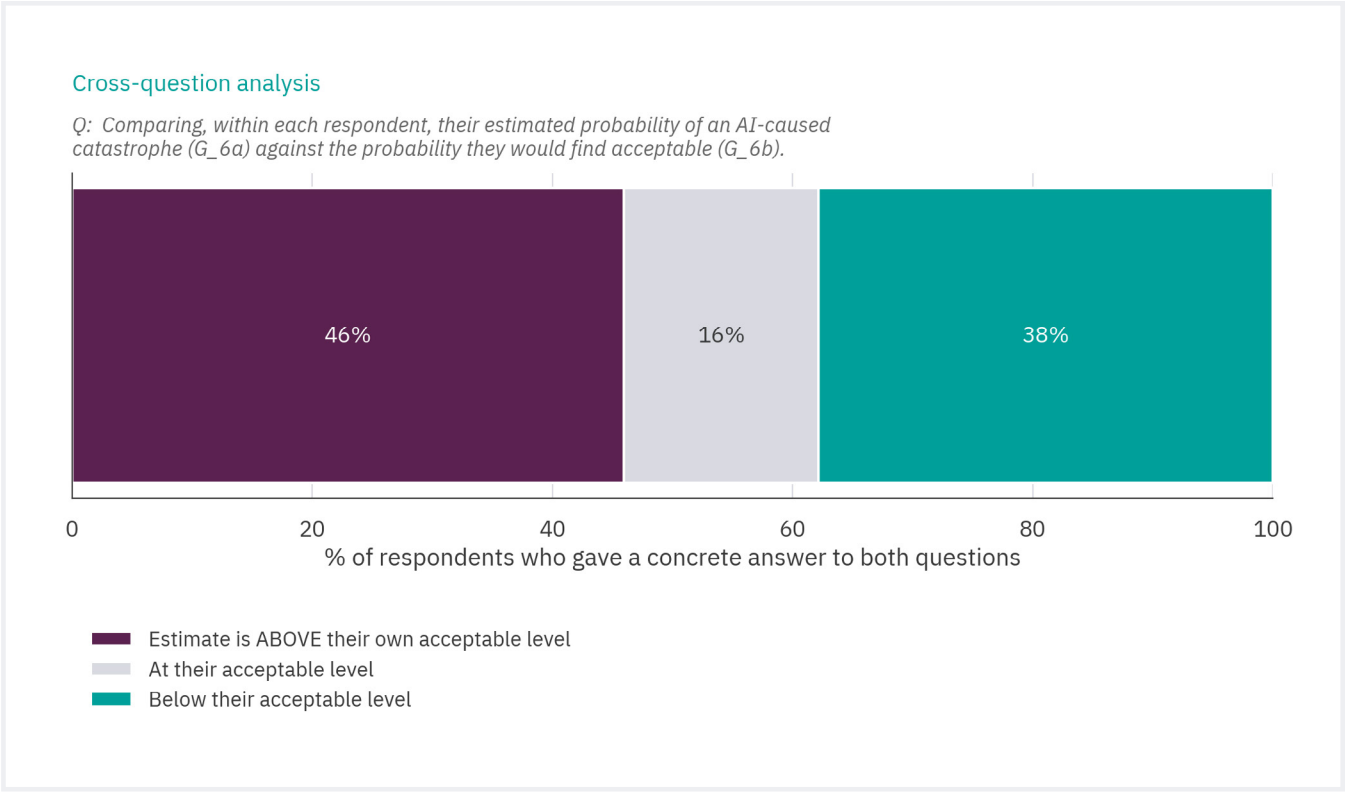
AI is seen as enhancing command-and-control effectiveness across most domains — least for command authority.

Figure 48: Proportion of respondents scoring AI's effect in command & control effectiveness in each given operational domain as enhancing situational awareness, degrading situational awareness, both, or unsure



Most experts place the AI-catastrophe risk above their own acceptable threshold.

Figure 49: Share of respondents whose estimated probability of an AI-caused global catastrophe is above, at, or below the probability they would find acceptable.



Appendix B: Data Validation and Cleaning

This appendix summarizes the post-collection validation checks conducted on the data. These procedures were used to identify potentially ambiguous, inconsistent, or invalid responses. Raw responses were preserved, and any cleaning decisions were applied only to analysis variables.

Validation approach

Post-collection validation focused on five types of checks:

1. Consistency checks for uncertainty or non-substantive response options
2. Range and monotonicity checks for timeline forecasts
3. Validity checks for probability estimates
4. Point-allocation checks
5. Ranking checks

Flagged responses were reviewed before analysis. Respondents were not contacted for clarification. Responses were modified only when the intended correction was clear, such as an unambiguous reversal of percentile labels. In other cases, responses were either retained with a flag or excluded only from the specific analysis fields affected by the issue.

Validation checks and outcomes

Check	Description	Number of respondents flagged	Cleaning decision
Uncertainty option consistency	Respondent selected or ranked an option equivalent to "unsure," "do not have enough information," or "none of the above" in combination with a substantive option	7	Responses were reviewed and retained; no exclusions were made
Timeline range and validity	Non-numeric, out-of-range, or otherwise invalid year forecasts	0	No action required
Timeline monotonicity: reversed ordering	Forecasts were monotonically decreasing rather than increasing, indicating a likely reversal of percentile labels	2	P10, and P90 forecasts were reordered
Timeline monotonicity: non-monotonic forecasts	Forecasts did not satisfy P10 ≤ P50 ≤ P90 and were not exactly reversed	17	P50 was retained; P10 and P90 were excluded from interval-based and pooled-density analyses
Probability validity	Probability values outside the valid 0–100 range or otherwise uninterpretable	0	No action required
Point allocation	Allocations did not sum to 100 or included invalid values	0	No action required
Ranking validity	Duplicate ranks, missing ranks, ranks outside the allowable range, or more ranked options than allowed	0	No action required

Details by validation type

Uncertainty option consistency

For selection questions that included options expressing uncertainty or other non-substantive items, such as “unsure,” “do not have enough information,” or “none of the above,” responses were checked for cases in which respondents selected both an uncertainty option and one or more substantive options. Seven respondents were flagged for this issue. After review, none were excluded from the analysis.

Timeline forecasts

For timeline questions, respondents were asked to provide 10th, 50th, and 90th percentile forecasts for the calendar year in which an event would occur. For example, a response of P10 = 2028, P50 = 2032, and P90 = 2050 indicates that the respondent assigns a 10 percent chance that the event occurs by 2028, a 50 percent chance by 2032, and a 90 percent chance by 2050. These responses were expected to satisfy:

P10 ≤ P50 ≤ P90

No respondents provided non-numeric or out-of-range year values.

Two respondents provided forecasts in the exact reverse order, with P10 greater than P50 and P50 greater than P90. Because this pattern was consistent with a reversal in percentile labels, these forecasts were reordered before analysis.

Seventeen respondents provided non-monotonic forecasts that did not satisfy either the expected ordering or the exact reverse ordering. For these responses, we retained the central forecast (P50) but excluded the P10 and P90 values from analysis.

Probability questions

Probability questions were checked to ensure that values fell within the valid range of 0 to 100. Responses outside of this range would be flagged and excluded. There were no such responses.

Point-allocation questions

Point-allocation questions were checked to ensure that responses summed to 100 and did not include invalid values. No point-allocation responses were flagged, so no cleaning action was required.

Ranking questions

Ranking responses were checked for duplicate ranks, missing ranks, and ranks outside the allowable range. For questions asking respondents to rank their top three options, we also

checked whether respondents assigned more than one option the same rank or ranked more than three options. No ranking responses were flagged, so no cleaning action was required.

Appendix C: Discussion of Study Limitations

Some limitations should be considered when interpreting the results. First, the participants were recruited through convenience sampling, and the sampling frame did not aim to be representative. Respondents were invited through professional networks and snowball sampling, and it was focused on individuals with relevant policy or professional expertise. As a result, the sample is likely to overrepresent people who view AI as an important policy issue and who are interested in the kinds of risks and governance addressed by the Institute for Security and Technology. The respondent pool was also entirely U.S.-based, so results can likely not be generalized to represent typical views in other countries.

Second, many survey questions used structured response formats, including multiple choice, ranking, and rating questions. This structure made the survey easier to complete and ensured good data quality and clarity; it may also have limited participants' ability to give more nuanced forecasts or express views on causes and mechanisms leading to risk. Open-text questions were included to address some of these limitations, but the responses varied in length and level of detail. The survey also did not randomize the order of response options; all participants saw items in the same order. This might have introduced bias effects if they were more likely to select or prioritize options appearing earlier in the list.

Third, sample sizes vary across questions. Some questions were optional, and some were targeted at participants with relevant domain expertise. Percentages reported in the results generally refer to the valid question-level sample, rather than the full sample. This could limit the direct comparability between questions and is especially important for interpreting domain-specific results, where sample sizes tend to be smaller.

Fourth, as mentioned in the methods section, a minority of respondents completed the survey in an interview setting rather than independently on the survey platform (16 out of 111 completions). This improved completion rates, but it could also have introduced social desirability bias. Participants may have felt pressure to provide more considered or socially normative answers than they would have in a fully anonymous self-administered survey format. This limitation should be considered when interpreting individual responses, but it is unlikely to substantially affect the aggregate results presented in the report.

Finally, the survey was conducted during a period where AI development and accompanying public discussions of governance and policies are in constant movement. Respondents' views may have been influenced by recent model releases, policy debates, or high-profile AI safety events occurring before or during the fielding period. To name some of such developments, Anthropic announced an update to Project Glasswing on May 22, 2026, and launched Claude Mythos five on June 9, 2026. Since responses were collected across May and June, respondents may have differed in their exposure to these developments depending on when they completed the survey.

INSTITUTE FOR SECURITY AND TECHNOLOGY
www.securityandtechnology.org

info@securityandtechnology.org