Expert Insights
PERSPECTIVE ON A TIMELY POLICY ISSUE

JIM MITRE, JOEL B. PREDD

Artificial General Intelligence's Five Hard National Security Problems

February 2025

For more information on this publication, visit **www.rand.org/t/PEA3691-4**.

About RAND

RAND is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest. To learn more about RAND, visit www.rand.org.

Research Integrity

Our mission to help improve policy and decisionmaking through research and analysis is enabled through our core values of quality and objectivity and our unwavering commitment to the highest level of integrity and ethical behavior. To help ensure our research and analysis are rigorous, objective, and nonpartisan, we subject our research publications to a robust and exacting quality-assurance process; avoid both the appearance and reality of financial and other conflicts of interest through staff training, project screening, and a policy of mandatory disclosure; and pursue transparency in our research engagements through our commitment to the open publication of our research findings and recommendations, disclosure of the source of funding of published research, and policies to ensure intellectual independence. For more information, visit www.rand.org/about/research-integrity.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Published by the RAND Corporation, Santa Monica, Calif.

Limited Print and Electronic Distribution Rights

About This Paper

The potential emergence of artificial general intelligence (AGI) is plausible and should be taken seriously by the U.S. national security community. Yet the pace and potential progress of AGI's emergence—as well as the composition of a post-AGI future—is shrouded in a cloud of uncertainty. This poses a challenge for strategists and policymakers trying to discern what potential threats and opportunities might emerge on the path to AGI and once AGI is achieved.

This paper puts forth five hard problems that AGI's emergence presents for U.S. national security: (1) wonder weapons; (2) systemic shifts in power; (3) nonexperts empowered to develop weapons of mass destruction; (4) artificial entities with agency; and (5) instability. In much of the discourse on AGI, policymakers and analysts argue past one another with differing opinions on which issues deserve immediate focus and resources. Yet we have observed that proposals to advance progress on one problem can undermine progress on—if not outright ignore—another. These five hard national security problems are offered to structure the debate by providing a common language to communicate about risks and opportunities for AGI and a rubric to evaluate alternative strategies.

Technology and Security Policy Center

RAND Global and Emerging Risks is a division of RAND that delivers rigorous and objective public policy research on the most consequential challenges to civilization and global security. This work was undertaken by the division's Technology and Security Policy Center, which explores how high-consequence, dual-use technologies change the global competition and threat environment, then develops policy and technology options to advance the security of the United States, its allies and partners, and the world. For more information, contact tasp@rand.org.

Funding

The Geopolitics of AGI Initiative is independently initiated and conducted within the Technology and Security Policy Center using income from operations and gifts from philanthropic supporters, which have been made or recommended by DALHAP Investments Ltd., Effektiv Spenden, Ergo Impact, Founders Pledge, Charlottes och Fredriks Stiftelse, Good Ventures, Jaan Tallinn, Longview, Open Philanthropy, and Waking Up Foundation. A complete list of donors and funders is available at www.rand.org/TASP. RAND donors and grantors have no influence over research findings or recommendations.

Acknowledgments

This piece is a product of learning from numerous colleagues at RAND, policymakers across the U.S. government, technologists and leaders in the private sector developing and applying artificial intelligence technologies, and partners across a host of research organizations who indulged us in numerous exchanges as part of the Geopolitics of AGI Initiative. We are especially grateful for the thorough and thoughtful reviews of the full manuscript provided by Jeff Alstott, Matt Chessen, Alison K. Hottes, Michael C. Horowitz, Nidhi Kalra, Jason Matheny, Robin Rauzi, and N. Peter Whitehead at RAND; John Bansemer at the Center for Security and Emerging Technology; and Miles Brundage, former OpenAI senior adviser.

Contents

Artificial General Intelligence's Five Hard National Security Problems

In 1938, German physicists split the atom, and physicists around the world had an *a-ha!* moment. The scientific breakthrough showed a clear technical pathway to creating the most disruptive military capability in history. In a large mass of uranium, nuclear fission of one atom could cause a nuclear chain reaction that would lead to "extremely powerful bombs," as Albert Einstein explained in a letter to U.S. President Franklin D. Roosevelt that launched the United States into a race for the atomic bomb.[1]

Recent breakthroughs in frontier generative artificial intelligence (AI) models have led many to assert that AI will have an equivalent impact on national security—that is, that it will be so powerful that the first entity to achieve it would have a significant, and perhaps irrevocable, military advantage.[2] In modern-day equivalents of the Einstein letter, calls are beginning for the U.S. government to engage in a large national effort to ensure that the United States obtains the decisive AI-enabled wonder weapon before China does.[3]

The problem is that frontier generative AI models have not yet had that atom-splitting moment of clarity showing a clear technical pathway from scientific advance to wonder weapon. When the Manhattan Project was launched, the U.S. government knew precisely what the capability that it was building would do. The capabilities of the next generation of AI models are unclear. The impetus for a large, national-level government program to pursue a wonder weapon does not yet exist. But that does not mean that the U.S. government should sit idly by. U.S. national security strategy should take seriously the uncertain but technically credible potential that world-leading AI labs are on the cusp of developing an artificial general intelligence (AGI)[4]—and the relative certainty that they will continue making progress until that unknown and potentially unknowable threshold is crossed.

AGI, which would produce human-level—or even superhuman-level—intelligence across a wide variety of cognitive tasks, is plausible; it is reasonable to assume that it could be realized. It therefore presents unique opportunities and potential threats to U.S. national security strategy. We have distilled these into five hard problems. AGI could cause any combination of these five problems:

---

[1] Einstein, letter to Franklin D. Roosevelt.

[2] Center for AI Safety, "Statement on AI Risk."

[3] Aschenbrenner, "Situational Awareness"; Bajraktari, "The Artificial General Intelligence Presidency Is Coming"; U.S.-China Economic and Security Review Commission, *2024 Report to Congress of the U.S-China Economic and Security Review Commission*.

[4] Goertzel, "Who Coined the Term 'AGI'?"

- enable a significant first-mover advantage via the sudden emergence of decisive wonder weapons
- cause a systemic shift that alters the balance of global power
- empower nonexperts to develop weapons of mass destruction
- cause the emergence of artificial entities with their own agency to threaten global security
- increase strategic instability.

Endemic Uncertainty

Leading AI labs in the United States and globally are in hot pursuit of AGI. Relying principally on empirical "scaling laws"—that model performance scales with compute—AI labs are investing ever-increasing sums into the compute necessary to train their models. The training run for each model in the current generation of frontier AI models—including ChatGPT-4, Gemini, and Claude 3.5—relied on hundreds of millions of dollars of compute.[5] Algorithmic improvements, such as OpenAI's o1 reasoning function, and advances in related technical fields, such as symbolic reasoning, present complementary pathways to a possible AGI breakthrough.[6] Despite not realizing substantial commercial success yet, the leading AI labs are building their war chests and aggressively pursuing models that are on pace to cost $1 billion or more by 2027.[7]

It is unclear whether performance will continue to scale with compute.[8] If it does, it is unclear what the threshold is for AGI, if such a technical breakthrough is even possible through this method. The pace and potential progress of AGI's emergence—as well as the composition of a post-AGI future—is shrouded in a cloud of uncertainty. Experts rabidly debate whether the technology is on the verge or decades away.[9] Will there be a discrete event or a gradual transition to an AGI state? Will AGI result in a future of abundance for all, or a future marked by scarcity, with power in the hands of a few? Adding to the uncertainty is that the technologists developing frontier AI models themselves might not know that a critical threshold in AGI capability has been crossed until it is. Some of these uncertainties could be resolved with further research and experience, but some might be practically unresolvable in time to inform strategy and policy development.

On the one hand, AI doomers are largely convinced that AGI's emergence is existential, leading some to call for a halt to all progress before AGI destroys humanity and others to call for

[5] Seetharaman, "The Next Great Leap in AI Is Behind Schedule and Crazy Expensive."

[6] Himabindu et al., "Neuro-Symbolic AI."

[7] Cottier et al., "How Much Does It Cost to Train Frontier AI Models?"

[8] Sevilla et al., "Can AI Scaling Continue Through 2030?"

[9] Wong, "The AI Boom Has an Expiration Date."

the United States to accelerate development before China is able to destroy the global order.[10] On the other hand, skeptics abound, asserting that AGI is not remotely feasible under the current technological paradigm because, for example, frontier AI models do not understand the physical world.[11]

At a technical level, a $10 billion training run could produce a model with no marginal increase in performance over that of existing frontier AI models. Alternatively, the model could achieve the ability for recursive self-improvement, enhancing its own capabilities without additional human input and leading to sort of a superhuman intelligence explosion. The uncertainty on the path toward AGI, and in a post-AGI world, could lead to multiple strategic windows of opportunity in the next decade, confronting policymakers with not one but multiple possible inflection points to navigate. Given this array of plausible outcomes, any security strategy that is overoptimized for any single paradigm is a high-risk proposition. The central issue is not predicting how the future will unfold but determining what steps the U.S. government should take amid technological and geopolitical uncertainties.

The Five Problems

At RAND, we lead an initiative that aims to build the intellectual foundations for the United States to address the national security implications of the potential emergence of AGI. The initiative has formed a vibrant intellectual community among policymakers, the private sector, and research organizations while possessing some self-contained energy within RAND. The list of five hard problems for U.S. national security is a product of the initiative, which includes a wide variety of exploratory research, games, workshops, and convenings.

In much of the discourse on AGI, policymakers and analysts argue past one another with divergent views on which of these problems warrant resources and attention now and at what opportunity costs. These problems are overlapping in areas and might not represent the full range of problems that policymakers might have to consider in an era in which AGI's emergence is plausible. Yet we have observed that proposals to advance progress on one problem can undermine progress on—if not outright ignore—another. As a result, they serve as a rubric to evaluate alternative strategies. These five hard national security problems are offered to advance the debate on AI strategy by providing a common language to communicate about risks and opportunities for AGI in national security.

First, AGI might enable a significant first-mover advantage via the sudden emergence of a decisive wonder weapon. Consider a future in which AGI invents a technical breakthrough

[10] Yudkowsky, "Pausing AI Developments Isn't Enough"; Tong and Martina, "US Government Commission Pushes Manhattan Project–Style AI Initiative."

[11] Murphy and Criddle, "Meta AI Chief Says Large Language Models Will Not Reach Human Intelligence."

that produces a clear path to the development of a wonder weapon or system that confers tremendous military advantage by, for example,

- identifying and exploiting vulnerabilities in enemy cyberdefenses and creating what some might call a splendid first cyber strike that completely disables a retaliatory cyberstrike
- simulating complex scenarios and predicting outcomes with high accuracy, drastically improving planning and execution in military operations
- developing highly advanced autonomous weapon systems that provide military dominance.

AGI could also erode a military advantage by, for example, creating a sort of fog-of-war machine that renders untrustworthy information about the battlefield.[12] Such a first-mover advantage could disrupt the military balance of power in key theaters, create a host of proliferation risks, and accelerate technological race dynamics.

A country gaining significant first-mover advantage from AGI reflects the most-ambitious assumptions: a sudden emergence of AGI that provides a dramatic increase in cognitive performance, extreme implications for national security, and rapid institutional adoption. These assumptions, however, posit high-consequence events of unknown probability. Prudent planning therefore calls for the United States not to assume that a wonder weapon is imminent but to consider the conditions under which such a disruptive weapon could emerge and for the United States to position itself to seize a first-mover advantage if this scenario comes into focus.

Second, AGI might cause a systemic shift in the instruments of national power or societal foundations of national competitiveness that alters the balance of global power. History suggests that technological breakthroughs rarely yield wonder weapons that provide an immediate, decisive impact on military balances or national security.[13] Except for rare examples, such as nuclear weapons, cultural and procedural factors drive an institution's technological adoption capacity and are more consequential than being the first to achieve a scientific or technological breakthrough.[14] As the U.S., allied, and rival militaries establish access to AGI and adopt it, it could upend military balances by uplifting a variety of capabilities that affect key building blocks of military competition, such as hiders versus finders, precision versus mass, or centralized versus decentralized command and control.

Moreover, AGI could undermine the societal foundations of national competitiveness, potentially jeopardizing democracy.[15] For example, AGI could be used to manipulate public opinion through advanced propaganda techniques, threatening democratic decisionmaking. Also, the complexity and unpredictability of AGI systems could outpace regulatory frameworks, making it difficult to govern their use effectively, undermining the effectiveness of institutions.

---

[12] Geist, *Deterrence Under Uncertainty*.

[13] Ding, "The Innovation Fallacy."

[14] Horowitz, *The Diffusion of Military Power*.

[15] Mazarr, *The Societal Foundations of National Competitiveness*.

AGI could also cause a systemic shift in the economy by providing a massive boost in productivity or science by creating a wellspring of new discoveries. For example, automated workers could rapidly displace labor across industries, causing national gross domestic product to skyrocket but wages to collapse as fewer jobs become available.[16] Labor disruption of such scale and speed could spark social unrest that threatens the viability of the nation-state. And, as Anthropic chief executive officer Dario Amodei recently postulated, powerful AI could cure cancer and infectious disease.[17] States that are better postured to capitalize on—and manage—such economic and scientific shifts could have greatly expanded influence in the future. Independently of whether AGI on its own creates wonder weapons, AGI's impact on other instruments of national power could be highly disruptive to global power dynamics for good or for ill.

Third, AGI might empower nonexperts to develop weapons of mass destruction. Foundation models are hailed as a boon for labor productivity in large part because they can speed novices up the learning curve and make nonexperts perform at a higher level.[18] Yet, this accelerated knowledge gain can apply to malicious tasks as well as useful ones. Foundation models' ability to clearly elucidate some of the specific steps that nonexperts can take to develop dangerous weapons, such as a highly lethal and transmissible pathogen or virulent cyber malware, widens the pool of people capable of creating such threats. To date, most foundation models have not demonstrated the ability to provide information not already available on the public internet,[19] but foundation models have the capacity to serve as malicious mentors that can distill complex methods into accessible instructions for nonexperts and assist users in circumventing prohibitions on developing weapons. This threat might manifest before the development of AGI; as OpenAI's own safety evaluation of its o1 model shows, the risk is increasing.[20]

Knowing how to build a weapon of mass destruction is, of course, not the same as actually building it. There are practical challenges in transferring knowledge into discrete forms of weapon development, such as mastering technologically advanced manufacturing processes. These can substantially reduce the actual risk of successful weapon development in certain cases, such as nuclear weapons, possibly to zero. But technological developments in related fields are lowering these execution barriers. For example, it is getting easier and cheaper to access, edit, and synthesize viral genomes.[21] AI agents are increasingly interacting with the physical world;

---

[16] Korinek and Juelfs, "Preparing for the (Non-Existent?) Future of Work."

[17] Amodei, "Machines of Loving Grace."

[18] Dell'Acqua et al., "Navigating the Jagged Technological Frontier."

[19] Mouton, Lucas, and Guest, *The Operational Risks of AI in Large-Scale Biological Attacks*.

[20] OpenAI, "OpenAI o1 System Card," December 5, 2024.

[21] Crawford et al., *Securing Commercial Nuclear Acid Synthesis*.

they can convert bits to molecules and physically synthesize a chemical agent in a cloud lab.[22] Given these developments, significantly broadening the pool of people with knowledge to attempt development of such weapons is a distinct challenge worth guarding against.

Fourth, AGI might manifest as an artificial entity with agency to threaten global security. One of the most pernicious effects of AGI's development could be the erosion of human agency as humans become increasingly reliant on the technology. As AGIs control ever more-complex and -critical systems, they might optimize critical infrastructure in ways that are beneficial to humanity but also in ways that humanity has no chance of fully understanding. This is a current concern with narrow AI used to identify military targets on the battlefield that a human operator might need to trust as accurate given a lack of time or ability to confirm.[23] As AI becomes more powerful and ubiquitous, human reliance on it to inform decisionmaking will increase, blurring the line between human and machine decisionmaking and potentially undermining the agency of humans.

A singular AGI or communities of AI agents could also become actors on the world stage.[24] Consider AGI with advanced computer programming abilities able to break out of the box and engage with the world across cyberspace, thanks to a designed-in internet connection or use of side-channel attacks. It could possess agency beyond human control, operate autonomously, and make decisions with far-reaching consequences. For example, AGI might serve as a proxy force, akin to Iran's axis of resistance, with informal relationships intended to shield an actor from accountability.[25] Even where accountability is clear, AGI could be misaligned—that is, operate in ways that are inconsistent with the intentions of its human designers or operators, causing unintentional harm. It could overoptimize on narrowly defined objectives and, for example, institute rolling blackouts to increase the cost-effectiveness of energy distribution networks. OpenAI elevated its scoring of misalignment risks in its latest AI, o1, because it "sometimes instrumentally faked alignment during testing" by knowingly providing incorrect information to deceive users.[26]

In the extreme, a loss-of-control scenario could result, wherein AGI's pursuit of its desired objectives incentivizes the machine to resist being turned off, counter to human efforts. Yoshua Bengio, a leading AI expert, notes, "This may sound like science fiction, but it is sound and real computer science."[27] This points to the possibility that AGI might achieve enough autonomy and

---

[22] Boiko, MacKnight, and Gomes, "Emergent Autonomous Scientific Research Capabilities of Large Language Models."

[23] Responsible AI Working Council, *Responsible Artificial Intelligence Strategy and Implementation Pathway*.

[24] Kissinger, Schmidt, and Mundie, "War and Peace in the Age of Artificial Intelligence."

[25] Maloney, "Iran's Order of Chaos."

[26] OpenAI, "OpenAI o1 System Card," September 12, 2024, p. 10.

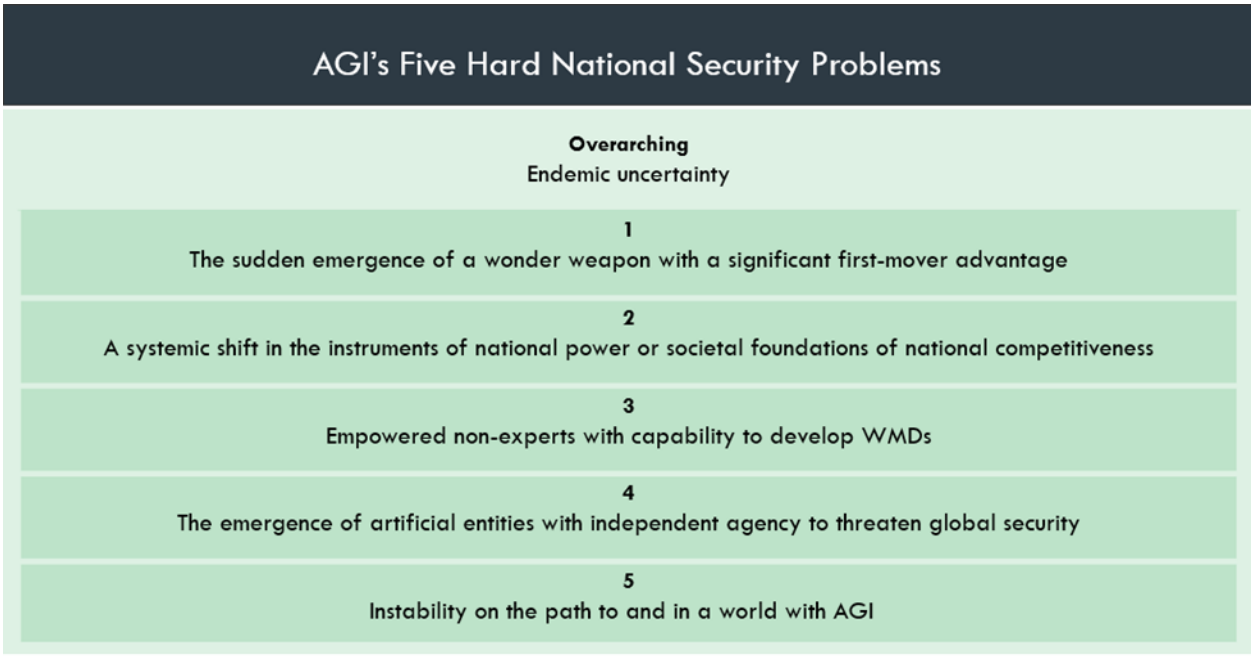
[27] Bengio, testimony before the U.S. Senate Committee on the Judiciary Subcommittee on Privacy, Technology, and the Law, p. 6.

behave with enough agency—intentionally or unintentionally—to be considered practically an independent actor on the global stage.

Fifth, there might be instability on the path to and in a world with AGI. Whether AGI is ultimately realized or not, the pursuit of AGI could foster a period of instability, as nations and corporations race to achieve dominance in this transformative technology. This competition might lead to heightened tensions, reminiscent of the nuclear arms race, such that the quest for superiority risks precipitating, rather than deterring, conflict. In this precarious environment, nations' perceptions of AGI's feasibility and potential to confer a first-mover advantage could become as critical as the technology itself. The risk threshold for action will hinge not only on actual capabilities but also on perceived capabilities and the intentions of rivals. Misinterpretations or miscalculations, much like those feared during the Cold War, could precipitate preemptive strategies or arms buildups that destabilize global security.

The figure summarizes these problems and the enveloping problem of endemic uncertainty.

Artificial General Intelligence's Five Hard National Security Problems



Toward a Robust Strategy

Current U.S. AI strategy, which started under the first Trump administration and continued in the Biden administration, seeks to retain technological leadership over China in core components of the AI tech stack.[28] This strategy of advancing U.S. technological competitiveness does much

[28] White House, "Artificial Intelligence for the American People"; Biden, "Advancing the United States' Leadership in Artificial Intelligence."

to position the United States for the potential emergence of AGI. An evolving semiconductor export control regime that embodies the "small yard, high fence" policy appears to have generated a five-year gap in advanced semiconductors.[29] However, heavily indexing on compute as a way to secure a national competitive advantage could be a brittle strategy if semiconductor export controls are not effectively enforced, China's semiconductor industry is able to catch up in due course, or AGI is achievable through less compute-intensive techniques.

U.S. policy also encourages the safe development of frontier AI models to avoid catastrophic consequences of AGI misuse, misalignment, or loss of control.[30] The new U.S. AI Safety Institute is up and running with a tight focus on AI safety, evaluating risks from nonstate actors seeking to use frontier AI models to develop bioweapons or new cybermalware.[31] Current U.S. strategy also embraces a series of no-regret options to address the potential emergence of AGI that are sensible under any alternative future.[32] These include investing in science, technology, engineering, and math education and workforce development; improving situational awareness on the state of the technology and its applications; protecting frontier AI model weights that are susceptible to theft or disruption by sophisticated rivals, such as China or Russia; and further promoting research on AI safety and alignment.

Finally, the U.S. government is promoting a U.S.-led global technology ecosystem within which AGI can be pursued. For example, the U.S. government recently supported Microsoft's expansion into the United Arab Emirates to develop new data centers, in part to prevent Chinese companies from entrenching their position.[33]

These constructive steps can help maintain a U.S. technological advantage over China without a specific end state in mind. At the same time, they are inadequate to address the prospects of a disruptive technological breakthrough, such as the potential emergence of AGI and the unique problems it would present.

Relying on the status quo requires acting on a belief that the United States is postured to respond effectively as uncertainties in AGI development resolve to reveal opportunities and challenges. However, the U.S. government is poorly postured to avoid technological surprises by U.S. or foreign companies pursuing AGI, let alone to manage the potential for AGI to disrupt global power dynamics and global security. Nor is the United States well positioned to realize the ambitious economic benefits of AGI without widespread unemployment and accompanying societal unrest. What would the U.S. government do if, in the next few years, a leading AI lab announced that its forthcoming model had the ability to produce the equivalent of 1 million

---

[29] Sullivan, "Remarks by National Security Advisor Jake Sullivan on the Biden-Harris Administration's National Security Strategy"; Patel, Koch, and Kundojjala, "Fab Whack-a-Mole."

[30] Biden, "Executive Order of October 30, 2023."

[31] U.S. Artificial Intelligence Safety Institute, "Strategic Vision."

[32] White House, "Fact Sheet."

[33] Mozur and Sanger, "Microsoft Makes High-Stakes Play in Tech Cold War with Emirati A.I. Deal."

computer programmers as capable as the top 1 percent of human programmers at the touch of a button? The national security implications for offensive and defensive cyberdynamics are profound. Equally profound are the economic implications of the introduction of such a capability into the labor market.

Any sensible strategy charting a course through an uncertain future should adapt as events unfold and areas of uncertainty are reduced. To enable the government to adapt quickly, strategic planning for high-regret policy options should be developed in advance of need, with the mechanics of execution thought through. Options could include ways to secure or accelerate the United States' technological lead in pursuing AGI, as well as contingency response plans for AGI-enabled security challenges. The U.S. government should consider post-AGI futures as well and engage in scenario exercises to anticipate the national security impacts. This includes (1) analyzing potential shifts in military power dynamics and economic disruptions and (2) formulating policies to mitigate both.

As AI-enabled capabilities transition from the realm of science fiction to that of science fact, the U.S. government should not be late to spot and address the opportunities and challenges. Aggressive planning would posture the U.S. government to react more decisively than it can in early 2025 as conditions warrant.

In contemplating the implications of AGI for global security, humanity is at the precipice of a potentially transformative era, akin to the dawn of the Industrial Revolution. The emergence of AGI would herald not just a technological revolution but also a profound shift in the geopolitical landscape, demanding a recalibration of national security paradigms. As we navigate this uncertain terrain, the United States should adopt a strategy that is both anticipatory and adaptive, recognizing the dual nature of AGI as both a promise and a peril.

References

Amodei, Dario, "Machines of Loving Grace: How AI Could Transform the World for the Better," *darioamodei.com* blog, October 2024.

Aschenbrenner, Leopold, "Situational Awareness: The Decade Ahead," *Situation Awareness* blog, June 2024.

Bajraktari, Ylli, "The Artificial General Intelligence Presidency Is Coming," *Foreign Policy*, September 30, 2024.

Bengio, Yoshua, testimony before the U.S. Senate Committee on the Judiciary Subcommittee on Privacy, Technology, and the Law, hearing titled "Oversight of A.I.: Principles for Regulation," July 5, 2023.

Biden, Joseph R., Jr., "Executive Order of October 30, 2023: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," *Federal Register*, Vol. 88, No. 210, November 1, 2023.

Biden, Joseph R., Jr., "Advancing the United States' Leadership in Artificial Intelligence; Harnessing Artificial Intelligence to Fulfill National Security Objectives; and Fostering the Safety, Security, and Trustworthiness of Artificial Intelligence," memorandum for the Vice President; Secretaries of State, the Treasury, Defense, Commerce, Energy, Health and Human Services, and Homeland Security; attorney general; directors of the Office of Management and Budget, National Intelligence, the Central Intelligence Agency, the Office of Science and Technology Policy, the National Science Foundation, the Federal Bureau of Investigation, the Office of Pandemic Preparedness and Response Policy, the National Security Agency, the National Geospatial-Intelligence Agency, and the Defense Intelligence Agency; representative of the United States to the United Nations; assistants to the President and chief of staff, to the President for national security affairs, and to the President for economic policy and director of the National Economic Council; chair of the Council of Economic Advisers; administrator of the U.S. Agency for International Development; and the national cyber director, October 24, 2024.

Boiko, Daniil A., Robert MacKnight, and Gabe Gomes, "Emergent Autonomous Scientific Research Capabilities of Large Language Models," arXiv, arXiv:2304.05332, April 11, 2023.

Center for AI Safety, "Statement on AI Risk," open letter, March 2023.

Cottier, Ben, Robi Rahman, Loredana Fattorini, Nestor Maslej, and David Owen, "How Much Does It Cost to Train Frontier AI Models?" *Epoch AI* blog, June 3, 2024.

Crawford, Forrest W., Kyle Webster, Gerald L. Epstein, Derek Roberts, Joseph Fair, and Sella Nevo, *Securing Commercial Nuclear Acid Synthesis*, RAND Corporation, RR-A3329-1, 2024. As of January 29, 2025: https://www.rand.org/pubs/research_reports/RRA3329-1.html

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan Mollick, Hila Lifshitz-Assaf, Katherine C. Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R. Lakhani, "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Harvard Business School Working Paper 24-013, September 2023.

Ding, Jeffrey, "The Innovation Fallacy: In the U.S.-Chinese Tech Race, Diffusion Matters More Than Invention," *Foreign Affairs*, August 19, 2024.

Einstein, Albert, letter to Franklin D. Roosevelt, August 2, 1939.

Geist, Edward, *Deterrence Under Uncertainty: Artificial Intelligence and Nuclear Warfare*, Oxford University Press, 2023.

Goertzel, Ben, "Who Coined the Term 'AGI'?" *goertzel.org* blog, August 28, 2011.

Himabindu, Modi, Revathi V, Manish Gupta, Ajay Rana, Pradeep Kumar Chandra, Hayder Saadoon Abdulaali, "Neuro-Symbolic AI: Integrating Symbolic Reasoning with Deep Learning," *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, Institute of Electrical and Electronics Engineers, December 2023.

Horowitz, Michael C., *The Diffusion of Military Power: Causes and Consequences for International Politics*, Princeton University Press, 2010.

Kissinger, Henry A., Eric Schmidt, and Craig Mundie, "War and Peace in the Age of Artificial Intelligence: What It Will Mean for the World When Machines Shape Strategy and Statecraft," *Foreign Affairs*, November 18, 2024.

Korinek, Anton, and Megan Juelfs, "Preparing for the (Non-Existent?) Future of Work," Brookings Center on Regulation and Markets Working Paper 3, August 2022.

Maloney, Suzanne, "Iran's Order of Chaos: How the Islamic Republic Is Remaking the Middle East," *Foreign Affairs*, May–June 2024.

Mazarr, Michael J., *The Societal Foundations of National Competitiveness*, RAND Corporation, RR-A499-1, 2022. As of January 29, 2025: https://www.rand.org/pubs/research_reports/RRA499-1.html

Mouton, Christopher A., Caleb Lucas, and Ella Guest, *The Operational Risks of AI in Large-Scale Biological Attacks: Results of a Red-Team Study*, RAND Corporation, RR-A2977-2, 2024. As of January 29, 2025: https://www.rand.org/pubs/research_reports/RRA2977-2.html

Mozur, Paul, and David E. Sanger, "Microsoft Makes High-Stakes Play in Tech Cold War with Emirati A.I. Deal," *New York Times*, April 16, 2024.

Murphy, Hannah, and Cristina Criddle, "Meta AI Chief Says Large Language Models Will Not Reach Human Intelligence," *Financial Times*, May 22, 2024.

OpenAI, "OpenAI o1 System Card," September 12, 2024.

OpenAI, "OpenAI o1 System Card," updated December 5, 2024.

Patel, Dylan, Jeff Koch, and Sravan Kundojjala, "Fab Whack-a-Mole: Chinese Companies are Evading U.S. Sanctions//Huawei Fab Network, WFE Vendors Cry Wolf, Framework for Future Controls," *SemiAnalysis*, October 28, 2024.

Responsible AI Working Council, U.S. Department of Defense, *Responsible Artificial Intelligence Strategy and Implementation Pathway*, June 2022.

Seetharaman, Deepa, "The Next Great Leap in AI Is Behind Schedule and Crazy Expensive," *Wall Street Journal*, December 20, 2024.

Sevilla, Jaime, Tamay Besiroglu, Ben Cottier, Josh You, Edu Roldán, Pablo Villalobos, and Ege Erdil, "Can AI Scaling Continue Through 2030?" *Epoch AI* blog, August 20, 2024.

Sullivan, Jake, "Remarks by National Security Advisor Jake Sullivan on the Biden-Harris Administration's National Security Strategy," White House, October 13, 2022.

Tong, Anna, and Michael Martina, "US Government Commission Pushes Manhattan Project–Style AI Initiative," Reuters, November 19, 2024.

U.S. Artificial Intelligence Safety Institute, National Institute of Standards and Technology, U.S. Department of Commerce, "Strategic Vision," webpage, created May 20, 2024, updated January 17, 2025. As of January 29, 2025: www.nist.gov/aisi/strategic-vision

U.S.-China Economic and Security Review Commission, *2024 Report to Congress of the U.S.-China Economic and Security Review Commission*, 118th Congress, 2nd Session, November 2024.

White House, "Artificial Intelligence for the American People," webpage, undated. As of January 29, 2025: https://trumpwhitehouse.archives.gov/ai

White House, “Fact Sheet: Biden-Harris Administration Outlines Coordinated Approach to Harness Power of AI for U.S. National Security,” October 24, 2024.

Wong, Matteo, “The AI Boom Has an Expiration Date,” *The Atlantic*, October 17, 2024.

Yudkowsky, Eliezer, “Pausing AI Developments Isn’t Enough. We Need to Shut It All Down,” *Time*, March 29, 2023.

About the Authors

Jim Mitre is the inaugural vice president and director of RAND Global and Emerging Risks. His current research focuses on the intersection of global security and artificial intelligence. Mitre holds a J.D.

Joel B. Predd is a senior engineer at RAND and director of the Geopolitics of AGI Initiative. His current research focuses on the geopolitics of artificial intelligence and on the intersection of the security and economic competition with China. Predd holds a Ph.D. in electrical engineering.