

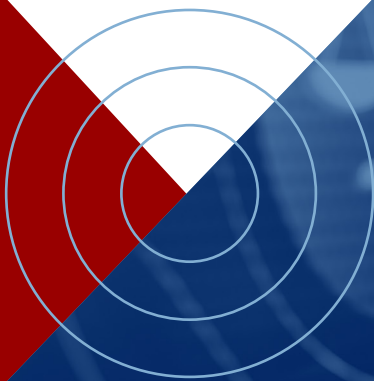


PERRY  
WORLD  
HOUSE  
UNIVERSITY of PENNSYLVANIA

POST-CONFERENCE REPORT

# The Future of Artificial Intelligence Governance and International Politics

NATALIA HENRY  
THOMAS J. SHATTUCK



## ABOUT PERRY WORLD HOUSE

Perry World House is a center for scholarly inquiry, teaching, research, international exchange, policy engagement, and public outreach on pressing global issues. Perry World House's mission is to bring the academic knowledge of the University of Pennsylvania to bear on the world's most pressing global policy challenges and to foster international policy engagement within and beyond the Penn community.

Located in the heart of campus at 38th Street and Locust Walk, Perry World House draws on the expertise of Penn's 12 schools and numerous globally oriented research centers to educate the Penn community and prepare students to be well-informed, contributing global citizens. At the same time, Perry World House connects Penn with leading policy experts from around the world to develop and advance innovative policy proposals.

Through its rich programming, Perry World House facilitates critical conversations about global policy challenges and fosters interdisciplinary research on these topics. It presents workshops and colloquia, welcomes distinguished visitors, and produces content for global audiences and policy leaders, so that the knowledge developed at Penn can make an immediate impact around the world.

## ABOUT THE AUTHOR

**Natalia Henry** is a PhD student in Political Science at the University of Pennsylvania.

**Thomas J. Shattuck** is a senior program manager at Perry World House at the University of Pennsylvania.

*This workshop is made possible in part by Carnegie Corporation of New York and the Schlein Emerging Technologies and Global Politics Initiative Fund.*

# Introduction

<< Amidst a period of intensifying geopolitical competition, the governance of artificial intelligence (AI) has emerged as a defining question for the future of international security, prosperity, and cooperation. >>

Amidst a period of intensifying geopolitical competition, the governance of artificial intelligence (AI) has emerged as a defining question for the future of international security, prosperity, and cooperation. The United States, People's Republic of China, and other powers are rapidly advancing AI capabilities across the civilian, commercial, and military domains. In particular, the pursuit of Artificial General Intelligence (AGI)—a model that could be equivalent to or better than human experts across a broad range of tasks and has the ability to improve itself—adds further urgency, raising fundamental questions about global stability. Overall, the rapid development and deployment of AI systems—from generative models to autonomous platforms—is raising difficult trade-offs between innovation and safety, as well as competition and cooperation.

To examine these dynamics, Perry World House (PWH), the University of Pennsylvania's hub for global affairs, convened a conference on October 6-7, 2025, with experts, scholars, and policymakers on AI technology, international relations, and national security. This event built on previous efforts by PWH as part of its Emerging Technologies and Global Politics Project, including prior conferences on artificial intelligence and

global security, a Penn-wide conference on AI policy, the Democracy and Emergent Technology series in cooperation with the Andrea Mitchell Center for the Study of Democracy, and recent academic collaborations between PWH and the RAND Corporation.<sup>1</sup> Evidence of the impact of these policy efforts includes the Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, which emerged out of conversations from PWH's research on emerging technologies.<sup>2</sup>

At this conference, discussions focused on AI racing dynamics, implications for military security and effectiveness, governance strategies, and the specter of AGI. This report provides a synthesis of the conference, distilling the main themes, key takeaways, and actionable policy recommendations. The goal was to illuminate how AI might reshape global politics and to chart pragmatic pathways for governance that can harness benefits while mitigating risks. Throughout two keynote conversations, a policy roundtable, and four panels, participants assessed key challenges, risks, and opportunities for international cooperation.

During a policy roundtable, experts focused on the question **"What is next for research on artificial**

1 Jim Mitre, Michael C. Horowitz, et al., *The Artificial General Intelligence Race and International Security* (Perry World House and RAND, 2025), <https://www.rand.org/pubs/perspectives/PEA4155-1.html>.

2 "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy," *United States Department of State*, n.d., accessed October 17, 2025, <https://www.state.gov/political-declaration-on-responsible-military-use-of-artificial-intelligence-and-autonomy-2/>.

intelligence and international politics?” They discussed opportunities for empirical, social science work on the geopolitical consequences of AI. Key themes included the need to move from general “AI” talk to application-specific analysis, the lack of evidence-based benchmarking and evaluation frameworks, and the prevalence of vague definitions and shifting goalposts for AGI. Interesting research opportunities included automation bias in different countries, governance of data center energy consumption and critical mineral supply chains, AI diffusion across private and public sectors, and AI-induced uncertainty in international relations.

Following the roundtable, four panel discussions commenced, featuring a diversity of practitioner, academic, and policy viewpoints.

### **AI, Chips, and U.S.-China Competition**

focused on the two countries’ different development and adoption of AI capabilities. Participants criticized the “AGI race” as ill-defined with unclear timelines, metrics, and theories of victory. The United States leads in frontier model development, compute infrastructure, and capital investment, yet lags in diffusion, adoption, and public trust in AI.<sup>3</sup> While American firms seem enthralled by the pursuit of AGI, China has prioritized integrating existing models into the economy and government.<sup>4</sup> During this conversation, industrial policy, supply-chain security, and compute capacity emerged as key policy priorities.

### **Artificial Intelligence in the Military**

**Domain** emphasized that military AI is not a single technology, with applications ranging from decision-support to fully autonomous weapons. Human control, automation bias, and dehumanization in use of force decisions were all central concerns. The wars in Ukraine and Gaza illustrate lessons on bottom-up and top-down adoption of AI tools by militaries. Deterrence was also hotly debated, yet the consensus was that nuclear deterrence still dominates great-power security. AI will be an important force multiplier and a factor in military competition but is unlikely to negate nuclear second-strike or suddenly allow total victory.

### **Artificial General Intelligence: Consequences for International Security**

discussed the strategic competition surrounding AGI. AGI will likely emerge gradually and reinforce, rather than disrupt, existing power and security dynamics. The acceleration of military decision-cycles<sup>5</sup> and operational tempo, propaganda and disinformation operations, and preemptive strike logic will remain key concerns as the technologies further develop. However, despite technological uncertainties, parallels to cyber- and network-centric warfare suggest that the AGI hype may ultimately give way to politics and warfare as usual.<sup>6</sup> Experts recommended extensive scenario planning and integrating AGI into existing strategic frameworks rather than inventing entirely new theories.

---

3 Sebastian Elbaum and Adam Segal, “What If China Wins the AI Race?,” *Foreign Affairs*, June 13, 2025, <https://www.foreignaffairs.com/united-states/what-if-china-wins-ai-race>; Colin H. Kahl and Jim Mitre, “The Real AI Race,” *Foreign Affairs*, July 9, 2025, <https://www.foreignaffairs.com/united-states/china-real-artificial-intelligence-race-innovation>; Paul Scharre, “America Can Win the AI Race,” *Foreign Affairs*, April 4, 2023, <https://www.foreignaffairs.com/united-states/ai-america-can-win-race>; and Radha Iyengar Plumb and Michael C. Horowitz, “What America Gets Wrong About the AI Race,” *Foreign Affairs*, April 18, 2025, <https://www.foreignaffairs.com/united-states/what-america-gets-wrong-about-ai-race>.

4 Elbaum and Segal, “What If China Wins the AI Race?”

5 Zachary Burdette et al., *An AI Revolution in Military Affairs? How Artificial Intelligence Could Reshape Future Warfare* (2025), [https://www.rand.org/pubs/working\\_papers/WRA4004-1.html](https://www.rand.org/pubs/working_papers/WRA4004-1.html); and Benjamin Jensen and Matthew Strohmeier, *Agentic Warfare and the Future of Military Operations* (Center for Strategic and International Studies, 2025), <https://www.csis.org/analysis/rethinking-napoleonic-staff>.

6 Michael C. Horowitz and Lauren Kahn, “The Cost of the AGI Delusion,” *Foreign Affairs*, September 26, 2025, <https://www.foreignaffairs.com/united-states/cost-delusion-artificial-general-intelligence>; and Lance Menthe et al., *Understanding the Limits of Artificial Intelligence for Warfighters: Volume 1, Summary* (2024), [https://www.rand.org/pubs/research\\_reports/RR1722-1.html](https://www.rand.org/pubs/research_reports/RR1722-1.html).

### **Opportunities for International Cooperation**

focused on identifying legal and ethical frameworks by which countries and firms can govern the AI landscape. Panelists emphasized the need to specify what types of AI and what types of applications they aim to govern or regulate, rather than try to establish general, all-AI policies. Meanwhile, collaboration between private and public sectors, as well as military and civilian actors remains critical. Since AI models originate in the private sector, and the technology bridges civilian and military domains, compartmentalized dialogues or governance would be counterproductive.<sup>7</sup>

Experts recommended starting cooperation where mutual interests (e.g., catastrophic risk avoidance, system reliability) are strongest. This is where confidence-building measures will be most likely to succeed. Yet, the need to navigate different risk tolerances and priorities across states and firms will continue. Ultimately, international cooperation must accommodate varying standards for acceptable risk and use cases.

---

<sup>7</sup> Jane Vaynman and Tristan A. Volpe, "Competition and Collusion: How the AI Arms Race Can Motivate Governance," in *The Artificial General Intelligence Race and International Security*, ed. Jim Mitre et al. (Perry World House and RAND, 2025), <https://www.rand.org/pubs/perspectives/PEA4155-1.html>.

# Key Takeaways

<< Policymakers and stakeholders must navigate the fundamental dilemmas and opportunities as they shape the future of AI governance. >>

Across the many topics discussed—from supply chains to military applications of AI—several tensions, insights, and recommendations emerged. Policymakers and stakeholders must navigate these fundamental dilemmas and opportunities as they shape the future of AI governance. We highlight a few of the most salient here:

## The Geopolitics of Compute and Capacity

While frontier models and labs receive disproportionate media and political attention, leadership in AI is increasingly defined less by algorithms and more by control over the foundations that make advanced AI possible: compute, energy, data, robotics, and talent. Rather than the government, a tight cluster of private labs, chipmakers, utilities, and cloud ecosystems are making multi-hundred-billion-dollar bets on AI infrastructure, such as massive data centers, specialized accelerators, networking, and power agreements. These dynamics demonstrate how AI expansion is not just a tech race, but an infrastructural realignment. The strategic picture that emerges from this industrial, technological, and geopolitical competition is U.S. leadership at the

algorithmic frontier, with China pressing advantages in diffusion, robotics, and adoption.

To secure U.S. and allied positions in the AI landscape, the following recommendations emerged:

- **Build Advantages beyond Compute:** While the United States currently has a structural edge in compute, this advantage is insufficient on its own. Chips and packaging, energy, data guidelines, and robotics create vulnerabilities.<sup>8</sup> While export controls can slow adversaries' access to frontier chips, allied capacity-building in design, advanced packaging, and maintenance will help to avoid brittle points of failure. The chip industry's narrow supply chain, with chokepoints in lithography, critical minerals, and advanced packaging, also produce significant risks. Meanwhile, energy is another security dependency. Data-center growth will yield massive energy requirements, creating power, water, and location bottlenecks.<sup>9</sup> If AI infrastructure becomes critical to public services and defense, energy availability and grid resilience will matter as much, or more, than model sophistication. "Digital embassies" and geo-distributed backups can serve as models for national

8 Jonathan D. Caverley, "So What? Reassessing the Military Implications of Chinese Control of Taiwan," *Texas National Security Review*, June 20, 2025, <https://tnsr.org/2025/06/so-what-reassessing-the-military-implications-of-chinese-control-of-taiwan/>.

9 Brian Deese and Lisa Hansmann, "The Coming Electricity Crisis," *Foreign Affairs*, September 9, 2025, <https://www.foreignaffairs.com/united-states/coming-electricity-crisis-data-brian-deese>.

resilience. On the privacy side, U.S. copyright regimes constrain some data pipelines relative to China's unconstrained data access. This limits content for model training but can strengthen legitimacy and international trust if well leveraged by protecting intellectual property.

- **Prioritize AI Diffusion:** Beyond model development, panelists converged on the importance of AI diffusion. The winner of this geopolitical competition will likely be the country that best diffuses AI and leverages it for economic, governance, and security gains. On this metric, China is ahead, with higher public trust in AI, more pervasive digital services, and top-down incentives that push adoption across ministries and industry.<sup>10</sup> By contrast, the United States shows systemic underinvestment in diffusion—especially in government operations and critical services—despite unmatched frontier research and capital expenditure.<sup>11</sup> Diffusion is also critical to ensure that economic gains from the AI transformation are more evenly distributed. The uneven distribution of economic benefits could result in domestic disruptions that undermine geopolitical gains.<sup>12</sup> To achieve this, the United States needs an action plan to develop public trust in AI and facilitate diffusion in a way that balances efficiency gains with human risks.
- **The Short-Term “AGI Race” is Not the Top Priority:** Prevalent narratives about the “AGI Race” and analogies to the Manhattan Project distract from the need to establish

secure energy and resilient infrastructure, and diffuse current AI capabilities. At present, the “AGI race” is ill-defined—timelines and metrics shift, and “victory” is unclear—yet the larger competitive narrative influences industrial policy and market expectations. Several speakers noted corporate and capital incentives intensifying this rhetoric, as happened with cyber a decade ago. A more sober perspective would focus on measurable capacities: trained floating-point operations per second (FLOPs) per month, secured energy per megawatt of compute, and verified access to critical minerals.

Overall, compute advantages are necessary but not sufficient. To build a credible and resilient AI strategy, the United States and allies should couple existing advantages with secure energy, resilient networks, and redundant cloud topologies, and then prioritize diffusion across the real economy. To maintain its edge, the United States must develop a broader strategy that encompasses the entire industry ecosystem from the ground up to ensure a sustainable model.

## AI and AGI in Defense and Deterrence

How will AI and AGI shape militaries, conflict, and the nature of deterrence? While headlines fixate on “killer robots” and hypothetical superintelligence, the military reality is more prosaic and urgent. Rather than a singular technology, “military AI” is an ecosystem of applications that each present distinct risks, timelines, and evaluation

---

10 Elsa B. Kania, “Artificial Intelligence and Chinese Power,” *Foreign Affairs*, December 5, 2017, <https://www.foreignaffairs.com/articles/china/2017-12-05/artificial-intelligence-and-chinese-power>; Elbaum and Segal, “What If China Wins the AI Race?”; James Kynge and Sun Yu, “China and Big Tech: Xi’s Blueprint for a Digital Dictatorship,” *The Big Read, Financial Times*, September 7, 2021, <https://www.ft.com/content/9ef38be2-9b4d-49a4-a812-97ad6d70ea6f>; “In the Struggle for AI Supremacy, China Will Prevail,” *The Economist*, September 27, 2018, <https://www.economist.com/books-and-arts/2018/09/27/in-the-struggle-for-ai-supremacy-china-will-prevail>.

11 Colin H. Kahl, “America Is Winning the Race for Global AI Primacy—for Now,” *Foreign Affairs*, January 17, 2025, <https://www.foreignaffairs.com/united-states/america-winning-race-global-ai-primacy-now>.

12 Yasmin Green and Gillian Tett, “AI and the Trust Revolution,” *Foreign Affairs*, July 7, 2025, <https://www.foreignaffairs.com/united-states/artificial-intelligence-and-trust-revolution-technology-transforming-human-connections>.

methods.<sup>13</sup> From intelligence, surveillance, and reconnaissance (ISR) fusion to logistics analysis, decision-support, autonomous vehicles, and information operations, these technologies can change how forces sense, decide, move, and sustain. Four framing points should be emphasized:

***First, speed and opacity are near-term concerns.***

Machine-speed can compress decision-making cycles and complicate signaling, creating risks for escalation and deterrence. Adversaries may misread automation as either overwhelming resolve (escalatory) or low resolve (risk-tolerant, because fewer troops are at risk).<sup>14</sup> Fast loops make deterrence, credibility, and resolve harder to communicate, particularly when effects propagate through opaque models and autonomous platforms.

***Second, “human judgment” is necessary but under-specified.*** A human who is overloaded by model outputs and prone to automation biases becomes a rubber stamp. The meaningful question is interpretability and role clarity. These assignments must be paired with domain-specific confidence thresholds. For instance, what are appropriate performance and risk criteria for ISR triage versus lethal strikes?

***Third, AI magnifies infrastructure risk more than it creates nuclear-style first-strike incentives.*** Participants generally agreed that preemptive strikes against data centers developing AGI are unlikely. Instead, cyber and gray-zone degeneration of AI infrastructure is the real risk. If militaries become more dependent on AI ecosystems, then data centers and energy infrastructure may

become targets during conflict. However, regardless of AGI’s emergence, deterrence fundamentals endure. AI will not likely alter the logic of nuclear deterrence. While AI will be an important force multiplier and factor in military competition, it will neither negate second-strike nor suddenly allow total victory.<sup>15</sup> That said, AI could subtly stress deterrence stability by introducing new uncertainties, adding speed and frictions to crisis management, and expanding deception, propaganda, and coercive information operations.

***Fourth, militaries are generally conservative organizations that change slowly.*** Like past waves of technology (tanks, airpower, cyber), AI integration may be uneven and contested by service cultures and procurement pathways. As with the general economy, the decisive competition will not be who fields a novel AI model, but who can reliably evaluate, integrate, and sustain AI across doctrine, training, and logistics, all without eroding legitimacy or control of force.<sup>16</sup>

Considering these points, participants provided the following recommendations:

- **Ensure Human Responsibility:** Just as compute leadership requires more than hardware, effective human responsibility and control requires institutional depth beyond slogans. Most militaries endorse human responsibility over the use of force, which is sometimes described as having a human “in the loop.” For high-risk applications such as target selection or autonomous engagement, just as with all uses of force, confidence

---

13 Menthe et al., *Understanding the Limits of Artificial Intelligence for Warfighters*; Burdette et al., *An AI Revolution in Military Affairs?*; Schmidt and Grant, “The Dawn of Automated Warfare.”

14 Todd S. Sechser et al., “Emerging Technologies and Strategic Stability in Peacetime, Crisis, and War,” *Journal of Strategic Studies* 42, no. 6 (2019): 727–35, <https://doi.org/10.1080/01402390.2019.1626725>; Sam Bresnick, “Could AI Lead to the Escalation of Conflict? PRC Scholars Think So,” *Lawfare*, September 20, 2024, <https://www.lawfaremedia.org/article/could-ai-lead-to-the-escalation-of-conflict--prc-scholars-think-so>.

15 James D. Fearon, “One Does Not Simply Dismiss the Nuclear Revolution,” in *The Artificial General Intelligence Race and International Security*, ed. Jim Mitre et al. (Perry World House and RAND, 2025), <https://www.rand.org/pubs/perspectives/PEA4155-1.html>.

16 Jon R. Lindsay, *Information Technology and Military Power* (Cornell University Press, 2020).

thresholds should be created, verified, and subject to monitoring for model drift. For lower-stakes domains (logistics, maintenance), automated execution can be broader, but audit trails should remain mandatory. The Pentagon and national militaries have favored framing the obligation of responsible militaries as appropriate human judgment rather than meaningful human control because control is not a standard that exists for non-AI enabled weapon systems.

To ensure legitimacy, human-machine teaming must be treated as a design discipline by developing interfaces that foreground uncertainty, requiring “explainability drills” in training, and embedding accountability in command structures. Critically, nuclear systems must remain AI-free zones—hard-coded by doctrine and procurement.

- **Institutionalize Joint Testing and Red-Teaming:** AI military evaluation and assurance will increasingly resemble the geopolitics of compute: a distributed, multi-actor challenge defined by coordination and shared baselines. No single nation or service can anticipate all failure modes or deception tactics in military AI. Accordingly, alliances should institutionalize continuous joint testing, evaluation, and red-teaming—not as ad hoc exchanges, but as standing multilateral programs comparable to joint exercises or nuclear verification regimes.

Regional and transnational groupings (NATO, AUKUS, the Quad) could establish combined testbeds and shared datasets for evaluating AI systems under contested conditions, including spoofing, adversarial perturbation, and data corruption. Such collaboration would mirror chip supply-chain resilience initiatives: pooling

insights, closing validation gaps, and reducing duplicated effort.

Institutionalized red-teaming would also help standardize incident reporting and build shared playbooks for AI deception, analogous to cybersecurity information-sharing frameworks. Over time, this could mature into a “safety alliance”: a trusted ecosystem of testing labs, simulation environments, and AI-range facilities that ensure systems meet verifiable performance and safety criteria before deployment. Shared testing standards would harden collective resilience and set de facto norms for responsible military AI use.

- **Invest in Resilient Command and Control:** The command-and-control (C2) backbone will define operational AI resilience. Military dependence on AI services—data centers, communication networks, and cloud-based inference—creates new strategic choke-points. A degraded model or compromised link could disrupt entire command chains, making C2 resilience as critical as nuclear hardening once was.

Militaries should therefore treat digital infrastructure as a core component of force readiness. The services would need to invest in redundant geo-distributed data centers, establish “digital embassies” in allied territories, and conduct regular exercises simulating degraded AI or contested spectrum environments.<sup>17</sup> Cloud dependencies should have air-gapped, offline fallbacks capable of manual override and continuity of command even in the absence of AI inputs.

Resilient C2 also extends to energy and network resilience. Just as compute geopolitics ties AI leadership to secure energy per

---

<sup>17</sup> Karl P. Mueller, “Averting Attacks Against AGI Development: Three Strategic Approaches,” in *The Artificial General Intelligence Race and International Security*, ed. Jim Mitre et al. (Perry World House and RAND, 2025), <https://www.rand.org/pubs/perspectives/PEA4155-1.html>.

megawatt of compute, military AI viability depends on secure bandwidth per operation and grid stability during conflict. Embedding redundancy, diversity, and reversibility in command networks—through mesh communications, hardened energy grids, and analog backups—translates the lessons of compute resilience directly into defense posture.

AI will sharpen the edges of deterrence, not by overturning nuclear logic, but by accelerating tempo, thickening the fog of war, and raising the premium on evaluation and control. These recommendations collectively aim to ensure that AI strengthens rather than destabilizes deterrence. Codifying human control for nuclear weapons and human judgment for all uses of force, institutionalizing collaboration, and investing in resilience are critical steps. The strategic winners of the military AI era will not be those who simply field the fastest algorithms, but will be those who can evaluate, integrate, and sustain AI systems under pressure, with transparent human oversight, verified assurance, and enduring control of force.

## Challenges and Opportunities for International Cooperation

Amid escalating competition over AI capabilities, international cooperation remains both essential and fraught. The global AI landscape is defined by divergent regulatory philosophies, asymmetric risks, and an unprecedented degree of private-sector control. The United States relies on a market-driven, export-control-heavy approach. Conversely, the European Union foregrounds precaution and regulation, while China operates in a top-down and centrally controlled fashion, integrating AI into state priorities for rapid diffusion.

These structural differences make broad, universal treaties nearly impossible.

However, History shows that even adversaries can cooperate on shared, verifiable interests—whether in arms control, nonproliferation, or crisis communications. For AI, that cooperation must be narrow, concrete, and empirically testable. A confidence-building measure approach is most likely to be successful in the military arena.<sup>18</sup> Rather than a universal treaty for “all AI,” policymakers should target specific systems and applications with measurable standards and operational practices. Since commercial firms lead the development and deployment of frontier capabilities and most military applications mirror civilian functions, AI cooperation must bridge private-public and civil-military divides from the outset.<sup>19</sup> In sum, the future of AI governance will depend on an *infra-structure* of verification, evaluation, and the collaborations that build multi-stakeholder trust.

After discussing salient barriers to cooperation, participants recommended the following priorities:

- **Establish Clear Standards and Commitments:** Just as compute leadership depends on reliable and transparent supply chains, durable AI cooperation depends on standards that are concrete, interoperable, and verifiable. Broad treaties about “safe AI” tend to collapse under definitional ambiguity. Instead, stability will emerge from shared commitments that can be measured, tested, and monitored. Building verifiable standards means agreeing not just on values, but on definitions and levels of autonomy, risk tiers, and assurance claims that can travel across institutions and jurisdictions. Shared benchmarks could be codified through established institutions (e.g.,

18 Michael C. Horowitz and Lauren Kahn, “Leading in Artificial Intelligence through Confidence Building Measures,” *The Washington Quarterly*, 44(4), 2021: pp. 91-106, <https://doi.org/10.1080/0163660X.2021.2018794>.

19 Vaynman and Volpe, “Competition and Collusion: How the AI Arms Race Can Motivate Governance”; Jane Vaynman and Tristan A. Volpe, “Dual Use Deception: How Technology Shapes Cooperation in International Relations,” *International Organization*, 77(3), 2023: pp. 599-632.

International Organization for Standardization (ISO), International Electrotechnical Commission (IEC), International Telecommunication Union (ITU),<sup>20</sup> Organisation for Economic Co-operation and Development (OECD), and the Global Partnership on Artificial Intelligence (GPAI)<sup>21</sup> and defense coalitions like NATO and the Quad. The goal is to create standards that travel—frameworks that enable mutual understanding even amid strategic rivalry.

In practice, this requires narrow, clearly defined commitments. For instance, prohibiting AI involvement in nuclear command and control, establishing peacetime non-targeting norms for AI infrastructure, and constraining mass model poisoning or large-scale data manipulation. These pledges would parallel long-standing arms-control norms. Moreover, verification should be technical as well as diplomatic, with audit trails, independent evaluations, post-incident coordination through hotlines, and shared playbooks for model failure or disinformation events.

To ensure credibility, these commitments must be embedded in systems and laws—from design to deployment—so that compliance and auditability is a built-in feature, not an afterthought. Just as export controls and inspection regimes hardened norms in the nuclear and semiconductor domains, interoperable standards for AI assurance could form the backbone of a verifiable and enforceable governance architecture.

- **Collaborate Across Borders:** AI assurance requires shared testing infrastructure and collective evaluation capacity. States and companies alike benefit from collaboration on verification, red-teaming, and post-incident

analysis, even when full transparency is impossible.

A pragmatic starting point is common testing methodologies and incident taxonomies, allowing each nation to learn from others' near misses without revealing proprietary details. Mutual recognition pilots—where audit artifacts such as datasets, test harnesses, and confidence metrics are standardized—could enable a system vetted in one jurisdiction to be understood elsewhere.<sup>22</sup>

Beyond testing, cooperation must extend to capacity building. Shared AI centers offering compute credits, open datasets, and technical training—especially in under-resourced regions and the Global South—can democratize access while reinforcing trust and responsible norms. Such inclusion mirrors the global push to diversify chip supply chains and energy sources: it builds legitimacy by making the AI order more equitable.

Track-2 exchanges among universities, labs, and defense technologists can rebuild epistemic trust, especially around verification practices and crisis management. Similarly, energy and infrastructure cooperation—on grid resilience, cooling, and water use—would give credibility to commitments not to target AI infrastructure in peacetime. Just as nations compete on reliable compute and deployment quality rather than coercion, they can compete through rights-respecting and secure AI deployments that signal stability rather than provocation.

- **Bridge the Gaps:** The general purpose nature of AI means that military, commercial, and civil technologies are now interdependent. Most of the same algorithms that drive

---

20 WSC – World Standards Cooperation, n.d., accessed October 18, 2025, <https://www.worldstandardscooperation.org/>.

21 "The OECD Artificial Intelligence Policy Observatory," accessed October 18, 2025, <https://oecd.ai/en/>.

22 Vaynman and Volpe, "Competition and Collusion: How the AI Arms Race Can Motivate Governance."

military logistics and ISR triage also underpin civilian applications in commerce and governance. Effective cooperation therefore depends on bridging structural divides between defense, industry, and civil society, creating unified lanes of problem-solving rather than parallel, siloed debates. The challenge is not simply to “deconflict” them, but to integrate so that institutions can share and manage risks jointly.

This dynamic means developing joint forums where militaries, regulators, and technology firms tackle shared challenges like model evaluation, misuse mitigation, and data provenance. Governments can use procurement as a strategic lever—requiring explainability, auditability, and post-deployment monitoring in public-sector tenders. These requirements can gradually pull industry toward interoperable assurance norms that allies can mirror, just as defense industrial bases once harmonized quality and safety standards across borders.

Where politically feasible, open-source stewardship offers another venue for cooperation. Joint work on misuse-mitigation tools—such

as watermarking, provenance systems, or safety filters—can reduce global externalities without freezing innovation. Transparency mechanisms should be graduated and calibrated. For instance, capability summaries, incident reports, and test coverage disclosures can help build trust while safeguarding intellectual property and operational security.

Ultimately, bridging these gaps is about institutional diffusion—translating safety and accountability norms from elite labs into procurement contracts, battlefield software, and public infrastructure. Resilient AI governance depends on coordination across military, civil, and commercial domains.

Overall, the most durable form of AI cooperation will be narrow, verifiable, and application-specific, not universal or abstract. Establishing credible standards, collaborating across borders, and bridging civil-military divides can reduce systemic risks while reinforcing responsible competition. The strategic winners of the AI era will be those who can embed trust, verification, and accountability into the architecture of their AI ecosystems—transforming competition into stability through institutional design.

# Conclusion

<< Artificial intelligence has become an organizing principle of power that cuts across industry, energy, finance, and force. >>

Artificial intelligence has become an organizing principle of power that cuts across industry, energy, finance, and force. *The Future of Artificial Intelligence Governance and International Politics* conference delivered a clear message: leadership in AI will be determined less by a single breakthrough than by the ability to secure infrastructure, diffuse capability responsibly, preserve human judgment in high-risk settings, and cooperate narrowly where interests align. Unlike the Atomic Age, when a few governments monopolized a single technology, today's AI revolution is diffuse—driven by private actors, layered supply chains, and cross-cutting social systems. This diffusion makes governance both more challenging and more essential. The discussions synthesized here suggest that building a stable and beneficial AI order will require disciplined integration, verifiable cooperation, and evidence-based policymaking grounded in empirical research.

***First, the center of gravity is infrastructure.***

Compute, energy, packaging, data, and robotics now shape national leverage as much as algorithms do. The United States holds a real—but brittle—edge in frontier models and segments of the chip stack, while China presses advantages in diffusion and embodied automation. Security therefore hinges on pairing export controls with allied capacity-building, hardening grids and clouds, and ensuring that the benefits of AI reach beyond a handful of firms and metros. Without diffusion into public services and critical sectors, frontier gains risk producing concentrated rents and political

backlash rather than durable advantage.

***Second, nuclear deterrence endures.*** Militaries will adopt AI unevenly, and nuclear second-strike logic still anchors great-power risk calculus. The near-term dangers come from speed, opacity, and infrastructure dependence: compressed decision cycles, shallow evaluations, and vulnerable data or energy networks. The appropriate response is not technological maximalism but disciplined integration—codified human control, mission-specific confidence thresholds, resilient C2, and joint red-teaming that identifies failure modes before adversaries do.

***Third, cooperation is possible when it is narrow, verifiable, and tied to shared interests.***

The fastest path runs through application-specific confidence-building measures: prohibiting AI automation of nuclear launch, establishing peacetime non-targeting norms for AI infrastructure, standardizing incident taxonomies and evaluation methods, and embedding auditability throughout the system lifecycle. Because AI is both dual-use and private-sector-led, durable cooperation must bridge public-private and civil-military lanes to ensure shared accountability.

***Fourth, policy must be informed by evidence, not assumptions.***

Participants underscored the need to move from rhetorical debates toward testable, data-driven analysis. A rigorous research agenda is essential to support sound governance and risk management.

Taken together, these insights translate into a

practical program:

- **Secure the Stack:** Diversify and harden compute, energy, and networks with allied capacity and clear resilience standards.
- **Win on Adoption:** Use procurement and workforce pipelines to diffuse trustworthy AI across the real economy and the public sector.
- **Bound Military Use:** Institutionalize human control, confidence thresholds, and red teaming; cordon AI from nuclear launch authority.
- **Cooperate Where It Counts:** Pursue narrow, verification-ready arrangements and shared evaluation tooling; fund regional capacity so standards travel.
- **Invest in Evaluation and Research:** Build open test suites, diffusion scorecards, and escalation playbooks that policymakers and commanders can use.

The window for shaping AI's trajectory is open now. If states, firms, and civil society move with purpose, AI can reinforce stability and prosperity rather than erode them. If they do not, path dependencies in compute, energy, and platform power will harden, brittle deployments will invite incidents, and zero-sum narratives will crowd out the narrow cooperation that reduces catastrophic risk.

The spirit of this conference was cautiously optimistic. We know enough to act, even amid uncertainty. By coupling ambition with governance, speed with safeguards, competition with cooperation, and prediction with human judgment, the international community can build an AI future defined not by fear and rivalry, but by stability, prosperity, and collective stewardship.

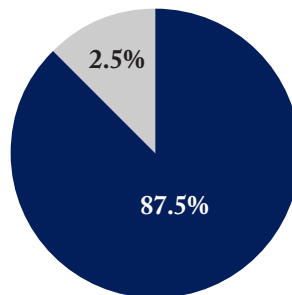
# Appendix: Pre-Conference Survey Results

Perry World House asked participants to fill out a short survey on key issues related to the conference themes. The following figures are based on participants' responses. Not all participants answered all questions, and these charts should not be interpreted to represent any individual panelist's view.

**Q:**

*Which country's companies are ahead in designing and building artificial intelligence models at the frontier?*

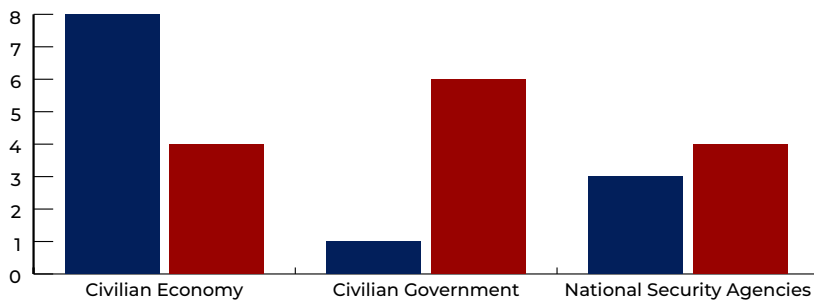
- United States
- Other



**Q:**

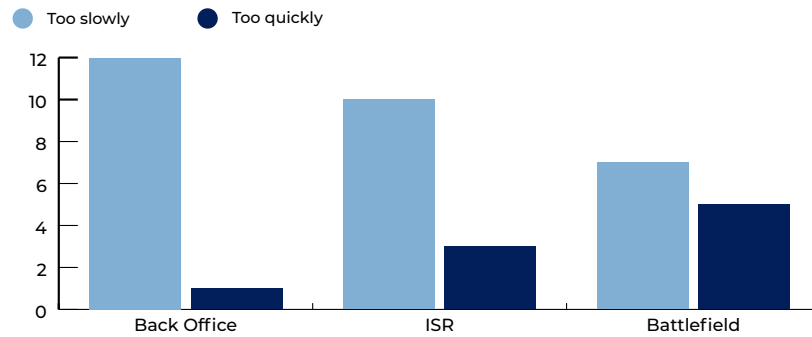
*Which country is ahead in adopting artificial intelligence for use cases in the civilian economy, civilian government, and national security agencies?*

- United States
- China



**Q:**

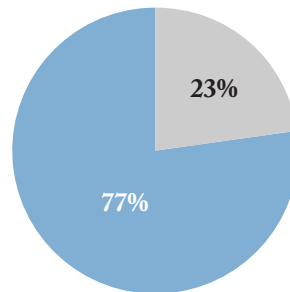
Do you think the U.S. military is moving too quickly or too slowly when it comes to adopting artificial intelligence?



**Q:**

How likely is a country to field a new lethal autonomous weapon system (LAWS) acknowledged as such over the next 5 years?

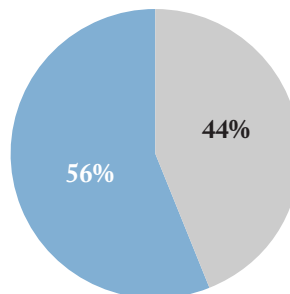
- Probable
- Improbable



**Q:**

Five years from now (2030), will major militaries use AI decision support tools for a majority of their operational planning?

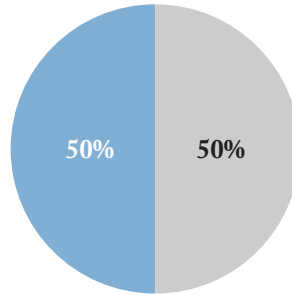
- Yes
- No



Q:

Should international dialogues on safety issues associated with non-military AI be combined with military AI dialogues?

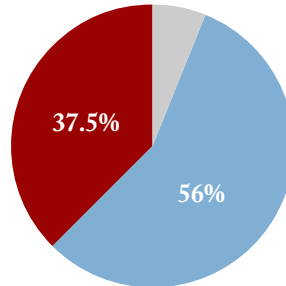
- Yes
- No



Q:

How long will it take for a company to successfully develop something widely recognized as artificial general intelligence?

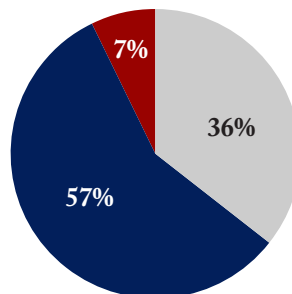
- Never
- 4-5 years
- 2-3 years



Q:

Which country will be the first to successfully develop AGI?

- United States
- China
- Other





UNIVERSITY OF PENNSYLVANIA | PERRY WORLD HOUSE  
3803 LOCUST WALK, PHILADELPHIA, PA 19104  
215.573.5730

@PERRYWORLDHOUSE  
FACEBOOK.COM/PERRYWORLDHOUSE

[GLOBAL.UPENN.EDU/PERRYWORLDHOUSE](http://GLOBAL.UPENN.EDU/PERRYWORLDHOUSE)