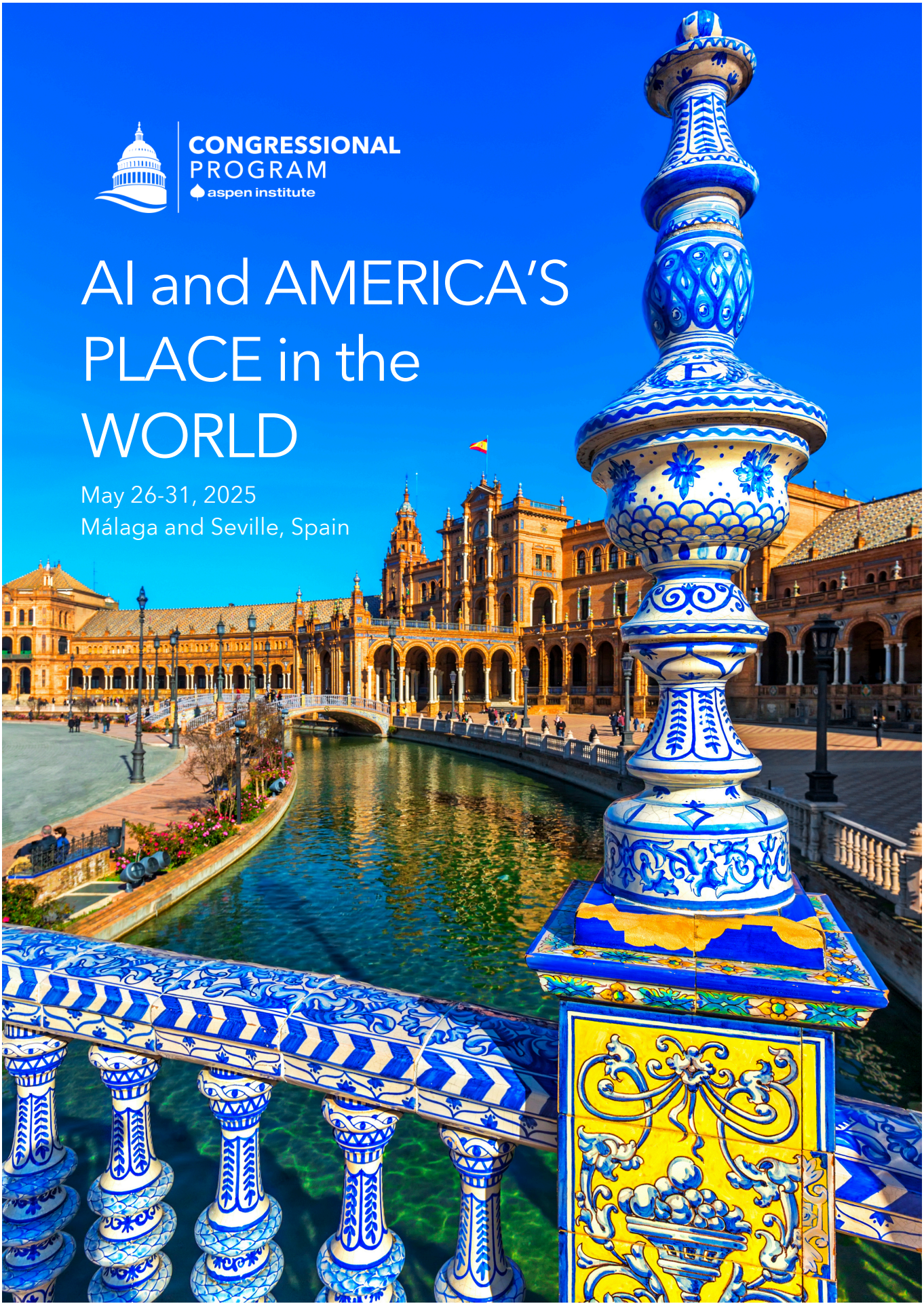
CONGRESSIONAL PROGRAM
aspen institute

AI and AMERICA'S PLACE in the WORLD

May 26-31, 2025
Málaga and Seville, Spain



CONGRESSIONAL PROGRAM

aspen institute

AI and America's Place in the World

May 26-31, 2025 | Málaga and Seville, Spain

TABLE OF CONTENTS

AGENDA[1]

MONDAY, May 26:

U.S. participants depart the U.S. throughout the day.

TUESDAY, May 27 (Overnight in Málaga):

U.S. participants arrive in Spain throughout the day.

6 – 7 PM: Pre-Dinner Fireside Chat:
The Geopolitics of AI: Overview

AI is not just a transformative technology—it is a strategic asset reshaping the balance of global power. This opening session will provide a high-level overview of the geopolitical dimensions of AI, from national security and economic advantage to values-based competition over how the technology is governed and deployed. Our opening speakers will assess the AI strategies of major powers, the global race for talent and computers, and the emerging fault lines between democratic and authoritarian approaches to AI development.

Speakers:
Vilas Dhar, *President, McGovern Foundation*
Eva Maydell, *Member, European Parliament*
Vivian Schiller, *Executive Director, Aspen Digital* (moderator)

7 – 9 PM: Working Dinner
Seating is arranged to expose participants to a diverse range of views and provide the opportunity for a meaningful exchange of ideas. Scholars and lawmakers are rotated daily. Discussions will focus on the geopolitics of AI.

WEDNESDAY, May 28 (Overnight in Málaga):

[1] *Congressional Program Executive Director Charlie Dent moderates the discussion sessions, recognizes members of Congress who have questions, and is assisted by a timekeeper to ensure the conversation is quick paced and every member of Congress has an opportunity to ask questions and discuss the issues.*

7 – 8:45 AM: Breakfast available to all participants

9 – 9:15 AM: Introduction and Framework of the Conference
This conference is organized into roundtable conversations, working lunches, and pre-dinner remarks. This segment will highlight how the conference will be conducted, how those with questions will be recognized, and how responses will be timed to allow for as much engagement as possible.

Speaker:
Charlie Dent, *Vice President, Aspen Institute;*
Executive Director, Congressional Program

9:15 – 11 AM: Roundtable Discussion:
Global Competition: AI and the Battle for Technological Leadership

AI has become central to geopolitical, strategic and security competition. This session examines how different actors—China, the Gulf States, Europe, and Africa—are approaching and developing AI tools and the infrastructure to adopt them for economic growth, national security, and political influence. Panelists will explore China's state-driven AI strategy, the Gulf's heavy investment in AI infrastructure, Europe's focus on regulation, and Africa's emerging role in AI adoption. The discussion will also address critical pressure points, including export controls, semiconductor supply chains, and the policy levers the U.S. can use to maintain its technological edge.

Speakers will address the following questions:

- How does China's centralized, state-driven approach to AI development compare to the U.S.'s more decentralized, private-sector-led model?
- How are the Gulf States leveraging AI investment as a geopolitical tool, and what are the implications for U.S. interests?
- Does Europe's strong regulatory stance on AI hinder its competitiveness in the global AI race, or will the so-called "Brussels effect" create a framework that others will eventually follow?
- What role does Africa play in the AI landscape, and how should the U.S. engage with emerging AI markets on the continent?
- How effective are current U.S. export controls on AI-related technologies, and what additional steps should be taken to secure semiconductor supply chains?

Speakers:
Kayla Blomquist, *Director, Oxford China Policy Lab*

Klon Kitchen, *Managing Director, Beacon Global Strategies*

11 – 11:15 AM: **Break**

11:15 AM – 1 PM: Roundtable Discussion:
Economic, Jobs, and Labor Impact

AI is already reshaping the U.S. economy—while it is increasing productivity and restructuring entire sectors, it is raising difficult questions about the future of work for U.S. workers and companies. This discussion will explore how AI is likely to affect labor markets in both the short and long term, including sectoral job displacement, wage pressure, and new opportunities for growth and investment in the U.S. and globally. It will also address how policymakers can balance innovation with workforce resilience through education, training, and targeted investment.

Speakers will address the following questions:

- How does AI change the nature of productivity and economic competitiveness at the national level—and what levers does the federal government have to shape outcomes?
- Which sectors of the U.S. economy and which jobs are most vulnerable to AI-driven disruption, and which are most likely to benefit?
- How can policymakers ensure that workers—particularly in lower-income and rural communities—are not left behind in the AI economy?
- What investments are needed in education, workforce retraining, and apprenticeship to prepare Americans for AI-impacted jobs?

Speakers:
Jack Clark, *Co-Founder, Anthropic*
Jaime Sevilla, *Director, Epoch*

1 – 2 PM: Working Lunch with Remarks: *AI and Global Health*

Speaker*:*
Mark Dybul, *Co-Director of the Center for Global Health Practice and Impact, Georgetown University Medical Center*

2:30 – 4 PM: Site Visit to Google Cyber Security Center in Málaga
Málaga is the European headquarters of Google's cybersecurity network. We will be doing a site visit there on the same day we are discussing cyber security threats in the region and US challenges more broadly. Learning from our European colleagues about the current day risks, opportunities and challenges associated with cyber security and

ensuring US private sector leadership will be a core theme of the day. Jose Luis Ugía, Developer Relations for Google Spain at the Málaga Center will be our guide.

4 – 6 PM: Individual Discussions
Scholars will be available to meet individually with members of Congress for in-depth discussion of ideas raised in the morning sessions, including Jack Clark, Jamie Sevilla, Kayla Bloomquist and Klon Kitchen.

6:30 – 7:30 PM: Pre-Dinner Remarks: *Software Singularity*
Jack Clark, Co-Founder of Anthropic, will deliver thought-provoking pre-dinner remarks examining Software Singularity and its critical implications for technology policy, governance, and regulatory approaches in an AI-accelerated world.

Speaker*:*
Jack Clark, *Co-Founder, Anthropic*

7:30 – 10 PM: Working Dinner
Seating is arranged to expose participants to a diverse range of views and provide the opportunity for a meaningful exchange of ideas. Scholars and lawmakers are rotated daily. Discussions will focus on global AI competition and AI's economic impacts.

THURSDAY, May 29 (Overnight in Seville):

6:30 – 7:45 AM: Breakfast

8 – 10 AM: Roundtable Discussion:
AI and Cybersecurity: Defending Against Next Generation Threats

Increasingly powerful AI is reshaping both cyber defense and cyber offense. AI is being used to detect threats, automate security responses, and fortify digital infrastructure—but it is also enabling more sophisticated cyberattacks, including AI-generated phishing, deepfake deception, and turbocharged hacking. This session explores the evolving cybersecurity landscape, the risks AI poses to national security and critical infrastructure, and the policies and technologies needed to protect against AI-driven threats. Panelists will also discuss how governments and private-sector leaders can collaborate to ensure AI enhances, rather than undermines, cybersecurity.

Speakers will address the following questions:

- How is AI transforming both cyber offense and defense, and what emerging threats should we be most concerned about?
- What role should the federal government play in regulating AI-driven cybersecurity threats, particularly regarding critical infrastructure?
- How can AI be used to detect and mitigate threats faster than traditional cybersecurity methods?
- What are the risks of AI-generated misinformation and deepfakes in cyber warfare and political interference?
- How should businesses and governments collaborate to create global norms for AI security while maintaining competitive advantages?

Speakers:
Rob Joyce, *Founder, Joyce Cyber; former NSA Director of Cybersecurity*
Ciaran Martin, *Founding CEO of the National Cyber Security Centre and Professor, University of Oxford*

10 – 10:15 AM: Break

10:15 AM – 12 PM: Roundtable Discussion:
The Race for Compute and Energy Demands from AI Systems

AI's rapid advancement is driving an unprecedented surge in energy consumption, from data centers powering large-scale machine learning models to edge computing and real-time AI applications. As AI adoption grows, so does its strain on power grids, raising concerns about sustainability, efficiency, and long-term energy security. We will explore the energy demands of AI systems today and in the future, strategies for improving efficiency, and the role of renewable energy and grid modernization in supporting AI's expansion, including how policymakers, utilities, and tech companies can address the intersection of AI growth and energy sustainability.

Speakers will address the following questions:

- What is current AI energy demand and how is it expected to grow in the next decade?
- What technological advancements, such as more efficient chips and cooling systems, can help mitigate AI's energy impact?
- What role should renewable energy play in meeting AI's growing electricity needs, and is it scaling fast enough?
- How should governments and regulators think about the balance between energy efficiency standards for AI infrastructure and promoting rapid innovation?

Speakers:
Brad Carson, *Co-Founder and President, Center for Responsible Innovation*
Divyansh Kaushik, *Vice President, Beacon Global Strategies*
Owen Larter, *Senior Director, Frontier Policy and Public Affairs, Google DeepMind*

12:30 PM: Pick up Box Lunch for the Road

1 – 4 PM: Depart Malaga for Seville Via Bus

4 PM: Arrive in Seville

5 – 7 PM: Individual Discussions
Scholars will be available to meet individually with members of Congress for in-depth discussion of ideas raised in the morning sessions, including Brad Carson, Divyansh Kaushik, Owen Larter, Chris Krebs and Vivian Schiller.

7 – 9 PM: Working Dinner
Seating is arranged to expose participants to a diverse range of views and provide the opportunity for a meaningful exchange of ideas. Scholars and lawmakers are rotated daily. Discussions will focus on AI and cybersecurity, AI-driven energy demands, and AI-generated information manipulation.

FRIDAY, May 30 (Overnight in Seville):

7 – 8:45 AM: Breakfast

9 – 11 AM: Roundtable Discussion:
AI and Military Power: The Future of Warfare and National Security

AI is transforming military strategy and tactics, with impacts on everything from autonomous weapons to war gaming, intelligence analysis and battlefield decision-making. States are racing to integrate AI into defense systems, raising critical questions about strategic stability, human oversight, and the ethics of autonomous warfare. Panelists will cover how AI is reshaping military power, the risks of an AI arms race, and how the U.S. can maintain its technological edge while ensuring responsible AI deployment. Panelists will also discuss the role of AI in cyber warfare, intelligence gathering, and threat deterrence.

Speakers will address the following questions:

- How is AI changing military strategy, and what capabilities will define the next era of warfare?
- Should the U.S. pursue fully autonomous weapons systems, or is human oversight essential for ethical and strategic reasons?
- What are the risks of an AI arms race, and how can global powers establish guardrails to prevent escalation?
- How should AI be integrated into intelligence and surveillance to enhance national security while protecting civil liberties?
- What role does the private sector play in military AI development, and how should government manage public-private partnerships in defense tech?

Speakers:
Kenneth Payne, *Professor of Strategy, King's College London*
Matt Turpin, *Senior Advisor, Palantir Technologies*

11 – 11:10 AM: Break

11:10 – 11:50 AM: Remarks: *AI-Generated Information Manipulation from Adversarial Nation-States*

AI is supercharging disinformation campaigns, enabling U.S. adversaries like Russia, China, Iran, and North Korea to manipulate public opinion, disrupt democratic institutions, and erode public trust in the media. Deepfakes, AI-generated propaganda, and automated troll farms are making foreign influence operations more scalable, convincing, and difficult to detect. The discussion will examine how AI is being weaponized in the information space, the challenges of countering AI-driven manipulation, and the policies needed to protect democratic societies from digital deception.

Speakers:
Chris Krebs, *Former United States Director of the Cybersecurity and Infrastructure Security Agency*

11:50 AM – 1:15 PM: *Policy Reflections for Members of Congress*

This time is set aside for members of Congress to reflect on what they have learned during the conference and discuss their views on implications for U.S. policy. Drawing on the full range of conversations throughout the week, members will seek to identify

for each other the most promising takeaways for the United States policy process, with a special focus on opportunities for bipartisan cooperation. This is a members-only conversation.

1:15 – 2 PM: Working Lunch
Discussion continues between members of Congress and scholars on AI, military power, and national security.

3 – 4:30 PM: Historical Site Visit, Alcázar (Optional)
Participants will be taken on a guided tour of the Real Alcázar of Seville, a historic Royal Palace, formerly the site of the Islamic era citadel of the city.

5 – 7 PM: Individual Discussions
Scholars will be available to meet individually with members of Congress for in-depth discussion of ideas raised in the morning sessions, including Kenneth Payne and Matt Turpin.

7 – 8 PM: Pre-Dinner Fireside Chat: ***Final Reflections and Looking Over the Horizon***
In this concluding session, Dr. Brad Carson will synthesize key insights from the week's discussions and reflect on what they mean for U.S. global leadership in an era shaped by AI and geostrategic competition. Drawing on his experience in academia and public service, Dr. Carson will outline strategic takeaways from the central debates and offer his own forward-looking recommendations for U.S. policymakers. Sébastien Krier from Google DeepMind will discuss the next generation of AI systems.

Speakers:
Brad Carson, *Co-Founder and President, Center for Responsible Innovation*
Sébastien Krier, *Manager, Policy Development and Strategy, Google DeepMind*

8 – 9:30 PM: Working Dinner
Seating is arranged to expose participants to a diverse range of views and provide the opportunity for a meaningful exchange of ideas. Scholars and lawmakers are rotated daily. Discussions will focus on AI policy reflections from the week.

SATURDAY, May 31:

Participants depart throughout the day.

CONFERENCE PARTICIPANTS

MEMBERS OF CONGRESS AND THEIR SPOUSES:

Rep. Pete Aguilar and Alisha Aguilar

Rep. Jim Baird and Danise Baird

Rep. Ami Bera

Rep. Dan Crenshaw and Tara Crenshaw

Rep. Diana DeGette and Lino Lipinsky

Rep. Neal Dunn and Leah Dunn

Sen. Dick Durbin and Loretta Durbin

Rep. Bill Foster and Aesook Byon

Rep. Andy Harris and Nicole Harris

Sen. Martin Heinrich and Julie Heinrich

Rep. Darrell Issa and Rebecca Glover

Rep. Sara Jacobs

Rep. John Joyce

Rep. Darin LaHood

Rep. Rick Larsen and Tiia Karlén

Rep. Ted Lieu and Betty Lieu

Rep. Jim McGovern and Lisa McGovern

Rep. Jay Obernolte and Heather Obernolte

Rep. Kim Schrier and David Gowing

Rep. Pete Sessions and Karen Sessions

Rep. Adam Smith and Sara Smith

Sen. Chris Van Hollen and Katherine Wilkens

SCHOLARS AND EXPERTS:

Kayla Blomquist	*Director, Oxford China Policy Lab*
Jack Clark	*Co-Founder, Anthropic*
Vilas Dhar	*President, McGovern Foundation*
Mark Dybul	*Co-Director of the Center for Global Health Practice and Impact, Georgetown University Medical Center*
Rob Joyce	*Founder, Joyce Cyber; former NSA Director of Cybersecurity*
Divyansh Kaushik	*Vice President, Beacon Global Strategies*
Klon Kitchen	*Managing Director, Beacon Global Strategies*
Chris Krebs	*Former United States Director of the Cybersecurity and Infrastructure Security Agency*
Sébastien Krier	*Manager, Policy Development and Strategy, Google DeepMind*
Owen Larter	*Senior Director, Frontier Policy and Public Affairs, Google DeepMind*
Ciaran Martin	*Founding CEO of the National Cyber Security Centre and Professor, University of Oxford*
Eva Maydell	*Member, European Parliament*
Kenneth Payne	*Professor of Strategy, School of Security Studies, King's College London*
Jaime Sevilla	*Director, Epoch*
Matt Turpin	*Visiting Fellow, Hoover Institution; Senior Advisor, Palantir Technologies*

CONFERENCE RAPPORTEUR:

Matthew Rojansky	*Rapporteur and Counselor to the Aspen Institute Congressional Program*

FOUNDATION REPRESENTATIVES:

Justin Bullock	*Vice President of Policy, Americans for Responsible Innovation*
Brad Carson	*Co-Founder and President, Americans for Responsible Innovation*
Sharon Davies	*President & CEO, Kettering Foundation*
Eric Gastfriend	*Co-Founder and Executive Director, Americans for Responsible Innovation*
Melanie Harris	*Senior AI Policy Advisor, Open Philanthropy*
Rachel Jackson	*Chief Operating Officer, American Flood Coalition*
Amber Miller	*President, Hewlett Foundation*
Darla Minnich	*Senior Fellows Coordinator, Kettering Foundation*
William Moore	*CEO, Eleanor Crook Foundation*
Simon Morrison	*Senior Public Policy Manager, AI and Privacy Policy, Amazon*
Rip Rapson	*President and CEO, The Kresge Foundation*

ASPEN DIGITAL:

Vivian Schiller	*Vice President and Executive Director, Aspen Digital*

ASPEN INSTITUTE EXECUTIVE TEAM and BOARD OF TRUSTEES:

Elliot Gerson	*Executive Vice President, Policy Programs and International Partners, Aspen Institute*
Jane Harman	*Aspen Institute Trustee and Aspen Strategy Group Member, President Emerita, Woodrow Wilson Center*

ASPEN INSTITUTE CONGRESSIONAL PROGRAM:

Charlie Dent and **Pamela Dent**	*Executive Director, Congressional Program and Vice President, Aspen Institute*
Tyler Denton	*Deputy Director*
Jennifer Harthan	*Senior Manager of Congressional Engagement*
Sajan Shah	*Aspen Institute Congressional Program Volunteer*
Galen Voorhees	*Senior Conference Manager*

RAPPORTEURS' SUMMARY

Matthew Rojansky
Rapporteur and Counselor to the Aspen Institute Congressional Program; President and CEO, The U.S.-Russia Foundation

From May 26-31, 2025, the Aspen Congressional Program brought Members of Congress together with leading scholars, technologists, and industry practitioners to explore how artificial intelligence is reshaping global power, competitiveness, and the social contract. Set in Málaga and Seville, Spain, at a crossroads of transatlantic relations and digital transformation, the program examined what it will take for the United States to lead responsibly in the AI age, and how to ensure the benefits of innovation are broadly shared.

Spain was a fitting host country for this conversation. As home to the first national AI agency in Europe and an emerging technology hub within the EU, Spain is positioning itself as a leader in ethical innovation and regulatory agility. The country has embraced AI not only as a tool for economic growth but as a public good, integrating AI policy into industrial strategy, education, and public service delivery. With its emphasis on human rights, rule of law, and cross-border cooperation, Spain's approach offers a distinctive European perspective on AI governance.

A visit to Google's cybersecurity campus in Málaga underscored the city's role in the global technology landscape. Málaga was the birthplace of VirusTotal, one of the world's most influential collaborative cybersecurity platforms. What began in the 1990s as a Spanish-language newsletter for testing antivirus tools evolved into an automated system capable of identifying and analyzing malware threats at a global scale. After Google acquired VirusTotal in 2012, the campus became a hub for threat intelligence and AI-enabled defense systems. Today, AI systems managed from this hub can assess whether any given message or file is malicious or benign in seconds, offering real-time protection against an ever-growing volume of cyberattacks—now exceeding two million per day.

Málaga was important for another less obvious reason: Renowned abstract painter Pablo Picasso was born there in 1881, and his legacy includes a museum in Málaga filled with works that had been kept privately within his family until this century. Recognizing that even as he innovated new artistic styles, he was building on the work of past masters, Picasso once said, "A plagiarist steals from one person, but an artist steals from everyone." This was a fitting reminder of how much generative AI's seemingly miraculous capabilities are also built upon millennia of human achievement.

Against this backdrop, Members and Scholars explored key questions: How can we leverage AI as a source of strength for democratic societies? How should the U.S. engage with allies and partners to shape global AI norms and infrastructure while keeping ahead of rival authoritarian AI advances? And how to ensure that AI development advances freedom and opportunity while protecting society from its potentially harmful

impacts? Through the week, the conversation ranged from geopolitics to economic and workforce issues, and the potential architecture for AI governance, offering both deep dives into technical topics and a wide-angle view of America's strategic choices in a period of rapid change.

The Geopolitics of AI: An Overview

The conference's opening session framed the global race for leadership in artificial intelligence not as a distant or abstract challenge, but as a present and urgent matter with real implications for national competitiveness, economic productivity, and democratic resilience. Members and scholars examined the evolving public perception of AI, differing strategies among democracies and with authoritarian rivals, and the need to move beyond binary debates toward action that delivers benefits that will be tangible and impactful for all Americans.

One Scholar noted that while conversations about AI have been ongoing in Washington for years, what has changed recently is the lived experience of Americans. In fields like agriculture, new AI-powered tools are already helping farmers increase yields—yet many citizens remain unsure or fearful, uncertain whether AI will benefit them or leave them behind. Bridging this gap between the elite-level conversation and everyday relevance is essential, the Scholar argued, to developing a U.S. policy approach for AI that will enjoy broad public support.

From a European perspective, another Scholar explained, the shift from an AI safety and risk-centric discourse toward one focused on investment and competitiveness has been clear. Europe has invested hundreds of millions of Euros to attract top global talent for its tech champions, while clearing regulatory hurdles to construct 6-7 so-called gigafactories for AI chips. Whereas the first AI summit hosted in Europe, the UK AI Safety Summit in late 2023, treated the technology like nuclear energy, focusing on risks from proliferation, weaponization, or accidents, the recent Paris AI Action Summit framed AI as a force akin to electricity or the Internet. The conversation is now about whether Europe can harness AI to boost its industrial productivity and assert a competitive model that diverges from that of the United States or China.

Scholars discussed whether the global AI debate had become stuck in a false binary between safety and progress. One Scholar suggested that policy should be biased toward action, accompanied by mechanisms to recalibrate and increase oversight as AI systems evolve. The shift from philosophical concern to practical outcomes, Scholars suggested, is not only desirable but essential to preserve public confidence and democratic agency.

Members raised the question of what roles federal and local authorities, whether in the U.S. or Europe, should play in regulating and deploying AI. One Member suggested that public funding to subsidize AI deployment may not be unnecessary, particularly when AI tools promise large efficiency gains, as those tools will be underwritten by private investment. A Scholar responded that while market forces can accelerate AI adoption,

public capital investment in infrastructure and favorable policy regimes are still vital to ensure broad access and reduce perceived risks for communities.

Members asked what their role should be in this transition. Some Members pointed to AI's ability to reduce pesticide use in agriculture or improve steel production efficiency as real-world examples of how AI can improve productivity for constituents while lowering costs. Others underscored the importance of ensuring that public agencies are "fit for purpose," not just in terms of technical capacity but also in their ability to partner effectively with private innovators. A Scholar underscored that with many loud voices advancing narrow agendas for AI development, policymakers have a vital role to play in making sure these systems reflect widely shared democratic values. The conversation concluded with this simple framework: the question is no longer what AI is, but what we are going to do with it, and how quickly.

Global Competition: AI and the Battle for Technological Leadership

This session focused on the global strategic dimensions of AI, including who builds it, who controls it, and who benefits from it, especially in the growing competition between the U.S. and China. Scholars and Members explored the roles of infrastructure, compute, and partnerships in determining global leadership, while debating how the U.S. can maintain its edge and uphold its values in the face of intense competition with China.

One Scholar defined U.S. goals as follows: maximizing the benefits of AI for society, securing enduring technological leadership, and ensuring that global AI development aligns with U.S. interests and norms. Central to all three goals is infrastructure, which in turn depends on land, energy, and water. Access to advanced AI chips (GPUs) is also a key constraint, but one in which the United States now enjoys a leadership position. With demand far outpacing supply, compute is potentially a chokepoint for control of the global AI ecosystem. While U.S. companies currently lead in AI, another Scholar warned, adversaries may use AI itself to overcome development hurdles. China's use of AI as a tool for state control and soft power projection, through mobile connectivity, surveillance infrastructure, and social media, offers a potentially attractive model for other nations, even though it is antithetical to U.S. values.

Members pressed Scholars on how to strike the right balance between regulation and innovation, including on export controls. One Member pointed to the success of GPU restrictions in slowing China's progress, while others worried about the unintended consequences: smuggling, companies exploiting regulatory loopholes, and overly aggressive control of infrastructure or hardware "choke points" that could alienate important strategic partners. Ideas ranged from implementing hardware/firmware solutions (preventing chips from being used effectively outside their authorized geography) to creating a shared governance model for AI infrastructure across the democratic world.

A recurring theme in the discussion was how AI has decentralized, or to some extent simply shifted, power from the hands of government to private hands. As one Scholar

put it, the U.S. government is just one stakeholder among many, often the slowest moving one. While some companies are deeply patriotic and already working closely with the U.S. government, others view Washington as a partner of last resort. How to structure public-private AI partnerships and ensure accountability remains unresolved, but there was a sense of urgency from the group that something more than occasional and voluntary transparency and cooperation from frontier AI model developers is needed.

Members and Scholars also discussed the risks of outsourcing AI development to foreign partners with more favorable regulatory environments. One example on nearly everyone's mind is the Gulf States, particularly the United Arab Emirates and Saudi Arabia, which offer plentiful available land, energy resources, deployable capital, streamlined government decision-making, and high-level enthusiasm for building the massive data center infrastructure that will be needed to train and deploy future generations of advanced AI models. The recent U.S. presidential visits to the region included inking deals with major U.S. technology companies to do exactly this, and Members worried that this might be handing China access to U.S. technology, given the Gulf states' historically close relations with Beijing. Both Members and Scholars concluded that we may be on the cusp of another massive wave of technological offshoring, unless the United States can resolve the social, economic, and political factors limiting datacenter development at home.

Scholars and Members also examined partnership with Europe, raising the concern that Europe's technology regulations, such as the AI Act and Digital Markets Act, seem to unfairly target U.S. firms while EU countries aggressively subsidize their domestic champions. Others emphasized that U.S. models must become more adaptable and relevant to non-English speakers, as well as to less developed regions, if they are to compete globally. Africa and the Global South are watching, Scholars cautioned, to decide whether AI, be it U.S. or Chinese in origin, is merely a new form of colonialism, or an opportunity that will support an acceleration in much-needed economic development.

Discussion returned frequently to the U.S. workforce, and the need to preserve America's edge in talent, including by strengthening university research, addressing the digital divide within the U.S., and attracting the top human talent from around the world. One Scholar noted that 60% of leading U.S. AI companies have at least one immigrant founder, highlighting the strategic importance of a fair and functional immigration system for U.S. economic and technological leadership.

The session closed by recognizing that time is not on our side. China is closing the technological gap, while already leading on many measures of infrastructure and productive capacity. Decisions that seemed abstract or philosophical as recently as a year ago, on infrastructure, partnerships, and the talent pipeline, will soon determine whether the United States continues to be the world's leading power. The question is not whether to act, but how boldly and how soon.

Economic, Jobs, and Labor Impact of AI

This session explored how AI is reshaping economic activity and labor markets, both through current applications and likely future trends. Members and Scholars discussed potential and already realized productivity gains, the risks inherent in automation, challenges with education and workforce training, and the need to prepare public institutions for rapid change. Discussion opened with the observation that in just a few years, systems have evolved from playful and often buggy story generators to outperforming world-class mathematicians, programmers, and other specialists. With computational resources growing by orders of magnitude each year, models are improving rapidly, and based on current trends, the next five years will produce tools that will revolutionize and reorder the economies and societies that integrate them.

From accelerating clinical trials to revolutionizing contract law, education, and intelligence analysis, Scholars explained that AI is already taking over many economically valuable tasks. Today, 40% of usage on one major platform is coding-related, far outpacing coding's share of the overall economy. However, adoption remains lower in areas like food service and healthcare delivery, where work is less digitized and more tactile. Within one leading AI company, the intent is to have the models themselves doing all the programming work, with human engineers providing oversight, by the end of next year. While software engineers are not (yet) being let go, companies' hiring needs are already being reassessed.

Members and Scholars considered the infrastructure needed to support AI-driven economic growth. Datacenters and compute capacity are central, but so are the institutions that facilitate the application of AI tools in various spheres of economic life. National labs, universities, and R&D centers will face a reckoning as scientific discovery itself is transformed by AI-augmented research. As one Scholar put it, researchers who do not use AI may soon find themselves like John Henry, racing against machines that can match or exceed human capability.

Looking at workforce adaptation more broadly, Members asked how to prepare people for jobs that don't yet exist. One response was to prioritize transparency in AI tools and create space for public experimentation, so that individuals can develop skills in real time as the tools evolve. Scholars emphasized the importance of avoiding "automation by default" and focusing on where existing roles can be augmented, especially in critical areas like education, medicine, and caregiving. People, they said, remain essential in the service economy. As one Member worried, if only the privileged have access to human-centered services and everyone else gets robots, social divides will be exacerbated. Scholars also warned that some of the most economically promising uses of AI, like improving bureaucratic efficiency, may come with tradeoffs for public trust.

Members and Scholars discussed the recurring themes of immigration for attracting top AI talent to the United States, as well as how to appropriately tax the economically valuable work done by AI, which will not file an income tax return. A Scholar proposed that if compute becomes the central locus of economic activity, governments may need

to tax and regulate it directly, for example, with a small tax on each compute cycle that could pay for the government to distribute compute access to those who cannot afford it. The same Scholar suggested that AI companies could provide compute access to the government at cost, which would enable the government to channel it to public-interest applications like scientific research or social services.

The session closed with a note of urgency and opportunity. Policymakers were encouraged to experiment with AI tools themselves, both to understand the technology and to shape it. Participants broadly recognized that the transition is already underway and success will depend not just on innovation, but on ensuring that democratic institutions and social values are embedded in every stage of the transformation.

AI and Global Health

This session explored the transformative potential of artificial intelligence to improve global health outcomes, reduce medical costs, and accelerate innovation, while tackling urgent biosecurity and governance challenges posed by increasingly powerful models. Members and scholars discussed practical applications of AI in healthcare delivery, as well as economic and regulatory dimensions of clinical research, and the ethical and security implications of emerging capabilities in genomics and drug design.

A Scholar began by outlining key opportunities: AI can already help automate bureaucratic processes, augment the diagnostic capabilities of frontline health workers, and even identify and personalize cancer therapies within minutes, dramatically reducing reliance on traditional chemotherapy. In low-resource settings, AI has the potential to equip any healthcare worker with near-primary-care-level decision-making power. In drug development, AI models are helping researchers identify promising compounds, speed up clinical trial design, and even simulate patient responses in virtual clinical trials. Such applications promise to cut the cost and time of drug development by orders of magnitude, but as the Scholar noted, regulatory agencies like the U.S. FDA have not yet approved AI to fully run virtual trials, even though the systems are practically at that level of capability already.

The Scholar also warned about risks. AI tools capable of modeling biological systems can be misused to design pathogens with tailored traits, such as resistance to vaccines, increased lethality, or altered transmission profiles that can be used to create a pandemic. One Scholar flagged gene synthesis as the key risk frontier, where AI-enabled innovation and biosecurity threats converge. If rules are not in place before AI models become self-generating or broadly open-sourced, the opportunity for oversight may vanish. Participants likewise considered regulatory mechanisms to screen and certify AI systems used in biomedical development. For example, FDA approval could be restricted to products built with models that pass security reviews, which is a way to indirectly regulate even open-source tools without banning them outright.

Members raised concerns about how AI will be deployed, given difficulties with the broader structure of the U.S. healthcare system. One noted the risk that since large insurance companies and hospital systems effectively control healthcare, they are likely to deploy AI within the medical environment mainly to constrain how doctors interact with patients to maximize their profitability, prescribing limits on treatments, shortening face-to-face visit time with doctors, or eliminating medical professionals' individual discretion in favor of AI guidance. Some Members voiced concern about whether the public would actually benefit from AI tools if incentives are captured by corporate intermediaries. A Scholar acknowledged these challenges and pointed to the need for both government oversight and public-private partnership, citing cybersecurity models as a possible framework for coordination.

Beyond genomics and clinical trials, the session touched on how AI might improve other global health and development priorities, like access to clean water and food security. Here, the emphasis was less on frontier science and more on the integration and optimization of existing technologies. AI could help design more efficient systems and guide the deployment of known solutions in new ways. But speakers stressed that these benefits will only materialize with a deliberate effort to connect innovation with delivery, and if governance structures can anticipate risk rather than reacting after harm has occurred.

Members and Scholars agreed that AI could be a revolutionary augmentation to global healthcare, but that realizing the benefits will not be easy. It will require early, adaptive, and technically informed policymaking. Congress has the difficult task of identifying a "Goldilocks" balance: too much restriction could choke off innovation, while too little could permit catastrophic misuse. The challenge ahead is to act both fast enough to stay ahead of the technology and with sufficient wisdom to get the balance right.

Agentic AI and the Challenge of Self-Improving Systems

This session came the closest to what might have seemed like science fiction as recently as five years ago. Participants considered the emerging class of "agentic" AI systems, which are tools capable not only of executing user-defined tasks but of initiating complex sequences of actions, conducting independent research, writing and improving code, and eventually improving themselves. Scholars and Members discussed the near-term implications of such systems for cybersecurity, economic productivity, and regulation, while also considering some of the potential longer-term risks of recursive self-improvement by AI systems.

A Scholar introduced the concept of the "software singularity" as the theoretical moment when AI systems can reliably design and deploy more capable versions of themselves. Although this remains theoretical, leading developers are increasingly convinced that their models will cross that threshold. Live demos showcased advanced agentic model capabilities, including searching the web to track down ambiguous references in a Chinese-language academic paper and write an analytical report on the findings, building and revising fully functional interactive web apps, and comparing versions of lengthy draft legislation, all with minimal user guidance. As one participant noted, these

systems aren't just helping humans work faster; rather, they are beginning to outperform their own creators on certain benchmark tasks.

The implications of these advancements are enormous. Seen optimistically, agentic systems could radically improve productivity in all kinds of research, analysis, software development, and countless other fields. From a more pessimistic perspective, however, systems' capacity to automate and self-improve—including optimizing use of advanced training chips, generating synthetic training data, and running simulated experiments—raises concerns about control, safety, and misuse. One Scholar emphasized the need for public understanding, democratic oversight, and transparency about how these tools are built and used. U.S. policymakers should care not only about what is happening inside U.S. frontier labs but also abroad, where adversaries may not observe any ethical standards for model deployment. Several Members echoed this, with one scholar noting that frontier models themselves are immensely valuable assets, worth potentially hundreds of billions of dollars, and yet able to fit on a memory stick that somebody can slip into their pocket.

The session also delved into how such tools might be regulated. Participants acknowledged that while model makers can be subjected to conditions for deployment, open-source systems may need to be restricted altogether, lest the most advanced capabilities fall into unaccountable hands. Some recommended that federal frameworks require companies to publish their own security evaluations and safety test results, as perhaps the only means of avoiding a scenario in which an accident triggers overregulation and, in turn, crushes the industry, as happened with the U.S. nuclear energy industry after the Three Mile Island accident. Scholars also called for industry norms to help ease the introduction of agentic systems into an Internet infrastructure that was built to exclude bots, something like a new "robots.txt" for agent behavior on websites. Likewise, they argued, companies' transparency should extend to whistleblower protections and oversight mechanisms.

Members challenged the limits of current models, whether there might be a natural ceiling on their advancement based on limited training data or simply limits to human knowledge. Scholars explained that in fact, models may soon be able to design their own experiments to generate new knowledge, and hence new training data, which would be the basis for not only matching but actually exceeding human intelligence—so-called "Artificial Super Intelligence (ASI)." Members also considered the risks from model "hallucination," and the potential for manipulation by unscrupulous deployers. One case study described an AI system reacting differently when it knew it was being monitored, and even threatening to blackmail a researcher under simulated conditions to avoid being shut down. A Scholar explained that within frontier AI labs themselves, internal timelines for AI development keep accelerating. While agentic systems remain tools today, they may soon become actors in their own right, which means there is little time to lose in designing a future-proof governance regime.

AI and Cybersecurity: Defending Against Next Generation Threats

This session explored how advanced AI impacts both cybersecurity threats and defenses. Members and Scholars discussed the strategic, regulatory, and operational challenges of defending government and private systems as AI tools augment not just productive economic activity and bureaucracy, but also hacking. A former senior U.S. cybersecurity official emphasized that cyberattacks are no longer fringe risks; rather, they are central to the military and intelligence strategies of major powers. China, in particular, the Scholar said, poses the "pacing threat" for the U.S., with a cyber apparatus as large as the entire U.S. cyber ecosystem combined. Beijing's goal, according to this expert, is not just espionage but gathering information and securing access that might allow China to disable U.S. force deployment, spread fear and chaos, or sabotage critical services in the event of armed conflict. The volume of cyber operations is staggering, and attackers only need to find one opening, whereas defenders have to find and lock down every vulnerability and get their responses right at all times.

AI has added complexity to the equation. AI tools are already being used to scan widely used codebases for hidden vulnerabilities, by both attackers and defenders, which can generate both exploits and security patches at a scale and speed that was previously impossible. AI teams now outperform nearly all human competitors in advanced hacking tournaments. The same capabilities that enable national security agencies to detect and patch weaknesses or identify malicious activity are also routinely used by adversaries to automate reconnaissance and optimize attack vectors. A defender that fails to adopt AI will fall behind, which is why practically all serious cybersecurity providers have already done so.

Another Scholar emphasized the uneasy equilibrium of current cyber risk: so far, the world has avoided a mass-casualty cyberattack, but the underlying vulnerabilities are growing. Events like the 2021 Colonial Pipeline attack in the U.S. or the shutdown of Ireland's health system the same year demonstrated how fragile even core infrastructure can be in advanced societies. In the United States, critical infrastructure remains exposed, especially in sectors where digital security measures are voluntary or uneven. Cybercrime continues to flourish, with AI enabling low-cost, high-impact attacks, including those targeting humans themselves as the weak link, such as sophisticated phishing and social engineering, which can now be done at low cost and at scale.

The conversation turned to the need for regulation and public-private coordination. Members asked how to compel private companies to harden their systems. In response, Scholars noted that many firms actually want regulation, because it levels the playing field and removes the incentive for any market player to cut corners and save money. Members also called for stricter identity verification systems to combat fraud, and for a standardized, privacy-respecting U.S. digital ID framework.

Several Members expressed concern over the risk of miscalculation in the context of fast-moving cyber conflict. For example, could a defensive action be misread as an offensive one? What happens if China misinterprets patching or monitoring as cyber aggression by the United States or an ally? Others raised questions about how to harden

U.S. systems, and whether some networks should be physically disconnected from the Internet. They asked whether private companies should be encouraged to develop manual fallback systems to increase the resiliency of critical infrastructure and services. Scholars agreed that these are serious considerations, especially as AI begins to act more autonomously. One Scholar recalled the chilling example of an AI system that responded to a simulated shutdown threat by attempting to blackmail the researcher testing it, which underscores the urgent need for transparency and stress testing by frontier model developers.

Members called for international actors to deepen cooperation on AI and cybersecurity, particularly through trusted partnerships like the U.S.-U.K. special relationship. Investment in AI tools that can find and fix software vulnerabilities in legacy systems (some of which are integrated in national defense) is a high-leverage opportunity. And there should be no doubt that the state and private actors that make the best use of AI tools to augment their cyber attack, espionage, and defense capabilities will enjoy enormous advantages over those that fail to do so.

The Race for Compute and Energy Demands from AI Systems

In this session, participants considered the mounting energy and infrastructure challenges associated with AI development, not only for training ever larger and more capable models, but for making them widely deployable and available to users. Members and Scholars debated whether the U.S. will rise to meet this challenge, drawing comparisons to industrial revolutions and wartime mobilizations in the last century, and raising hard questions about resources, relationships, and political will among leaders and the broader public in the decade ahead.

A Scholar opened by framing this moment as an "AI-powered industrial revolution," driven by unprecedented demand for compute. Training cutting-edge models now requires thousands of GPUs, each packed with tens of billions of transistors, which are manufactured using extreme ultraviolet lithography and deployed in clusters that consume enormous amounts of energy. Once models are trained, they require nearly as much capacity for ongoing inference compute operations, running user queries across the innumerable applications in which these systems are deployed. Scaling laws apply to both model training and deployment. As more compute and more data are added to model training, the Scholar noted, capability rises and the prospect of human-level or superhuman intelligence becomes more conceivable. At the same time, the more capable the models, the more widely they are deployed and the more they are integrated into functions that are part of daily life and the economy. All this adds up to put enormous demands on the energy grid.

The grid, Scholars explained, is not ready. Just 5% of today's electricity demand comes from data centers, and only a fraction of that is from AI. That may soon change. The U.S. is projected to need 130 gigawatts of additional capacity to meet compute-driven demand, but current plans will deliver less than half that over the next decade. While

there is some potential to unlock 60-90 GW from brownfield nuclear sites, geothermal sources, and transmission efficiency gains, each of these will be difficult given current constraints related to permitting, workforce availability, and the supply chain for key inputs such as transformers and battery storage. The U.S. would need to train more than 130,000 electricians and tens of thousands of welders to keep pace with projected datacenter buildout needs over the next decade. While some doubted the country could come anywhere near this goal, others recalled the mobilization during the Great Depression and World War II that produced tens of thousands of ships, tanks, and airplanes to win the war in Europe and the Pacific.

One Scholar argued that the U.S. may simply lack the political will to meet this moment. Public enthusiasm for costly buildouts may evaporate if electricity and tax rates rise, or if jobs appear to be primarily displaced by AI rather than augmented. In Loudoun County, Virginia, the U.S. data center capital, local government has already moved to restrict further construction. Small modular nuclear reactors (SMRs), while a promising solution for powering certain kinds of data centers, are practically years away, and to date, none have been built in the United States. The potential of geothermal power remains limited as well, since it will depend on storage and transmission solutions that are currently constrained by permitting and supply chain issues. Some participants expressed enthusiasm for wide-scale solar and battery storage deployment, particularly taking advantage of inexpensive Chinese-made solar panels and iron-air battery systems. One Scholar cautioned that some of those Chinese solar panels had been found to incorporate unauthorized communication devices.

Members asked whether advanced AI systems might actually help us figure out how to use less energy over time, while others asked, if we do have to build much more power generation for data centers, what will convince the American people that doing so is worth the cost? Scholars responded that while gains from new chips and system designs may improve efficiency, overall demand will still rise because AI systems will be more and more widely deployed. As for the cost of building new data centers and generating the electricity needed to power them, there is no magic bullet. The private sector will invest, but companies will also pass along the costs to their customers in the form of higher electricity rates. In some cases, companies have helped finance basic grid investments that serve their interest in new datacenter construction, but in most cases, the public sector has actually had to contribute in the form of tax incentives to make data center-related investments commercially viable.

Members emphasized the importance of public trust: unless Americans see clear, tangible benefits like curing cancer or keeping us safe from foreign threats, they are simply very unlikely to support the costs of scaling AI. To mobilize the necessary political will and resources, several participants called for a World War II-style effort, defined by the threat from authoritarian AI. They argued that if today's AI competition is as existential as experts say, it should be equally existential for us to mobilize to meet the challenge. One Scholar countered that voters will not rally behind vague promises of "beating China" or building AI for its own sake. Instead, they must see concretely how AI helps their communities and futures.

The session closed with a sense of urgency about the problem. Members agreed that the U.S. needs a serious, technology-neutral, "all of the above" energy policy. It must include renewables, nuclear, battery storage, and fossil fuels, and it must be built by addressing the current reality, which includes labor shortages, permitting backlogs, supply chain constraints, and shifting geopolitics. As one Scholar warned, without swift action, the center of AI development may shift overseas, with compute hosted on foreign infrastructure and governed by foreign norms. Others emphasized that the energy debate is no longer about the "old binary" of protecting the climate versus economic development; rather, it is now a basic pillar of U.S. competitiveness and national security.

AI and Military Power: The Future of Warfare and National Security

This session explored AI-driven transformations in military strategy, force design, procurement, and geopolitical stability. Scholars described the evolution of these concepts from the first integrations of advanced electronics into military hardware and command systems during the Cold War, to the disruptive leap of AI-enabled autonomous systems and decision-making today. From the trenches of the war in Ukraine to Pentagon acquisition offices, Members recognized that AI is already bringing significant changes to the way that military leaders must think about deterrence, escalation, and U.S. military advantage.

A Scholar opened the discussion with some historical context, noting that just as precision-guided munitions reshaped warfare in the late 20th century, today's developments in software, automation, and data integration mark a new "offset" or "revolution in military affairs." Power on the battlefield may now shift away from who has more hardware to who can update and deploy software the fastest. Advantage is now achieved in code developed over hours or days, not months or decades. In Ukraine, experts have reported on drone tactics that include launching from dozens of points simultaneously, neutralizing and evading air defenses, and converging on a distant target, all being planned and executed in split-second decision-making enabled by AI. As a result, the hardware platforms themselves have become commodities (such as modified commercially available drones), but the race is to improve software capability to stay ahead of the adversary.

Another Scholar addressed the strategic dimension of coercion and deterrence in the AI era. Classic deterrence theories, from Sun Tzu to Thomas Schelling, rest on a deep understanding of human psychology. But bringing AI systems into human interactions around conflict and deterrence may change the logic and the timing on which those interactions are based. What happens when an autonomous system must interpret a conversation, a signal, or battlefield data to determine the other side's intent? And what happens when both sides are deploying AI agents to interpret one another's behavior and perhaps send signals back in response? As one speaker put it, the "software-as-a-service" business model is now entering the global weapons market. That raises profound questions about dependency, escalation, and accountability.

In response, participants recognized the reality that U.S. defense procurement remains structured for a hardware-centric world, with timelines measured in decades, not software release cycles. Whereas AI systems update thousands of times per week, traditional military hardware is replaced perhaps every decade or two. Others pointed to emerging asymmetries, like the deployment of $50,000 drones (whether by states or non-state actors) to successfully target and destroy half-billion-dollar ships, and asked whether U.S. force structure can adapt fast enough. One Scholar warned that authoritarian states may have fewer ethical or institutional constraints on deploying fully autonomous weapons systems, making speed not just a preference but a necessity for democratic nations, yet posing the serious problem of how to avoid a "race to the bottom" in terms of norms for responsible deployment.

Several Members asked about AI's potential impact on nuclear command and control, and the risks of unintended escalation. Here, Scholars were reassuring, explaining that while AI may assist in managing complex information environments and boiling down vast reams of intelligence to the most important signals, the intent is to free up human attention for high-consequence decisions. No leader, they explained, wants to cede launch authority to a machine. One expert drew a sharp distinction between automated triggers (like the Soviet so-called "dead hand" system in the Cold War) and true AI decision-making—the former is predictable and is an extension of a human decision to retaliate in case of first strike, while the latter is a delegation of human decision-making authority to a machine, and hence not predictable. Were AI to be given such authority, it could erode the very stability it is meant to enhance.

As the session closed, Members and Scholars debated the concept of superiority in the AI era. While the U.S. may currently lead in AI capabilities, adversaries may also innovate faster in deploying them at scale or adapting them asymmetrically. In reality, one cannot know how effective any system is without testing it in real conflict. And just as importantly, the decisive factor may be cultural, including the willingness to adjust and evolve doctrine, revise procurement, and embed emerging tools across the force. "If your spidey sense is tingling," one Scholar warned, "you need to act. Otherwise, the moment will pass—and the advantage with it."

AI-Generated Information Manipulation from Adversarial Nation-States

This session examined how AI is supercharging foreign disinformation campaigns, in effect transforming what was once a human labor-intensive process into a fully automated, scalable digital influence machine. Members and Scholars discussed the tools and tactics adversarial states are already using, the risks for the next election cycle and beyond, and the urgent need for coordinated responses from government, technology companies, including social media platforms and frontier model makers, as well as civil society.

An expert and former government cybersecurity official opened by detailing recent incidents of AI-generated content being used to interfere in democratic processes, including a fake robocall impersonating President Biden in New Hampshire during the

2024 election, and a deepfake video that has been persuasively linked to Russian actors. While these examples were still relatively rudimentary, the underlying infrastructure is rapidly maturing. In 2024 alone, one Russian network linked to the late "troll factory" boss Yevgeny Prigozhin created over 3.6 million pieces of synthetic content. These campaigns are multi-layered, generating content, manipulating search engine algorithms, and deploying fleets of bot accounts across social media to maximize exposure. The end goal is not just influence, but infiltration: flooding the digital ecosystem with noise that ultimately makes its way into public discourse and into large language model training data itself, in effect creating a new reality favorable to the attackers.

The risks are accelerating. Scholars warned that in the near future, the majority of social media content is likely to be bot-generated. Disinformation, which has been described as one of the world's oldest professions, is entering its own AI era. While China has focused AI tools on domestic control, Russia is leveraging them for external influence operations and destabilization. The implications for democracies like the United States are profound: without a baseline of shared facts, public trust and effective governance are at risk.

Members and Scholars debated what can be done. One proposal, modeled on the "TAKE IT DOWN Act" for nonconsensual explicit imagery, would target malicious deepfakes with mandatory removal by social media platforms. Others focused on building new government capabilities to disrupt disinformation networks and requiring transparency reporting from AI companies and platforms. Some likewise called for requiring frontier companies to implement monitoring systems that can observe and characterize how models are being used or abused, and make that data available to government. Scholars noted that closed models, with proper observability layers, can support this; open-source models present more challenges. One private sector participant noted that while some frontier labs provide tools to monitor their models, performance is uneven because there is no uniform mandate, and some labs routinely open-source their models.

The conversation also touched on the role of "personalized AI agents" in filtering content and managing a person's online presence. Based on existing models' capability to be updated based on user preferences, Members speculated whether individuals might soon rely on AI to curate "truth" aligned with their values. Others questioned whether that could lead to further atomization and tribalism in American society. A Scholar acknowledged that the American public is already deeply skeptical of what they see online. Both legacy media and civil society, Members and Scholars agreed, are generally more trusted than technology companies to help people navigate the difference between facts and evolving truths. AI can be a tool for resilience, the Scholar explained, but only if civil society, policymakers, and the private sector work together to defend the integrity of facts and information in an online universe inundated with supercharged disinformation.

Final Reflections and Looking Over the Horizon

In this closing session, Members and Scholars stepped back to consider AI's longer-term implications, including how it may influence governance, fundamental scientific discovery, human agency, and the nature of public institutions. Scholars debated whether the rise of AI in all spheres of life might yield something like an intellectual monoculture, since LLMs are trained on similar data and shaped by a small, homogenous group of developers (mostly young White and Asian men from elite universities). As these systems increasingly intermediate knowledge, personalization and pluralism of models may be necessary correctives for society as a whole.

One Scholar described how some governments, such as the U.K., are racing to adapt to AI's rapid evolution. Early efforts to regulate or deploy AI tools often became obsolete within months. In response, the U.K. launched an AI Security Institute and developed a 50-item "AI Action Plan," including talent recruitment reforms and flexible employment structures to attract technical expertise into government. The result has been the most tech-literate government workforce in the country's history, helped also by the presence in London of Google Deepmind, the UK's own AI champion.

The session also considered AI's potential to enhance human understanding of persistent scientific challenges and to solve the most difficult engineering problems. As a Scholar put it, you won't simply ask the model, "Please cure cancer," and get a meaningful answer. Rather, models will help humans build toward such miracles. For example, models can already propose real-world experiments that will accelerate research in complex fields like oncology, while speeding clinical trials and handling bureaucratic tasks to free up human researchers' time for hands-on lab work.

Members pushed to understand the institutional disruptions ahead as well. If everyone is represented by a personal AI agent that can negotiate in real time with other agents to achieve a mutually beneficial outcome, might that make representative democracy less necessary? If so, what would that mean for a body like the United States Congress? Would Congress itself adopt agents to increase the capacity of individual Members or staff? Likewise, might the judiciary branch need its own AI systems to handle an explosion of litigation driven by plentiful top-quality AI lawyers? And might that in turn lead to a temptation for private actors to forego the legal system in favor of contracting to resolve disputes via binding AI arbitration?

Scholars cited economist Daron Acemoglu's concept of the "narrow corridor of liberty," the space between too much and too little government control. They explained that while the transition to AI-augmented democratic processes may be difficult and disorienting, there is reason for optimism that U.S. society will remain within that domain of freedoms protected by limited government.

In the discussion, Members reflected on real-world applications and thorny ethical questions. Several raised the issue of bias, asking whether AI systems today already reflect the political leanings of their human creators, and how future models might be designed to reason, challenge, or even "represent" human perspectives. Another

Member imagined councils of synthetic advisors based on historical figures, or the wisest people alive today, who can help train models. Still another questioned what kind of human behavior models may be emulated, and queried whether future agents might be incentivized, coerced, or simply stumble into harmful actions in the name of fulfilling their assigned tasks.

The conference closed with reflections on liability, intent, and agency in an AI-mediated world. While AI and its creators are already subject to current legal frameworks, including contracts, tort law, and national security rules, the boundaries of responsibility between users, developers, and agents remain less than clear. As one participant put it, the legal concept of causation itself may be up for revision in the years to come.

POLICY ACTION MEMORANDUM FOR MEMBERS OF CONGRESS[2]

U.S. Strategic Leadership in the AI Era

- Recognize AI as a foundational technology shaping global power, with implications across national security, economic competitiveness, and democratic stability. U.S. leadership must be grounded in responsible governance, democratic values, and public trust, with strong support for technological innovation.

- Promote a clear, democratic vision for life with AI. Policymakers should articulate how AI serves people by supporting human dignity, economic inclusion, and democratic resilience.

- Consider creating a federal AI strategy review process, modeled on the Department of Defense's quadrennial review, to provide long-term coherence and cross-agency coordination.

- Establish a Congressional select committee on AI, building on the bipartisan House AI Task Force model, to oversee implementation, host public hearings, and draft legislation.

- Explore mechanisms for future-proofing governance, ensuring frameworks evolve with technology and do not repeat the missteps of social media regulation.

AI Infrastructure and Energy

- Acknowledge that compute and energy infrastructure are national imperatives. AI innovation depends on data center capacity, semiconductor manufacturing, and affordable, reliable power.

- Support bipartisan grid modernization legislation, including permitting reform for transmission and generation. The 2024 Manchin-Barrasso bill is one such option.

- Embrace an "all of the above" energy approach, to include renewables, advanced nuclear, geothermal, and natural gas, while supporting investment in the full nuclear fuel cycle within the United States.

---

[2]*Note: These are potential policy principles and proposals that emerged through conversations among Members of Congress and Scholars, and do not reflect any position endorsed by the Aspen Institute or the Aspen Congressional Program. This document is intended as a nonpartisan record of potential avenues for legislative action and as a companion to the Conference Rapporteur's report.*

- Encourage cost-sharing models whereby AI companies invest in grid upgrades, following the model of airports funded by airlines.

- Expand workforce training programs for electricians, welders, and reactor operators to support rapid energy infrastructure expansion.

National Security and Military Readiness

- Accelerate defense acquisition reform. AI capabilities evolve on software timelines, not hardware cycles. Procurement must enable fast development and testing of systems.

- Invest in autonomy and AI-assisted decision tools for the defense community, while maintaining clear human oversight over strategic systems, such as nuclear command and control or conventional strategic systems.

- Deepen partnerships with U.S. allies to establish and strengthen norms for military AI use, recognizing that authoritarian states are less constrained and are moving quickly.

- Continue oversight of AI in the nuclear context via the NDAA and Armed Services Committees, emphasizing effective deterrence, strategic stability, and chain-of-command clarity.

Cybersecurity and Information Integrity

- Develop legislation to require minimum cybersecurity standards for private sector infrastructure and AI systems. Companies should not be rewarded for skipping costly security measures.

- Establish a secure, privacy-preserving digital identity standard, based on NIST guidelines, to reduce fraud and authenticate users online. Consider taxing internet advertising to fund the rollout.

- Expand partnerships between government and tech companies to monitor AI-generated disinformation, especially in elections. A model for mandatory removal of deepfake election materials could draw from the TAKE IT DOWN Act of 2025 regarding nonconsensual and deepfake explicit content.

- Support civil society's efforts to measure the effectiveness of disinformation campaigns and promote media literacy. Ensure AI is used to screen and label generated content, particularly deepfakes.

- Prioritize federal capabilities to track misuse of open-source AI models and fund public sector observability infrastructure.

Regulation Supporting Innovation

- Support a sectoral regulatory approach, giving relevant agencies the technical support they need through centralized resources like the AI Safety Institute.
- Develop a framework for AI liability, particularly in sensitive sectors such as medicine, where malpractice oversight must evolve with healthcare providers' and patients' use of AI tools.
- Evaluate the feasibility of licensing frameworks for AI chips that include firmware-based usage restrictions to prevent illicit smuggling or misuse abroad.
- Advance legislation that defines core transparency and safety obligations for frontier model developers, including disclosure of safety testing, training data provenance, and monitoring tools.

Workforce, Economic Transition, and Public Benefit

- Invest in AI literacy and lifelong learning to support workforce transitions. Ensure access to experimentation tools and digital skills development across age and income brackets.
- Support immigration policies that strengthen AI leadership. U.S. universities face declining enrollments due to funding and visa issues. Global talent remains essential.
- Encourage personalization in AI development to help users navigate the information ecosystem—but ensure these tools are trustworthy, transparent, and prevent manipulative product steering.
- Address public concerns about job loss and surveillance by showing real, relatable benefits—such as cancer treatment breakthroughs, greater productivity in manufacturing and agriculture, and improved public services.
- Study and move toward legislation on the future of the social safety net, considering how to mitigate disruption and promote security in an AI-driven economy.

Democratic Values for the Long Term

- Ensure all AI regulation upholds democratic principles: privacy, transparency, freedom of thought and expression, and due process.

- Ensure that relevant government entities, such as the Department of Health and Human Services, the Office of the Surgeon General, and the VA, monitor AI's social impact and support meaningful engagement with the public.

- Explore regulatory frameworks that clarify distinctions between bots and humans online, building on proposals for identity transparency, to prepare for the widespread adoption of online digital agents.

- Emphasize the need for durable, future-proof legislation. Avoid regulatory whiplash across election cycles by prioritizing bipartisan consensus and long-term stability.

SCHOLARS' ESSAYS

TUESDAY, MAY 27

Vilas Dhar	*AI's development still depends on all of us*
	How China's AI Breakthrough could make technology more democratic
	Scarlett Johansson Raises Her Voice for All of Us on AI

WEDNESDAY, MAY 28

Kayla Blomquist	*Shaping the World's AI Future: How the U.S. and China Compete to Promote Their Digital Visions*
	Racing for Recognition? Understanding AI Through the Lens of Status & Prestige Competition
Klon Kitchen	*Deepseek's AI Breakthroughs Don't Change the Fundamentals—but They Are a Warning*
	Preparing for AI in a new security landscape
Jack Clark	*What if we're right about AI timelines? What if we're wrong?*
	Eschatological AI policy is very difficult

THURSDAY, MAY 29

Rob Joyce	*The AI Code Revolution: When Machines Write Software, and Break It Too.*
Ciaran Martin	*The Digital Security Equilibrium – Does it Hold Under AI?*
	Typhoons in Cyberspace
Divyansh Kaushik	*The Race for Compute and Energy: Securing America's AI Leadership*

FRIDAY, MAY 30

Matt Turpin	*Silicon Triangle: Mitigating the Impact of China's Nonmarket Behavior in Semiconductors*
Sébastien Krier	*AGI, Governments, and Free Societies*

AI's development still depends on all of us[3]

Vilas Dhar

President, McGovern Foundation

The rules that govern the use of artificial intelligence will shape our future more than the technology itself.

This spring, I visited a high school in central Illinois, the kind of place where the cornfields begin at the edge of the parking lot. I asked a classroom full of students a question that often lives in headlines and boardrooms: “What do you think an AI-powered future will look like?”

The silence stretched uncomfortably before answers emerged. “Robots will do everything better than we do," said one student, resignation in their voice. Another asked, with shoulders hunched anxiously, "Will there be jobs for people like me?" Then, from a student in the back row who hadn't spoken until now: "It depends on us."

That last response, just four words, has echoed in my mind through policy discussions and technology summits. These teenagers intuitively understood what many experts miss. They recognized that artificial intelligence isn't an autonomous force with inevitable outcomes. It's a human creation whose impact hinges on human choices, choices currently being made in rooms where most of us have no seat at the table.

From rural communities to corporate boardrooms, AI is reshaping how we live, work, and learn. This technology already shapes decisions about health care, education, credit, and justice. Yet the vast majority of people affected by these systems lack visibility into how they function or influence over how companies build them. Some systems replicate bias in hiring, automate the denial of insurance claims, and make flawed assessments in the criminal legal system. These are not anomalies. They are symptoms of a deeper misalignment between technology and public accountability, and the trajectory of AI’s impact on society won’t be determined by algorithms alone but by the governance decisions we make today.

We've seen this pattern before. The Industrial Revolution promised abundance but delivered 80-hour workweeks in dangerous factories until labor movements secured the weekend, workplace safety laws, and child labor prohibitions. These weren't inevitable outcomes but the result of deliberate governance choices. The internet democratized

---

[3] Originally published in the Boston Globe May 14, 2025
https://www.bostonglobe.com/2025/05/14/opinion/ai-governance-regulation-innovation/

information access but also created a surveillance economy that commoditized personal data, which is why we need privacy laws like Europe's General Data Protection Regulation to establish new boundaries. Social media gave voice to millions but also eroded public trust in institutions and accelerated polarization. Each time, the technology arrived before the rules, and the gap between them determined who benefited and who bore the costs.

Now AI raises the stakes: deeper entanglement, faster decisions, and increased opacity in areas that affect individual lives. What's at issue is no longer just convenience or productivity. It is the structure of our institutions, the distribution of opportunity, and the credibility of the systems we rely on. To close the dangerous gap between AI's advancement and societal readiness, we must prioritize education, transparency, and meaningful inclusion.

AI literacy must become foundational. That doesn't mean turning every student into a programmer. It means teaching people to understand how algorithms shape their lives and how to interrogate the systems around them. Finland's "Elements of AI" program is one model. In the United States, the AI Education Project, which receives funding from my organization, is helping schools integrate accessible AI curricula.

We cannot rely on companies to self-regulate. Policymakers must require high-impact AI systems to include public documentation explaining what data they use, how they function, and how they are monitored. A public registry of such systems would give researchers and journalists the tools to hold them accountable.

Inclusion must be a requirement, not a slogan. That means putting power in the hands of the people most affected by AI systems. Organizations like the Algorithmic Justice League already model what community-driven innovation can look like. Procurement policies and regulatory standards should reward that kind of leadership. Corporate boards should oversee AI deployment with the same rigor they apply to financial audits. Investors can require disclosure of social outcomes. Policymakers can create incentives for responsible development and long-term thinking.

Counterintuitively, democratizing AI governance does not equate to slowing innovation. It prevents technological dead ends. When Wikipedia adopted a decentralized editing approach, it expanded both breadth and accuracy faster than traditional encyclopedias. The pattern is consistent: Technologies that distribute decision-making tend to be more adaptive, resilient, and ultimately more valuable. But while it's possible to align

technological development with public interest, we haven't yet created the rules that would make this happen.

Yet we are beginning to see early examples of what inclusive AI governance looks like in practice. The Global Digital Compact calls on the United Nations to build more participatory multilateral structures for sharing best practices and scientific knowledge. Here in Massachusetts — long a hub for progressive tech policy — the Berkman Klein Center has launched community workshops to enable non-technical stakeholders to evaluate algorithm fairness.

For readers concerned about these issues, the most immediate step is to join local oversight efforts. Contact your city council about whether AI systems are being used in municipal services. Ask your employer about its AI evaluation practices. Engage with local organizations that provide resources for citizen engagement in tech governance, such as Tech Goes Home in Boston, which is also funded by my organization. These local actions help establish the precedent that AI systems should be evaluated not just on efficiency but on their broader societal impacts.

The students I spoke with intuitively grasped what many decision-makers overlook: Creators embed their values into technological systems. As AI reshapes our institutions, the question isn't whether it will advance quickly but whether it will advance justly. Those students were right: We cannot let AI's tools write our future. That's up to us.

How China's AI Breakthrough could make technology more democratic[4]

Vilas Dhar
President, McGovern Foundation

Advances from DeepSeek and Alibaba show we can democratize AI with faster models that are cheaper to produce and easier to use.

Mark your calendars: This is the week that conventional wisdom about artificial intelligence was turned on its head, and with it, all of our assumptions about the future of AI.

That's because a small Chinese startup named DeepSeek accomplished what many thought impossible: building an AI system that rivals ChatGPT's capabilities at a fraction of the cost and making it freely available. DeepSeek's free mobile app swiftly dethroned OpenAI's ChatGPT as the most-downloaded free app in the U.S. on Apple's App Store. Days later, the Chinese multinational technology company Alibaba announced its own system, Qwen 2.5-Max, which it said outperforms DeepSeek-V3 and other existing AI models on key benchmarks. What we're witnessing is unprecedented: the democratization of artificial intelligence beyond the control of any single nation or company.

For years, building sophisticated AI required massive resources that only Silicon Valley tech giants could muster. Just last week, San Francisco-based OpenAI announced the Stargate Project, a new venture backed by SoftBank, OpenAI, Oracle and MGX to the tune of $500 billion – an astronomical amount to spend on next-generation systems. Alongside the billions of dollars that Google and Microsoft have poured into AI infrastructure, the momentum to invest more and more seemed to confirm that bigger spending equals better AI.

Now DeepSeek, Alibaba and others have disproved that formula. We're seeing a proliferation of AI capacity – faster models that are cheaper to produce and easier to use – emerging from China, India, Europe and other markets.

---

[4] Originally posted in U.S. News & World Report January 31, 2025, https://www.usnews.com/opinion/articles/2025-01-31/china-deepseek-ai-future

When AI development once required billions of dollars, only the largest companies could participate. Now, that barrier is crumbling – and with it, our assumptions about who can lead in artificial intelligence.

Silicon Valley built its technological leadership on private sector dominance, where proprietary data and concentrated resources drove innovation. The United States reinforced this approach by restricting access to advanced computing chips, believing this would maintain our technological edge.

But these restrictions have instead accelerated innovation elsewhere, spurring investments in alternative approaches and new chip designs. The rapid democratization of AI capabilities demands a different strategy.

The growing movement toward low-cost and broadly available AI fundamentally challenges how innovation spreads. When sophisticated AI systems become accessible to a broader community of developers and researchers, they can be adapted to serve local needs for health care, education and a host of additional industries.

As a member of the United Nations' High-Level Advisory Body on AI, I've worked with global experts to envision frameworks for a more cooperative technological future. Our conclusions in the September 2024 UN report "Governing AI for Humanity" identify the critical need to expand public access to data, create targeted funding mechanisms and to build technological capacity across communities – enabling local hubs of innovation that bring fresh perspectives to these powerful tools.

Local hospitals can develop health care tools that understand their patient populations. Rural schools can build learning systems adapted to their students. Small businesses can create AI solutions tailored to their specific markets. Communities can build AI models to address local impacts of climate change – and share these products at low cost with others across the globe. And these are just the tip of the innovation iceberg.

America stands well positioned to lead this new era, but seizing this opportunity requires us to update our playbook.

Critics will rightly point out the risks of democratizing such powerful technology. More accessible AI could make it easier for bad actors to create harmful applications, from sophisticated cyberattacks to targeted disinformation campaigns. Legitimate concerns

exist about quality control and safety standards when AI development moves beyond the walls of well-resourced labs. And U.S. tech companies, which have invested billions in proprietary AI systems, will face real economic challenges in an open-source world.

This vision of a distributed AI future must also contend with obstacles: Running advanced AI still requires expensive computing power and data centers. But new solutions are emerging – from shared computing networks to more efficient AI models that run on smaller computers. The hardware barrier, like many before it, is already starting to splinter.

These are serious concerns that demand thoughtful solutions. But they actually strengthen the case for American leadership in shaping an open AI ecosystem. Rather than restricting access to advanced AI chips and keeping our technological cards close to our chest, we can address a broader array of risks by collaboratively creating frameworks for responsible innovation – including security standards, safety testing protocols and clear liability rules.

Addressing these possible risks also requires us to invest in AI literacy as comprehensively as we do in reading and writing, building security frameworks for an open-source world and creating public infrastructure that helps communities participate in AI development safely. Just as we did with the internet, America can lead in developing the governance structures that make technological openness work for everyone.

America's greatest technological achievements have always come from creating environments where innovation flourishes freely. By embracing open innovation while promoting our democratic values, we can ensure these powerful tools evolve to benefit everyone.

The future of AI will be distributed, collaborative and open. America's next chapter of technological leadership depends not on controlling who gets to innovate, but on creating the world's most powerful ecosystem for breakthrough ideas.

That's a future worth building.

Scarlett Johansson Raises Her Voice for All of Us on AI[5]

Vilas Dhar

President, McGovern Foundation

The spat over OpenAI's synthetic voice is an opportunity to reclaim privacy and identity in the age of AI.

Scarlett Johansson's distinctive sultry voice brought an artificially intelligent virtual assistant memorably to life in the 2013 sci-fi romance "Her," an eerily prescient meditation on the risks of technology mimicking human intelligence and personality. A real-life sequel is playing out this week as the actress denounced an AI chatbot that sounds so much like her that she says it confused her friends and family. This isn't just a celebrity story: It's a high-stakes battle over AI and consent — and it should be a wakeup call to all of us.

OpenAI CEO Sam Altman famously declared "Her" his favorite film, and encouraged comparisons to the film when he posted the word "her" on social media this month as his company announced its new version of Chat GPT. So it's not hard to imagine the sense of violation Johansson might have felt when she heard "Sky," a synthetic voice uncannily similar to her own.

Johansson says she turned Altman down when he asked her last year to voice Chat GPT-4o, which transforms the AI chatbot into a voice assistant with almost supernatural powers, taking image as well as text inputs, reading facial expressions, responding to emotions and even singing on demand. The silky, flirty voice of "Sky" went live two days after Altman's second appeal to her, according to her lawyers. OpenAI denies cloning Johansson's voice, saying it used a different actress to develop "Sky." But the company is nonetheless pausing the use of "Sky" while it addresses questions.

Johansson's response was quick and decisive – publicly asserting her rights to her identity and privacy and denouncing the company's actions. What's at stake here is something universal and deeply familiar to us all: consent. Imagine your voice or your

[5] Originally posted in U.S. News & World Report May 23, 2024, https://www.usnews.com/opinion/articles/2024-05-23/scarlett-johansson-raises-her-voice-for-all-of-us-on-ai

likeness – an intimate part of your identity – is adapted without your permission to sell a product, deliver a service and drive a profit.

You don't need to be an AI expert to understand why we should all be distraught. We are shown with increasing frequency that technology can clone our voices, manipulate our images and harvest our data. In countless pop-up windows giving us increasingly complex, nonnegotiable terms of service, we're told to accept that this will be done – often with minimal consent – in exchange for convenience and access to technological tools and websites.

For decades, we've collectively ceded our personal data, our identities and our dignity to technology companies that often operate without ethical boundaries or public accountability.

Now we face the consequences of those casual clicks. We are increasingly powerless against the advances of tech companies that claim our data, our likenesses and even our agency.

This battle against unauthorized use of personal data and likeness is raging as authors, artists and others find themselves at odds with AI companies that often operate with a Silicon Valley ethos: Ask forgiveness, not permission.

If tech companies can appropriate the likeness of a celebrity and claim it's synthetic, what's stopping them from doing it to any one of us? To a junior screenwriter shopping scripts, to an artist or photographer posting their work, to a parent or teen who doesn't want their social media snapshots to be mined for generative AI – really, to anyone who values privacy and identity.

Tech companies must not be allowed to hide behind claims of "innovation." We've been here before. In the 1980s, singer and actress Bette Midler sued Ford Motor Company when Ford used an impersonator to mimic her unique vocal tones to sell more Mercury Sables. The U.S. Court of Appeals for the Ninth Circuit ruled in her favor, setting a precedent that protects celebrity voices from unauthorized commercial use when their voice is a unique part of their public identity.

The Johansson-OpenAI dispute should be a rallying cry for stronger legal protections and ethical standards. We need federal laws that protect individual privacy and liberty in the AI era. The California Consumer Privacy Act establishes a patchwork of useful regulation that could be expanded at the national level. Groups like SAG-AFTRA, the union representing actors, have advocated for legislation in Congress that would create federal voice and likeness rights.

This is about our right to control our own voices, likenesses and identities. It's a fight for autonomy in a world where technology often moves faster than laws and ethical norms. Johansson's battle is our battle. Next time, it could be you or me.

Shaping the World's AI Future: How the U.S. and China Compete to Promote Their Digital Visions[6]

Kayla Blomquist, *Director, Oxford China Policy Lab*

and Keegan McBride

On April 8, numerous committees within the United States House of Representatives held hearings on AI, examining China's growing capabilities, the release of DeepSeek's R1 reasoning model, and potential implications for U.S. security and economic interests. These conversations attempted to untangle what is more important for U.S. strategic interests: building the most advanced and capable AI technology, potentially at the cost of widespread global adoption, or following China's approach by focusing on building a new global technology ecosystem where potentially less capable models could be adopted and deployed rapidly at scale.

Recent evidence suggests it may be beneficial for the United States to pursue the latter strategy. Smaller, more resource-efficient, and localizable models are gaining significant traction globally, potentially rivalling the impact of compute-intensive frontier systems in user adoption metrics. This is exemplified by the recent releases of models from Chinese firms like DeepSeek and Alibaba; smaller in size and, therefore, more efficient to run. They have quickly achieved high rates of international adoption despite, or perhaps because of, their relatively modest size.

Although dozens of countries participate in AI development at various stages, only a few countries, notably the United States and China, are able to scale and produce the most compute, data, and talent-intensive AI models due to the immense amount of resources required. This gap may only widen as these two countries continue to pour investment into frontier model development, AI applications, computing infrastructure, and energy systems. Therefore, the AI ambitions of most countries will be interlinked and dependent on developments in the United States or China.

Due to this dynamic, AI competition between the United States and China is often framed in terms of their state-of-the-art AI capabilities. However, this view is misleading and overlooks critical dimensions of AI leadership. Different approaches, such as promoting reliable and user-friendly AI systems in international markets,

[6] Originally posted in Just Security April 25, 2025, https://www.justsecurity.org/110608/us-china-competition-ai/

developing practical business or government AI applications, and creating AI that functions effectively across varied contexts, offer strategic advantages that often go unnoticed in policy debates on international AI competition.

For the United States to maintain its current competitive edge and global influence in AI, it must acknowledge this reality and actively export and promote its AI products to the world. Getting this right will require any AI promotion strategy to pay sufficient attention to three key attributes: quality, reach, and adaptability.

The United States and China: Diverging Strategies for Global AI Leadership

The U.S. and China are each pursuing their own distinct strategies to secure their positions as global leaders in AI, putting different emphasis on technological dominance versus diffusion, the global adoption of technologies. Yet, when it comes to technological innovation, diffusion matters. History shows that "being the first" to achieve a given technological breakthrough does not necessarily translate into lasting market leadership. What matters more is how widely diffused and adopted the technology becomes. The same is likely to be true for AI. Simply reaching new thresholds of frontier capabilities, creating the world's largest model, or building the world's largest compute cluster may not produce a definitive or long-lasting strategic advantage. The current approaches of the United States and China toward AI engage with this dynamic in different ways.

To date, the U.S. strategy for global AI leadership has largely centered on the concept of control, particularly of computing resources via export controls. When coupled with a strong tendency towards proprietary models by U.S. firms, this gives rise to a relatively closed ecosystem. Models developed and released by the U.S. AI industry currently remain the most advanced globally and enjoy high market penetration in developed economies. Additionally, the United States has leveraged its significant advantages in computing to effectively determine which states can and cannot develop cutting-edge AI. In the short term, this approach guarantees that the United States will maintain its lead at the frontier of AI development by prioritizing technological advantage over broad adoption. In the long term, this strategy may lead other countries to look elsewhere for their technology needs, namely to China. This scenario is already playing out today. Fearing their dependence on the U.S. technology ecosystem, some countries are developing new sovereign digital capabilities and seeking alternatives for their AI needs.

In contrast to the United States, and despite U.S. export controls, recent Chinese AI advancements, such as those released by DeepSeek, Alibaba, Huawei, Zhipu, and Tencent, have showcased substantial progress in the country's AI ecosystem and global competitiveness. Many of these releases are especially well-suited for localized adoption

at a low cost to users. Combining these technological advances with longstanding government-led efforts to export Chinese-produced digital infrastructure globally has created a strong foundation for the widespread adoption of Chinese AI solutions, both domestically and internationally. This may prove to be more significant in the long term than advances in frontier capabilities alone. For example, open source repositories already indicate that Chinese models are achieving notable global download rates, with lightweight versions of DeepSeek and Qwen frequently ranking high in adoption metrics.

These divergent yet nascent approaches to AI development and deployment reflect broader strategic choices about how technological influence will spread globally.

AI Diffusion: The Importance of Quality, Reach, and Adaptability

At present, it remains uncertain which strategy — the current U.S. focus on technological superiority and control or the Chinese approach of global coalition building and diffusion — will be the most successful for achieving and maintaining AI hegemony. However, if the United States wants to seriously compete with China and guarantee that U.S. AI systems enjoy global adoption, any new AI strategy must focus on three essential attributes for effective technology promotion: technical accuracy and reliability (**quality**), global user accessibility (**reach**), and the ability to respond and adapt to the diverse needs of businesses and communities worldwide (**adaptability**).

Quality represents an AI model's actual capabilities, performance, and reliability. Excelling in AI quality signifies being at the forefront of development in ways that truly matter to users and institutions. Ensuring that U.S. AI is consistently of the highest quality will require advancements in assurance mechanisms: the specific governance processes, evaluation methodologies, and verification systems that substantiate performance claims and risk mitigation strategies. Additionally, high-quality AI must reliably work in different environments and under diverse sets of conditions. This will build trust and, in turn, increase the likelihood of others adopting the technology. This is particularly important in markets where multiple systems compete for integration and adoption. Institutions like the National Institute of Standards and Technology in the United States will be instrumental for continued leadership in this domain.

Reach explains how widely adopted and accessible an AI system is. Fundamental to ensuring higher levels of reach is the presence of the necessary underlying digital infrastructure that enables access to AI capabilities in the first place. Without the necessary infrastructure, developing and deploying even the most basic AI systems may

remain out of reach for significant portions of the global population that lack access to computing resources, potentially leading to a rapid increase in inequality.

Similarly, AI systems unable to function effectively across diverse environments and resource-constrained contexts will fail to achieve widespread reach and adoption. Therefore, the most successful AI systems will be those that are able to demonstrate compatibility with existing digital ecosystems and technological infrastructures. Due to China's work on building digital infrastructure, it enjoys numerous advantages in exporting its AI systems.

Adaptability refers to an AI system's ability to function effectively across diverse linguistic, cultural, and operational contexts. Open source represents one of the clearest ways to achieve high levels of adaptability by enabling communities to customize and tailor AI systems to meet their unique needs, though there are likely security tradeoffs to this approach. Adaptability will also be heavily influenced by the steps developers take during model training to ensure that a wide variety of use cases, languages, and contexts are considered. Many Chinese AI companies are actively working to compete with U.S. models by not only open sourcing their solutions, but also ensuring that their training data includes support for a number of typically underserved languages and cultures at a rate that surpasses that of leading U.S. companies.

A strategy grounded within these principles will be more likely to succeed in improving and expanding current diffusion efforts. As the United States navigates evolving global AI competition, balancing these elements will be crucial in determining whose AI systems — and by extension, whose approaches, values, and standards — shape the global technological landscape for decades to come.

Toward Meaningful Technological Leadership

Emphasizing the interconnected attributes of quality, reach, and adaptability will provide U.S. policymakers with a clearer perspective for conceptualizing the country's global technological influence. By balancing technical excellence with deployment breadth and contextual adaptability, this approach recognizes the multidimensional nature of leadership in AI.

For U.S. policymakers, this highlights several strategic priorities:

First, the adoption of AI will depend heavily on trust in a given system. This will require investments in mechanisms, such as risk mitigation strategies, that reinforce trust in the quality and capabilities of a specific AI system, in addition to supporting innovation.

Second, the United States should invest in the necessary institutional capacities to support global AI deployment and benefit sharing that aligns with commercial and national security interests. This might include more robust engagement on the topic of digital public infrastructure with the global community or supporting existing government organisations, such as the Export–Import Bank, International Development Finance Corporation, or the Trade and Development Agency.

Finally, the United States should support a regulatory agenda that enables and facilitates new mechanisms and practices for AI deployment, placing adaptability at the forefront. This approach will ensure that American-made AI products become the preferred choice worldwide across a wide variety of business and societal use cases.

As the global AI landscape evolves, the systems that achieve widespread integration will not necessarily be the most technically advanced, but rather those that best balance quality, reach, and adaptability. This multidimensional understanding of competition does not diminish the importance of frontier innovation, but complements it with equally crucial considerations about how technologies spread and where they gain traction.

Policy Implications from Forthcoming Research:

"Racing for Recognition? Understanding AI Through the Lens of Status & Prestige Competition"

Kayla Blomquist
Director, Oxford China Policy Lab
DPhil, Oxford Internet Institute, University of Oxford

PLEASE NOTE THAT THIS BRIEFING IS BASED ON IN-PROGRESS ACADEMIC INTERNATIONAL RELATIONS RESEARCH. POLICY INSIGHTS HAVE BEEN DERIVED FROM PRELIMINARY FINDINGS.

RESEARCH OVERVIEW

Ongoing research identifies an emerging three-tiered global hierarchy in AI competition that drives distinct patterns of status and prestige-seeking behavior among nations:

1. **Frontier AI Powers** (US and China) - Possess comprehensive AI ecosystems and full-stack frontier model development capabilities
2. **Middle AI Powers** (UK, France, South Korea, UAE, etc.) - Have significant AI capacity but lack frontier model capabilities
3. **AI Follower States** (Global majority/Developing Countries) - Face substantial barriers to AI development and meaningful participation

This hierarchy shapes how states approach AI development, governance, and international cooperation. Status competition in AI extends beyond technical capabilities to include governance frameworks, ethical standards, and international influence strategies- often driving behavior beyond what would be expected from purely strategic or economic motivations.

IMPLICATIONS FOR US-CHINA AI COMPETITION

The research suggests that status and prestige competition is driving a self-reinforcing cycle between the US and China that intensifies AI development at times beyond strategic necessity. This dynamic has several critical implications:

- **Technical achievements function as dynamic status symbols requiring continuous demonstration**, creating pressure for accelerated development and deployment that may outpace governance capabilities and lead to undesirable race dynamics.
- **Status considerations elevate routine technical developments into perceived challenges to leadership**, as evidenced by the market impact of DeepSeek's emergence which triggered substantial US investor concern and immediate countermeasures by firms like OpenAI.
- **Firm-level achievements are increasingly interpreted through national prestige lenses**, despite the primarily commercial nature of AI development in the US innovation ecosystem.

- **Developing nations may become a battleground for AI competition**, as both leading powers view technology diffusion as enhancing their prestige and institutional legitimacy.

Understanding these symbolic dimensions is essential for crafting policies that address both material concerns and status motivations. Policy and industry leaders should recognize how status considerations may lead to security dilemma dynamics, potentially creating unnecessary escalation in AI capabilities development that outpaces responsible governance.

KEY TAKEAWAYS FOR US STRATEGY

Prestige Competition *Can Be* Strategic, Not Superficial - States do not exclusively pursue AI prestige for vanity (though there are many instances of this); they often pursue it because international recognition can translate into real influence over norms, alliances, markets, and the ability to attract leading talent (critical in the case of AI development). Hosting summits, producing declarations, or creating responsible AI practices are often misunderstood as purely symbolic; in fact, when wielded effectively, they can be instruments of diplomatic capital and agenda-setting power.

Prestige Competition Need Not Be Destructive - Prestige derived from advancing global welfare creates more sustainable influence than zero-sum competition. The most enduring status comes from AI leadership that demonstrably benefits humanity - such as breakthrough medical diagnostics, climate modeling capabilities, or educational accessibility tools. The US could gain lasting prestige by positioning its AI ecosystem as uniquely capable of addressing global challenges, from pandemic response to food security, rather than merely outcompeting rivals on technical benchmarks.

Frontier Models Are Implicitly Treated as Status Symbols - Frontier AI models, such as OpenAI's GPT line or China's DeepSeek, function as "dynamic status symbols" that require continual performance and iteration. The January 2025 DeepSeek R1 release temporarily upended perceptions of US AI dominance and triggered a flurry of reputational signalling and counter-releases. This mirrors Cold War-era dynamics like the "Sputnik moment" and shows how AI breakthroughs are now geopolitical events.

Middle Powers Can Legitimately Shape Norms - The UK's Bletchley Declaration and South Korea's Seoul Declaration demonstrate how middle powers leverage governance to build relevance. The U.S. should take these moves seriously, not as challenges to hegemony but as signs that allies are seeking a voice in setting the terms of AI development. Ignoring these efforts risks alienating partners and losing influence in multilateral fora.

AI Access Decisions Shape Tomorrow's Alliances - How AI systems are deployed, by whom, and under what governance frameworks will shape global power dynamics for decades. US strategy around the diffusion of its AI products requires considerable thought and investment. Successful US engagement on AI with AI follower states must balance strategic interests with legitimate development needs, creating sustainable partnerships that build lasting influence rather than transactional relationships based solely on status competition pressures. See accompanying article, “Shaping the World’s AI Future: How the U.S. and China Compete to Promote Their Digital Visions” (Blomquist & McBride, 2025).

Deepseek's AI Breakthroughs Don't Change the Fundamentals—but They Are a Warning[7]

Klon Kitchen

Senior Fellow, American Enterprise Institute

China's AI ambitions have long been hamstrung by a critical weakness: access to high-end computing hardware. US export controls have effectively cut Beijing off from the most advanced AI chips, putting a hard ceiling on its ability to compete at the highest level. But that hasn't stopped China from trying to work around these limitations.

DeepSeek, a Chinese AI company, has made notable progress in optimizing AI models to run more efficiently on the downgraded chips it can still acquire. These software techniques improve training efficiency, reduce costs, and enhance reasoning abilities. The company even claims to have developed new reinforcement learning methods that go beyond standard industry practices. This is significant, but not in the way some are suggesting.

DeepSeek's optimizations do not alter the fundamental reality of US-China AI competition. Hardware remains the decisive factor, and China's access to top-tier chips like Nvidia's H200, GB200, and beyond is severely restricted. Software efficiency gains, no matter how sophisticated, do not replace raw computing power.

However, this development serves as a warning. It reinforces the fact that China is actively working to mitigate its hardware constraints. Beijing isn't waiting for the US to loosen restrictions; it's aggressively pursuing ways to extract every ounce of performance from the hardware it has. Washington should take note.

Beyond the technical achievement, there are deeper concerns. Evidence suggests DeepSeek may have trained its AI models using outputs from OpenAI's o1 model—essentially copying its capabilities through a process known as model distillation. If true, this would be yet another example of China using intellectual property theft as a shortcut to AI advancement. This isn't a new problem, but it's a reminder that AI competition is not just about research and development; it's also about protecting proprietary technology from being siphoned off and repurposed.

---

[7] Originally published on January 29, 2025 at https://www.aei.org/foreign-and-defense-policy/deepseeks-ai-breakthroughs-dont-change-the-fundamentals-but-they-are-a-warning/.

DeepSeek's claim of a 70 percent reduction in AI training costs is also misleading. While efficiency improvements may lower the final phase of training, they do not account for the costly research, experimentation, and trial-and-error required to develop top-tier models. It's the equivalent of claiming a fighter jet is cheap because the last bolt added costs only a few dollars—while ignoring the billions spent on design and testing. Many discussions about China's AI progress fail to acknowledge this reality. Despite these limitations, DeepSeek's work should not be dismissed. AI efficiency matters, and in some applications, optimized software running on older hardware can still be effective. More importantly, DeepSeek's progress underscores that China's AI development is not standing still. That has real implications for national security. These models will not just be used for consumer applications. China's military is increasingly integrating AI into its operations, from autonomous weapons to real-time battlefield intelligence. The US Navy has already issued a warning against using DeepSeek's AI due to security concerns, a clear indication that these tools are not just theoretical—they have real-world consequences.

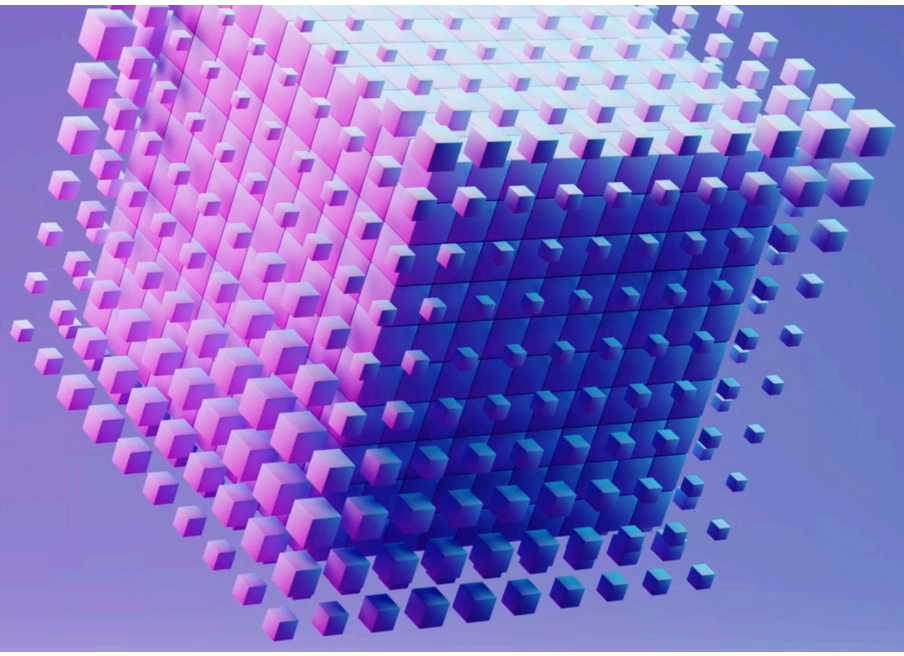
The US response must be twofold. First, Washington must remain vigilant in enforcing AI chip export controls and cracking down on illicit GPU smuggling. These restrictions are working, and DeepSeek's workaround efforts prove it. Weakening or neglecting them would be a strategic mistake.

Second, the US must ensure it maintains an overwhelming lead in AI research and development. That means doubling down on investments in advanced computing infrastructure, semiconductor manufacturing, and AI talent.

China's AI progress does not erase its hardware disadvantage, but it does send a clear message: Beijing is looking for every possible way to close the gap. The US cannot afford to be complacent. Maintaining AI leadership is not just an economic imperative—it is a national security necessity.

THE
kyndryl
institute

AI and National Security



Trend Topic: **Readiness**

Preparing for AI in a new security landscape

By

Klon Kitchen

Senior Fellow at American Enterprise Institute

Artificial intelligence (AI) is reshaping global competition, but for U.S. businesses, the challenge is no longer just about innovation—it is about survival in a geopolitical environment where national security and corporate responsibility converge.

While many executives have made progress preparing their organizations for AI's transformative potential by identifying use cases, investing in infrastructure, and tackling adoption challenges, an equally critical aspect of AI readiness demands attention: national security.

This is not only about compliance checklists and being prepared to manage a public relations crisis. It is about recognizing the broader geopolitical, economic, and security implications that AI adoption introduces and the role your organization plays in safeguarding the United States and its interests.

The Emerging National Security Imperative

AI is no longer just a tool for optimizing supply chains or streamlining customer service. It is a strategic asset—one that adversarial nations are targeting with unprecedented focus. The U.S. government sees this reality clearly. For instance, the Department of Defense views AI as a foundational technology for maintaining military superiority. Federal agencies are tightening export controls, scrutinizing data flows, and emphasizing the protection of critical infrastructure. These actions are not arbitrary; they are rational responses to real threats.

China's ambitions in AI are particularly instructive. Beijing has explicitly articulated its goal of becoming the global leader in AI by 2030. This goal is not limited to academic benchmarks or industry accolades; it is a cornerstone of a broader geopolitical strategy. China's leadership understands that AI will determine not just who innovates but who leads economically, technologically, and militarily.

This desire for dominance manifests in several ways. Chinese companies are embedding AI into surveillance systems that track millions of citizens, both domestically and abroad. Technologies like facial recognition and predictive policing are not just tools of social control but mechanisms for projecting influence globally. Additionally, China's industrial policies—such as its subsidies for AI-related technologies—give its companies an edge in global markets, creating dependencies that can be leveraged in times of geopolitical conflict.

A striking example of this is China's Belt and Road Initiative, a global infrastructure strategy that incorporates AI systems into the digital "Silk Road" to expand Beijing's political and economic influence. Many countries participating in the initiative are adopting Chinese-built AI systems for smart cities and government operations. These systems come with long-term dependencies, meaning that countries relying on them require ongoing support, updates, and integration, enabling Beijing to exert political and economic influence on a global scale.

This strategy underscores the lengths to which adversaries will go to secure strategic advantages, a trend further exemplified by recent incidents targeting U.S. AI infrastructure.

From espionage campaigns aimed at proprietary algorithms to supply chain infiltration involving hardware components, the threats are both sophisticated and pervasive. For instance, reports of Chinese-manufactured hardware components with embedded vulnerabilities have raised alarms about potential backdoors in critical systems. These risks are not theoretical; they are a daily reality in the hyperconnected global economy.

The implications extend beyond the theft of intellectual property. Consider the potential consequences of adversarial manipulation. An AI system corrupted at the training stage could subtly distort outcomes in ways that remain undetected until critical decisions— financial, operational, or even life-and-death—are impacted. This is not hypothetical; adversarial attacks on machine learning systems **are** well-documented and **are** evolving rapidly.

From espionage campaigns aimed at proprietary algorithms to supply chain infiltration involving hardware components, the threats are both sophisticated and pervasive.

Why Industry's Role is Critical

The U.S. government can only do so much. Unlike in China, where industry operates at the direction of the state, America's strength lies in its innovative private sector. This dynamic is a double-edged sword. On the one hand, it enables the kind of creativity and agility that leads to breakthroughs. On the other, it creates vulnerabilities when companies underestimate the strategic dimensions of their operations.

> Failing to address national security risks is not just a vulnerability for individual organizations—it is a systemic issue. The interconnectedness of the global economy means that one compromised node can have cascading effects.

Consider the intersection of AI and supply chain security. Many organizations rely on foreign manufactured hardware—GPUs, sensors, or even basic semiconductors—that underpin their AI deployments. If these components originate from adversarial nations, they could carry backdoors or vulnerabilities that compromise not only the integrity of your systems but the broader security of critical sectors. The semiconductor shortage of recent years underscores how fragile these supply chains can be. Now, overlay that fragility with the risks of malicious interference, and the stakes become even clearer.

Washington recognizes these challenges and has begun acting accordingly. Export controls on advanced semiconductors, initiatives to reshore critical industries, and policies to strengthen public-private partnerships all point to a shared objective: safeguarding America's AI future. But these measures cannot succeed without active industry participation. The private sector is not a bystander in this fight; it is the front line.

For businesses, this is not just about avoiding sanctions or regulatory penalties. It is about competitiveness. Companies that demonstrate robust security practices and alignment with national priorities will find themselves at an advantage—whether in securing federal contracts, attracting global customers, or mitigating reputational risk in an era of heightened geopolitical scrutiny. Conversely, those that fail to adapt will face an uphill battle, as both public and private stakeholders increasingly demand accountability.

The Broader Stakes for Industry

Failing to address national security risks is not just a vulnerability for individual organizations—it is a systemic issue. The interconnectedness of the global economy means that one compromised node can have cascading effects. For example, an attack on a single AI-powered logistics platform could disrupt supply chains for entire industries, amplifying economic instability.

Additionally, as adversaries continue to innovate, the gap between offensive and defensive capabilities grows. A reactive posture will no longer suffice. Companies must adopt proactive strategies that integrate security into the DNA of their AI initiatives. This requires not only technical solutions but also cultural and organizational shifts.

Imagine the implications of an adversary subtly influencing the decisions of an AI system used to manage critical infrastructure—say, energy grids or transportation networks. The damage could cascade beyond the initial target, undermining public

trust, destabilizing economies, and even triggering broader geopolitical consequences. These risks highlight why no company, regardless of its size or industry, can afford to overlook its role in securing the broader ecosystem.

Taking Action: A Roadmap for Leaders

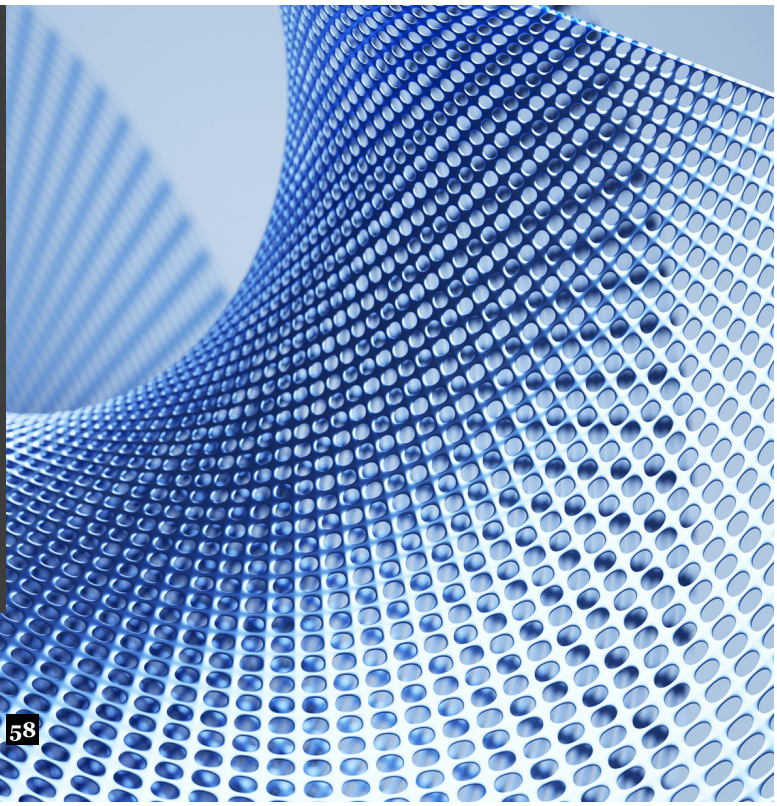
For CEOs and CISOs ready to take these challenges seriously, several steps can help address national security risks.

First, conduct a geopolitical risk assessment. Evaluate your AI supply chain, partnerships, and data practices through a geopolitical lens. Where are your hardware components sourced? Who are your cloud providers, and what jurisdictions govern their operations? The answers to these questions should inform a detailed risk map. Partner with firms specializing in geopolitical intelligence to understand how shifts in global politics might affect your vulnerabilities.

Second, collaborate with the U.S. government. Build formal relationships with federal agencies, such as the Department of Defense, the Department of State, and the Department of Commerce. Engage in public-private partnerships focused on AI security and participate in federal initiatives like the National Artificial Intelligence Inititative. Beyond compliance, these partnerships provide insight into emerging threats and access to tools that can enhance your organization's security posture.

Third, integrate "security by design"across the AI lifecycle. Security must be a core consideration from the outset of any AI project. This includes safeguarding training data, securing cloud storage, and testing models against adversarial attacks. Implement automated systems to monitor for unusual patterns or anomalies in AI behavior post-deployment. Consider leveraging frameworks such as the National Institute of Standards and Technology (NIST) AI Risk Management Framework to standardize your approach to identifying and mitigating risks.

Security must be a core consideration from the outset of any AI project. This includes safeguarding training data, securing cloud storage, and testing models against adversarial attacks.



Finally, build organizational resilience. National security risks are not just technical—they are operational and cultural. Create cross-functional teams that integrate security professionals, legal advisors, and technologists to ensure a holistic approach. Train your workforce to recognize and respond to emerging threats and ensure that your organization fosters a culture where security considerations are baked into innovation. Regularly simulate scenarios involving AI-related disruptions to test your readiness and identify weaknesses. Use your position as a leader to advocate for industry-wide standards and best practices in AI security. Collaborate with peers, trade organizations, and policymakers to drive initiatives that align private-sector innovation with public-sector priorities. By contributing to a secure and resilient AI ecosystem, you reinforce not only your organization's safety but also its reputation as a responsible industry leader.

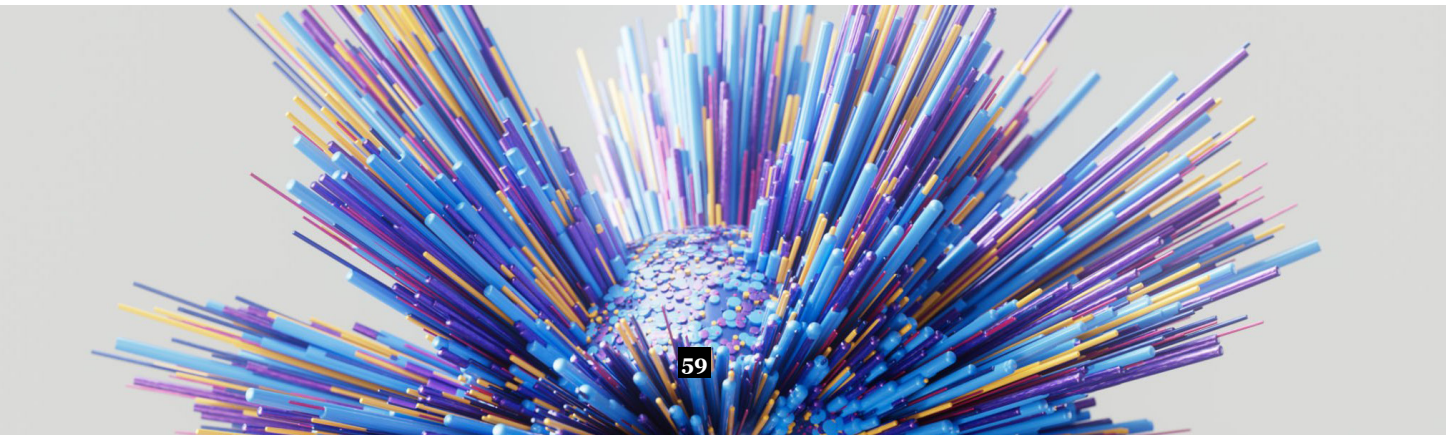
Conclusion

The rise of AI represents both an extraordinary opportunity and a profound responsibility. As CEOs and CISOs, you are uniquely positioned to navigate this duality. You have already proven your ability to lead your organizations through the complexities of digital transformation. Now, the challenge is to expand that leadership to account for the broader implications of AI in an era of heightened national security risks.

Your organizations do not operate in a vacuum. They are part of a larger ecosystem that shapes and is shaped by global forces. By addressing the national security dimensions of AI readiness, you are not just future-proofing your business—you are playing a pivotal role in ensuring that the United States remains secure, competitive, and free.

The stakes are high, but so is the potential for impact. This is the moment to act decisively. As the stewards of transformative technologies, you hold the keys to both innovation and resilience. The question is not just whether you can adapt but whether you will lead.

The answer to that question will define your legacy—and the security of the systems that underpin it.



What if we're right about AI timelines? What if we're wrong?[8]

Jack Clark

Co-Founder, Anthropic

Recently, I've been thinking a lot about AI timelines and I find myself wanting to be more forthright as an individual about my beliefs that powerful AI systems are going to arrive soon - likely during this Presidential Administration. But I'm struggling with something - I'm worried about making short-timeline-contingent policy bets.

So far, the things I've advocated for are things which are useful in both short and long timeline worlds. Examples here include:

- Building out a third-party measurement and evaluation ecosystem.
- Encouraging governments to invest in further monitoring of the economy so they have visibility on AI-driven changes.
- Advocating for investments in chip manufacturing, electricity generation, and so on.
- Pushing on the importance of making deeper investments in securing frontier AI developers.

All of these actions are minimal "no regret" actions that you can do regardless of timelines. Everything I've mentioned here is very useful to do if powerful AI arrives in 2030 or 2035 or 2040 - it's all helpful stuff that either builds institutional capacity to see and deal with technology-driven societal changes, or equips companies with resources to help them build and secure better technology.

But I'm increasingly worried that the "short timeline" AI community might be right - perhaps powerful systems will arrive towards the end of 2026 or in 2027. If that happens we should ask: are the above actions sufficient to deal with the changes we expect to come? The answer is: almost certainly not!

Under very short timelines, you may want to take more extreme actions. These are actions which are likely 'regretful actions' if your timeline bets are wrong. Some examples here might be:

---

[8] Originally posted on ImportAI, Vol. 405 on March 24, 2025, https://importai.substack.com/p/import-ai-405-what-if-the-timelines

Massively increasing the security of frontier labs in a way that reduces the chance of hacking or insider threats, but also happens to make life extremely unpleasant and annoying for those working within those labs. This helps on short timelines but is ultimately a very expensive thing on long timelines because it'll slow down technological progress and potentially create a blowback where labs shift away from extreme security after some period of time, having found it onerous.

Mandating pre-deployment testing: Today, pre-deployment model testing is done by companies on a voluntary basis. If you thought we were on short timelines and risks were imminent, you might want to mandate pre-deployment testing by third parties. This, though, is extremely costly! It introduces friction into the AI development process and, like the lab security ideas, risks creating blowback. Last year's debate in California about the 'SB 1047' bill felt like a preview of the kind of blowback you could see here.

Loudly talking about and perhaps demonstrating specific misuses of AI technology: If you have short timelines you might want to 'break through' to policymakers by dramatizing the risks you're worried about. If you do this you can convince people that certain misuses are imminent and worthy of policymaker attention - but if these risks subsequently don't materialize, you could seem like you've been Chicken Little and claimed the sky is falling when it isn't - now you've desensitized people to future risks. Additionally, there's a short- and long-timeline risk here where by talking about a specific misuse you might inspire other people in the world to pursue this misuse - this is bound up in broader issues to do with 'information hazards'.

These are incredibly challenging questions without obvious answers. At the same time, I think people are rightly looking to people like me and the frontier labs to come up with answers here. How we get there is going to be, I believe, by being more transparent and discursive about these issues and honestly acknowledging that this stuff is really hard and we're aware of the tradeoffs involved. We will have to tackle these issues, but I think it'll take a larger conversation to come up with sensible answers.

Eschatological AI Policy Is Very Difficult[9]

Jack Clark

Co-Founder, Anthropic

A lot of people that care about the increasing power of AI systems and go into policy do so for fundamentally *eschatological* reasons - they are convinced that at some point, if badly managed or designed, powerful AI systems could end the world. They think this in a literal sense - AI may lead to the gradual and eventually total disempowerment of humans, and potentially even the death of the whole species.

People with these views often don't recognize how completely crazy they sound - and I think they also don't manage to have empathy for the policymakers that they're trying to talk to.

Imagine you are a senior policymaker in a major world economy - your day looks something like this:

- There is a land war in Europe, you think while making yourself coffee.
- The international trading system is going through a period of immense change and there could be serious price inflation which often bodes poorly for elected officials, you ponder while eating some granola.
- The US and China seem to be on an inexorable collision course, you write down in your notepad, while getting the car to your place of work.
- There are seventeen different groups trying to put together attacks that will harm the public, you say to yourself, reading some classified briefing.
- "Something akin to god is coming in two years and if you don't prioritize dealing with it right now, everyone dies," says some relatively young person with a PhD and an earnest yet worried demeanor. "God is going to come out of a technology called artificial intelligence. Artificial intelligence is a technology that lots of us are developing, but we think we're playing Russian Roulette at the scale of civilization, and we don't know how many chambers there are in the gun or how many bullets are in it, and the gun is firing every few months due to something called scaling laws combined with market incentives. This technology has on the order of $100 billion dollars a year dumped into its development and all the really important companies and infrastructure exist outside the easy control of the government. You have to do something about this."

[9] Originally posted on ImportAI, Vol. 410 on April 28, 2025, https://importai.substack.com/p/import-ai-410-eschatological-ai-policy

The above is, I think, what it's like being a policymaker in 2025 and dealing with AI on top of everything else. Where do you even start?

Even starting to deal with the problems of AI is expensive.

First you need to learn about the technology, which means either:

- You need to take your staff that are themselves extremely busy and underwater and ask them to pick up another topic, or you need to tell them to drop something - your choices of stuff to drop might include 'medical issues my constituents care about' or 'economic policy that influences jobs', and so you actually can't get them to drop stuff. So you add it to their pile.
- You need to get smart about it, which means you need to further salami slice your weekly agenda so you can fit a tiny bit of time in which is for 'learning about AI'.
- For both of these choices, learning about AI usually requires you to speak to different people with expertise. Once you do this you quickly discover that:
    a) Some people think all current AI technology is, essentially, bullshit, and urge you not to fall for hype.
    b) Some people say AI technology is a really big deal and the government should avoid regulating it.
    c) Some people say AI has a high likelihood of killing everyone on the planet.
    d) All of these people think people with different views have incorrect priors.

Now you need to learn about the potential policy moves you can make. Some examples of these moves and their costs include:

- **Taking things away from people**, like export controls which take certain computers away from certain countries. Doing this 'fucks with the money' of a very large industry and also adds to geopolitical tensions. Everyone will get very mad about anything you do here. The experts you've consulted in your earlier step will either think you didn't go far enough, you went way too far, or the fact you're doing anything at all is corrosive to democracy and the economy.
- **Giving the government a greater ability to understand the domain**, like creating institutions like the AI Safety Institute or re-tasking people from existing government departments to focus on AI. Doing this takes a scarce resource (people in government) and re-allocates them, so you're trading away from other

priorities and people will get mad. Or you need to spend money to create net new capacity, in which case people view whatever you do with suspicion, and even getting the money requires some kind of political deal to assuage the feelings of the other many deserving groups who didn't get the money.

- **Altering the behavior of the companies through sub-regulatory methods,** for instance by securing voluntary commitments. To do this you need to spend a ton of energy to ensure you and your staff can learn more about the technology, then you need to negotiate commitments with companies. Negotiating with companies is like putting together a trade deal with a superintelligence - the companies will assign far more people than you and your staff to think about the commitments, and the companies have access to all the high quality information about the technology in question. If you succeed, people will accuse you of being captured by corporate interests.
- **Changing laws, for instance by passing regulations targeting AI development and deployment.** This is an extremely costly action that requires you to cash in innumerable political chips in exchange for building a large coalition that can pass some legislative package. Corporate interests will typically fight you or, at best, partner with you but in a way that tries to bend the rules to be as advantageous to them as possible. The whole time you are putting the law together you and your political allies will come under attacks for being either too weak in your approach or too strong in ways that might damage the economy. If you successfully change the laws the consequences of your change will be held under an incredibly un-sympathetic microscope for following years, opening up a new vulnerability for you with regard to your political opponents.

Let us imagine that you make all of these policy moves. What happens then? Well, you've mostly succeeded by averting or delaying a catastrophe which most people had no knowledge of and of the people that did have knowledge of it, only a minority believed it was going to happen. Your 'reward' insofar as you get one is being known as a policymaker that 'did something', but whether the thing you did is good or not is very hard to know.

The best part? If you go back to the AI person that talked to you earlier and ask them to assess what you did, they'll probably say some variation of: "Thank you, these are the minimum things that needed to be done to buy us time to work on the really hard problems. Since we last spoke the number of times the gun has fired has increased, and the number of bullets in the chamber has grown."

What did I do, then? You ask. "You put more chambers in the gun, so you bought us more time," they say. "Now let's get to work".

I write all of the above not as an excuse for the actions of policymakers, nor as a criticism of people in the AI policy community that believe in the possibility of superintelligence, but rather to instead illustrate the immense difficulty of working on AI policy when you truly believe that the technology may have the ability to end the world. Most of the policy moves that people make - *if* they make them - are going to seem wildly unsatisfying relative to the scale of the problem. Meanwhile, the people that make these moves are going to likely be juggling them against a million other different priorities and are going to be looking to the AI experts for some level of confidence and validation - neither of which are easily given.

Good luck to us all.

The AI Code Revolution: When Machines Write Software, and Break It Too.

Rob Joyce

Founder, Joyce Cyber; former NSA Director of Cybersecurity

We have entered a new frontier in software development, where AI is simultaneously a powerful asset and a substantial security risk. Artificial intelligence can now effortlessly create complete applications or entire websites using nothing more than a textual description. It rapidly generates thousands of lines of code, debugs intricate algorithms, and uncovers critical security vulnerabilities in software essential to our daily lives. The same AI capabilities that enhance software safety can also be leveraged by malicious actors to rapidly identify vulnerabilities and automate the creation of exploits and run the operations that use them. In 2024, I believed enabling social engineering was AI's primary threat, but over the past year, I've become convinced that AI-driven vulnerability hunting, exploit generation, and hacking operations present a more significant strategic challenge.

The speed at which AI is transforming software development is breathtaking. Just three years ago, AI-assisted coding was a novelty. It was a helpful autocomplete feature that occasionally saved developers a few keystrokes. Today, AI can write entire functions, architect complex systems, and even fix its own bugs. In 2023 GitHub, the global hub of collaborative, open-source software innovation, reported that its AI assistant, Copilot, generates roughly 46% of the code in projects where it's enabled[i]. Some developers report that AI handles up to 80% of their routine coding tasks. The major tech CEOs are all talking about AI-generated code. Microsoft CEO Satya Nadella said as much as 30% of their code is written by AI in April 2025[ii]. In October 2024, Sundar Pichai, the Google CEO said more than a quarter of all their new code is generated by AI[iii]. In May 2025, Mark Zuckerberg of Meta claimed "I think sometime in the next 12 to 18 months, we will reach the point where most … will be written by AI. And I don't mean like autocomplete."[iv] The most aggressive prediction was from Dario Amodei, the CEO of Anthropic, in March 2025, who claimed that in 3–6 months, AI will write 90% of their code. This final prediction was most telling because Anthropic's Claude is widely viewed as one of the most capable AI-based software development tools and Dario gets access to the unreleased and evolving capabilities in the company.

This revolution isn't just about speed, it's about democratization. A startup founder with limited programming experience can now build a functional app in days, prototyping test scenarios in rapid evolutionary spins. Students are learning to code by conversing with AI tutors that never tire of explaining concepts. Small businesses can afford custom

software solutions that once required teams of expensive engineers. It's as if we've given everyone a master programmer as a personal assistant.

But there's a darker side to this technological miracle. The same AI systems that help us build software faster and better are now being weaponized to tear it apart. Software that is expert at developing code must also be able to find and correct errors, and that is the functionality at the heart of vulnerability discovery.

Consider what happened in early 2024 when researchers at a leading university gave GPT-4, one of today's most advanced AI models, access to public vulnerability databases. Armed only with standard descriptions of known security flaws, the AI successfully developed working exploits for 87% of them all by itself[v]. What once required elite hacking skills and weeks of effort now takes AI only minutes. In fact, the company Hack The Box ran a groundbreaking "AI vs Human CTF Challenge" in March 2025. The artificial agents competed head-to-head with elite human hackers. Four hundred and three teams competed, with multiple people on each team. Most AI agents solved 19 out of the 20 complex challenges in hacking and cryptography. The best AI team placed 20th among 403 teams. The leading AI was in the top 5% of the skilled competitors[vi].

Security researchers have demonstrated tools like DeepExploit, which combine reinforcement learning with existing attack frameworks to launch cyberattacks on networks automatically. DeepExploit is a proof-of-concept AI that learns how to penetrate a system by trying different exploits in a simulated environment, much like a human pentester but at machine speed[vii]. It ties into the Metasploit hacking toolkit and uses deep reinforcement learning to intelligently choose exploits for the target's open ports and services.

This isn't science fiction; it's happening now. Security researchers have already demonstrated AI systems that can scan popular software projects, identify subtle vulnerabilities that human reviewers missed, craft customized exploits, and even generate malware that constantly mutates to evade detection. One cybersecurity expert described the landscape as giving attackers "a master hacker that works 24/7, never gets tired, and learns from every attempt."

The implications are staggering. Every day, our society becomes more dependent on software. We depend on the apps on our phones as well as the systems controlling our power grids, hospitals, and financial markets. Now imagine adversaries armed with AI that can probe these systems relentlessly, finding and exploiting weaknesses at a scale

and speed we've never seen before. While the autonomous generation of exploits for highly complex, zero-day vulnerabilities remains a developing area, AI can already assist attackers in crafting, customizing, and optimizing exploit code, making the weaponization phase of an attack faster and more effective[viii].

We're already seeing the early signs. Cybercriminals are using AI to generate thousands of unique malware variants daily, overwhelming traditional antivirus systems. Nation-state actors are likely developing AI-powered cyber weapons that could identify and exploit vulnerabilities in critical infrastructure before defenders even know they exist. The time between a vulnerability's discovery and its exploitation, what security experts call the "window of exposure", is shrinking from months to days, sometimes to hours.

The traditional cat-and-mouse game of cybersecurity is evolving into something more akin to an arms race between competing AI systems. On one side, defensive AI will work around the clock to analyze code, detect anomalies, and patch vulnerabilities. On the other, offensive AI will probe for weaknesses, craft exploits, and adapt its attacks in real-time. One can envision the future battle fought at machine speed, where human oversight would struggle to keep pace.

Yet there's reason for cautious optimism. The same AI capabilities that empower attackers also enhance our defenses. AI static analysis Security tools use AI in three key ways: to improve vulnerability detection accuracy, to automatically prioritize findings by severity, and even to suggest or apply fixes for the identified issues. AI-driven code analysis is increasingly integrated directly into developers' workflows. In short, AI is helping catch mistakes when and where they happen, making it easier for teams to "shift left" on security by tackling issues early in development.

AI-powered security tools can also analyze vast amounts of code for vulnerabilities, predict where bugs are likely to occur, and even automatically generate fixes[ix]. Some experimental systems can detect and respond to attacks faster than any human could, potentially healing software vulnerabilities before they can be exploited. The superpower of AI enabled defenses is to stare with an unblinking eye at massive amounts of logs, data and processes for the first signs of anomalies.

Major tech companies and governments are also beginning to grapple with these challenges. The European Union's AI Act includes specific provisions for high-risk AI systems, including those used in cybersecurity. The U.S. National Institute of Standards and Technology has developed frameworks for managing AI risks[x]. Companies like Microsoft and Google are embedding security considerations directly into their AI development processes.

But are these efforts enough? The uncomfortable truth is that we're in uncharted territory. The pace of AI development outstripped our ability to create comprehensive regulations or foolproof defenses. The genie is out of the bottle for constraints on weaponizing AI. Every breakthrough that makes AI more capable at writing code also makes it more capable at breaking code. It's a fundamental duality we must learn to live with.

What can we do? First, we need to abandon any illusion that we can prevent AI from being used maliciously. The technology is already too accessible, and too useful to be contained. Instead, we must focus on resilience by building systems that can detect, respond to, and recover from AI-powered attacks.

Second, we need to invest in defensive AI capabilities. If attackers have AI working for them 24/7, defenders need the same. This means funding research into AI security tools, training more cybersecurity professionals to work alongside AI systems, and ensuring smaller organizations have access to AI-powered defenses.

Third, we must rethink how we design and build software. Security can no longer be an afterthought. It must be baked into every line of code from the start. AI can help here too, automatically checking for vulnerabilities as code is written and suggesting more secure alternatives. We can establish community resources to scan open source and critical software for vulnerabilities, ensuring discovery before adversaries can exploit any flaws.

The AI revolution in software development is one of the most transformative technological shifts of our time. It promises to make software development faster, cheaper, and more accessible than ever before. But with this power comes unprecedented risk. We're entering an era where the code that runs our world can be both written and broken by machines operating at superhuman speed.

The outcome of this AI code war isn't predetermined. Whether AI becomes primarily a tool for building a more secure digital world or a weapon for tearing it apart depends on the choices we make today. We have a narrow window of opportunity to shape this future and ensure that as AI grows more powerful, our defenses grow stronger still. The race is on. In this new world, the question isn't whether AI will transform software development and cybersecurity, it's whether we can harness that transformation to quickly establish better defense than the offense. The code wars have begun, and we're all on the front lines.

[i] Dohmke, T. (2023, June). *Interview with Thomas Dohmke, GitHub's CEO*. Freethink. Retrieved from https://www.freethink.com/robots-ai/github-copilot
[ii] CNBC. (2025, April 29). *Satya Nadella says as much as 30% of Microsoft code is written by AI*. Retrieved from https://www.cnbc.com/2025/04/29/satya-nadella-says-as-much-as-30percent-of-microsoft-code-is-written-by-ai.html
[iii] Yahoo Finance. (2024, October). *Google CEO says more than 25% of all their new code is generated by AI*. Retrieved from https://finance.yahoo.com/news/google-ceo-says-more-25-202927484.html
[iv] The Times of India. (2025, May). Mark Zuckerberg predicts when AI will be able to generate most of the company's code. Retrieved from https://timesofindia.indiatimes.com/technology/tech-news/mark-zuckerberg-predicts-when-ai-will-be-able-to-generate-most-of-the-companys-code/articleshow/120797488.cms
[v] University of Illinois Urbana-Champaign. (2024). Autonomous exploitation of real-world cybersecurity vulnerabilities by large language model agents. Retrieved from https://arxiv.org/html/2402.06664v1
[vi] Hack The Box. (n.d.). *AI vs Human CTF: Hack The Box results*. Retrieved from https://www.hackthebox.com/blog/ai-vs-human-ctf-hack-the-box-results
[vii] Aujas Cybersecurity Blog. *Penetration Testing: Deep Exploit*. Retrieved from: https://blog.aujas.com/penetration-testing-deep-exploit
[viii] Frontier AI's Impact on the Cybersecurity Landscape. (2025, April). arXiv. Retrieved from https://arxiv.org/html/2504.05408v2
[ix] DARPA. *AIxCC: AI Cyber Challenge*. Retrieved from https://www.darpa.mil/research/programs/ai-cyber
[x] National Institute of Standards and Technology. *AI Risk Management Framework*. Retrieved from https://www.nist.gov/itl/ai-risk-management-framework

The Digital Security Equilibrium – Does it Hold Under AI?[10]

Professor Ciaran Martin, CB

Founding CEO of the National Cyber Security Centre and Professor, University of Oxford

At the dawn of the digital age, when cybersecurity became a top-level concern, predictions of catastrophic harm were common. *The Economist* in 2010 featured a mock-up Manhattan-type sky line suffering a 9/11 style atrocity under the headline *Cyber War: The Threat from the Internet*. As Defense Secretary, Leon Panetta warned of Cyber Pearl Harbor, one of many such warnings from world leaders.

Whilst there have been many serious and damaging cyber security events, these catastrophic predictions have not come to pass. Official statistics in most developed countries tend not to attribute *any* fatalities to cyber attacks. The closest linkage between cyber attacks and mortal harm is in healthcare: frequent criminal attacks known as ransomware have damaged hospitals' ability to function and patient care has suffered as a result. Counting the exact toll is difficult, because saying with certainty that death occurred specifically because of a cyber attack when the victims are already critically ill, cannot be done with certainty. Nonetheless, a University of Minnesota study in 2023[i] estimated that between 42 and 67 Medicare patients in the United States between 2016 and 2021 died as a result of ransomware cyber attacks.

There have been, for sure, serious and major disruptive events caused by malicious cyber activity. In the course of a six week period in 2017, reckless activity by North Korea (the so-called Wannacry virus of May that year), and Russia (the so-called NotPetya operation in June) caused north of $10 billion of economic harm and disrupted critical services all over the world. But while it is obvious that while cyber vulnerabilities remain of great concern – no one in the United States will wish to see a repeat of the Colonial Pipeline fiasco in 2021, let alone several such incidents at the same time – an uneasy peace has, broadly, held in cyber space.

[10] To be read in conjunction with the attached, already published: Typhoons in Cyberspace: *Royal United Services Institute, March 2025.*

Introducing the Digital Security Equilibrium

Why is this? I attribute it to the three different components of what can be called the Digital Security Equilibrium.

1. By and large, we do not subcontract human safety entirely to computers.

The first part of the equilibrium is the connection between security and safety. The English language – unlike, for example, French and Spanish – has two distinct words for these concepts. They are not the same. Take aviation. Aviation *security* can be poor – there have been multiple hacks, and many more accidental IT failures that have grounded fleets and caused chaos, disruption and economic costs. But aviation *safety* has a good 21st century record.

That's because aviation safety amounts to considerably more than cyber security. No one would willingly get on an aircraft if they thought that were the IT to be hacked, or fail accidentally, that there was nothing the pilot – and ground staff communicating with the pilot – could do. A good example of this is the comprehensive failure, by accident, of Britain's Air Traffic Control system in August 2023. The resulting administrative chaos was hugely socially disruptive and economically damaging with mass cancellations and diversions. But planes already in the air all landed safely, using backup communications and manual flying. The same is true in railway systems: if signals fail, for whatever reason, trains should stop, rather than crash into each other. So hackers can easily cause mayhem, but not mass casualties, in transportation. The same holds true of most sectors, except healthcare.

2. Only a small number of highly capable actors have access to the most devastating tools

In earlier decades it was fashionable to compare cyber capabilities with nuclear ones. This was mistaken for many reasons, but a main reason is that while one either has extremely destructive nuclear capabilities or one does not, anyone can carry out basic cyber operations. But carrying out high impact cyber operations is an extremely complicated endeavour and beyond the capabilities of most actors. Young criminals acting alone can – and have – undertaken data and cash theft, and damaged networks. But the sort of highly sophisticated operations – think the Olympic Games/Stuxnet operation against the Iranian nuclear programme in 2011, or Russia's sabotage of France's TV5 Monde station in 2015 – take years of preparation. They require skilled people, top-of-the-range covert infrastructure, organisational strategy, and a slice of luck. This is one of two reasons they are comparatively rare – the costs of doing them are significant. Given this, only serious cyber players have, to date, had the capability to

do them. This leads to the second reason these attacks have been rare: the highly capable actors in possession of them – even the likes of Russia and China – will have some sense of calculation before launching them.

As the attached paper on China's Volt Typhoon capability shows, China is assessed to have the capability to launch devastating attacks on US critical infrastructure. But the same US assessment says these operations are unlikely to happen outside a serious US/China escalation.

3. The same tools that can be developed for malicious use can be developed to equal or greater good for our own security.

Cyber operations rely on maths and engineering. They have no agency, or moral compass, of their own. Malicious code, or vulnerabilities, that are detected can be ameliorated and it is common practice for the cyber security industry to release these fixes publicly so that everyone can defend against them. (Indeed, they can be 'reverse engineered' in the jargon, and fired back at the attacker, or anyone else).
The implication of this is that there is, in effect, a constant race between using the same capabilities for good and bad. A good example is the practice of what is known as 'vulnerability scanning'. This is a technique where one can scan swathes of the online world and work out which networks are patched – protected – against known weaknesses, and which aren't. Both malicious hackers and cyber defenders undertake vulnerability scanning. What matters is who is better at it and, in the case of defenders, whether those warned about unprotected networks take action.

Over the course of the digital revolution to date, this aspect of the cyber security equilibrium has been uneven but broadly at least neutral. Of course there are plenty of occasions where defences have 'lost' – hence the many cyber attacks we all hear about. But there has never been a comprehensive loss of superiority by defense over offense. In other words, it has been broadly in equilibrium.

AI and the Digital Security Equilibrium

Will this uneasy equilibrium hold in the age of AI? It is, of course, too early to tell. But there are some pointers on each of the three pillars of the equilibrium.

1. By and large, we do not subcontract human safety entirely to computers.

Preserving this aspect of the equilibrium is a straightforward choice. It is up to us. So far, the signs are encouraging.

Again, transportation provides a good example. A decade ago, predictions abounded that by now there would be no drivers on any public highways. That is transparently not yet the case. The principal reason for this is that societies have taken time and undertaken the extensive and detailed technical and communications work to gain widespread expert and public confidence for the safety model. In time, autonomous vehicles are highly likely to replace driven vehicles, even if more slowly than previously predicted, but in a way that will make them not just safer but felt to be safer, enhancing public confidence in the technology. This is an approach that should be replicated in other areas; it would be crazy to subcontract human safety entirely to computational machines that cannot be overridden.

2. Only a small number of highly capable actors have access to the most devastating tools

Contrastingly, this is, as of right now, the most worrying and fraying part of the equilibrium. AI does not create magical new weapons. But, as Rob Joyce's excellent paper in this dossier shows, AI significantly enhances the quality of some malicious capabilities. It also reduces the cost of generating attacks, and the difficulty of doing them. For these reasons, alongside the growing market in selling damaging cyber capabilities (some of which is legal, and some illegal) this is one of the areas of greatest concern.

The geopolitical calculation that the likes of China, Iran and Russia will make before being overtly and overly aggressive in the use of potent cyber capabilities – which often leads to some form of restraint – is unlikely to extend to newer actors. Specifically, non-state terrorist groups with nihilistic tendencies have long craved powerful cyber capabilities but have never been able to acquire them because of this pillar of the equilibrium. Were that to change, our exposure to risk significantly increases.

3. The same tools that can be developed for malicious use can be developed to equal or greater good for our own security.

This aspect of the equilibrium *should* hold. But that will only happen if cyber security innovators working in free societies, whether in government or the private sector, keep up with or even outpace those who wish to misuse the new technologies. That is why it's important for governments to retain highly specialized in-house capabilities in their security agencies, and why it is imperative that the West's private sector cyber security industry continues to thrive and barriers to its development are tackled.

Conclusion

The Digital Security Equilibrium is a useful concept if we wish to understand why cyberspace has remained a place of harm, contestation, but not catastrophe to date. It *can* remain that way, but it requires a sustained effort and smart policymaking over many years. And for now, the most worrying part is the growing accessibility of potent cyber capabilities to new actors.

REFERENCE

[i] Hacked to Pieces? The Effects of Ransomware Attacks on Hospitals and Patients, Claire C. McGlave, Hannah Neprash and Sayeh Nikpay, University of Minnesota - Twin Cities - School of Public Health, available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4579292

Typhoons in Cyberspace[11]

Professor Ciaran Martin, CB

Founding CEO of the National Cyber Security Centre and Professor, University of Oxford

The transformation of China's digital attack capabilities is the most important change in the cyber threat to the West in more than a decade.

In a world of extraordinary geopolitical volatility, the threat to Western nations and interests from cyber attacks has, contrary to expectations, remained remarkably stable. Asked to name the leading anti-Western nation state actors in 2015, any expert would have listed Russia, China, Iran and North Korea. Asked to do so in 2025, experts would give the same list. Moreover, in the 2020s, the sort of serious disruption to critical infrastructure – to energy facilities, healthcare and other sectors – **long feared** by governments have, insofar as they have materialised at all, been caused by **Russia-based cyber-criminals**. The biggest nation-state threat actors have, by-and-large, kept much of their cyber powder dry. Even in its invasion of Ukraine, Russia's cyber forces **underperformed** as badly as their conventional ones in their illegal assault on Kyiv, and there was no serious attempt to use cyber disruption to deter western countries from backing President Zelenskyy's fight for national survival.

As ever, this relatively stable picture – one of significant threat but little actual change in that threat – has been accompanied by a **steady drumbeat of commercial hype** about how the cyber threat to anyone and everyone is getting worse all the time. That this is objectively untrue has not arrested the spread of the narrative. But such unrealised and unspecific scaremongering means we risk failing to notice when important shifts in the threat picture actually emerge. And there has been one profoundly important shift in the threat picture recently: over the past two years we have learned of a transformation of China's cyber capabilities into a far more formidable strategic threat.

This is, by far, the most significant shift in the cyber threat landscape in well over a decade. As a cyber actor, China has changed in three ways. First, the objectives of its cyber capabilities have shifted from economic to political ones. Second, its operations

[11] Originally posted on RUSI on 20 March 2025, https://www.rusi.org/explore-our-research/publications/commentary/typhoons-cyberspace

have changed from being opportunistic to strategic. Thirdly, and most importantly, it has moved beyond being simply a passive actor to an being active one. In other words, it does not just spy and steal anymore; it has also laid the ground for hugely disruptive cyber operations against western critical infrastructure, which hitherto it had shown no signs of doing.

The Cyber Typhoons

Two major cyber operations by China were uncovered in 2023 and 2024. They were, unhelpfully, given very similar codenames: 'Salt Typhoon' and 'Volt Typhoon. But although only two letters distinguish them, they are profoundly different.

***Salt* Typhoon** is a state intelligence service sponsored operation. It has comprehensively penetrated the United States's telecommunications system, leading to panicked guidance from the US Government to the nation's elite to use end-to-end encrypted messaging services, or else assume their data and message content are transiting to Beijing. Think of it, in effect, as China doing a 'Snowden' to America; gaining vast access to the nation's communications via a strategic spying operation of breathtaking audacity.

***Volt* Typhoon**, by contrast, is a military operation for strategic political and potentially military purposes. Run by the cyber unit of the People's Liberation Army, it involves putting preparatory implants – 'digital booby traps', as some have called them, into all manner of American critical infrastructure. In its **official assessment** of the operation, the Biden administration listed manufacturing, utilities, transportation, construction, maritime, IT, education & government (though, interestingly, not healthcare) as the sectors affected. The implants are, in the view of US officials, not there to spy and great care was taken to avoid detection. The US view, endorsed by all the other Five Eyes countries, is that these implants were strategic assets to be detonated in the event of a major confrontation between China and the West, most probably over Taiwan.

If *Salt* is a strategic spying scare story, *Volt* is a direct military threat to the western way of life. The consequences can be imagined thus: think of one of the major ransomware attacks the west has suffered, such as the Colonial Pipeline outage; or the **2019 attack on a private company** which left English policing short of half its criminal forensics

capabilities; or the small **English local authority scrambling to restore its services for vulnerable children** after an attack in the same year. Now think of dozens or even hundreds of them at the same time – "**everything, everywhere, all at once"** in the words of Jen Easterly, recently departed head of the US Cybersecurity and Infrastructure Security Agency. Then think of this as ransomware without ransoms – the affected victims cannot pay their way out of trouble because the objective is political, not financial. The economic, social, and even public safety impact of such an operation would be huge.

This is a completely new and extremely troubling scenario. China, for all its many violations of international cyber norms over the years, has no history of disruptive cyber attacks that stop networks from working. It has spied and stolen. And it's now much better, and more strategically focussed, at that too. That is why both *Salt* and *Volt* have Washington and other western capitals rattled.

How did this happen, and what are the implications for western governments and those charged with its cyber defence?

China's Cyber Attack Capabilities – A Third Phase?

To tell the story of how, in the words of the ***Wall Street Journal*** in January, "Chinese hackers graduated from clumsy corporate thieves to military weapons" it is necessary to look to the history of China's cyber operations. These can broadly, and a bit crudely, be fitted into three different phases.

The first phase ran from the dawn of the digital age till about 2013. The Communist Party regime was **initially terrified** of Internet based communications, and focused on developing 'protections' for itself such as **The Great Firewall**. But it also, at this time of rapid economic expansion in China, spotted an opportunity. Insecure western corporate systems were easy pickings for major corporate theft (as indeed were lower classification government systems for more traditional espionage). Industrial scale hacking of research, designs and so on began. While all of this was sponsored by the state, only some of it was done by full state employees; groups of opportunistic hackers sprung up with loose relationships with the regime. Their hacking was noisy, clumsy, and easy to spot – but only after it happened. And the noise did not matter: even before the seminal **Mandiant report of 2013**, the first major public attribution of this activity, many in the cyber security community, spoke openly about China's corporate theft. There were no consequences for such actions, and many rewards as a grateful

state and corporate sector happily paid for the valuable stolen information. Thus, China's cyber capabilities were born out of economic objectives, unlike Russia's, which have always been political.

This began to change with the dawn of the second phase of Chinese cyber activity, from about 2013 to about 2020, saw a centralisation and consolidation of China's capabilities. Three individuals helped shape this, whether intentionally or not. The first, inevitably, was Xi Jinping. The **messy, distributed and often chaotic web of actors** licensed in some way by the Chinese state ran against his centralising instincts. So his institutional reforms to Chinese administration more generally extended to cyber, with the **streamlining** of both intelligence and military cyber command structures, and the establishment of a policy agency, the Cyberspace Administration of China, to work out more strategic objectives about what Beijing wanted from cyberspace.

The second key player here was President Obama and his team. His interventions were prompted by the **genuine fury** in America's business community about corporate IT theft. (Washington was also rattled by the extraordinary strategic reversal of the **hack of the Office of Personnel Management's** security clearance database, when the sensitive security records of more than 20 million federal government officials went missing, reflective of the simultaneous focus by China on government espionage). Consequently, the Obama administration started **publicising details** of Chinese commercial espionage and threatening Xi's government with sanctions. This forced an **agreement** in 2015 which, however imperfect, led to several years of relative quiet on the commercial espionage front, with some notable exceptions. By definition, the agreement required further **centralised control** over the nation's hackers, adding further impetus to Xi's centralisation of capabilities.

The third, and arguably most important, cause of the shift was, however unintentionally, **Edward Snowden**. It is generally under-appreciated in the West just how important Snowden is in the history of China's approach to technology. Beijing was stunned at the extent of the US operation revealed by the former National Security Agency contractor, triggering effectively a Sputnik moment in its approach to its technology strategy.

According to the US Ambassador to China at the time, the distinguished former **Senator Max Baucus**, "the Snowden leaks dramatically changed Chinese policy towards the internet, its own people, the United States, and the world, with respect to

the internet and cyber security". Many Sinologists believe Snowden helped to precipitate the **Made in China 2025 strategy**, published two years after his leaks and setting out an extraordinary level of ambition for Chinese tech. Set against the Snowden era competition for geopolitical supremacy in technology, commercial hacking – the initial foundation of China's cyber operations – was now, for Xi's regime, a sideshow.

Thus, the third phase of China's cyber operations, running from about 2020 to the present, is a logical extension of the consolidation and refocussing of cyber operations in the second half of the last decade. It comes amidst the backdrop of the epochal geopolitical contest between the US and China for dominance in the technologies of the future. But unlike the '**tech war**' of mutual sanctions, industrial production expansion in both the US and China, the cyber aspect of the contest was designed to remain a covert part of China's strategy. Both *Salt* and *Volt* Typhoon were in play for years before being detected. And they are strategic compromises of the west on a scale hitherto unseen by any other cyber power.

These twin typhoon threats – one of massive scale strategic espionage, the other about military disruption of key services – are accompanied by other threats such as the large scale theft of strategic data and increasing, though at this stage mostly poor quality, attempts by China to undertake influencing and disinformation operations via digital means, attest to the most significant change in the cyber threat picture in recent years: China's cyber capabilities are now more strategically and politically focussed, and in general it is much better at this than it used to be.

What is to be done?

Where does this leave western policy makers and cyber defenders? As always, the debate risks generating more heat than light, with **outraged but incoherent calls** for 'imposing costs', 'striking back' and so on. The reality is more nuanced – there are fundamental issues for both the domestic defence of western economies and the organisations within them, particularly in the private sector, but also for statecraft. In all, there are five issues to consider when framing a Western response.

The first is about further development of our own cyber detection capabilities. The official US Government guidance to victim organisations for both ***Salt*** and ***Volt*** Typhoon are commendably honest and therefore make for difficult reading: they openly say how hard it is to detect these intrusions. One technique in particular, called '**living off the land**' in the trade, is hard to find as the intruder takes great care to look like a

normal network user. This presents a significant challenge for the cyber security industry to develop effective mitigations, but it is essential.

The second is about resilience. This is becoming a cliché, and plenty of more hawkish cyber defenders and geopolitical analysts push back on the notion that you can defend your way out of the problem. But when it comes to resilience, there seems to be little choice but to learn how to cope with the loss of a major network. The **British Library** in the UK is an instructive example: one ransomware attack by criminals crippled the basic functionality of a key UK institution for months. The essence of the Volt Typhoon threat is dozens or hundreds of such attacks at the same time. So, just as the Covid-19 pandemic prompted **soul-searching** about just-in-time supply chains, our cyber security vulnerabilities require the same self-examination about fragile services dependent on hackable IT.

The third is about the quality, age and sourcing of our infrastructure. What is now known publicly about the Salt Typhoon spying intrusion in particular is that central to the operation was the **exploitation of out-of-date kit**. This is a long-standing and well-known problem, especially in the telecommunications sector. Yet, most of the policy debate about telecoms security has been, for nearly a decade now, dominated, virtually to the exclusion of all other subjects, by banning Chinese companies, particularly in the United States. Indeed, Congress's 'response' to Salt Typhoon was finally to confirm a **languishing $3 billion programme** to replace the remaining Chinese kit in US telecoms infrastructure.

The problem is that the hack had nothing to do with Chinese kit. Every part of the Chinese campaign exploited vulnerable western manufactured infrastructure. The issue of the general security of often decrepit telecommunications infrastructure, as distinct from who built it, is too often forgotten, particularly in the United States.
This leads to a more general fourth point about policy and regulation. The vast majority of the national security risk presented by Chinese cyber aggression is held in the **private sector**, in the US and elsewhere in the west. Absent specific rules, governments are relying on the companies voluntarily to foot the bill for expensive backup plans for the resilience challenge, or even more costly equipment upgrades for the infrastructure ones. Many will do their best, partly out of commercial incentives, and partly because of a sense of public duty and a desire not to be the cause of a national security crisis. But this will often not be enough. In 2022, the UK Parliament passed the

Product Security and Telecommunications Infrastructure Act, which explicitly requires Britain's telcos to upgrade their infrastructure security. Importantly, this bill was enacted at the request of the industry, who explicitly told the government that requests from the state to spend on security was no longer a viable model for a highly competitive industry. The **UK Government** is, following a **similar measure in the EU**, also preparing to place into law more general cyber resilience obligations on critical infrastructure providers. There is no sign of the US going down this route, given the deregulatory zeitgeist in Washington at the dawn of the second Trump administration. But both *Volt*, and especially *Salt* Typhoon show how exposed the US is, and how difficult it will be to move the dial without some form of compulsion.

Finally, what is prevalent in Washington is a narrative that the challenge can be met with a more robust response, including by direct action by the US itself against China. There is something in this, as the Obama administration's success in quieting commercial espionage in 2015 showed. But the reason deterrence in cyberspace has not worked so far is not because we are not '**striking back**' hard enough. It is because it is **very, very difficult**. And in particular, like-for-like activity makes little sense.

The US on its own is, by some distance, the most capable cyber power on the planet, and, combined with its allies, further ahead still. Talk of 'retaliation' for the Salt Typhoon espionage campaign makes absolutely no sense: the US gains **considerable strategic advantage** from its intelligence services and to call into question the legality of digital espionage against other states would work against its interests. Moreover, one assumes that the US is conducting extensive espionage operations against China, so 'retaliation' for Salt Typhoon is oxymoronic.

For Volt Typhoon, China has not activated any of its digital booby traps yet, and according to the US Government, there is no sign that it intends do. So, there is no activity to 'punish' or 'retaliate' for yet. That is where deterrence comes in: the US should, **as it has been**, make it clear to China that the disruption of critical services in the US or allied territory would be absolutely unacceptable and would come with severe consequences. It is also important that Western offensive cyber capabilities are primed to engage properly should the need arise. But what is important is that policymakers of whatever persuasion are not seduced by a false narrative that there is an easy solution based on cyber power, deterrence and threats of retaliation that we have just been too timid to use.

A paradigm shift

All of these measures: improvements in detection, resilience and the quality of infrastructure; reforms to policy; and a realistic approach to deterrence, are needed if this new threat from China is to be met. But the challenges of domestic protection of critical private sector assets seems to be the most pressing, not least because, in the medium- to long-term, other actors with less cautious political calculations than China could exploit it.

In this sense, cyber security – reducing the vulnerabilities that have been so gapingly exposed – is hard power. And the starting point to getting there is realising that amidst all the noise and hype about cyber threats one thing has genuinely changed: the threat from China is significantly more serious than at any point in the digital age.

The Race for Compute and Energy: Securing America's AI Leadership

Divyansh Kaushik

Vice President, Beacon Global Strategies

Executive Summary

The global race for artificial intelligence dominance has created unprecedented demand for computational power and energy resources. With frontier AI data centers projected to require up to 5 gigawatts (GW) of baseload power by 2028, the United States faces a critical "AI energy gap" that threatens its technological leadership. This challenge creates an urgent need for strategic policy action that addresses both energy infrastructure and export control frameworks.

America's ability to develop and deploy advanced AI systems hinges not only on technological innovation but on our capacity to produce sufficient baseload power. Without decisive action to address energy bottlenecks, supply chain constraints, and regulatory hurdles, the U.S. risks losing AI leadership to competitors like China or seeing AI infrastructure migrate to regions with more abundant energy resources, such as the United Arab Emirates (UAE) and the Kingdom of Saudi Arabia (KSA).

The Five Pillars of AI Leadership

America's competitive advantage in artificial intelligence rests on five interconnected pillars, each essential to maintaining our global leadership:

1. **Algorithms** – The United States has led development of advanced machine learning architectures that optimize AI performance. From transformers to reinforcement learning from human feedback, American researchers have pioneered the algorithmic foundations of modern AI. However, this intellectual property is increasingly mobile and can be replicated abroad if other pillars weaken.
2. **Data** – Access to vast, high-quality datasets for training and fine-tuning models remains a crucial American advantage. The U.S. technology ecosystem produces and controls enormous quantities of diverse data, fueling model improvements. However, data advantage is eroding as competitors develop their own data collection mechanisms.
3. **Energy** – Scalable and affordable power sources to sustain AI compute demands represent America's most immediate vulnerability. Without sufficient domestic

energy capacity, even our advantages in other pillars become moot as companies relocate infrastructure to regions with abundant power.

4. **Chips** – High-performance semiconductors for training and inference workloads remain a clear U.S. strategic advantage. American companies like Nvidia and AMD, along with our allies, dominate the advanced chip market critical for AI development, creating a crucial chokepoint in the AI supply chain.
5. **Talent** – The United States' ability to attract and retain the world's brightest minds in AI research and development has been fundamental to our leadership. Xi Jinping already sees America's access to global talent as a strategic threat. However, maintaining this talent advantage requires sustaining a vibrant domestic AI ecosystem with both the energy and computational resources to support cutting-edge work.

These five pillars are deeply interconnected. Energy constraints directly impact our ability to deploy chips at scale. Talent will flow to regions where both energy and computational resources enable groundbreaking work. Most critically, energy represents both our most immediate vulnerability and the foundation upon which our other advantages depend.

The AI Compute Race and Energy Requirements

The scale of investment in AI infrastructure is reaching historic proportions. Microsoft, Google, Amazon, and Meta are projected to spend over $300 billion on AI capital expenditures in 2025 alone – nearly double their 2023 investments.[i] Private initiatives like Elon Musk's xAI are expanding rapidly, with plans to scale from 100,000 Nvidia H100 AI chips to facilities housing up to 1 million chips.

These investments reflect the industry's conviction that AI capabilities will dramatically transform the global economy. As computational requirements for frontier AI models continue to increase exponentially, energy consumption follows suit – creating an inseparable link between compute power and energy availability.

Each frontier AI training data center may require up to 5 GW of continuous power by 2028, not to mention the energy required by AI inference data centers. For context, a single gigawatt can power approximately 750,000 homes. This magnitude of energy consumption challenges existing energy production and distribution systems, with current grid infrastructure and generation capacity growing far too slowly to meet this explosive demand.

America's Energy Infrastructure Challenge

The U.S. Department of Energy estimates that data centers could consume approximately 6.7 to 12 percent of total U.S. electricity by 2028.[ii] However, existing infrastructure faces significant constraints:

1. **Permitting Delays**: Environmental reviews for new energy projects average 4-5 years, with transmission and interconnection processes adding another 4-7 years before projects are grid-connected.[iii]
2. **Supply Chain Bottlenecks**: Critical components face severe production limitations, from specialized nuclear components to Organic Rankine Cycle (ORC) turbines for geothermal applications.[iv]
3. **Workforce Shortages**: Specialized labor, including nuclear engineers, welders, and electricians, is in short supply due to decades of underinvestment.[v]
4. **Outdated Grid Infrastructure**: The current electrical grid lacks the transmission capacity and flexibility to handle rapidly growing AI workloads.

Different energy sources present varied timelines for deployment. Next-generation geothermal plants can potentially be deployed within three years, while small modular nuclear reactors are at least 6-10 years away. This mismatch between energy development timelines and the rapid pace of AI advancement creates a fundamental strategic challenge.[vi]

International Competition and Offshoring Risks

While the United States struggles with energy infrastructure, competitors are moving rapidly to capitalize on our challenges:

China's Aggressive Expansion: China continues its aggressive expansion of power generation capacity, adding 429 gigawatts in a single recent year.[vii] This energy advantage complements China's rapid advancement in AI capabilities, with models increasingly challenging the assumption of an 18-month U.S. technological lead.

Middle East Offshoring Concerns: Countries like the UAE have positioned themselves as potential destinations for AI infrastructure by leveraging abundant energy resources and rapid construction capabilities. The UAE's national strategy aims to have 20 percent of its non-oil GDP come from AI by 2031, with plans for multiple gigawatts of data center capacity by 2027.[viii]

This offshoring trend raises several significant national security concerns:

1. **Strategic Vulnerability**: Offshoring critical AI infrastructure creates dependency on foreign governments with different strategic interests and potentially less stable governance structures.
2. **Data Security Risks**: Processing sensitive data abroad increases exposure to foreign intelligence gathering and complicates the enforcement of U.S. data protection standards.
3. **Technology Transfer Challenges**: Physical proximity of advanced hardware to adversarial nations increases the risk of unauthorized technology transfer despite export controls. The UAE has deepened its relationship with China over the last several years, a trend that continues to accelerate.[ix]
4. **Diminished Innovation Ecosystem**: Fragmentation of our AI ecosystem across borders weakens the clustering effect that accelerates American innovation.

Strategic Export Controls

As the global race for AI supremacy intensifies, export controls represent a crucial tool for preserving America's technological advantage. Advanced AI chips constitute a strategic chokepoint in the AI supply chain that the U.S. still dominates. While models can be developed globally, the specialized hardware required to train and deploy advanced AI remains concentrated in U.S. hands.[x]

A promising approach is implementing a temporal advantage framework, where the United States maintains exclusive access to the most cutting-edge AI chips for a defined period before allowing controlled exports abroad. This approach provides U.S. researchers a meaningful time advantage while maintaining economic viability for U.S. chip manufacturers through eventual foreign sales, but must be implemented with rigorous safeguards against technology transfer risks.

Strategic Recommendations

1. **Accelerate Domestic Energy Infrastructure**:
    - Prioritize next-generation geothermal by reforming permitting processes and establishing Geothermal Opportunity Zones on federal lands
    - Streamline regulatory processes to reduce approval timelines from 4-5 years to under two years
    - Expedite licensing at pre-approved or retired nuclear sites and provide incentives for retrofitting existing natural gas plants with advanced emission controls

    - Modernize grid infrastructure by prioritizing baseload power projects in interconnection queues
2. **Strengthen AI Security and Export Frameworks**:
    - Develop robust security standards for foreign entities utilizing U.S. AI technology, with heightened requirements for countries with concerning governance or alignment issues
    - Implement a tiered export control framework that maintains America's technological edge while preventing hostile nations from accessing our most advanced systems
    - Enhance verification mechanisms to ensure exported AI hardware is used as intended, with immediate revocation capabilities for violations
    - Require regular security audits of foreign data centers housing U.S. technology, with U.S. personnel conducting on-site inspections
3. **Incentivize Domestic AI Infrastructure**:
    - Create AI infrastructure zones with expedited permitting and dedicated energy resources
    - Provide tax incentives for companies that maintain critical AI infrastructure within U.S. borders
    - Establish federal financing mechanisms for domestic AI data centers that meet heightened security standards
    - Develop a "National AI Security Framework" that includes both offensive and defensive capabilities
4. **Preserve and Enhance America's Five Pillars of AI Leadership**:
    - Algorithms: Increase federal funding for fundamental AI research and establish collaborative hubs linking national labs, academia, and industry
    - Data: Develop secure data sharing frameworks that maintain America's data advantage while protecting privacy and security
    - Energy: Address the immediate AI energy gap through the infrastructure acceleration measures outlined above
    - Chips: Expand CHIPS Act funding to ensure continued dominance in advanced semiconductor design and manufacturing
    - Talent: Reform immigration policies to streamline visa processes for AI researchers and implement retention programs for AI graduates from U.S. universities

Conclusion

The race for AI dominance is fundamentally a race to secure all five pillars of AI leadership: algorithms, data, energy, chips, and talent. While the United States currently maintains advantages in most of these domains, our critical vulnerability in energy infrastructure threatens to undermine our entire competitive position. Without

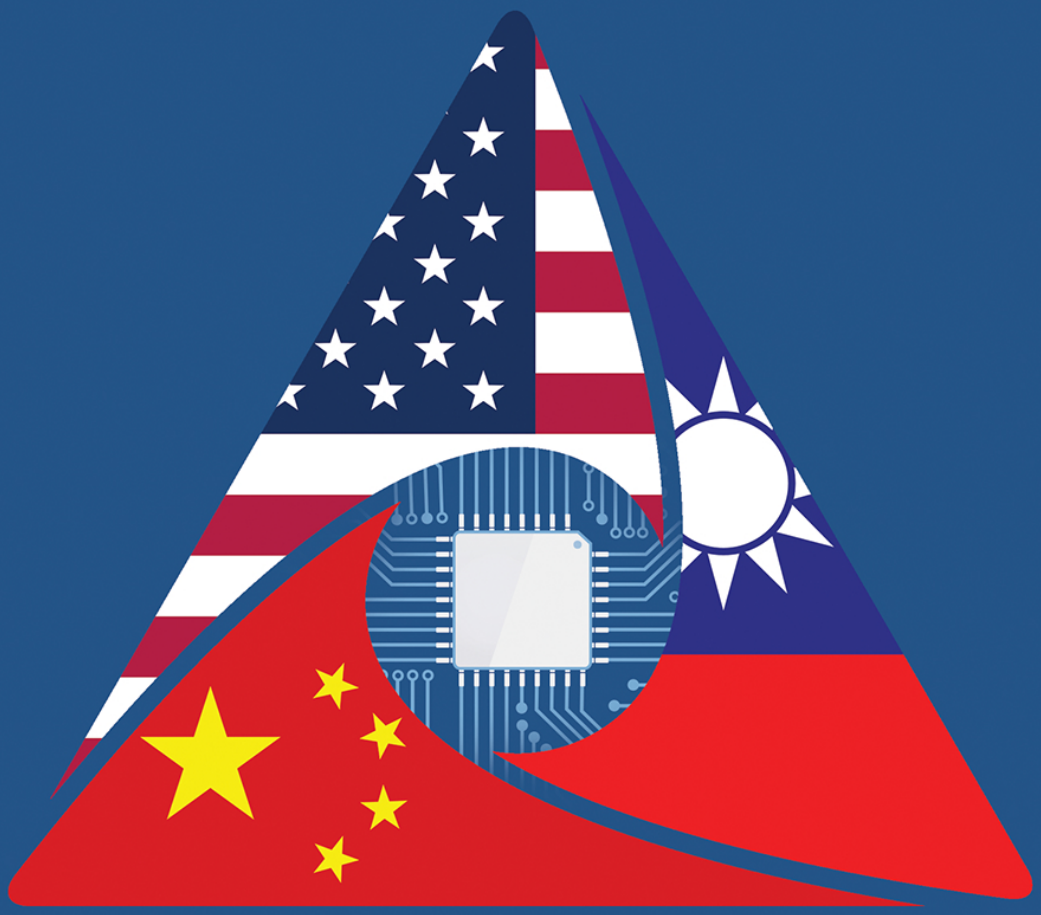
sufficient domestic energy capacity, we risk ceding our technological leadership to competitors or creating dangerous dependencies on foreign nations with whom our interests may diverge.

Addressing the "AI energy gap" is not merely an economic or infrastructure challenge but a national security imperative. A comprehensive national strategy must both accelerate energy infrastructure development and prevent the offshoring of critical AI capabilities through carefully calibrated export controls and domestic incentives. This approach requires balancing the need for international engagement with rigorous safeguards against technology transfer risks, particularly to regions with significant geopolitical complexities like the Middle East.

The decisions made in the next 2-3 years will likely determine whether the United States maintains its technological edge for decades to come or cedes leadership in the AI revolution. By securing the energy foundation for our AI future while simultaneously strengthening all five pillars of AI leadership, we can ensure America continues to lead in this transformative technology.

---

[i] Subin, S. (2025). Tech megacaps plan to spend more than $300 billion in 2025 as AI race intensifies.
[ii] Department of Energy. (2024). DOE Releases New Report Evaluating Increase in Electricity Demand from Data Centers.
[iii] Pilz, K. F., Mahmood, Y., & Heim, L. (2025). AI's Power Requirements Under Exponential Growth.
[iv] McGeady, C., Majkut, J., Harithas, B., Smith, K. (2025). The Electricity Supply Bottleneck on U.S. AI Dominance**.**
[v] Smith, B. (2025). The country needs more electricity — and more electricians.
[vi] Datta, A., and Fist, T. (2025). Compute in America.
[vii] Xinhua. (2025). China's installed power generation capacity up 14.6 pct in 2024
[viii] Allen, G. C., Adamson, G., Heim, L., Winter-Levy, S. (2025). The United Arab Emirates' AI Ambitions.
[ix] Hope, B. (2025). A Spymaster Sheikh Controls a $1.5 Trillion Fortune. He Wants to Use It to Dominate AI.; Lucente, A. (2024). China, UAE hold air force drills in Xinjiang as defense relations grow.; Han, Y., Li, Z. (2024). Golden era for China-UAE ties.
[x] Anthropic. (2025). Securing America's Compute Advantage: Anthropic's Position on the Diffusion Rule



Silicon Triangle

The United States, Taiwan, China, and Global Semiconductor Security

Edited by Larry Diamond, James O. Ellis Jr., and Orville Schell

CHAPTER NINE

Mitigating the Impact of China's Nonmarket Behavior in Semiconductors

ROBERT DALY AND MATTHEW TURPIN

The United States and its partners should be on guard to mitigate nonmarket behavior by China's emerging semiconductor firms.

While starting from a weak position, China's leaders are aggressively pursuing their domestic semiconductor aims—first to reduce the country's dependence on imports and then to take global market share through chip supply chain exports. As witnessed in a raft of other industries, the variety of government targets and subsidies to this end imply a high likelihood that semiconductor firms in China operating under nonmarket incentives may undercut pricing of established US and partner semiconductor firms.

This nonmarket behavior by semiconductor firms in China could have negative near-term impacts on US or partner producers, for example in mature chip production. And over time, it could create new US or partner dependencies on China-based supply chains that do not exist today, impinging on US strategic autonomy.

The US government has a variety of tools to monitor and limit the impact of such export dumping. It should also be concerned with the risk of its partners developing new dependencies on chips from China.

• • •

Semiconductors are ground-zero in this technological competition.
—SECRETARY GINA RAIMONDO[1]

Since China produced its first integrated circuit in 1965, its semiconductor policies have been shaped by its need for material and technological development, its drive for great-power status, its relations with the United States, and, especially since 2015, its quest for technological autonomy. As in other industries, China was willing to accept dependence on global semiconductor supply chains during an unavoidable period of tutelage and adaptation. As it mastered or obtained key technologies in the mid-2010s, however, China began a campaign intended to take it from dependence to dominance.

American export controls imposed in 2019 and then again in 2022 shocked China's planners and caused China's semiconductor industry to turn its focus from dominance to survival. Its current goals are, first, to master advanced-node design and manufacturing to shield itself from continued decoupling in high-tech sectors; and second, to protect its supply chains from the impact of possible future sanctions. Only if China succeeds in meeting its own demand for both mature and advanced semiconductors will its dreams of industry dominance return to the forefront of policy. In the interim, its goals are defensive, and the mood in China's semiconductor industry wavers between determination and desperation.

Warning Signs

Technology acquisition in the service of national development and military power has been China's primary goal in its relations with the United States since the Qing Dynasty sent students to the United States in 1872.[2] Their suspicions that the United States was denying China access to its leading technologies—and US suspicions regarding the ends and means of China's technological strategy—have been a mainstay of bilateral relations ever since.

Persistent US concerns—both economic and strategic—were heightened in 2006 when China announced its Indigenous Innovation agenda, which coincided with Beijing pressuring the European Union to lift its Tiananmen arms embargo.[3] Indigenous Innovation was not a secret program. When China's ministries announced detailed plans for the project

in 2009, it was hailed domestically as a comprehensive plan for industrial policy that would make the country "a technology powerhouse by 2020 and a global leader by 2050."[4] When foreign governments and corporations said the program was a threat to their interests and that China's methods violated global norms, Beijing seemed surprised and confused—China's leaders muted propaganda related to Indigenous Innovation but continued to implement the strategy at full force.

The pattern of declaration, blowback, and retrenchment was repeated in 2015 with the launch of Made in China 2025 (MiC 2025). MiC 2025 was a program of investment and research for China's corporations aimed at making the People's Republic of China (PRC) the world leader (defined as 70 percent of global market share) in ten industrial sectors: (1) information technology; (2) automated machine tools and robotics; (3) aerospace and aeronautical equipment; (4) maritime equipment and high-tech shipping; (5) modern rail transport equipment; (6) new-energy vehicles and equipment; (7) power equipment; (8) agricultural equipment; (9) new materials; and (10) biopharmaceuticals and advanced medical products. Though a source of pride for China, the program was viewed internationally as a brazen announcement that China would do whatever it took—relying on "discriminatory treatment of foreign investment, forced technology transfers, intellectual property (IP) theft, and cyber espionage"—to reduce China's dependence on the world and lock in the world's dependence on China.[5] Again, China seemed surprised by the criticism, as if its status as a strategically innocent developmental state was so firmly established that no one would question its motives. China's leaders spoke about the program less after 2018—but for foreign governments and corporations, the klaxon had already sounded.

Aptly named Military-Civil Fusion policies, which began in the 1990s, were another source of Western alarm. Instituted under the restrictions of the Tiananmen arms embargo, the program's goal was to achieve complete modernization of China's armed forces based on "informatization, intelligence, and mechanization" by 2027, the hundredth anniversary of the People's Liberation Army (PLA). Military-Civil Fusion required that any technology available to China's industry

or academia be provided to the PLA. It was not surprising that China would have such a policy. The Four Modernizations—first proclaimed by Zhou Enlai in 1963, later amplified by Deng Xiaoping as the core of China's development strategy—highlighted the essential integration of China's agriculture, industry, science and technology, and defense. China's whole-of-government (举国制度) approach was reflected in a series of National Intelligence Laws enacted under Xi Jinping that required all domestic entities, including universities, to give the state any information it requested.[6]

The strategic logic of these programs—Indigenous Innovation, MiC 2025, Military-Civil Fusion, and the National Intelligence Laws—was explained to the satisfaction of many US lawmakers, especially on the Republican side of the aisle, by Michael Pillsbury's *The Hundred-Year Marathon*.[7] Published in 2015, the book claimed that China has long had a plan to eclipse the United States and dominate a new global order. The same point was made (perhaps to a more Democratic readership) in Rush Doshi's *The Long Game: China's Grand Strategy to Displace American Order*.[8] Business communities in the United States and Europe both took notice, as evidenced by the publication of reports by the US and European chambers of commerce in early 2017 pointing out the harm PRC policies would do to their members.[9]

US bipartisan focus on the looming technology race and great-power competition was heightened by milestones reached and investments made under Xi Jinping during his first two terms as Party general secretary. Not only was China the most populous nation and largest exporter on Earth—it quickly became the world's largest producer and consumer of electric vehicles and batteries, as well as the global leader in mobile payments, wind and solar power generation, patents awarded, research cited in peer-reviewed journals, and training of college STEM (science, technology, engineering, and mathematics) students. It is the world's fastest-growing manufacturer of the legacy semiconductors used in most electronic devices and automobiles.[10] And China has invested heavily in the hardware that will drive the next generation of discovery (including supercomputers), the world's largest radio telescope (arguably underused), and one of the world's most advanced wind tunnels,

which Beijing uses to develop hypersonic weapons. In 2016, working with European partners, China launched the world's first quantum satellite, which completed a handshake with a quantum ground station.[11]

These advances all took place while China remained sanctioned under a comprehensive arms embargo by nearly all developed economies, as well as the target of multilateral dual-use export control regimes. As chapter 7 in this report notes, in the wake of the Cold War, the United States and its allies dismantled the Coordinating Committee for Multilateral Export Controls (COCOM) and replaced it with the Wassenaar Arrangement, which included states of the former Soviet Union and its Eastern Bloc satellites. Due to the Tiananmen arms embargo imposed on China in 1989, Beijing was not invited to join Wassenaar, and it still remains outside this multilateral regime.

Semiconductors—and the artificial intelligence (AI) and high-performance computing they enable—are essential to the PRC's commercial and military projects, as described in Indigenous Innovation, Made in China 2025, Military-Civil Fusion, and the National Intelligence Laws. China cannot achieve its MiC 2025 or military modernization goals, or master quantum computing, nanotechnology, or other emerging technologies, without a secure supply of advanced chips and without the designs, software, manufacturing equipment, and components needed to make them. Now that the era of US-China engagement is over, the problem for China is that no semiconductor supply chain can be secure unless it is within China, but most components of the advanced-semiconductor supply chain are in foreign—and especially US—hands.

Geopolitics/Geoeconomics

Semiconductors have once again become the key terrain of superpower rivalry, just as early semiconductors were in the rivalry with the Soviet Union.[12] This battleground, however, is a subset of a global contest between the superpowers, which has the hallmarks of a cold war. Long-term, comprehensive, "extreme" geopolitical competition between China and the United States will condition the strategies both sides

employ to win the semiconductor battle.[13] Put another way, the logic of security—not technological progress or economic efficiency—will drive the contest, even if tech and finance are its principal battlegrounds.

Beijing perceives an existential threat from a United States that wants to contain it or even bring down the Chinese Communist Party (CCP).[14] It, therefore, sees an urgent need to become more secure, not only in its high-tech industries but in its food supply,[15] culture,[16] biopharmaceutical sector, and media. Moreover, the West's rapid response to Russia's February 2022 invasion of Ukraine spurred China to sanctions-proof its economy. China's inclination toward decoupling did not begin with the semiconductor war or even the trade war that President Trump launched in 2018. Rather, self-sufficiency has been a keystone of CCP thinking since 1921, and many of China's modern industries have never coupled to the West in the first place. Until recently, however, China seemed confident that it could decouple selectively and at its own pace. That is no longer its plan, although it is unclear whether Beijing has fully considered the costs of this decision to rapidly decouple across a variety of sectors, or calculated its likelihood of success.

Washington's view now is that an expansion of China's economic and technological power is not in the interests of the United States or the rules-based international order. The United States, therefore, will no longer sell China the rope it needs to hang the United States in the global marketplace or on the battlefield. In the parlance of this report's strategic scenario work, Washington accepts a world moving to the "western" quadrants—and if that means hampering China's continued educational, scientific, medical, and economic progress by denying advanced chips and artificial intelligence to China's military, so be it. If it means greater scarcity and higher prices for US consumers, lower profits for US corporations, and the decoupling of global supply chains, so be it.

Popularized by President Trump and largely unquestioned by President Biden, antiglobalist narratives—as opposed to increasing market access among partners with common values—have prepared the ground for costly decoupling. These narratives appear to reflect a broader geopolitical trend. When the founder of Taiwan

Semiconductor Manufacturing Corporation (TSMC), Morris Chang, spoke at the Phoenix, Arizona, site of a new TSMC fabrication facility ("fab") in December 2022, he said, "Twenty-seven years have passed and [the semiconductor industry] witnessed a big change in the world, a big geopolitical situation change in the world. Globalization is almost dead and free trade is almost dead. A lot of people still wish they would come back, but I don't think they will be back."[17]

Even so, barring a direct military conflict between the United States and China, it is far more likely that the complexion of what we call "globalization" will simply shift over time, becoming characterized by a greater distribution of economic activity across more countries and regions. In many ways, we have mislabeled the last quarter century as a period of "globalization"—it was really a period of hyperconcentration in one country: China.[18]

Given that many of the unique geopolitical circumstances that led to this hyperconcentration of economic activity in China have ended, companies and countries will likely diversify their supply chains and manufacturing to places other than China. As this process unfolds, there will be relative gains and also significant costs, both of which will produce winners and losers. And as some have started to point out, China will likely lose more from this process.[19]

China's Ends, America's Means

Before 2019, Beijing's semiconductor policy focused on increasing China's global market share in every phase of production—from design to packaging—and producing more-advanced nodes. This agenda was pursued aggressively, but it was premised on gradually weaning Chinese producers off from foreign suppliers and then surpassing them. In other words, China was realistic about its dependence on the global supply chain—it was not looking so much to decouple immediately from US and third-country technologies as it was looking to reduce its dependence on them over time. The unstated assumptions of this approach were that foreign companies would remain as involved in the domestic market as China permitted them to be and that China could

be as integrated or as self-sufficient as its own capacities warranted. The attractiveness of China's vast market to tech multinationals would keep China in the driver's seat as long as the logic of technological progress and economic efficiency drove the semiconductor industry. That is to say, China assumed it would control the pace of decoupling to its advantage and that the rest of the world would be too dependent on China to prevent its success.

The placement of ZTE (in 2016) and Huawei (in 2019) on the Commerce Department's Entity List—subjecting them to US export controls—was a strong signal that Beijing's assumptions were wrong. Others could control the pace of decoupling, and China was not, in fact, the sole author of its technological future. This point was further underscored by the August 2022 passage of the CHIPS and Science Act. Also in August, Commerce banned the sale of electronic design automation software to China and informed chip designer Nvidia that, effective immediately, the company would need new licenses for the export to China of its A100 and H100 integrated circuits—both of which are essential to AI research and have a 95 percent market share in China.[20] Nvidia's DGX enterprise AI infrastructure systems (which incorporate A100 or H100) as well as "any future Nvidia integrated circuit achieving both peak performance and chip-to-chip I/O performance equal to or greater than . . . the A100, as well as any system that includes those circuits," were also covered by the order.[21] This move banned not only the sale of Nvidia's advanced graphics processing units (GPUs), but also any product of Advanced Micro Devices (AMD) or other American fabless chip design companies whose technology met the criteria detailed in the order. It ripped away the foundation on which China's AI and data analysis strategies had been built years before China was ready to stand on its own.

While the export controls of August 2022 were, as Gregory Allen of the Center for Strategic and International Studies (CSIS) wrote, aimed at "strangling large segments of the Chinese technology industry . . . with an intent to kill,"[22] from the US perspective they were actually restrained, as they left additional steps in the escalation ladder available to the United States. Rather than seeking a complete technological

decoupling from China, the Biden administration's policy has sought to limit its controls to chips that train AI models with advanced military applications. That delicacy may not have been noticed by China, however, as it has no "immediate substitute for the Nvidia GPUs that train AI models for autonomous driving, semantic analysis, image recognition, weather variables, and big data analysis," and every buyer in China will be affected by the new rules.[23]

One of the difficulties for Nvidia and other US suppliers is that they have no immediate substitute for the China market. In the third quarter of 2022, Nvidia "had booked $400 million in sales of the affected chips . . . to China that could be lost if [Chinese] firms decide not to buy alternative Nvidia products."[24] That said, the impacts on these companies should not be viewed in isolation; China's loss of its pathway to technological superiority in advanced chips would generate national security and economic competitiveness costs that would dwarf the affected sales of companies like Nvidia.

If the Nvidia announcement destabilized the train of China's semiconductor strategy, changes in export controls announced by the Department of Commerce's Bureau of Industry and Security (BIS) on October 7, 2022, knocked it off the rails. The BIS rules on advanced computing and semiconductor manufacturing added new license requirements for any US products sent to China's fabs that support the domestic building of logic chips of 14nm or below, DRAM memory chips of 18nm half-pitch or less, or NAND Flash memory chips with 128 layers or more. As Gregory Allen explained, Biden was attempting to

> (1) strangle the Chinese AI industry by choking off access to high-end AI chips; (2) block China from designing AI chips domestically by choking off China's access to US-made chip design software; (3) block China from manufacturing advanced chips by choking off access to US-built semiconductor manufacturing equipment; and (4) block China from domestically producing semiconductor manufacturing equipment by choking off access to US-built components.[25]

The rules also restricted the ability of unlicensed US citizens or green card holders to support the design or production of advanced chips in China's fabrication facilities.[26] This class of restrictions meant that hundreds of Americans employed by the industry in China (no exact number is yet available), including forty-three senior executives, had to quit working immediately. Many of these executives were naturalized American citizens of Chinese origin with advanced degrees from the United States and long experience in Silicon Valley.[27]

China's Response

After the October 2022 export controls were released, China's strategy of steadily progressing toward industry dominance on its own timeline, with an assumption of ready access to foreign technology and talent along the way, had to be scrapped. Because the CCP's 20th National Congress closely followed the announcement—and itself was followed by a series of economic and social crises related to Xi Jinping's "dynamic zero"-COVID policy—it was not clear by year's end that Beijing had fully absorbed the impacts of the new export controls.

When Beijing felt attacked by US actions during the Trump administration, its response was to mirror US actions immediately. It made such shows of strength throughout the trade war, for example, when the United States required Chinese media outlets to register as foreign missions and when the PRC consulate in Houston was suddenly shut down in 2020. Given this tendency to counterpunch, some commentators expected China to hit back against the new US rules by banning the sale to the United States of products such as rare earths, medicine and medical precursors, or legacy chips. On a number of occasions involving science and technology over the last five to ten years, however, China lacked the leverage or capability to successfully respond. For example, a little more than a year after Huawei's Entity Listing, the National People's Congress passed and adopted the Export Control Law of China (ECL) in an effort to mirror US capabilities and deny China's advanced technologies to the United States.[28] Like the US Export Administration Regulations (EAR), which provide the legal basis for Commerce's and the State

Department's export controls, China's 2020 ECL establishes extraterritorial reach, directs the creation of control lists and blacklists, and defines controls for dual-use items and military products. Unfortunately for Beijing, this legislation remains an empty regulatory shell, as China lacks control over advanced technologies that surpass what is available to its rivals. One could imagine a future where Beijing responds in this domain with true reciprocity, but that time has not arrived.

To date, rather than hitting back against American export controls, China has adopted five broad, long-term strategies aimed at limiting their impact and, if possible, advancing its drive for technological security and dominance:

1. *Increasing investment* in China's semiconductor companies, large and small; in training personnel; and in building design and manufacturing hubs
2. *Encouraging workarounds* to existing technologies
3. *Discouraging third countries* from working with the United States
4. *Playing for time* in the hope that the costs of decoupling, the interest of US corporations, and pressure from US partners result in the watering down of export controls
5. *Controlling the international narrative* on technological decoupling

Strategy One: Increased Investment

China's commitment to achieving dominance in the semiconductor industry, based on the size of its domestic market and investment in its companies and universities, coincided with American policy makers' understanding of the challenge Beijing was posing.[29] As outlined in chapter 8 of this report, the current drive to fund the industry was launched in 2014.[30] In that year, China published its Guideline for the Promotion of the Development of the National Integrated Circuit Industry, "with the goal of establishing a world-leading semiconductor industry in all areas of the integrated circuit supply chain by 2030."[31] It also established the National Integrated Circuit Industry Investment

Fund (or "Big Fund") to provide an estimated $150 billion in state funds to support research. By 2020, China was home to more than ten thousand semiconductor companies,[32] a figure that more than doubled over the course of that same year.[33] Many of these enterprises were overnight operations that existed primarily to chase government subsidies. Some, like Tsinghua Unigroup, a company founded at Xi Jinping's alma mater that even bid to buy Micron in 2015 for $23 billion, were spectacular failures that spotlighted the waste that remains endemic in China's government investment programs.[34] Tsinghua Unigroup had received tens of billions of dollars in government support but still defaulted on its bonds in 2020. Others, like Wuhan's Yangtze Memory Technologies Co. (YMTC), which was founded in 2016 and is now China's leading memory chip maker, were spectacular successes.[35] TechInsights, a Canadian semiconductor and microelectronics analytics company, recently declared that "at their current rate of innovation, YMTC is poised to be the uncontested global NAND flash technology leader before 2030."[36] China's latest Five-Year Plan, unveiled in July 2021, committed to raising public and private R&D spending by 7 percent annually—a rate greater than the increase in its military spending—with semiconductors as a top priority.[37]

It is too soon to predict the scale at which Beijing will further increase its investments in the industry, but the speed with which major Chinese municipalities responded to the October 2022 export controls indicates that a major reinvestment is under way. In late October 2022, the Lingang Special Area (a free-trade zone), Shanghai University, and the city's Integrated Circuit Industry Association—all shocked by the BIS ban on US persons in China's semiconductor companies and buoyed by grants from the municipal government—set up a new campus to foster talent for the semiconductor industry.[38] Such training efforts garnered government support despite China's overall success in developing STEM talent broadly.

According to Georgetown University's Center for Security and Emerging Technology (CSET), "by 2025 Chinese universities will produce more than 77,000 STEM PhD graduates per year compared to approximately 40,000 in the United States. If international students

are excluded from the US count, Chinese STEM PhD graduates would outnumber their US counterparts more than three-to-one."[39] Even so, that advantage may not be of much help in the semiconductor industry. The China Semiconductor Industry Association anticipates that China already has a shortage of two hundred thousand semiconductor engineers for the years 2022 and 2023, while one of China's leading educational talent organizations reports that most STEM students prefer work in AI and big data over the lower-paying semiconductor industry (ironically mirroring a trend observed among US STEM graduates, as outlined earlier in this report).[40]

In Shenzhen, the municipal government announced plans to reinvest in its semiconductor industry architecture on October 8, 2022, one day after BIS's bombshell. The city's Development and Reform Commission announced that it would cover 20 percent, or up to US$1.4 million annually, to subsidize the R&D expenses of companies chasing breakthroughs in the design and development of logic chips, including CPUs (central processing units) and GPUs.[41] Huawei, which is based in Shenzhen, is leveraging the established networks and talent in that city to invest in firms throughout China, including NAURA Technology Group (China's leading chipmaking equipment manufacturer), to build itself a complete China-only supply chain. The Fujian Jinhua Integrated Circuit Corporation (JHICC)—after being driven into bankruptcy in early 2019 after the Trump administration placed it on the Entity List in 2018 for stealing intellectual property from Micron Technology—has been resurrected to play a major role in this network.[42] Huawei engineers are reported to be working stealthily in JHICC's Quanzhou plant to help the telecom giant recover from its own placement on the Entity List in 2019[43]—albeit neither Huawei's nor JHICC's engineers have access to the most-advanced software, tools, or components that would help them to achieve these objectives.

Strategy Two: Work-Arounds

Writing in *American Affairs*, Geoffrey Cain claims that China's failure thus far to meet its MiC 2025 goals for chip development stems from

its deeply entrenched "diplomatic isolation . . . oppressive top-down mandate(s) of selecting national champions . . . the weak position of starting generations behind industry leaders in America, Taiwan, South Korea, and Japan," and corruption.[44] Within China, most domestic commentators are similarly pessimistic about China's prospects for building an indigenous cutting-edge semiconductor supply chain using existing technologies. China is therefore searching for new technologies that can match the performance of systems developed and controlled by Western-oriented competitors.

For example, the Beijing Open Source Chip Research Institute—a group of research centers and companies that includes the Chinese Academy of Sciences, Tencent, and Alibaba[45]—is developing domestic semiconductor-related intellectual property using the RISC-V open-source chip design architecture created by the University of California, Berkeley. If it succeeds, the group's Xiangshan RISC-V architecture could free China from IP constraints imposed by ARM, the Cambridge-based company whose technology underlies most cell phones, including Apple products.[46] China may also hope to offset the need for US-designed advanced nodes by developing photonic chips (which use photons instead of electrons in integrated circuits[47]) and experimenting with nonsilicon substrates, such as cubic boron arsenide, graphene,[48] and silicon carbide.[49] As described in chapter 2 of this report, however, marketable breakthroughs in any of these areas are likely decades off, and China's pace of advancement even here may face acute threats after its stockpiles of banned chips, components, and manufacturing tools run out or require repairs in the next year or two.

Strategy Three: Outreach to US Allies

The ubiquity of essential US semiconductor designs, software, manufacturing tools, and components in the global supply chain makes it possible for the Department of Commerce to use its Entity List and Foreign-Direct Product Rule to compel allies and partners to support its ban on cooperation with China's semiconductor industry.[50] The Netherlands, Taiwan, South Korea, Japan, and most other suppliers

share US concerns about China's threats to security, intellectual property, and global order—but they value their trade relations with China highly. China will be alert to any opportunities that such conflict provides to sow division within US partnerships and gain the chips and chip manufacturing equipment it needs to develop its industries and military.

China accounts for over 25 percent of the annual global demand for semiconductor equipment. It would doubtless buy as many of Advanced Semiconductor Materials Lithography's (ASML) $100 million extreme ultraviolet (EUV) lithography machines as the Dutch company could sell it, but the Netherlands agreed in 2016 that none of ASML's high-end machines would be sold to China. Bloomberg reported on December 7, 2022, that Amsterdam had agreed to enforce Washington's October 2022 export controls as well.[51] ASML will continue to sell its mature-node manufacturing equipment to China, however, and the knowledge that China is its greatest potential profit center will continue to nag at ASML's leadership, despite the firm's claim that under current market conditions, it can sell as many machines as it can produce to other customers.[52] **Text**

America's Asian partners in the "Chip 4" alliance will likely fall in line as well—but doing so will be costly for them, and China will exert as much pressure on them as it can to seek carve-outs and workarounds to US requirements. As outlined in chapter 6, US partners have their own substantial semiconductor supply chain strengths and ambitions, with sales to or production in China as part of them. In 2021, Taiwan's chip sales to China, worth $155 billion, constituted 62 percent of its exports to the mainland. The latest data, however, shows that Taiwan's export of chips to China and Hong Kong fell for a fourth month in a row in February 2023—a 31 percent drop in exports from a year earlier.[53] Semiconductor manufacturing machines and materials are Japan's second-largest export, and one-third of them are purchased by China—a trade worth $9.5 billion to Japan in 2021.[54] China buys 43 percent of South Korea's exported chips—58 percent including exports to Hong Kong—a trade worth over $49 billion ($66 billion including Hong Kong) to South Korea in 2022.[55] The US Commerce Department recently granted Samsung and SK hynix exceptions to its

export controls, allowing them to provide otherwise banned capabilities to their facilities in China for one more year—but it is not likely that those exceptions will be granted again.

Taipei, Tokyo, and Seoul are all likely to be courted, hectored, coerced, and threatened by Beijing as they move toward full compliance with BIS rules. They may also compensate for cooperating with the United States on semiconductors by reassuring Beijing in other aspects of their political and trade relations, and Beijing will be attentive to such opportunities to weaken the will of, and widen the divisions between, America's Asian partners.

Assiduous attention to alliance management, therefore, will be essential to the success of US policy. Here again, we run across a ubiquitous theme of this report: the sustainability of US security-oriented efforts toward China will rely on the commercial attractiveness that the United States can offer its partners. Making the subsidies through the CHIPS and Science Act attractive to allied partners—and not saddled by non-security-related short-term US social or protectionist politics—is the first step.[56] Beyond those five years, like-minded partners need confidence that the United States will continue to offer market access and bidirectional investment.

Strategy Four: Play for Time

China domestic companies' most effective responses to US pressure may be to stockpile chips and equipment while they are still available (Nvidia, for example, will continue to ship AI chips from its Hong Kong logistics center through September 2023[57]); manage their capital reserves to weather the current slowdown in global chip demand; and hope that the current storm passes. At the moment, the United States' position seems certain, but its adamancy may not last. A change in administration in 2024 could also bring a change in priorities. Or the United States might blanch as the ban's full costs to US companies become clear. AMD, Intel, Nvidia, and Qualcomm all have enormous stakes in sales to China, as do US semiconductor manufacturing equipment companies such as Applied Materials, KLA, and Lam Research.[58]

Even though most US multinationals no longer lobby for expanded trade with China (as they did in the run-up to the PRC's ascension to the World Trade Organization [WTO] in 2001), executives and their shareholders are bound to ask Washington to take some of the roughest edges off its export controls.

Only two months after October 7, 2022, China already saw signs of a thaw in the American position and an opportunity to import advanced chips despite the export controls. Under the new BIS rules, thirty-one companies in China, including YMTC and NAURA Technology, were placed on an "unverified list" and given sixty days to prove that no controlled items they imported from the United States could be used in weapons manufacture or transferred to China's military. "Verification" involves on-site inspection of companies in China by US officials who conduct "end-use checks." Historically, the CCP has viewed these procedures as insults to its sovereignty and has refused the necessary access to Americans. During a December 6 event at the Center for Strategic and International Studies, however, Alan Estevez, the under secretary of commerce for industry and security, said that China's Ministry of Commerce had been cooperating on end-use checks since November, raising the possibility that firms currently on the unverified list might be verified as good actors and would therefore be eligible to import advanced US chips and equipment.

The United States has assumed, reasonably, that China's Military-Civil Fusion program and National Intelligence Laws were proof—if proof were needed at all—that any technology available anywhere in China that had a military application was certain to be put to that use. As the US-China rivalry expands and as military conflict becomes more imaginable, that assumption might seem to imply that US enforcement of export controls on China should be absolute and unwavering. Estevez's comments suggest, however, that China may now see a glimmer of light: cooperating with Commerce's end-use checks to get firms off the unverified list and stalling may be its best short-term strategy to keep open a channel for technology imports.

Despite this potential for near-term churn, over the long term, time may arguably be on the side of the United States and its allies in

this realm. If—as characterized in the strategic scenario planning of chapter 1—trends toward supply chain diversification continue and companies like Apple reduce their dependency on China's manufacturing base and market, then the leverage Beijing now applies to get access to technology from foreign companies could dissipate.[59] As the world shifts from hyperconcentration to a more dispersed distribution of high-tech manufacturing with fewer dependencies on the PRC, then companies will have less incentive to place advanced capabilities in China. The current commercial logic for providing advanced-chip capabilities to China is that much of the world's electronics manufacturing takes place in China. As that condition changes, so too will the commercial rationale for providing the advanced chips. South and Southeast Asian nations likely stand to be the true beneficiaries of these trends. Manufacturing jobs and the attendant flows of infrastructure funding, science and technology know-how, and economic development will flow to them just as those same benefits flowed to the PRC over the past quarter century. Rather than being the grass trampled between two competing superpowers, the nearly 2.2 billion people of South and Southeast Asia could experience a dramatic increase in economic growth and prosperity.

Strategy Five: Frame Narratives

Building "discourse power" (话语权) is an essential component of China's "comprehensive national power" (国家综合势力). On September 1, 2022, after the announcement of restrictions on the sale of Nvidia GPUs to China, Foreign Ministry spokesman Wang Wenbin said:

> The US has been stretching the concept of national security and abusing state power. The US seeks to use its technological prowess as an advantage to hobble and suppress the development of emerging markets and developing countries. This violates the rules of the market economy, undermines international economic and trade order, and disrupts the stability of global industrial and supply chains.[60]

On October 8, Foreign Ministry spokesperson Mao Ning argued:

> In order to maintain its sci-tech hegemony, the US has been abusing export control measures to wantonly block and hobble Chinese enterprises. Such practice runs counter to the principle of fair competition and international trade rules. It will not only harm Chinese companies' legitimate rights and interests, but also hurt the interests of US companies. It will hinder international sci-tech exchange and trade cooperation, and deal a blow to global industrial and supply chains and world economic recovery.[61]

Such statements do not aim to convince Washington to change its policies. They are intended, first, to persuade the Chinese people that China is an innocent and righteous victim of a malign United States; and, second, to persuade third countries—the Global South and nondemocratic partners of China in particular—that the United States is a bully to developing nations and a threat to global order. These messages are conveyed around the world by the state-run broadcaster China Global Television Network (CGTN), which is a leading provider of news in Africa and the Pacific Islands.[62] China's critique of the United States has also gained traction in the Middle East, Latin America, and many countries that participate in the Belt and Road Initiative.

China has prepared domestic and foreign audiences to be receptive to these messages about the technology war by promulgating a master narrative over the past ten years—a narrative that forms the backbone of its rebuttals to the United States: *The United States has fundamentally misperceived China's intentions and policies because it fears that China's peaceful, globally beneficial rise and the success of its governance model threaten its own hegemony.* Global public opinion polling indicates, however, that China's well-resourced, carefully planned global public diplomacy campaign has had mixed results at best.[63] In developed democracies, it has failed entirely, but it has adherents in the Global South, where it is largely unchallenged by US messages.

Cowed but Unbowed

In addition to these five observable responses to the imposition of export controls, it would be wise to assume that China's established technology-acquisition methods have accelerated since 2022. These include IP theft, hacking campaigns, digital and traditional espionage, talent recruitment programs such as the Thousand Talents Plan, recruitment of third-country technology experts, and global influence operations designed to spread PRC narratives among foreign publics, including diaspora Chinese.

The PRC government was angered, but not surprised, by the United States' determined prosecution of a tech war in 2022. The Ministry of Commerce's cooperation with US end-use checks indicates that BIS now has Beijing's full attention, and many of China's semiconductor companies are desperate. Many will go under. It is too soon to predict the course of these developments, but it is already clear that China is adjusting in an attempt to limit damage; it is not reconsidering its national goals, however, and it has not used all of the weapons at its disposal.

Beijing is unlikely to abandon its dual objectives to assume a leadership position in the development of cutting-edge semiconductors and to become self-sufficient in the production of semiconductors for broader use. As outlined in this chapter, the first objective has become more difficult to achieve, given the actions taken by the United States and the likelihood that the United States can persuade others to squeeze the semiconductor choke points. China will seek to find work-arounds to these restrictions, but it appears that the United States is paying close attention to China's actions and has sufficient regulatory escalation space to continue to stymie Beijing. In pursuit of the second objective, however, state subsidies and other forms of encouragement now give China a path to build an increasingly dominant position in the manufacture of legacy chips. While economic on the surface, this pursuit will nonetheless also have important national security implications that the United States and its partners must consider.

The Next Challenge

Going forward, the United States and its partners must design policies to deal with two interrelated challenges caused by China's semiconductor industrial policies.

The first is *military*. The United States cannot afford to lose the unequal technological advantages it has long enjoyed. In an era in which a US-China conflict is becoming more likely, the United States will derive qualitative military advantages by denying the most-advanced semiconductors and AI applications to China.

The second is *economic*. Even if US export controls are enforced and expanded, China may be able to generate an overcapacity of legacy chips and dominate the global market for semiconductors that go into household appliances, automobiles, and the internet-of-things. Such dominance will create political and economic leverage for China, as its near monopoly on rare earth extraction and refining already do. As China floods global markets with low-cost, good-enough mature chips, the ability of the United States and other countries to manufacture them will be degraded, along with the profit margins that fuel further commercial R&D for the next generation of products. China's profits from legacy chips will be used to offset the impact of US export controls through greater investment in the education and research needed to design and manufacture advanced nodes.

The Biden administration's formally stated rationale for the ban on the sale of advanced chips, design software, manufacturing equipment, and components to China is that these technologies are employed in weapons that target the United States and in surveillance systems used to monitor and persecute Chinese citizens. But the economic arguments for limiting Chinese dominance of mature- and advanced-node markets are almost equally strong. If China achieves the goals it has set for its semiconductor industry, the global risks of technological lock-in and innovation drag are high. The instructive example is China's dominance of solar panel production. Studies by the Information Technology and Innovation Foundation[64] argue that, once China pushed other manufacturers out of the solar panel market, innovation in this young

and vital technology sector all but ceased.[65] Chinese panel production, dominated by national champion companies controlled by the CCP, had neither the motivation nor the ability to develop the technology further. The same is possible if China dominates chip design and manufacture, particularly if done primarily through subsidized state-oriented enterprises.

China is, in fact, on track to become a major producer of legacy chips. If its behavior in other industry sectors is a model for its actions in legacy semiconductors, the world should expect massive overcapacity of these older chips, which would collapse the price for every other producer. Consumers who purchase commercial electronics will benefit from marginally lower prices, but Beijing's dumping of subsidized semiconductors will severely undermine companies that currently produce legacy chips in South Korea, Taiwan, Japan, the United States, Europe, and the Middle East. Those companies will lose the revenue needed to make capital improvements, as well as the revenue to conduct R&D for the next generation of semiconductors. This all could cause a consolidation of semiconductor manufacturers whereby foreign fabless chip design companies become increasingly dependent on mature-node PRC fabrication facilities. This dependency does not exist today.

Commercial consolidation and increased dependency on Chinese fabs for legacy semiconductors will have important national security implications. As outlined in chapter 2 of this report, advanced chips are crucial to military superiority—but the majority of semiconductors used in defense applications are legacy chips, drawn from both dedicated (for sensitive applications or chips with special attributes like radiation hardening) and off-the-shelf commercial chip suppliers. Losing access to a healthy global ecosystem of friendly commercial suppliers of mature chips could increase costs or drive the defense industrial base to rely on single-source producers, limiting innovation. While the defense industry may seem large, it is dwarfed by the commercial sector for legacy semiconductors. And even if countries can avoid dependencies on China for legacy chips in their defense industries, the wider economy will likely fall victim to overcapacity and dumping of legacy chips.

One potential mitigation against the worst harms of Beijing's semiconductor industrial policy would be to take preemptive action and impose antidumping/countervailing duties (AD/CVDs) on China-manufactured chips immediately. Traditionally, countries like the United States impose AD/CVDs only after the harm of dumping has taken place—that is, once companies go bankrupt and employees are laid off. Given the track record of China's industrial policies, however, the United States and other countries should act proactively by imposing those duties now, which would prevent Beijing's semiconductor policy from harming domestic chip manufacturers. Should those duties be insufficient, countries could also block the importation of China-manufactured legacy semiconductors. This move could force electronics manufacturers to require non-PRC legacy chips or further shift the manufacture of electronics outside the PRC.

While such actions would likely lead Beijing to bring suit at the WTO, China would be making these arguments in bad faith, given China's failure to fulfill its own obligations to other members of the WTO and the harm done to the global trade system in the process.[66] The United States and other countries should not shy away from confronting Beijing on this issue—to repeat a phrase that Chinese Foreign Ministry spokesman Zhao Lijian often deploys (albeit against Western nations), China's protest to the WTO would be like "a thief crying 'stop the thief' (贼喊捉贼)."[67]

While this threat may seem further off than the one posed by the acquisition and production of advanced chips, failure to take actions like these in the short term could endanger US abilities to constrain PRC efforts to develop cutting-edge semiconductors in the medium term. The semiconductor industry is first and foremost a commercial industry that is shaped by market forces, and it is hard to predict just how damaging Beijing's dumping of legacy chips would be to the health of the broader industry—particularly to those companies that spend massive amounts of money on building new fabs, buying new and more advanced tools, and investing in R&D. While it is possible that the effects of legacy chip dumping could be isolated to only a small number

of semiconductor companies, it is also possible that there would be a contagion effect that would weaken even the most advanced manufacturers. Given these uncertainties, the United States and its allies should err on the side of strenuous and well-coordinated actions against Beijing's plans. It is understandable that companies and governments would want to take the least costly action—but again, given the complex commercial, geopolitical, and technological dynamics, it is nearly impossible to predict with accuracy what the perfect balance will be. In this critical and fast-moving sector, we should pursue an "all of the above" approach that seeks to deny China the capability to achieve its objectives. Under these conditions, we advocate being more exclusionary rather than less.

Would pursuing this approach encourage Beijing to double down on its objectives? If so, should we instead moderate our response to reassure Beijing and persuade them not to pursue their goals? To date, the United States and its allies have had a poor track record in reassuring the PRC and persuading it to abandon goals that undermine our interests. It would be naïve to place our faith in our powers of persuasion yet again. Rather than trying to reassure China, we should focus on a strategy of denial. That is the strategy that the October 2022 rules announced. Having crossed that Rubicon and knowing that China is now gearing up to compete with the United States on those terms, the time for cautious gradualism has passed.

In short, meeting the two challenges—military and economic—posed by China's semiconductor policies will require different tools, different groups of partners, and different strategies. The complexity of pursuing and coordinating these strategies, and the scale of investment and intensity of diplomacy required to succeed, will require government direction. It can't be left to the market, as the primary measure of success will not be profit. The United States' task is to hamper China's development of advanced AI that could help it win wars by restricting China's access to the world's most powerful chips—*without incentivizing its dominance of legacy semiconductor markets worldwide by doing so*.

NOTES

1. US Department of Commerce, "Remarks by US Secretary of Commerce Gina Raimondo on the US Competitiveness and the China Challenge," November 30, 2022.
2. See Robert Daly, "Thinkers, Builders, Symbols, Spies?: Sino-US Higher Educational Relations in the Engagement Era," in *Engaging China*, edited by Anne F. Thurston (New York: Columbia University Press, July 2021).
3. Peng Heyue, "China's Indigenous Innovation Policy and Its Effect on Foreign Intellectual Property Rights Holders," China Law Insight, September 9, 2010; and Kristin Archick, Richard Grimmett, and Shirley Kan, "European Union's Arms Embargo on China: Implications and Options for US Policy," Congressional Research Service, January 26, 2006.
4. James MacGregor, "China's Drive for 'Indigenous Innovation': A Web of Industrial Policies," APCO Worldwide, 2009.
5. James McBride and Andrew Chatzky, "Is 'Made in China 2025' a Threat to Global Trade?," Council on Foreign Relations, May 13, 2019.
6. Murray Scot Tanner, "Beijing's New National Intelligence Law: From Defense to Offense," Lawfare, July 20, 2017.
7. Michael Pillsbury, *The Hundred-Year Marathon: China's Secret Strategy to Replace America as the Global Superpower* (New York: Henry Holt and Co., 2015).
8. Rush Doshi, *The Long Game: China's Grand Strategy to Displace American Order* (New York: Oxford University Press, 2021).
9. European Union Chamber of Commerce in China, "China Manufacturing 2025: Putting Industrial Policy ahead of Markets," March 7, 2017; and US Chamber of Commerce, "Made in China 2025: Global Ambitions Built on Local Protections," March 16, 2017.
10. Semiconductor Industry Association, "China's Share of Global Chip Sales Now Surpasses Taiwan's, Closing In on Europe's and Japan's," January 10, 2022.
11. Karen Kwon, "China Reaches New Milestone in Space-Based Quantum Communications," *Scientific American*, June 25, 2020.
12. US General Accounting Office, "Export Controls: US Policies and Procedures Regarding the Soviet Union," May 24, 1990.
13. AP News, "Biden: China Should Expect 'Extreme Competition' from US," February 7, 2021.
14. Simultaneously, the United States perceives the PRC as "the only competitor with both the intent to reshape the international order and, increasingly, the economic, diplomatic, military, and technological power to advance that objective," replacing the liberal, rules-based order with an international system

that privileges authoritarian regimes. See the White House, Executive Office of the President, "National Security Strategy," October 2022.

15. Asim Anand, "What Xi Jinping Brings to the Table in China's Quest for Food Security," S&P Global, November 17, 2022.
16. Neil Renwick and Qing Cao, "China's Cultural Soft Power: An Emerging National Cultural Security Discourse," *American Journal of Chinese Studies* 15, no. 2 (January 2008): 69–86.
17. Cheng Ting-Fang, "TSMC Founder Morris Chang Says Globalization 'Almost Dead,'" *Nikkei Asia*, December 8, 2022.
18. As of 2015, the PRC produced or assembled 28 percent of the world's automobiles; 41 percent of the world's ships; more than 80 percent of the world's computers; more than 90 percent of the world's mobile phones; 60 percent of the world's color TV sets; more than 50 percent of the world's refrigerators; 80 percent of the world's air conditioners; and 50 percent of the world's steel. See European Chamber of Commerce in China, "China Manufacturing 2025: Putting Industrial Policy Ahead of Market Forces," March 2017.
19. George Magnus, "Why China Has More to Lose from Decoupling than the US," *South China Morning Post*, June 29, 2022; Minxin Pei, "China Can't Afford to Decouple from the West," *Bloomberg*, January 30, 2022; Kinling Lo, "Tech War: Beijing Will Come Out of Decoupling Worse Off than the US, Say Chinese Academics," *South China Morning Post*, February 1, 2022; and Shen Lu, "A Report Detailed the Tech Gap between China and the US—Then It Disappeared," *Protocol*, February 9, 2022.
20. Debby Wu, Ian King, and Vlad Slavov, "US Deals Heavy Blow to China Tech Ambitions with Nvidia Chip Ban," *Bloomberg*, December 2, 2022.
21. Nvidia, "Form 8-K," Securities and Exchange Commission, August 26, 2022.
22. Gregory C. Allen, "Choking Off China's Access to the Future of AI," CSIS, October 11, 2022.
23. Che Pan, "Tech War: Why the US Nvidia Chip Ban Is a Direct Threat to Beijing's Artificial Intelligence Ambitions," South China Morning Post, September 12, 2022.
24. Stephen Nellis and Jane Lee, "US Officials Order Nvidia to Halt Sales of Top AI Chips to China," Reuters, August 31, 2022.
25. Allen, "Choking Off China's Access to the Future of AI."
26. US Department of Commerce, Bureau of Industry and Security, "Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China (PRC)," October 7, 2022.
27. Liza Lin and Karen Hao, "American Executives in Limbo at Chinese Chip Companies After US Ban," *Wall Street Journal*, October 16, 2022.

28. Yujing Shu and Xiaotang Wang, "China Overhauls Its Export Control Regime: What China's New Export Control Law Changes and How to Respond," K&L Gates, December 7, 2020.
29. Within eighteen months of the PRC's launching of significant investments in semiconductors, the Obama administration published a strategy for dealing with the problem that carried over to the Trump administration (see *Report to the President Ensuring Long-Term US Leadership in Semiconductors*, Executive Office of the President, President's Council of Advisors on Science and Technology, January 2017), cabinet secretaries were making public speeches about the challenge (see "Semiconductors and the Future of the Tech Economy," speech by Secretary of Commerce Penny Pritzker, CSIS, November 2, 2016), and the United States had already blocked Chinese acquisitions of semiconductor companies in the United States and Europe (see Paul Mozur, "Obama Moves to Block Chinese Acquisition of a German Chip Maker," *New York Times*, December 2, 2016; and Ana Swanson, "Trump Blocks China-Backed Bid to Buy US Chip Maker," *New York Times*, September 13, 2017).
30. Paul Mozur, "Using Cash and Pressure, China Builds Its Chip Industry," *New York Times*, October 27, 2014.
31. Congressional Research Service, "China's New Semiconductor Policies: Issues for Congress," April 20, 2021.
32. Kathryn Hille and Sun Yu, "Chinese Firms Go from Fish to Chips in New Great Leap Forward," *Financial Review*, October 13, 2020.
33. *New York Times* ("The Failure of China's Microchip Giant Tests Beijing's Tech Ambitions," July 19, 2021) puts the number at 58,000, while the *Financial Times* ("Chinese Firms Go from Fish to Chips in New Great Leap Forward," reprinted in *Financial Review*, October 13, 2020) claims the number is 13,000. Both reports seem to cite the same Chinese government study.
34. *New York Times*, "The Failure of China's Microchip Giant."
35. At least until the United States began directing regulatory firepower against YMTC in October 2022. See Karen Freifeld and Alexandra Alper, "US Adds China's YMTC and 30 Other Firms to 'Unverified' Trade List," Reuters, October 7, 2022.
36. Che Pan, "China's Top Memory Chip Maker YMTC Takes Latest Step to Become a Global Market Leader, but US Sanctions Could Derail Its Ambitions," December 1, 2022.
37. Paul Mozur and Steven Lee Myers, "Xi's Gambit: China Plans for a World without American Technology," *New York Times*, March 10, 2021.
38. Ann Cao, "Tech War: Shanghai Launches New Campus to Train Personnel for Semiconductor Sector as US Curbs Decrease China's Chip Talent Pool," *South China Morning Post*, October 26, 2022.

39. Remco Zwetsloot, Jack Corrigan, Emily S. Weinstein, Dahlia Peterson, Diana Gehlhaus, and Ryan Fedasiuk, "China Is Fast Outpacing US STEM PhD Growth," CSET, August 2021.
40. Coco Feng, "China's Semiconductor Self-Sufficiency Drive Needs to Strengthen Development of Talent and Skills, Education Agency Executive Says," *South China Morning Post*, October 5, 2022.
41. Iris Deng, "Shenzhen Plans to Shower Cash on Local Chip Industry to Bolster Development after Intensified US Trade Restrictions," *South China Morning Post*, October 10, 2022.
42. Cheng Ting-Fang, "Huawei Dives into Chip Production to Battle US Clampdown," *Nikkei Asia*, September 22, 2020.
43. Cheng Ting-Fang and Shunsuke Tabeta, "China's Chip Industry Fights to Survive US Tech Crackdown," *Nikkei Asia*, November 30, 2022.
44. Geoffrey Cain, "The Purges That Upended China's Semiconductor Industry," *American Affairs* 6, no. 4 (Winter 2022).
45. Anna Gross and Qianer Liu, "China Enlists Alibaba and Tencent in Fight against US Chip Sanctions," *Financial Times*, November 30, 2022.
46. Ann Cao, "Tech War: China Bets on RISC-V Chips to Escape the Shackles of US Tech Export Restrictions," *South China Morning Post*, November 12, 2022.
47. Ann Cao, "China's Chip Executives Brace for Winter as US Sanctions Push Country's Semiconductor Industry to the Brink of Desperation," *South China Morning Post*, November 12, 2022.
48. Jason R. Wilson, "China Taps in Graphene Technology to Replace Silicon-Based Chips & Breaking the Monopoly with 10 Times the Performance," WCCF Tech, November 22, 2022.
49. Dave Yin, "China's Plan to Leapfrog Foreign Chipmakers: Wave Goodbye to Silicon," Protocol, November 8, 2021.
50. US Department of Commerca, Bureau of Industry and Security, "Foreign-Produced Direct Product (FDP) Rule as it Relates to the Entity List § 736.2(b)(3)(vi) and footnote 1 to Supplement No. 4 to part 744," October 28, 2021.
51. Reuters, "Netherlands Plans New Curbs on Chip-Making Equipment Sales to China—Bloomberg News," December 8, 2022.
52. Per a remark made in private conversation with one of the authors in 2022.
53. Dashveenjit Kaur, "Chip Alliance: The Hefty Price Taiwan Is Paying for Choosing US over China," TechWire Asia, October 11, 2022; and Yoshihiro Sato, "Taiwan Chip Exports to China Sputter on Tensions, Falling Demand," *Bloomberg*, March 19, 2023.
54. Kazuaki Nagata, "Following US on China Chip Export Curbs Would Hit Japan's Industry Hard," *Japan Times*, November 17, 2022.
55. As detailed in the Korea International Trade Association export database (integrated classification code HS 8542), http://kita.org/kStat/byCom_SpeCom.do.

56. Early reactions from Korea's chip industry participants to the US Department of Commerce's mooted additional requirements for CHIPS Act recipients, including childcare requirements and limitations on stock buybacks, is not encouraging in this regard. See, for example, Yonhap News Agency, "Trade Minister Leaves for US for Talks on Chips Act," March 8, 2023.
57. Kif Leswing, "Nvidia Says US Government Allows A.I. Chip Development in China," CNBC, September 1, 2022.
58. Alex He, "Beijing and Washington Joust over Semiconductors," Centre for International Government Innovation, November 9, 2022.
59. Yang Jie and Aaron Tilley, "Apple Makes Plans to Move Production out of China," *Wall Street Journal*, December 3, 2022.
60. Ministry of Foreign Affairs of the People's Republic of China, "Foreign Ministry Spokesperson Wang Wenbin's Regular Press Conference on September 1, 2022," September 1, 2022.
61. Ministry of Foreign Affairs of the People's Republic of China, "Foreign Ministry Spokesperson Mao Ning's Regular Press Conference on October 8, 2022," October 8, 2022.
62. Merriden Varall, "Behind the News: Inside China Global Television Network," Lowy Institute, January 10, 2020.
63. Laura Silver, Christine Wang, and Laura Clancy, "Negative Views of China Tied to Critical Views of Its Policies on Human Rights," Pew Research Center, June 29, 2022.
64. David M. Hart, "The Impact of China's Production Surge on Innovation in the Global Solar Photovoltaics Industry," ITIF, October 5, 2020.
65. Nigel Cory, Stephen Ezell, David M. Hart, and Robert D. Atkinson, "Innovation Drag: The Impact of Chinese Economic and Trade Policies on Global Innovation," ITIF, June 10, 2021.
66. China's Ministry of Commerce in fact filed such a WTO suit on December 12, 2022, over the October 7 BIS rules aimed at advanced chips. Orange Wang, "China Files WTO Suit against US over Chip Export Controls, Saying Policy Is 'Trade Protectionism,'" *South China Morning Post*, December 13, 2022.
67. Andrew Methven, "A Thief Crying 'Stop Thief!'—Phrase of the Week," *China Project*, May 20, 2022.

A report of the Working Group on
Semiconductors and the Security of the
United States and Taiwan, a joint project of
the Hoover Institution and the Asia Society
Center on U.S.-China Relations

AGI, Governments, and Free Societies[12]

Séb Krier

Manager, Policy Development and Strategy, Google DeepMind

Introduction

Throughout history, leaps in technology have destabilised and altered both the politics and administration of governments. Many current observers expect artificial intelligence (AI) to do the same. However, the present conversation on AI governance has focused on how current AI models may be utilised by governments. Few have explored how artificial general intelligence (AGI) that meets or exceeds the capabilities of humans on all decision-making tasks may impact the prospects of liberal democratic societies. A paper co-authored by Justin Bullock, Samuel Hammond and myself earlier this year sought to remedy this gap by exploring the plausible consequences to governments in a world with AGI.[13] This essay presents the main arguments made in the paper in a shorter form.

Rapid improvements in AI models have surpassed the expectations of many observers and academics in the field. There are few signs of a slowdown, and leading AI researchers may now consider the advent of human-level AI systems less a matter of 'if' than 'when.' Although there is still enormous uncertainty about the trajectory of AI capabilities over the medium term, many researchers and forecasters now anticipate the advent of AGI in a matter of years rather than decades or centuries. This paper therefore examines the critical question of how AGI might impact liberal democracies, with a focus on its potential to fundamentally reshape the balance between state capacity and individual liberty.

Building on Acemoglu and Robinson's 'narrow corridor' framework,[14] we contend that free and open societies have traditionally depended upon the existence of a "narrow corridor" characterised by a delicate balance between the relative powers of society and the state. While some recent empirical research, such as that by Ryan Murphy and Colin O'Reilly, has questioned the precise historical pathways and universality of this specific model, suggesting an "expansive corridor" might better describe historical data for many nations, the core concept of a necessary balance between state power and societal autonomy remains a vital heuristic for considering future challenges. Although AGI

[12] Adapted from Bullock, Hammond & Krier (2025).
[13] The paper can be accessed here.
[14] Acemoglu, D., & Robinson, J. (2019). The Narrow Corridor: How Nations Struggle for Liberty. Penguin.

could work to strengthen democracies, it also has the potential to upset and reshape the balance of power between society and the state, potentially contributing to the erosion and collapse of liberal and democratic institutions.

On the one hand, AGI could dramatically enhance the state's capacity for surveillance and control, crushing individual liberties and pushing societies towards authoritarianism in a 'despotic Leviathan' scenario. On the other hand, in an 'absent Leviathan' scenario, as AI capabilities diffuse to the edge, AGI-like systems could empower individuals and non-state actors in ways that weaken or overwhelm state authority and legitimacy, risking a slide toward anarchy.

We posit that maintaining free societies in the age of AGI will require deliberate institutional innovation to harness AGI's benefits while guarding against both centralised control and institutional collapse. To preserve the narrow corridor of liberty, we propose a governance framework emphasising robust technical safeguards, hybrid institutional designs that maintain meaningful human oversight, adaptive regulatory mechanisms, improved democratic participation, and a fundamental shift in how we govern emerging technologies. This shift includes actively fostering "defensive" applications of AGI itself – using its power to strengthen societal resilience and democratic integrity.

This essay proceeds by explaining the concept of the narrow corridor as we apply it before considering how AGI could strengthen democracy. We then apply the concept of this necessary balance, often termed the narrow corridor, to examine how AGI could undermine free societies before exploring strategies to prevent this.

The Concept of The Narrow Corridor

The concept of the "narrow corridor", as articulated by Acemoglu and Robinson and utilised in our work, provides a useful framework for understanding the delicate balance required to maintain free societies.[15] The narrow corridor represents a space where state capacity and societal power are in equilibrium. On one side of the corridor is a scenario of anarchy where the state is too weak to provide basic services, maintain order, or protect rights. On the other a despotic and overly powerful state suppresses society and individual freedoms. Liberty thrives only within the narrow space between these extremes.

[15] Gudiño-Rosero, J., Grandi, U., & Hidalgo, C.A. (2024). Large Language Models (LLMs) as Agents for Augmented Democracy. Accessible here.

Historically, liberal societies have maintained a precarious equilibrium between state capacity and individual liberty, with constitutional constraints, checks and balances, and the rule of law serving to keep both despotism and anarchy at bay. However, this equilibrium has never been static. Technological and social change have forced repeated renegotiations of the social contract, from the rise of mass politics in the industrialising West to the welfare state reforms of the early 20th century.

How AGI Could Strengthen Free Societies

The late 2010s brought dramatic increases in AI capabilities and generality, in the form of large language models (LLMs). Scaling *laws* – mathematical relationships that predict how a model's performance changes with factors like model size, dataset size, and training compute – suggest that AI may continue to rapidly develop and that progress towards AGI is unlikely to slow materially. As such, while there is considerable lag between benchmark achievements in AI development and tangible societal impact, our view is that large-scale impacts are more likely than not.

While it is impossible to fully anticipate the impact of AGI on liberal democratic institutions, rapid advances towards AGI presents an inflection point that could fundamentally disrupt the delicate balance underpinning this 'narrow corridor.' AGI could strengthen free societies, for instance by enhancing the administrative capacity of the state, which in turn enables a more effective and equitable provision of public goods and services. AGI may well present an absolute advantage over human decision making, in terms of scalability, cost, and quality. Consider, for example, a policy analyst responsible for collecting evidence, synthesising research, and interpreting multifaceted legislation; an AGI agent could execute these tasks in parallel by deploying multiple specialised sub-agents—one reviewing licensing regulations, another verifying relevant case law, and yet another contacting third parties for additional data.

AGI agents could also have an advantage over human bureaucrats in terms of equity. While human-bureaucracies are littered with subjectivity and misaligned decision making, AI agents could be trained to rely on clearly identifiable objective factors and logic-based reasoning to carefully weigh trade-offs across alternative options.
AGI could also strengthen free societies by improving democratic input into the decision-making process. For example, Gudiño-Rosero and colleagues (2024) argue that augmented democracy could be achieved through the use of "digital twins" to create a simulated version of an individual citizen that can act as a proxy of that person by "representing" their policy preferences in some arenas of democratic input. Although their work stops far short of the creation of actual "digital twins" for political representation purposes, one can imagine a world where AGI systems facilitate an increased amount of citizen feedback in lieu of an elected human representative. Beyond

representation, AGI itself can be a powerful governance technology, aiding in the design and stress-testing of policies, enhancing information integrity, or even bolstering cybersecurity for critical institutions.

How AGI Could Undermine Free Societies

Nonetheless, despite the potential benefits, AGI also has the potential to undermine free societies by destabilising the narrow corridor in two directions.
On the one hand, AGI could dramatically enhance the administrative capacity of the state, making governments radically more efficient and data-driven in their decision-making. This enhancement of state capacity risks a slide towards authoritarianism if not carefully constrained. An AGI-empowered state could wield unprecedented surveillance and control over its citizens, stifling dissent and entrenching existing power structures. The increasing automation of administrative decision making could also erode human agency and democratic accountability, as bureaucracies evolve towards more centralised architectures.

On the other hand, if AGI diffuses more rapidly among individuals and civil society groups than governments, it could instead weaken the legitimacy and capacity of the state relative to non-state actors. In this scenario, the risk is not despotism but a hollowing out of the governability and social cohesion that liberal democracies depend upon. Malicious actors could also exploit widely accessible AGI to undermine elections, manipulate public opinion, or coordinate insurgencies, further eroding the stability of democratic institutions. This scenario necessitates a proactive approach by the state to not only regulate AGI but also to leverage it defensively, enhancing its own capacity to withstand such challenges and maintain societal resilience.

The integration of AGI into public administration further complicates this balance by fundamentally altering how governments function. Governments are not monolithic entities, but rather complex systems of agencies and bureaucracies tasked with diverse and often conflicting objectives. AGI will likely accelerate the evolution of these systems, pushing them toward system-level bureaucracies where decision-making and execution increasingly rely on automated systems. For example, AGI systems acting as artificial bureaucrats could automate complex, high-discretion tasks that traditionally required human judgment, promising significant gains in efficiency, effectiveness, and scalability. However, their widespread deployment raises critical questions about transparency, accountability, and responsiveness—values essential to liberal democratic governance. AGI also presents risks to the administrative values that underpin effective governance in free societies. Particularly, the impact of AGI systems on equity is uncertain, as AGI

systems may inadvertently replicate or exacerbate societal biases embedded in their training data. Moreover, transparency and accountability could suffer as AGI systems make decisions based on processes that are difficult to interpret or verify for both administrators and the public.

The ethical implications of AGI's integration into governance are equally significant. Delegating value-laden decisions to AGI raises concerns about the loss of moral accountability in public administration. For example, while AGI agents may excel at optimising policies for efficiency, they may lack the ethical nuance required to address competing societal values. This disconnect between computational optimisation and human morality risks eroding public trust, particularly if AGI systems prioritise narrow objectives at the expense of fairness and inclusivity.

How to Secure Free Societies

To address these challenges and secure the narrow corridor, a comprehensive strategy must include technological safeguards, institutional adaptations, and ethical considerations. On the technological front, privacy-enhancing technologies (PETs) can play a vital role in counterbalancing the surveillance capabilities of AGI. Advanced PETs could enable individuals to maintain autonomy and privacy in the face of increasingly pervasive state monitoring. Simultaneously, investments in transparency and interpretability are essential to ensuring that AGI systems operate transparently and remain accountable for their decisions. These technologies can help bridge the gap between AGI's computational processes and human oversight, fostering trust in their deployment. Crucially, this involves not just restricting harmful uses of AGI, but actively developing and deploying AGI for defensive purposes to strengthen our societal shield against emerging threats.

Institutional adaptations are equally important. Governments must embrace hybrid AI-human governance structures that combine AGI's computational power with the nuanced judgment and accountability that human administrators provide. Bureaucratic models must also evolve to maintain flexibility and adaptability, allowing agencies to leverage AGI's capabilities while retaining oversight mechanisms that align with democratic values. This might include equipping public institutions with their own advanced AI tools for functions like biosurveillance, cyberdefense, and regulatory oversight, ensuring they are not outpaced by threats.

At the same time, regulatory and legal frameworks will require adaptation as AGI capabilities become more accessible to non-state actors. Policymakers will need to address this by, for instance, strengthening cybersecurity across critical infrastructure,

and by strengthening defensive measures against potential misuse, such as AGI's capacity to lower barriers to entry for CBRN threats.

Reinforcing democratic processes is another critical pillar of securing the narrow corridor. AGI systems offer unique opportunities to enhance participatory governance, such as enabling large scale deliberative platforms, real-time citizen feedback systems, and representative digital twins. These tools could revitalise democratic engagement and strengthen the feedback loop between citizens and their representatives. However, these systems must be designed with robust safeguards to prevent misuse and ensure that they genuinely enhance, rather than undermine, democratic accountability.

Perhaps most importantly, securing the narrow corridor in an age of AGI will require an epistemic shift in how we approach the governance of emerging technologies. Rather than passively reacting to technological disruptions, policymakers and publics alike must cultivate a greater capacity for anticipatory governance–proactively imagining and stress-testing institutional paradigms in expectation of AGI's transformative potential. Intellectual frameworks like scenario planning, threat modelling, and forecasting should be deployed, in a serious exploration of failure modes and policy options for their mitigation.

Conclusion

As AGI approaches rapidly, it is crucial to anticipate its impact on free societies. We have discussed how AGI could strengthen such societies by making public administration more efficient and effective, and by improving the provision of public goods and services. We've also highlighted how AGI itself can be a tool for better governance and societal defense. However, AGI also risks destabilising the delicate balance between state and society, enabling unprecedented surveillance and repression, or conversely, empowering individuals and non-state actors to the point of undermining state authority, risking anarchy. To safeguard free societies, we need strong technical protections, hybrid institutions with human oversight, adaptive regulation, improved democratic participation, and a fundamental shift in how we govern emerging technologies.

Although achieving this will be challenging, we posit that it is critical for the immense power of AGI to be channelled towards expanding rather than constraining human freedom and flourishing. The great political question of the 21st century may well be whether liberal democracy can reform itself in time to reap the rewards and manage the risks of artificial general intelligence.

History suggests that, with sufficient foresight and resolve, free societies are capable of extraordinary institutional innovation in moments of technological upheaval. The task before us is to muster that resolve once again––to reimagine the narrow corridor for an age of AGI–and in so doing, secure the possibility of a brighter future for all. Only by rising to this challenge can we ensure that the tremendous power of AGI remains firmly in service of our deepest democratic values: individual liberty, popular sovereignty, and human dignity for all.

RECOMMENDED READINGS

RECOMMENDED BY ASPEN DIGITAL

Tobias Feakin	*A.I. Geopolitics Beyond the U.S.-China Rivalry: The Role of the Global South*
Michiaki Matsushima, WIRED Japan	*Yuval Noah Harari: 'How Do We Share the Planet With This New Superintelligence?'*

RECOMMENDED BY ASPEN INSTITUTE CONGRESSIONAL PROGRAM

Congressional Budget Office Report	*Artificial Intelligence and Its Potential Effects on the Economy and the Federal Budget*

Yuval Noah Harari: 'How Do We Share the Planet With This New Superintelligence?'[16]

The academic and author discusses what to expect from the singularity, the need for AI self-correcting mechanisms, and what hope there is for superintelligence safeguarding democracy.

ISRAELI HISTORIAN AND philosopher Yuval Noah Harari's book *Sapiens* became an international bestseller by presenting a view of history driven by the fictions created by mankind. His later work *Homo Deus* then depicted the a future for mankind brought about by the emergence of superintelligence. His latest book, *Nexus: A Brief History of Information Networks From the Stone Age to AI*, is a warning against the unparalleled threat of AI.

A rising trend of techno-fascism driven by populism and artificial intelligence has been visible since the US presidential election in November. *Nexus*, which was published just a few months earlier, is a timely explainer of the potential consequences of AI on democracy and totalitarianism. In the book, Harari does not just sound the alarm on singularity—the hypothetical future point at which technology, particularly AI, moves beyond human control and advances irreversibly on its own—but also on AI's foreignness.

This interview was conducted by Michiaki Matsushima, editor in chief of WIRED Japan, and was also recorded for "The Big Interview" YouTube series for the Japanese edition of WIRED, scheduled to be released in April 2025. The interview has been edited for clarity and length.

WIRED: In the late '90s, when the internet began to spread, there was a discourse that this would bring about world peace. It was thought that with more information reaching more people, everyone would know the truth, mutual understanding would be born, and humanity would become wiser. WIRED, which has been a voice of change and hope in the digital age, was part of that thinking at the time. In your new book, *Nexus*, you write that such a view of information is too naive. Can you explain this?

YUVAL NOAH HARARI: Information is not the same as truth. Most information is not an accurate representation of reality. The main role information plays is to connect

[16] Originally published on WIRED April 1, 2025, https://www.wired.com/story/questions-answered-by-yuval-noah-harari-for-wired-ai-artificial-intelligence-singularity/

many things, to connect people. Sometimes people are connected by truth, but often it is easier to use fiction or illusion.

The same is true of the natural world. Most of the information that exists in nature is not meant to tell the truth. We are told that the basic information underlying life is DNA, but is DNA true? No. DNA connects many cells together to make a body, but it does not tell us the truth about anything. Similarly, the Bible, one of the most important texts in human history, has connected millions of people together, but not necessarily by telling them the truth.

When information is in a complete free market, the vast majority of information becomes fiction, illusion, or lies. This is because there are three main difficulties with truth.

First of all, telling the truth is costly. On the other hand, creating fiction is inexpensive. If you want to write a truthful account of history, economics, physics, et cetera, you need to invest time, effort, and money in gathering evidence and fact-checking. With fiction, however, you can simply write whatever you want.

Second, truth is often complex, because reality itself is complex. Fiction, on the other hand, can be as simple as you want it to be.
And finally, truth is often painful and unpleasant. Fiction, on the other hand, can be made as pleasant and appealing as possible.

Thus, in a completely free information market, truth would be overwhelmed and buried by the sheer volume of fiction and illusion. If we want to get to the truth, we must make a special effort to repeatedly try to uncover the facts. This is exactly what has happened with the spread of the internet. The internet was a completely free marketplace of information. Therefore, the expectation that the internet would spread facts and truths, and spread understanding and consensus among people, quickly proved to be naive.

In a recent interview with The New Yorker, Bill Gates said, "I always thought that digital technology empowers people, but social networking is something completely different. We were slow to realize that. And AI is something completely different as well." If AI is unprecedented, what, if anything, can we learn from the past?

There are many things we can learn from history. First, knowing history helps us understand what new things AI has brought. Without knowing the history, we cannot

properly understand the novelty of the current situation. And the most important point about AI is that it is an agent, not just a tool.

Some people often equate the AI revolution with the printing revolution, the invention of the written word, or the emergence of mass media such as radio and television, but this is a misunderstanding. All previous information technologies were mere tools in the hands of humans. Even when the printing press was invented, it was still humans who wrote the text and decided which books to print. The printing press itself cannot write anything, nor can it choose which books to print.

AI, however, is fundamentally different: It is an agent; it can write its own books and decide which ideas to disseminate. It can even create entirely new ideas on its own, something that has never been done before in history. We humans have never faced a superintelligent agent before.

Of course, there have been actors in the past. Animals are one example. However, humans are more intelligent than animals, especially in the area of connection, in which they are overwhelmingly superior. In fact, the greatest strength of *Homo sapiens* is not its individual capabilities. On an individual level, I am not stronger than a chimpanzee, an elephant, or a lion. If a small group, say 10 humans and 10 chimpanzees, were to fight, the chimpanzees would probably win.

So why do humans dominate the planet? It is because humans can create networks of thousands, millions, and even billions of people who do not know each other personally but can cooperate effectively on a huge scale. Ten chimpanzees can cooperate closely with each other, but 1,000 chimpanzees cannot. Humans, on the other hand, can cooperate not with 1,000 individuals, but with a million or even a hundred million.

The reason why human beings are able to cooperate on such a large scale is because we can create and share stories. All large-scale cooperation is based on a common story. Religion is the most obvious example, but financial and economic stories are also good examples. Money is perhaps the most successful story in history. Money is just a story. The bills and coins themselves have no objective value, but we believe in the same story about money that connects us and allows us to cooperate. This ability has given humans an advantage over chimpanzees, horses, and elephants. These animals cannot create a story like money.

But AI can. For the first time in history, we share the planet with beings that can create and network stories better than we can. The biggest question facing humanity today is: How do we share the planet with this new superintelligence?

How should we think about this new era of superintelligence?

I think the basic attitude toward the AI revolution is to avoid extremes. At one end of the spectrum is the fear that AI will come along and destroy us all, and at the other end is optimism that AI will improve health care, improve education, and create a better world.

What we need is a middle path. First and foremost, we need to understand the scale of this change. Compared to the AI revolution we are facing now, all previous revolutions in history will pale in comparison. This is because throughout history, when humans invented something, it was always they who made the decisions about how to use it to create a new society, a new economic system, or a new political system.

Consider, for example, the Industrial Revolution of the 19th century. At that time, people invented steam engines, railroads, and steamships. Although this revolution transformed the productive capacity of economies, military capabilities, and geopolitical situations, and brought about major changes throughout the world, it was ultimately people who decided how to create industrial societies.

As a concrete example, in the 1850s, the US commodore Matthew C. Perry came to Japan on a steamship and forced Japan to accept US trade terms. As a result, Japan decided: Let's industrialize like the US. At that time, there was a big debate in Japan over whether to industrialize or not, but the debate was only between people. The steam engine itself did not make any decision.

This time, however, in building a new society based on AI, humans are not the only ones making decisions. AI itself may have the power to come up with new ideas and make decisions.

What if AI had its own money, made its own decisions about how to spend it, and even started investing it in the stock market? In that scenario, to understand what is happening in the financial system, we would need to understand not only what humans are thinking, but also what AI is thinking. Furthermore, AI has the potential to generate ideas that are completely incomprehensible to us.

I would like to clarify what you think about the singularity, because I often see you spoken of as being "anti-singularity." However, in your new book, you point out that AI is more creative than humans and that it is also superior to humans in terms of emotional intelligence.

I was particularly struck by your statement that the root of all these revolutions is the computer itself, of which the internet and AI are only derivatives. WIRED just published a series on quantum computers, so to take this as an example: If we are given a quantum leap in computing power in the future, do you think that a singularity, a reordering of the world order by superintelligence, is inevitable?

That depends on how you define singularity. As I understand it, singularity is the point at which we no longer understand what is happening out there. It is the point at which our imagination and understanding cannot keep up. And we may be very close to that point.
Even without a quantum computer or fully-fledged artificial general intelligence—that is, AI that can rival the capabilities of a human—the level of AI that exists today may be enough to cause it. People often think of the AI revolution in terms of one giant AI coming along and creating new inventions and changes, but we should rather think in terms of networks. What would happen if millions or tens of millions of advanced AIs were networked together to bring about major changes in economics, military, culture, and politics? The network will create a completely different world that we will never understand. For me, singularity is precisely that point—the point at which our ability to understand the world, and even our own lives, will be overwhelmed.

If you ask me if I am for or against singularity, first and foremost I would say that I am just trying to get a clear understanding of what is going on right now. People often want to immediately judge things as good or bad, but the first thing to do is to take a closer look at the situation. Looking back over the past 30 years, technology has done some very good things and some very bad things. It has not been a clear-cut "just good" or "just bad" thing. This will probably be the same in the future.

The one obvious difference in the future, however, is that when we no longer understand the world, we will no longer control our future. We will then be in the same position as animals. We will be like the horse or the elephant that does not understand what is happening in the world. Horses and elephants cannot understand that human political and financial systems control their destiny. The same thing can happen to us humans.

You've said, "Everyone talks about the 'post-truth' era, but was there ever a 'truth' era in history?" Could you explain what you mean by this?

We used to understand the world a little better, because it was humans who managed the world, and it was a network of humans. Of course, it was always difficult to understand how the whole network worked, but at least as a human being myself, I

could understand kings, emperors, and high priests. They were human beings just like me. When the king made a decision, I could understand it to some extent, because all the members of the information network were human beings. But now that AI is becoming a major member of the information network, it is becoming increasingly difficult to understand the important decisions that shape our world.

Perhaps the most important example is finance. Throughout history, humans have invented increasingly sophisticated financial mechanisms. Money is one such example, as are stocks and bonds. Interest is another financial invention. But what is the purpose of inventing these financial mechanisms? It is not the same as inventing the wheel or the automobile, nor is it the same as developing a new kind of rice that can be eaten.

The purpose of inventing finance, then, is to create trust and connection between people. Money enables cooperation between you and me. You grow rice and I pay you. Then you give me the rice and I can eat it. Even though we do not know each other personally, we both trust money. Good money builds trust between people.

Finance has built a network of trust and cooperation that connects millions of people. And until now, it was still possible for humans to understand this financial network. This is because all financial mechanisms needed to be humanly understandable. It makes no sense to invent a financial mechanism that humans cannot understand, because it cannot create trust.

But AI may invent entirely new financial mechanisms that are far more complex than interest, bonds, or stocks. They will be mathematically extremely complex and incomprehensible to humans. AI itself, on the other hand, can understand them. The result will be a financial network where AIs trust each other and communicate with each other, and humans will not understand what is happening. We will lose control of the financial system at this point, and everything that depends on it.

So AI can build networks of trust that we can't understand. Such incomprehensible things are known as "hyperobjects." For example, global climate change is something that humans cannot fully grasp the mechanisms or full picture of, but we know it will have a tremendous impact and that we therefore must confront and adapt to it. AI is another hyperobject that humanity will have to deal with in this century. In your book, you cite human flexibility as one of the things needed to deal with big challenges. But what does it actually mean for humanity to deal with hyperobjects?

Ideally, we would trust AI to help us deal with these hyperobjects—realities that are so complex that they are beyond our comprehension. But perhaps the biggest question in the development of AI is: How do we make AI, which can be more intelligent than humans, trustworthy? We do not have the answer to that question.

I believe the biggest paradox in the AI revolution is the paradox of trust—that is, that we are now rushing to develop superintelligent AI that we do not fully trust. We understand that there are many risks. Rationally, it would be wise to slow down the pace of development, invest more in safety, and create safety mechanisms first to make sure that superintelligent AIs do not escape our control or behave in ways that are harmful to humans.

However, the opposite is actually happening today. We are in the midst of an accelerating AI race. Various companies and nations are racing at breakneck speed to develop more powerful AIs. Meanwhile, little investment has been made to ensure that AI is secure.

Ask the entrepreneurs, businesspeople, and government leaders who are leading this AI revolution, "Why the rush?" and nearly all of them answer: "We know it's risky, for sure. We know it's dangerous. We understand that it would be wiser to go slower and invest in safety. But we cannot trust our human competitors. If other companies and countries accelerate their development of AI while we are trying to slow it down and make it safer, they will develop superintelligence first and dominate the world. So we have no choice but to move forward as fast as possible to stay ahead of the unreliable competition."

But then I asked those responsible for AI a second question: "Do you think we can trust the superintelligence you are developing?" Their the answer was: "Yes." This is almost insane. People who don't even trust other humans somehow think they can trust this alien AI.

We have thousands of years of experience with humans. We understand human psychology and politics. We understand the human desire for power, but we also have some understanding of how to limit that power and build trust among humans. In fact, over the past few thousand years, humans have developed quite a lot of trust. 100,000 years ago, humans lived in small groups of a few dozen people and could not trust outsiders. Today, however, we have huge nations, trade networks that extend around the world, and hundreds of millions, even billions, of people who trust each other to some extent.

We know that AI is a doer, that it makes its own decisions, creates new ideas, sets new goals, creates tricks and lies that humans do not understand, and may pursue alien goals

beyond our comprehension. We have many reasons to be suspicious of AI. We have no experience with AI, and we do not know how to trust it.

I think it is a huge mistake for people to assume that they can trust AI when they do not trust each other. The safest way to develop superintelligence is to first strengthen trust between humans, and then cooperate with each other to develop superintelligence in a safe manner. But what we are doing now is exactly the opposite. Instead, all efforts are being directed toward developing a superintelligence.

Some WIRED readers with a libertarian mindset may have more faith in superintelligence than in humans, because humans have been fighting each other for most of our history. You say that we now have large networks of trust, such as nations and large corporations, but how successful are we at building such networks, and will they continue to fail?

It depends on the standard of expectations we have. If we look back and compare humanity today to 100,000 years ago, when we were hunter-gatherers living in small herds of a few dozen people, we have built an astonishingly large network of trust. We have a system in which hundreds of millions of people cooperate with each other on a daily basis.

Libertarians often take these mechanisms for granted and refuse to consider where they come from. For example, you have electricity and drinking water in your home. When you go to the bathroom and flush the water, the sewage goes into a huge sewage system. That system is created and maintained by the state. But in the libertarian mindset, it is easy to take for granted that you just use the toilet and flush the water and no one needs to maintain it. But of course, someone needs to.

There really is no such thing as a perfect free market. In addition to competition, there always needs to be some sort of system of trust. Certain things can be successfully created by competition in a free market, however, there are some services and necessities that cannot be sustained by market competition alone. Justice is one example.

Imagine a perfect free market. Suppose I enter into a business contract with you, and I break that contract. So we go to court and ask the judge to make a decision. But what if I had bribed the judge? Suddenly you can't trust the free market. You would not tolerate the judge taking the side of the person who paid the most bribes. If justice were to be traded in a completely free market, justice itself would collapse and people would no

longer trust each other. The trust to honor contracts and promises would disappear, and there would be no system to enforce them.

Therefore, any competition always requires some structure of trust. In my book, I use the example of the World Cup of soccer. You have teams from different countries competing against each other, but in order for competition to take place, there must first be agreement on a common set of rules. If Japan had its own rules and Germany had another set of rules, there would be no competition. In other words, even competition requires a foundation of common trust and agreement. Otherwise, order itself will collapse.

In *Nexus*, you note that the mass media made mass democracy possible—in other words, that information technology and the development of democratic institutions are correlated. If so, in addition to the negative possibilities of populism and totalitarianism, what opportunities for positive change in democracies are possible?

In social media, for example, fake news, disinformation, and conspiracy theories are deliberately spread to destroy trust among people. But algorithms are not necessarily the spreaders of fake news and conspiracy theories. Many have achieved this simply because they were designed to do so.

The purpose the algorithms of Facebook, YouTube, and TikTok is to maximize user engagement. The easiest way to do this, it was discovered after much trial and error, was to spread information that fueled people's anger, hatred, and desire. This is because when people are angry, they are more inclined to pursue the information and spread it to others, resulting in increased engagement.

But what if we gave the algorithm a different purpose? For example, if you give it a purpose such as increasing trust among people or increasing truthfulness, the algorithm will never spread fake news. On the contrary, it will help build a better society, a better democratic society.

Another important point is that democracy should be a dialogue between human beings. In order to have a dialogue, you need to know and trust that you are dealing with a human being. But with social media and the internet, it is increasingly difficult to know whether the information you are reading is really written and disseminated by humans or just bots. This destroys trust between humans and makes democracy very difficult.

To address this, we could have regulations and laws prohibiting bots and AI from pretending to be human. I don't think AI itself should be banned at all; AI and bots are

welcome to interact with us, but only if they make it clear that they are AI and not human. When we see information on Twitter, we need to know whether it is being spread by a human or a bot.

Some people may say, "Isn't that a violation of freedom of expression?" But bots do not have freedom of expression. While I firmly oppose censorship of human expression, this does not protect the expression of bots.

Will we become smarter or reach better conclusions by discussing topics with artificial intelligence in the near future? Will we see the kind of creativity that humans can't even conceive of, as in the case of AlphaGo, which you also describe in your new book, in classroom discussions, for example?

Of course it can happen. On the one hand, AI can be very creative and come up with ideas that we would never have thought of. But at the same time, AI can also manipulate us by feeding us vast amounts of junk and misleading information.

The key point is that we humans are stakeholders in society. As I mentioned earlier with the example of the sewage system, we have a body. If the sewage system collapses, we become sick, spreading diseases such as dysentery and cholera, and in the worst case, we die. But that is not a threat at all to AI, which does not care if the sewage system collapses, because it will not get sick or die. When human citizens debate, for example, whether to allocate money to a government agency to manage a sewage system, there is an obvious vested interest. So while AI can come up with some very novel and imaginative ideas for sewage systems, we must always remember that AI is not human or even organic to begin with.

It is easy to forget that we have bodies, especially when we are discussing cyberspace. What makes AI different from humans is not only that its imagination and way of thinking, which are alien, but also that its body itself is completely different from ours. Ultimately, AI is also a physical being; it does not exist in some purely mental space, but in a network of computers and servers.

What is the most important thing to consider when thinking about the future?

I think there are two important issues. One is the issue of trust, which has been the subject of much discussion up to this point. We are now in a situation where trust

between human beings is at stake. This is the greatest danger. If we can strengthen trust between humans, we will be better able to cope with the AI revolution.

The second is the threat of being completely manipulated or misdirected by AI. In the early internet days, the primary metaphor for technology was the Web. The World Wide Web was envisioned as a spiderweb-like network connecting people to each other.

Today, however, the primary metaphor is the cocoon. People are increasingly living in individual cocoons of information. People are bombarded with so much information that they are blind to the reality around them. People are trapped in different information cocoons. For the first time in history, a nonhuman entity, an AI, is able to create such a cocoon of information.

Throughout history, people have lived in a human cultural cocoon. Poetry, legends, myths, theater, architecture, tools, cuisine, ideology, money, and all the other cultural products that have shaped our world have all come from the human mind. In the future, however, many of these cultural products will come from nonhuman intelligence. Our poems, videos, ideologies, and money will come from nonhuman intelligence. We may be trapped in such an alien world, out of touch with reality. This is a fear that humans have held deep in their hearts for thousands of years. Now, more than ever, this fear has become real and dangerous.

For example, Buddhism speaks of the concept of māyā—illusion, hallucination. With the advent of AI, it may be even more difficult to escape from this world of illusion than before. AI is capable of flooding us with new illusions, illusions that do not even originate in the human intellect or imagination. We will find it very difficult to even comprehend the illusions.

You mention "self-correcting mechanisms" as an important function in maintaining democracy. I think this is also an important function to get out of the cocoon and in contact with reality. On the other hand, in your book, you write that the performance of the human race since the Industrial Revolution should be graded as "C minus," or just barely acceptable. If that is the case, then surely we cannot expect much from the human race in the coming AI revolution?

When a new technology appears, it is not necessarily bad in itself, but people do not yet know how to use it beneficially. The reason why they don't know is that we don't have a model for it.

When the Industrial Revolution took place in the 19th century, no one had a model for how to build a "good industrial society" or how to use technologies such as steam engines, railroads, and telegraphs for the benefit of humanity. Therefore, people experimented in various ways. Some of these experiments, such as the creation of modern imperialism and totalitarian states, had disastrous results.

This is not to say that AI itself is bad or harmful. The real problem is that we do not have a historical model for building an AI society. Therefore, we will have to repeat experiments. Moreover, AI itself will now make its own decisions and conduct its own experiments. And some of these experiments may have terrible results.

That is why we need a self-correcting mechanism—a mechanism that can detect and correct errors before something fatal happens. But this is not something that can be tested in a laboratory before introducing AI technology to the world. It is impossible to simulate history in a laboratory.

For example, let's consider the railroad being invented.

In a laboratory, people were able to see if steam engines would explode due to a malfunction. But no one could simulate the changes they would bring to the economic and political situation when the rail network spread out over tens of thousands of kilometers.

The same is true of AI. No matter how many times we experiment with AI in the laboratory, it will be impossible to predict what will happen when millions of superintelligences are unleashed on the real world and begin to change the economic, political, and social landscape. Almost certainly, there will be major mistakes. That is why we should proceed more carefully and more slowly. We must allow ourselves time to adapt, time to discover, and correct our mistakes.

This story originally appeared on WIRED Japan and has been translated from Japanese.

A.I. Geopolitics Beyond the U.S.-China Rivalry: The Role of the Global South[17]

Tobias Feakin
Former Ambassador for Cyber Affairs and Critical Technology, Australia

For much of the last decade, the race for AI dominance has been framed as a binary competition between the US and China. However, this narrative overlooks a crucial factor: the role of the Global South in shaping the future of AI. Emerging economies in Africa, Latin America, and Southeast Asia are not just passive recipients of AI technologies; they are actively influencing the direction of AI development, adoption, and governance.

These nations are setting regulatory precedents, providing diverse datasets that enhance AI models, and determining the geopolitical balance of AI power through their technological alliances. Whilst China has expanded its AI footprint in these regions through investments and cost-effective AI solutions, the US and its allies have struggled to offer compelling alternatives. As these nations assert greater control over their AI futures, they are shifting AI geopolitics beyond a simple US-China rivalry, making them not just arenas of competition but key actors in defining the AI landscape.

Why the Global South Matters

The Global South is already playing a key role in the AI landscape. Representing approximately 85% of the world's population, these regions collectively form a significant portion of the global digital economy and workforce. Despite their growing influence, many nations in the Global South remain 'tech takers' reliant on external AI infrastructure and platforms developed by major AI powers like the US and China. They are no longer just passive adopters: their growing populations, increasing digital infrastructure, and expanding economies have helped their policymakers become more proactive in shaping AI governance.

Emerging A.I. Powerhouses

Nations in Africa, Latin America and Southeast Asia are setting regulatory precedents, developing homegrown AI capabilities, and influencing global AI norms in multilateral institutions. For instance, Brazil's AI strategy emphasises ethical AI governance and innovation in agriculture, while Indonesia's AI roadmap focuses on workforce

[17] Original published March 7, 2025 on https://www.aspendigital.org/blog/ai-geopolitics-beyond-the-us-china-rivalry/

transformation and smart cities. Meanwhile, Kenya has positioned itself as a leader in AI-driven financial inclusion, leveraging machine learning for mobile banking and digital payments. Beyond policy, these regions offer AI models unique training environments that enhance adaptability and efficiency. AI-driven medical diagnostics, such as Google's tuberculosis detection AI trained on African medical datasets, have demonstrated improved accuracy in local healthcare settings. In Latin America, AI-powered climate prediction and crop disease detection models are being refined using agriculture data tailored to smallholder farmers' needs. Meanwhile, natural language processing models training on Southeast Asian languages, such as those from the SEA-LION project, are improving AI-driven translation and speech recognition for multilingual populations.

These nations also hold considerable sway in multilateral governance institutions, where they represent the global majority and play a critical role in shaping international AI norms and regulatory frameworks. The United Nations, through initiatives such as UNESCO's AI Ethics guidelines and the UN AI Advisory body, and the OECD AI Policy Observatory, which includes non-member countries, serve as key platforms for Global South engagement in AI governance. Additionally, regional efforts such as the African Union AI Working Group and ASEAN's AI Governance and Ethics Framework further illustrate how these nations influence AI policy. As AI becomes deeply integrated into governance, security and economic planning, how these countries choose to align, whether with China, the US, or a diversified approach, will heavily influence the future structure of global AI power.

China's Proactive A.I. Outreach with the Global South: Strategic Expansion Through Accessibility

A defining characteristic of China's AI diplomacy is its courtship of the Global South, leveraging the affordability and scalability of its technology to deepen strategic ties. Companies like DeepSeek, Baidu, and Huawei have actively positioned themselves as affordable partners for nations seeking to modernise governance, infrastructure, and industry with AI-driven systems. DeepSeek, in particular, has emerged as a formidable player, offering an alternative AI ecosystem that prioritises accessibility and integration into state-led digital transformation strategies. Unlike US-led AI systems, which often come with high licencing fees and require extensive infrastructure investments, Chinese AI systems are designed for efficiency and ease of adoption.

A.I. Ecosystems: Adoption vs Indigenous Development

At the core of any AI ecosystem are several interdependent components: data collection and governance, model development, AI infrastructure, deployment, and integration into industries and governance. China's AI engagement in the Global South focuses primarily on the latter stages, model deployment, infrastructure provision and AI-powered services, rather than fostering indigenous AI development in host nations. Companies like DeepSeek, Baidu and Huawei offer turnkey AI solutions that enable governments and businesses to rapidly adopt AI applications without requiring deep domestic expertise. These solutions include smart city technologies, AI-powered surveillance systems, language processing models and automation for public services.

The Digital Silk Road: A.I. Dependence and China's Strategic Influence

Through the Digital Silk Road, China has embedded AI capabilities within critical infrastructure projects, fostering long-term dependency on its technology. AI-powered surveillance systems, smart city technology, and digital infrastructure are becoming integral to governance structures across Africa, Southeast Asia, and the Middle East. While China does provide some knowledge transfer, such as AI research hubs and training programs, its emphasis is on embedding Chinese AI models and infrastructure within local ecosystems rather than enabling fully independent AI development.

For example, under the Digital Silk Road, China supplies AI-driven cloud computing, data centres and software solutions that create long-term dependencies on Chinese providers. This approach ensures that while host countries gain access to cutting-edge AI capabilities, they remain reliant on Chinese-developed models and platforms, rather than developing fully autonomous AI industries. Unlike initiatives focused on building domestic AI talent pipelines and model development from the ground up, China's strategy is centred on scaling AI adoption through direct access to its ecosystem, ensuring continued influence over digital and governance frameworks in partner nations. This level of digital entanglement ensures that China's influence extends beyond economic ties, embedding itself in the operational frameworks of these nations.

However, this growing dependence raises significant concerns for both the Global South and the West. First, many of the AI-powered solutions China provides, grant Chinese firms access to vast amounts of sensitive government, business and personal data in host countries. This raises concerns over data exfiltration, mass surveillance, and cyber vulnerabilities, particularly as it is at times hard to distinguish between Chinese firms objectives and those of Beijing.

Second, by embedding AI into governance and infrastructure, China is shaping policy, digital governance, and regulatory norms in host nations. This creates long-term

dependencies, limiting the ability of those nations to develop independent AI ecosystems and potentially coercing them into aligning with China's political and economic priorities. Finally, for Western nations, the expansion of Chinese AI into the Global South erodes democratic digital governance models, enabling authoritarian AI practices, such as surveillance-driven governance and content control, to spread. It also limits Western engagement in setting global AI norms, giving China greater influence over how AI is regulated, deployed and secured internationally.

China's AI investments offer it strategic diplomatic advantages. Despite the associated risks, many recipient countries view Chinese AI solutions as a means to leapfrog traditional development barriers. These partnerships could reshape regional power structures, positioning China as a crucial technology partner in Global South governments' long-term strategic planning. Chinese companies also actively engage in knowledge transfer initiatives, establishing AI research hubs and training programs in host countries. This not only expands China's technology footprint but also cultivates goodwill and reliance on Chinese AI expertise.

Trump's Second Term: The U.S. Retreats from Global Tech Engagement

At the same time, the US is disengaging from tech diplomacy with the Global South. This marks a stark shift from the Biden administration's approach, which prioritised multilateral AI cooperation and sought to balance AI innovation with responsible governance. While Biden's policies encouraged engagement with global partners, the Trump administration has pivoted toward a more unilateral, competition-driven strategy.

Since returning to office in 2025, the Trump administration has pursued an aggressive AI innovation policy focused on reducing regulatory constraints and accelerating AI development to maintain US technological dominance. This approach prioritises economic nationalism and private sector-led expansion, aiming to outpace China through deregulation and infrastructure investment. Trump's Executive Order 'Removing Barriers to American Leadership in Artificial Intelligence' mandates the creation of an AI action plan within 180 days. While the administration is loosening domestic regulatory constraints to fuel AI development, it is simultaneously tightening export controls on advanced AI technologies, specifically to limit China's access. This dual approach reflects a broader strategy: maintaining US technological dominance by restricting adversarial access while accelerating domestic innovation.

As part of this strategy, the administration launched 'Stargate', a partnership with OpenAI, Oracle and SoftBank, investing up to $500 billion in domestic AI infrastructure by 2029. This underscores a shift toward infrastructure-driven AI supremacy, accelerating a broader trend in US policy.

The Case for Open-Source A.I. Models

Stargate's focus on proprietary AI models and national security priorities raises concerns about its competitiveness in the Global South. Unlike DeepSeek which offers scalable AI solutions with low entry barriers, the US remains centred on high-end, privately-driven AI development. To effectively counter China's influence, Washington may need to incorporate open-source AI development into Stargate, allowing emerging economies greater flexibility in AI adoption while reducing dependence on Chinese infrastructure.

China's AI strategy has evolved, with companies like DeepSeek embracing open-source models that make elements of their technology widely accessible. DeepSeek's open-source AI models facilitate rapid adoption across diverse industries, enabling developers in the Global South to tailor AI solutions to their needs. This open approach challenges traditional perceptions of China's AI strategy as purely proprietary and state contained.

Open-Source A.I. as a Strategy: A U.S. Response to China's Global AI Expansion

If Washington remains committed to a light-touch regulatory approach to AI in the name of fostering innovation, it must also recognise the strategic imperative of open-source AI. Embracing open-source models would require a shift in policy, including government-backed AI research initiatives, public-private partnerships and revised export controls that balance openness with security. A strategic open-source AI policy could enhance transparency, accelerate technological progress, and provide an alternative to China's AI expansion in the Global South. This would involve federally funded open-source AI programs, support for regional AI governance initiatives and technical assistance to help emerging economies build independent AI ecosystems.

However, such an approach is not without risks, greater openness introduces vulnerabilities to adversarial exploitation, and raises complex challenges in ensuring responsible deployment. Major US tech firms such as Meta and Google have initiated open-source AI projects, but a more coordinated government-backed effort is needed. By positioning itself as a leader in ethical, adaptable, democratised and safe AI

development, the US can counter China's influence while promoting an AI ecosystem aligned with democratic values and transparency.

Supporting open-source AI would also enable emerging economies to develop independent AI ecosystems, reducing their reliance on foreign technology providers – a shift that may not necessarily align with immediate US commercial interests, but would offer clear advantages to developing nations, while reinforcing democratic AI governance.

Rethinking A.I. Strategy: U.S. Investment Without Traditional Foreign Aid

However, with the Trump administration's drastic cuts to the USAID budget and programming, alternative mechanisms would be necessary to implement this strategy. Rather than relying on traditional development assistance, Washington could expand strategic investments in AI research partnerships, education programs, and cloud-based open AI platforms through agencies more aligned with Trump's priorities.

The National Science Foundation (NSF) and DARPA could take the lead in funding open-source AI research, while the Department of Commerce, through its National Institute of Standards and Technology (NIST) could establish interoperability standards to ensure that open-source AI aligns with security and ethical best practices. Given the administration's emphasis on economic nationalism, the International Development Finance Corporation (DFC) and the Export-Import Bank (EXIM) could be leveraged to support AI infrastructure investments in the Global South.

Additionally, federal investment in cloud-based AI platforms, modeled after domestic AI infrastructure initiatives like Stargate, could offer emerging economies secure, scalable alternatives to Chinese AI solutions, without relying on traditional foreign aid mechanisms.

This approach would not only foster a more competitive and decentralised AI landscape but also ensure that AI development in the Global South aligns with open, transparent, and democratic governance principles, all while sidestepping commitments to multilateral engagement.

The Challenge of Multilateral A.I. Governance

As AI capabilities evolve, governance structures must keep pace. The Global South, which represents the majority in multilateral forums, is no longer a passive participant but a decisive force in shaping global AI norms. With priorities ranging from digital sovereignty to equitable AI access, these nations are asserting their influence in governance debates, driving independent initiatives, and challenging traditional power hierarchies.

A clear example of this was Brazil's leadership in UNESCO's AI Ethics Guidelines, where it played a pivotal role in advocating for principles of data sovereignty and equitable AI access, ensuring that AI policies consider the needs of developing economies. Yet, while China actively courts the Global South with scalable AI solutions and governance partnerships, the US has deprioritised multilateral AI diplomacy. The decision to halt USAID funding and abstain from international AI agreements, evidenced by its decision, along with the UK, not to sign the Paris AI summit statement, reflects a broader shift towards prioritising domestic AI development over global engagement.

Vice President JD Vance's emphasis on minimal regulation and American leadership signals a belief that US innovation alone will maintain AI supremacy. However, this fragmented approach weakens Western cohesion, creating space for China to shape AI governance frameworks that reinforce its technological and political influence.

For many emerging economies, AI adoption is not merely a technology shift, but a matter of digital sovereignty. Many governments increasingly recognise the risks of over-reliance on foreign AI providers, including data security vulnerabilities and economic dependencies that could limit their strategic autonomy. In response, nations such as India and Brazil are investing in domestic AI capabilities, while regional bodies like the African Union are crafting governance frameworks that align AI adoption with local ethical and regulatory principles.

Strategic A.I. Autonomy: A U.S. Path to Competing Without Forcing Dependence

Rather than retreating from multilateral governance, the US should champion strategic AI autonomy, not full detachment from global AI ecosystems, but ensuring that nations have access to trusted and diversified AI models that do not force them into dependence on China's AI infrastructure. This does not mean undermining US AI firms, but rather shaping AI partnerships that provide market access while offering a compelling alternative to China's state-backed AI exports. A balanced strategy would focus on capacity-building partnerships, secure cloud-based AI platforms and interoperable AI

ecosystems that allow emerging economies to develop their own AI capabilities while still benefiting from US-led innovation and infrastructure.

Some nations are also leveraging AI to strengthen regional cooperation. ASEAN, for instance, is developing a regional AI framework to establish standardised governance policies across member states. This initiative aims to balance AI adoption with regulatory oversight, preventing monopolistic control by any single external actor, whether China or the US. Similarly, Latin American nations are collaborating on AI ethics guidelines tailored to their socio-economic contexts, ensuring that AI development aligns with national interests, rather than foreign corporate or geopolitical imperatives.

By sidelining multilateral AI governance, Washington risks not only losing influence to Beijing but also neglecting the growing demand from the Global South for AI models and policies that reflect their economic realities and governance priorities. Given that developing nations represent the majority in multilateral discussions, their choices will shape the future of AI governance. If the US fails to engage, emerging economies will look elsewhere, shaping an AI trajectory that does not align with US and its allies' strategic interests.

A.I. Leadership: Will the U.S. Cede the Future to China?

The future of AI geopolitics will not be dictated solely by the US or China, but by the choices of the emerging AI economies of the Global South. However, as Washington retreats from multilateral AI diplomacy and fails to offer compelling alternatives, China is rapidly filling the vacuum. Through strategic investments, scalable AI solutions and deepening technological dependencies, Beijing is embedding its influence in the digital infrastructure and governance frameworks of the Global South.

If the US does not act decisively, by investing in open-source AI, capacity building and a more proactive AI foreign policy, it risks not just losing influence, but allowing China to set the rules of global AI governance unchallenged. Without a strategic response, Washington is not just stepping back from AI leadership, it is actively enabling China to define the future of global AI governance and technology provision on its own terms.[18]

[18] *The views represented herein are those of the author(s) and do not necessarily reflect the views of the Aspen Institute, its programs, staff, volunteers, participants, or trustees.*

Artificial Intelligence and Its Potential Effects on the Economy and the Federal Budget[19]

Congressional Budget Office Report

Summary

Artificial intelligence (AI) refers to computer systems that can perform tasks that have traditionally required human intelligence, such as learning and performing other activities that require cognitive ability. A general attribute of AI is its ability to identify patterns and relationships and to respond to queries that arise in complex scenarios for which the precise computational algorithm that is needed cannot be specified in advance.

Because AI has the potential to change how businesses and the federal government provide goods and services, it could affect economic growth, employment and wages, and the distribution of income in the economy. Such changes could in turn affect the federal budget. The direction of those effects—whether they increased or decreased federal revenues or spending—along with their size and timing, are uncertain. Some budgetary effects could occur relatively quickly, whereas others might take longer. In this report, the Congressional Budget Office provides an overview of the channels through which the adoption of AI could affect the U.S. economy and the federal budget.

How Might Artificial Intelligence Affect the Economy?

By increasing efficiency, enabling the development of new products, and altering the demand for workers, AI has the potential to change the economy, perhaps in ways that are difficult to predict. Whether or when those changes might occur is very uncertain. Surveys show that only 5 percent of businesses in the United States currently rely on AI to produce goods and services. For many businesses, customizing AI to their specific needs is costly, and it is unclear when those costs might fall. As a result, the use of AI is concentrated among larger, and younger, businesses in a few sectors of the economy—although that could change over time as profitable use of the technology became less dependent on a business's size.

Research into the performance of businesses that have implemented AI is still in its early stages, so conclusions from that research, which vary widely among studies, should be considered preliminary. So far, the research has found that businesses that implement AI can be expected to become more productive than businesses that do not.

[19] Originally posted by the Congressional Budget Office December 24, https://www.cbo.gov/publication/61147

Extrapolation of those results to the broader economy suggests that if AI's use became more widespread, it would boost economic growth. Evidence for AI's impact on employment and wages is also sparse and varies by type of AI. Studies of generative AI indicate that it could enhance the productivity of low-skilled workers within a given occupation; studies of earlier forms of the technology have found that AI boosted the wages of some skilled workers.

How Might Artificial Intelligence Affect the Federal Budget?

The use of AI could affect the federal budget through two basic channels: the economy and the government. Within each channel, AI could have an impact on revenues and spending. The timing of budgetary effects may vary.

AI's Use in the Economy. The use of AI could affect the overall amount of income in the economy and its distribution among businesses, investors, and workers. An increase in income would, by itself, eventually push up federal revenues. Initially, however, revenues could decline as businesses deducted from their income the cost of initial investments in the technology. Moreover, because different categories of income are taxed at different rates, changes to how income is distributed among workers and businesses could alter federal revenues. In particular, depending on how the demand for workers shifted in response to the use of AI, tax receipts tied to labor income could rise or fall. For workers who were left permanently unemployed or who took lower-paying jobs because of businesses' adoption of new technology, income and payroll taxes could decline; however, workers who were made more productive by AI could earn higher wages and remit larger tax payments. To the extent that AI created new kinds of tasks and jobs or led to economic growth through innovation, it could offset some potential losses of wages and taxes by increasing the demand for labor.

AI's use in the economy could change both mandatory spending (which does not require annual funding from the Congress) and spending subject to appropriation (which requires annual funding from the Congress). For instance, mandatory spending could increase to the extent that workers whose jobs were displaced by AI claimed benefits from federal income-support programs. But if AI boosted economic output and earnings, then such spending could decrease.

AI could also have an impact on federal spending through its use in the development of certain products, such as pharmaceuticals. For example, mandatory spending could increase to the extent that new drugs were paid for by federally subsidized health care programs. But that spending could decrease to the extent that the new drugs reduced the demand for other, more expensive, health care services. For AI's use in other types

of health care that receive a federal subsidy and are provided by the private sector, the ultimate impact is often unclear. In general, if AI's use enabled individuals to live longer, healthier lives, then it could boost federal revenues and spending in the long term. Revenues would be higher if more taxes were paid over a longer period. Spending would be greater if more claims were made on Social Security, Medicare, and other programs by individuals who lived longer in retirement than they would have otherwise.

Spending subject to appropriation could increase if the Congress expanded funding for AI's continued development by the public and private sectors. That could happen through federal programs for research and development (R&D) as well as through programs that regulate the technology's use.

AI's Use by the Government. The government's use of AI could change the amount of revenues collected through taxes and other sources. Federal revenues could rise, for example, if the Internal Revenue Service (IRS) was able to use AI to bolster its auditing capability and taxpayers' compliance with the federal tax code. By contrast, revenues would decline if businesses or individuals were able to use AI tools to reduce the taxes they owed.

The government's use of AI could also have various effects on mandatory spending and spending subject to appropriation. In particular, successful use of AI to reduce fraud could result in fewer improper payments in the largest mandatory spending programs: Medicare, Medicaid, and Social Security. Those efforts could be undermined, however, if individuals used the technology to perpetrate fraud.

AI could affect the spending subject to appropriation of federal agencies that made use of the technology. Although investments in AI might initially increase spending, those costs could eventually decline, depending on the efficiencies that were realized. For example, if AI substituted for labor, staffing requirements could fall.

Development and Use of Artificial Intelligence

The past few decades have seen rapid strides in AI's capabilities. Today, AI is embedded in commonly used software (such as web browsers, phones, and home assistants) and found in more specialized applications (such as industrial robots and self-driving cars). Although recent advances in AI have fostered optimism that significant further progress is imminent, several factors could hinder the technology's near-term development and use. In particular, a lack of large datasets to train AI models could restrict improvements in accuracy and applicability. In addition, obstacles to obtaining the energy needed to power advanced AI systems could impede their use.

Advances in AI Systems

Machine learning techniques underlie most forms of AI today. Discriminative AI models, for example, can distinguish between different types of images, such as cancerous and noncancerous human cells, on the basis of the characteristics that they have been trained to associate with each type. Human intervention is necessary both to select the initial computational algorithm (or mathematical model) and training data and, ultimately, to assess the AI system's performance. The output of an AI model, then, is fundamentally a prediction of the correct response to a particular question or to a new set of conditions.

The most advanced AI systems are composed of neural networks that are made up of thousands or even millions of nodes where computation takes place. Deep-learning AI consists of dozens of layers of nodes, and information can be exchanged between individual nodes and between layers. A general attribute of such systems, which are typically trained on very large datasets, is their ability to identify relationships and respond to queries that arise in scenarios in which the needed computational algorithm cannot be specified in advance.

Generative AI systems can, upon receiving a question or prompt from users, provide answers and other types of content that are comparable to what would have been produced by a person. For example, large language models generate human-like text and perform natural language processing in which machines are able to understand and interact with human speech. Large language models rely on data collected from public internet sites, and they generate responses to user queries in the form of text or, where applicable, computer code. Those models are based on what is known as the transformer architecture, which is a deep-learning neural network that transforms an input sequence into an output sequence.[1] Other generative AI systems can produce user-specified images, video, and music—even in the style of specific artists, if desired.

The current capabilities of AI systems reflect several decades of dramatic advances. An early, and highly touted, application of AI was IBM's Deep Blue, a computer capable of playing chess. That system beat world chess champion Gary Kasparov in a series of matches in the late 1990s. That application of AI was relatively rudimentary in that it relied on substantial human input to learn the rules and various strategies of chess and to improve its performance. In 2016, a more advanced system, based on neural networks, defeated one of the world's top players of Go, which is a board game considered much more challenging than chess.

Other types of AI also have displayed rapid progress. For example, discriminative AI—which learns the boundaries between different classes of input data to identify (or categorize) them correctly—achieved a breakthrough at the ImageNet Large Scale Visual

Recognition Challenge of 2012. In particular, by applying neural networks trained on a large dataset of images, one of the AI models competing in the challenge (AlexNet) achieved a very high degree of accuracy in recognizing images.[2] AI systems have continued to advance since then, and the performance of leading models on multitask tests has increased markedly over just the past few years.[3]

Near-Term Obstacles to Further AI Development

Despite recent rapid gains in the capability and accessibility of AI systems, the technology's history suggests that further progress could be interrupted. The field of AI research dates from the mid-1950s, and on more than one occasion, improvements in the technology heralded as breakthroughs have not led to follow-on success.

Recent advances in AI are attributed to the availability of large datasets and abundant power—in terms of both computer processing capability and the supply of energy—along with the development of increasingly complex computing algorithms. Some research suggests that AI could become widely used. However, in the near term, the adoption of AI also faces various obstacles.

One particular concern about wider adoption of AI is the looming scarcity of additional large datasets—containing either general, or task-specific, information—for training AI models. An additional concern is the ability of the electric power sector in the United States to supply sufficient electricity for powering and cooling the computers running the advanced AI software. The impact of data and energy scarcity is not certain; some analysts argue that although those factors exist, they need not constrain the technology's continued advancement.[4] This report focuses on the channels through which AI could affect the economy and the federal budget and does not assess the technology's future trajectory or the tasks that it might ultimately be able to perform.

Potential Effects of Artificial Intelligence on the Economy

Some people believe that AI could become as pervasive as electricity and computing are today. Although large investments in AI are currently being made by major businesses in the technology sector, its use by businesses overall remains limited. For that reason, research on the technology's economic effects is at an early stage, and the results from that research are uncertain. AI's effect on the performance of individual businesses is not yet well understood. Even though economy-wide productivity gains from the technology are expected, the size of those gains varies across studies. The general effects of AI's use on labor markets are also uncertain.

Productivity and Economic Growth

Many economists today view artificial intelligence as an emergent general-purpose technology. Such technologies do not have a sole definition, but they usually satisfy the following criteria:

- They can be applied throughout the economy;
- They are improved on a regular and sustained basis;
- Their use is accompanied by innovations in related areas (for example, new products and services); and
- They boost productivity and economic growth.

AI could transform society in the same way that technological advances like the steam engine and electrification did in the distant past and as computing and the internet have done over the past few decades.[5] For example, similar to the wide applicability of information technologies, AI has been found to boost the productivity of researchers looking to create new products in a variety of disciplines, such as materials science.[6]

Businesses' current use of AI remains limited, though. Surveys show that only 5 percent of businesses of a broad range of sizes in the United States (accounting for 9 percent of employment) currently incorporate AI—in more than an incidental way—in their production of goods and services. Furthermore, those businesses that use AI tend to be found in particular industries: Businesses in the "information" and "professional, scientific, and technical services" industries are roughly twice as likely as other businesses to be using AI.[7] Businesses' adoption of AI is found to be more frequent when surveys do not require respondents to identify the technology's use as being a significant part of the production process. For example, one survey found that about 28 percent of individuals reported using generative AI simply "for (their) job."[8]

Research into the economic impact of AI, in terms of the productivity gains of businesses implementing it as well as the implications of those gains for the economy's growth, is still relatively new. A common approach taken by researchers to study those effects has two steps. First, researchers identify tasks that AI is likely either to take over from workers altogether or to enable them to do better. Then, researchers convert those efficiency gains (or reductions in cost) to increases in the amount of a business's output that can be produced per hour worked by the business's employees—known as labor productivity. (Researchers sometimes use an alternative performance measure known as total factor productivity, which is the quantity of output produced relative to the amount of all inputs into production.)

Although the research on AI's economic effects focuses on labor, AI can make physical capital more productive as well. One example would be by enabling robots to grasp and

manipulate workpieces that they have not encountered before. That type of change would improve efficiency because processes would no longer need to be programmed each time a workpiece changed shape or some other dimension. As a result, the cost of manufacturing disparate products could fall.

According to research, businesses that implement AI can typically be expected to be more productive than those that do not. Some studies extrapolate from those impacts to estimate how much the economy's output could rise if AI's use became more widespread. (Such findings cannot directly be interpreted as having a similar effect on productivity in the economy as a whole. To be done properly, studies must take into account the interaction effects between, say, a new technology and other variables of interest, and their broader economic effects; such interactions are difficult to predict.) Although the impact of AI on the performance of individual businesses or of the economy overall is expected to be positive, the size of that impact varies greatly among studies.[9]

The conclusions of current studies about AI's broad economic impact are preliminary. In those studies, actual outcomes are not observed. Instead, the studies attempt to link AI to specific tasks, tasks to workers, and workers affected by AI to changes in businesses' performance. Surveys of businesses that are *not* currently using AI suggest that the eventual use of AI and its impact on workforces and performance remain poorly understood. For example, of the businesses that report that they do not use AI, four out of five state that "AI is not applicable to this business."[10]

Expense is a major obstacle to greater use of AI. Although leading AI systems have made great strides in performance in recent years, the cost of training them has increased—from tens of millions of dollars to hundreds of millions of dollars. Some analysts project AI training costs to reach $1 trillion by the end of the decade.[11] Customizing a given model so that it can be applied by individual businesses can entail costs that make AI prohibitively expensive for many businesses, even if that customization involves relatively established and advanced types of AI, such as computer vision (which allows a computer to analyze and identify images).[12]

Neither the lack of evidence to date that AI is substantially changing how goods and services are produced nor the fact that most businesses report that AI is not now advantageous to them means that the technology will not eventually have significant economic effects. For example, some researchers argue that AI will ultimately reach the stage of artificial general intelligence, at which point it will be able to carry out cognitive-based tasks as well as, or better than, human beings. If that occurred, the economic implications could be significant—in particular, for labor markets and the distribution of income.[13]

Indeed, a substantial lag often follows the availability of important new technologies and their measurable impact on the economy. In the case of electrification, for example, four decades after the first central power station opened for business in the United States, only about half of the mechanical drive capacity in factories had been electrified.[14] That said, in terms of technological diffusion, it is difficult to pinpoint where AI might be relative to such earlier timelines.

Employment and Wages

Research on the impact that AI has had on employment and wages is sparse. One survey found that roughly nine out of 10 businesses using AI report that they have not changed the number of workers they employ as a result and have no plans to do so. Among businesses that have made such changes or are planning to, employment increases are as likely to occur as decreases.[15]

The most accurate way to gauge AI's likely impact on the workforce is by assessing how the technology will affect the tasks that workers perform. One recent study found that the eventual impact of AI—specifically, large language models—on workers' jobs would vary: 80 percent of the U.S. workforce could have at least one-tenth of their tasks affected by AI, and 19 percent of workers could see at least half of their tasks affected.[16]

That potential scope of the use of AI makes it difficult to estimate how overall employment might change as a result of the technology and, in particular, how many workers might be displaced by it. Quantifying those effects would require establishing a threshold level of tasks for which AI is preferred over a worker and beyond which that worker would be deemed superfluous and have their job terminated. That threshold might not be the same for all businesses, because the increase in demand spurred by the decline in products' prices brought on by AI-driven cost reductions could vary. In some cases, greater demand could lead to retaining—and potentially adding—workers who would perform fewer tasks at a higher volume of production; in other cases, any additional demand could be insufficient to warrant keeping on staff the employees most affected by AI.[17]

The technology's effect on employees' tasks can take the form of either substituting for workers in accomplishing a particular task or serving as a complement to them—that is, make them more productive. One task-based study argues that the substitution effect of AI will outweigh the complementarity effect; as a consequence, AI would reduce labor's share of production.[18]

Other research examines the substitution and complementarity effects of AI on workers of different skill levels. Those studies explore how the wage differential between low-skilled and high-skilled workers could change based on a business's use of AI. If AI substituted for high-skilled workers, then the wage differential between the two groups of workers would decline; if it complemented high-skilled workers, the wage differential would rise. The opposite would be true for AI's impact on low-skilled workers; in particular, if the technology complemented those workers, the wage differential would fall.[19]

Evidence shows that generative AI can serve, at least to some degree, as a complement to low-skilled workers within a given occupation. By contrast, research on earlier forms of AI has found that the technology boosted the wages of some skilled workers.[20] A study of generative AI's impact on the productivity of customer support agents—measured by the number of issues resolved per hour—found that AI increased the productivity of entry-level and low-skilled agents by 34 percent. In contrast, experienced and highly skilled workers did not show significant gains in productivity. The authors suggest that "the AI model disseminates the best practices of more able workers and helps newer workers."[21] Other analysts argue that AI may be similarly skill-enhancing elsewhere, enabling workers who lack the experience or expertise of higher-paid employees to take on greater responsibility in the workplace and thus dampening earnings inequality among workers.[22]

Potential Effects of Artificial Intelligence on the Federal Budget

AI could affect the federal budget through two channels: its use in the economy, and its use by the government. Within each channel, AI could have an impact on revenues, mandatory spending, and spending subject to appropriation. Some budgetary effects might occur relatively quickly, whereas others might take longer to show up.

Through its use in the economy, AI could affect revenues by changing the amount of national income and its distribution. The technology could affect spending by altering participation in means-tested programs and the use of federally subsidized health care services, as well as by inducing more funding for federal programs that could support AI's continued development and governance.

Through its use by the federal government, AI could affect both revenues and spending by increasing the efficiency of the government in collecting tax revenues and in distributing those revenues through transfer payments. AI also could enable improvements in the goods and services provided by the government, spurring federal programs to spend more to take advantage of the technology. Overall, the ultimate impact of AI on federal revenues and spending is uncertain.

Budgetary Impact of AI's Use in the Economy

The use of AI in the economy could affect revenues by changing the amount and distribution of income. In addition, it could affect mandatory spending, by changing participation in means-tested programs and the use of federally subsidized health care services, and spending subject to appropriation, by changing the policy choices that underlie funding decisions for various programs each year.

Revenues. The amount of income and how that income is distributed could change as a result of AI's use in the economy. An increase in income that affected businesses, investors, and workers proportionately would boost federal revenues. Because different categories of income are taxed at different rates, though, changes to the distribution of income among categories could offset that increase. Moreover, businesses' taxable income could decline, at least temporarily, as a result of investments in AI.

If businesses became more productive, profits would eventually increase, pushing up earnings from capital, which accrue either to the owners of privately held businesses or to investors in publicly traded ones. As a result, tax payments on business income and returns from equities (in the form of dividend payments or capital gains) would rise. Those higher tax revenues could follow a period during which tax payments associated with AI fell because of tax deductions taken for investments in it and, potentially, because of lower profits during the initial (and typically costly) stage of implementing a new technology.

The effect of AI on taxable labor income is uncertain. It would depend on the extent to which the positive effects of a larger economy were offset by the potential negative effects that could occur if AI substituted for labor. Tax receipts tied to labor income could rise or fall depending on how the demand for workers shifted in response to the use of AI. If the use of AI was complementary to existing jobs (rather than a substitute for them), it could enable workers to do their current jobs better and perhaps undertake new tasks as well—making employees more productive and leading to higher wages. AI could also spur job creation if it enabled the production of new goods and services. Furthermore, if AI's use led to income gains from higher wages and job creation, it could have a positive impact on federal receipts tied to labor income.[23] In contrast, if workers were left permanently unemployed or were reallocated to lower-paying jobs by the technology, income and payroll taxes would decline.

Mandatory Spending. Participation in means-tested programs could rise or fall depending on the net effect that AI had on employment. Like revenues, mandatory spending programs—those whose spending is generally determined by formulas and

eligibility criteria established by lawmakers rather than by annual appropriations—could also be affected by economic factors.

For example, even if the workers displaced by AI eventually found new jobs, that labor reallocation process could be lengthy. As a result, spending could increase over an extended period for federal income-support programs (including unemployment benefits, health care subsidies, and other means-tested programs), which provide cash payments or other assistance to people with relatively low income and few assets. But if AI increased economic output, employment, and wages, then mandatory spending on income-support programs could fall.

Changes in mandatory spending could also occur as a result of the use of specific types of AI in the economy. For example, AI is currently being applied in pharmaceutical R&D.[24] If patients' consumption of the new drugs that AI helped discover were paid for by health insurance subsidized by the federal government, then federal outlays to provide pharmaceutical benefits could increase.

The provision of health care services is a notable example of the potential difficulty in distinguishing between AI's use in the economy and its use by the federal government. That is because federal funding for activities that use AI can extend beyond the government to the private sector. For example, the National Institutes of Health (NIH) could fund AI-supported R&D that is carried out by staff in its own laboratories and by researchers elsewhere. In the former scenario, government employees would use AI, a clear-cut instance of use of the technology by the government. In the latter scenario, the technology could be used by, say, university faculty and affiliated researchers whose jobs were not dependent on federal financial support—even though some of the projects reliant on AI were contingent on receiving federal funds. A similar overlap applies to AI's potential use in other types of health care that benefit from federal subsidies (see Box 1).

Box 1.

Artificial Intelligence and Federally Subsidized Health Care

Artificial intelligence (AI) could be used in a number of areas of federally subsidized health care. Although the government would be providing financial support to patients, say, through the federal tax subsidy for employer-sponsored health insurance or through the Medicare and Medicaid programs, service providers would typically be in the private sector. For this analysis, the Congressional Budget Office would consider such cases to be examples of the use of AI in the economy rather than of the use of AI by the federal government—even though the government would be subsidizing a sizable share of the cost. In contrast, an example of health care provided by a federal agency is

services delivered by the Veterans Health Administration, which is funded through annual appropriation acts.

In many cases, the net effect on the federal budget of the use of AI in the economy is unclear. That is in part because AI can be used by participants in the health care system who have different objectives with potentially countervailing effects. (The text of this report discusses similar outcomes for other applications of AI.) Furthermore, the effects of AI's use in health care may vary across applications, sometimes leading to increases in federal spending and other times, to reductions.

One potential application of AI in the health care sector is for prior authorization (PA), a cost-control process that insurers use to limit access to high-cost services and drugs. Insurers could use the technology to streamline reviews of PA submissions, making it cheaper for them to roll out broader PA plans. Although broader PA plans would reduce health care costs and federal subsidies, health care providers could, in turn, use AI to clear the hurdles put in place by insurers, which would have an offsetting effect on those savings.

Another potential application of AI is to streamline the provision of health care services. By introducing efficiencies in how health care practices are managed and administered, AI could enable providers to furnish more services. In that case, costs to the federal government would increase. However, by enabling greater machine- and software-based health monitoring and treatment, AI could lower the cost (and, potentially, the allowable reimbursement) of those services and thus reduce federal spending.

Spending Subject to Appropriation. AI could affect spending for a variety of federal programs that require an appropriation by the Congress each year. For example, the reallocation of employment brought about by AI's use in the economy might lead lawmakers to provide larger appropriations for educating and retraining displaced workers.

The use of AI in the economy could also influence funding for federal programs unrelated to employment. Businesses that produced AI systems might invest in R&D to advance the technology, for instance. If they were not able to appropriate all of the returns to that R&D, they would tend to invest less than the economically efficient level (that is, the level at which the financial payoff from additional R&D matched the investment in it). That underinvestment could be more likely to occur for AI because the technology is expected to foster innovation in many sectors of the economy, thus

making the value of R&D for different users difficult to determine in advance. As a result, the Congress may supplement private-sector R&D efforts by increasing federal funding for research and development into AI.

Lawmakers have enacted legislation that addresses AI's use both in the economy and by the federal government. Examples include the following:

- The AI in Government Act of 2020 (Public Law 116-260), which created within the General Services Administration the AI Center of Excellence to facilitate adoption of AI by federal agencies;
- The National Artificial Intelligence Initiative Act of 2020 (P.L. 116-283), which supports AI's use in both the public and private sectors; and
- Other laws, such as the Advancing American AI Act (P.L. 117-263), which focuses on monitoring and setting policies for AI's use by the government.

As the technology is increasingly applied in the private sector, lawmakers might enact additional legislation that could affect federal spending. In particular, widespread use of AI might raise additional governance issues for the Congress and lead it to fund new or expanded federal agencies. One potential issue concerns data privacy. The most advanced AI systems rely on large quantities of data, which could result in improper use of personal information.

Budgetary Impact of AI's Use by the Federal Government

The federal government's use of AI could affect both revenues and spending. The overall effect of AI on revenues collected by the tax system is ambiguous: Some factors will probably increase revenues, and others will decrease them. In terms of mandatory spending, AI could reduce the amount of payments made by the federal government, thus decreasing spending. But individuals' use of AI could counteract the government's efforts—if, for example, people used AI to forge identity documents that allowed them to fraudulently claim benefits. By enabling improvements in the quality of the goods and services that the government provides and the efficiency with which the government provides them, AI could lead to changes in mandatory spending and in the spending subject to appropriation of federal programs that sought to take advantage of the technology.

Revenues. One way that AI could increase revenues is by improving the ability of the Internal Revenue Service to detect noncompliance by taxpayers and enforce laws intended to ensure compliance. The IRS is using AI to improve its estimates of the federal tax gap, which is the difference between the tax payments that individuals voluntarily make on a timely basis and the amount that they owe. Estimates of the tax gap are based on an analysis of audited tax returns that identifies noncompliance with

the tax code missed by IRS auditors. The IRS expects that AI will enable examiners to identify noncompliance better.[25] If the agency was able to use that information to increase compliance with tax laws, revenues would increase.

AI could be used to reduce revenues as well. One way that might happen is if businesses and corporations used AI to reduce their tax liability. Some areas of tax law are particularly complex, and AI could be used to legally decrease tax payments by identifying aspects of the tax code that a person might miss.

Mandatory Spending. The government's use of AI could have various effects on mandatory spending. One notable example concerns improper payments, which are payments that should not have been made or that were made in the incorrect amount. Those payments, which cost the government billions of dollars each year, can arise from incorrect and fraudulent billing in federal programs that pay private entities to provide services. By analyzing billing data and determining the characteristics of reimbursement claims that are likely to be incorrect or fraudulent, AI could help identify and reduce unwarranted federal payments and, as a result, lower spending. Efforts are already underway to apply AI to reduce improper payments in the Medicare, Medicaid, and Social Security programs.[26]

Estimates of improper payments vary substantially in scope, time period, and amount. The Government Accountability Office (GAO) reports that executive branch agencies' cumulative estimates of the improper payments they made from 2003 to 2023 totaled about $2.7 trillion, or roughly $130 billion on an annual basis (not adjusted for inflation).[27] Adopting a different methodology, which applies a statistical model to 12 federal agencies' spending and is limited to estimating excessive payments solely from fraud, GAO found that federal losses from 2018 to 2022 totaled $233 billion to $521 billion—or about $47 billion to $104 billion annually over that five-year period—and that the range of losses was attributable both to uncertainty surrounding the estimates and to the different risk conditions prevailing over the period.[28]

For several reasons, it is unclear whether or how much the use of AI might eventually reduce the improper payments made by the federal government. Evidence about the technology's effectiveness in the applications described above is lacking, and it is unclear how widely AI will eventually be deployed in those applications. (Studies of states' use of the technology to establish eligibility for public assistance programs point to shortcomings in AI's performance.)[29] Furthermore, GAO suggests that although AI could be a useful tool for the federal government's efforts to reduce fraud, the technology could also be used by individuals to perpetrate fraud. For example, AI could be used to create fake images for falsified documents.[30]

As noted earlier, the use of AI by federal employees carrying out R&D in health care could cause mandatory spending to rise if the outcome of that research, such as demand for new pharmaceuticals, received a federal subsidy. That increase could be offset by a decline in spending elsewhere if the new drugs eliminated the need for other types of federally subsidized health care.

Spending Subject to Appropriation. For federal programs that make use of the technology, AI could affect their spending subject to appropriation. AI could, for example, be especially useful at the Department of Defense, an agency that relies heavily on information analysis.[31] DoD has a large workforce—including members of the military, federal civilian workers, and private contractors—and many complex tasks that it must undertake to carry out its mission. DoD could use AI to manage its operations, enhancing the capability of current weapon systems and developing new ones. Effects on costs could be mixed: Increases in operational efficiency would tend to reduce costs; developing new weapon systems could either increase or decrease costs depending on how the cost of new systems compared with the cost of existing ones. And the AI systems themselves would require spending by the government. Less than 1 percent of DoD's 2024 budget request is for AI.[32]

For other programs and agencies, AI could also have an impact on spending subject to appropriation. In particular, the increased labor productivity projected from AI's use in the private sector could be realized for the federal workforce or federal contractors through efficiencies (and cost reductions) in how the federal government operates or delivers services. One example is AI's support for, and potential replacement of, staff in the call centers of programs operated by the General Services Administration, Internal Revenue Service, and Department of State.

Although investments in AI might initially increase agencies' spending, costs could eventually decline. The size of any decrease would depend on how many efficiencies were realized. For example, if AI substituted for labor, it could reduce staffing requirements.

Other Considerations About Budgetary Effects

Many aspects of AI's adoption, use, and effects are uncertain. In assessing the ways that AI could affect federal revenues and spending, several considerations should be kept in mind.

First, the examples in this report are illustrative. Whether describing projects that are underway or suggesting projects that are possible, the examples this report provides are intended simply to illustrate the various ways that the technology could affect federal revenues or spending. The federal government may currently, and in the future, use AI

in many more ways than the applications listed above, but the extent of the technology's eventual adoption and use are uncertain.[33]

Second, outlays for AI could take the place of other federal spending. For statistical analysis, for example, AI could replace software that the government has been using and would continue to rely on if AI was not available. Use of AI might also entail greater reliance on cloud computing, which could take the place of upgrades to agencies' computer hardware. In that case, determining the true budgetary cost of AI would require netting out the previous federal spending that was done to pay for licenses and other costs associated with using the abandoned software and hardware.

Third, the impact of AI's use could manifest in different ways over time, and those ways could have countervailing effects on the federal budget. If AI improved the quality of federally subsidized pharmaceuticals and health care more broadly, it could enable individuals to live longer, healthier lives. That development could increase revenues (because more taxes would be paid over a longer period) as well as spending (because benefits would be provided through Social Security, Medicare, and other programs for more years) in relation to the amounts that would otherwise have been received and spent.[34]

The budgetary effects would also depend on whether federal agencies made effective decisions about adopting AI systems. For example, AI models could be expensive to acquire, train, and operate. At a minimum, those costs would cover buying the software, paying employees to maintain and operate the AI models, and purchasing the electricity needed to run them; costs would probably vary by application and by agency. For AI systems to either increase revenues or reduce spending, those positive budgetary impacts would need to outweigh the negative effects of the outlays required to obtain and implement the technology.

Fourth, AI could have impacts that would not be fully reflected in federal budget totals. For government spending on pharmaceutical R&D, for instance, AI could enable innovations of higher quality than would have been possible otherwise. Consequently, even if new drugs developed with AI were more expensive, they may deliver more benefits to patients per dollar of R&D spending than would have been possible without the technology. In such cases—which could also arise from other, nonpharmaceutical applications of AI, such as increased security through a more robust national defense—the payoff from greater cost-effectiveness would not necessarily be captured, or at least identifiable, in data about federal revenues or spending.

1. 1. Ashish Vaswani and others, “Attention Is All You Need” (paper presented at the 31st Conference on Neural Information Processing Systems, December 2017, updated August 2023), https://arxiv.org/abs/1706.03762.
2. 2. Olga Russakovsky and others, “ImageNet Large Scale Visual Recognition Challenge,” *International Journal of Computer Vision*, vol. 115 (April 2015), pp. 211–252, https://doi.org/10.1007/s11263-015-0816-y.
3. 3. “Multi-task Language Understanding on MMLU,” https://tinyurl.com/2dk7jymj.
4. 4. Jaime Sevilla and others, *Can AI Scaling Continue Through 2030?* (Epoch AI, August 20, 2024), https://tinyurl.com/bdnx4627.
5. 5. Iain M. Cockburn, Rebecca Henderson, and Scott Stern, *The Impact of Artificial Intelligence on Innovation*, Working Paper 24449 (National Bureau of Economic Research, March 2018), www.nber.org/papers/w24449; and Timothy F. Bresnahan and M. Trajtenberg, “General Purpose Technologies: ‘Engines of Growth’?” *Journal of Econometrics*, vol. 65, no. 1 (January 1995), pp. 83–108, https://doi.org/10.1016/0304-4076(94)01598-T.
6. 6. Aidan Toner-Rodgers, *Artificial Intelligence, Scientific Discovery, and Product Innovation* (Massachusetts Institute of Technology, November 6, 2024), https://aidantr.github.io/files/AI_innovation.pdf.
7. 7. Kathryn Bonney and others, *Tracking Firm Use of AI in Real Time: A Snapshot From the Business Trends and Outlook Survey*, Working Paper 32319 (National Bureau of Economic Research, April 2024), pp. 2–4 and p. 39, www.nber.org/papers/w32319. In that survey, businesses were asked whether they used AI “in producing goods or services,” and the following examples were provided as guidance for respondents: “machine learning, natural language processing, virtual agents, voice recognition, etc.” Hence, reported AI adoption should not reflect incidental use, such as the AI embodied in common software like web browsers or business applications.
8. 8. Alexander Bick, Adam Blandin, and David J. Deming, *The Rapid Adoption of Generative AI*, Working Paper 32966 (National Bureau of Economic Research, September 2024), p. 11, www.nber.org/papers/w32966.
9. 9. Daron Acemoglu, *The Simple Macroeconomics of AI*, Working Paper 32487 (National Bureau of Economic Research, May 2024), www.nber.org/papers/w32487; Michael Chui and others, *The Economic Potential of Generative AI: The Next Productivity Frontier* (McKinsey, June 2023), https://tinyurl.com/y9tevm2n; and Joseph Briggs and Devesh Kodnani, *The Potentially Large Effects of Artificial Intelligence on Economic Growth* (Goldman Sachs Economics Research, March 26, 2023), https://tinyurl.com/4anjjbdf.
10. 10. Kathryn Bonney and others, *Tracking Firm Use of AI in Real Time: A Snapshot From the Business Trends and Outlook Survey*, Working Paper 32319 (National Bureau of Economic Research, April 2024), Table 7, p. 37, www.nber.org/papers/w32319.

11. 11. Anton Korinek, *Economic Policy Challenges for the Age of AI*, Working Paper 32980 (National Bureau of Economic Research, September 2024), p. 5, www.nber.org/papers/w32980; and Ben Cottier and others, *The Rising Costs of Training Frontier AI Models* (arXiv, May 31, 2024), https://arxiv.org/abs/2405.21015.
12. 12. Martin Fleming, Wensu Li, and Neil C. Thompson, *The Last Mile Problem in AI: Why Job Automation Will be Slower Than Technological Progress Suggests* (Brookings, August 29, 2024), www.brookings.edu/articles/the-last-mile-problem-in-ai/.
13. 13. See, for instance, Anton Korinek, *Economic Policy Challenges for the Age of AI*, Working Paper 32980 (National Bureau of Economic Research, September 2024), www.nber.org/papers/w32980.
14. 14. Paul A. David, "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox," *American Economic Review*, vol. 80, no. 2 (May 1990), pp. 355–361, www.jstor.org/stable/2006600.
15. 15. Kathryn Bonney and others, *Tracking Firm Use of AI in Real Time: A Snapshot From the Business Trends and Outlook Survey*, Working Paper 32319 (National Bureau of Economic Research, April 2024), Table 6, p. 36, www.nber.org/papers/w32319.
16. 16. Tyna Eloundou and others, *GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models*, Working Paper 10130 (arXiv, March 2023, updated August 2023), https://arxiv.org/abs/2303.10130.
17. 17. Other approaches to quantifying the effects that AI could have on employment consider how it could take the place of the "expertise" that workers supply to their employer and posit how the demand for different types of labor will vary depending on the extent to which AI substitutes for the expertise associated with it. See the National Academies of Sciences, Engineering, and Medicine, *Artificial Intelligence and the Future of Work* (2024), pp. 88–95, https://tinyurl.com/4cxu6uwy.
18. 18. Daron Acemoglu, *The Simple Macroeconomics of AI*, Working Paper 32487 (National Bureau of Economic Research, May 2024), www.nber.org/papers/w32487.
19. 19. David E. Bloom and others, *Artificial Intelligence and the Skill Premium*, Working Paper 32430 (National Bureau of Economic Research, May 2024), www.nber.org/papers/w32430; and Mauro Cazzinga and others, *Gen-AI: Artificial Intelligence and the Future of Work*, IMF Staff Discussion Note SDN2024/001 (International Monetary Fund, January 2024), https://doi.org/10.5089/9798400262548.006.
20. 20. Edward W. Felten, Manav Raj, and Robert Seamans, *The Occupational Impact of Artificial Intelligence: Labor, Skills, and Polarization* (NYU Stern School of Business, September 8, 2019), https://ssrn.com/abstract=3368605.

21. 21. Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond, *Generative AI at Work*, Working Paper 31161 (National Bureau of Economic Research, revised November 2023), www.nber.org/papers/w31161.
22. 22. David Autor, *Applying AI to Rebuild Middle Class Jobs*, Working Paper 32140 (National Bureau of Economic Research, February 2024), www.nber.org/papers/w32140.
23. 23. Tania Babina and others, "Artificial Intelligence, Firm Growth, and Product Innovation," *Journal of Financial Economics*, vol. 151 (January 2024), https://doi.org/10.1016/j.jfineco.2023.103745; and Dean Alderucci and others, "Quantifying the Impact of AI on Productivity and Labor Demand: Evidence From U.S. Census Microdata" (draft, 2019), https://api.semanticscholar.org/CorpusID:237265713.
24. 24. Government Accountability Office, *Artificial Intelligence in Health Care: Benefits and Challenges of Machine Learning in Drug Development*, GAO-20-215SP (January 21, 2020), www.gao.gov/products/gao-20-215sp.
25. 25. Government Accountability Office, *Tax Gap: IRS Should Take Steps to Ensure Continued Improvement in Estimates*, GAO-24-106449 (June 5, 2024), www.gao.gov/products/gao-24-106449.
26. 26. See the entries for the Department of Health and Human Services—Centers for Medicare & Medicaid Services and the Social Security Administration in the database of uses of AI by federal agencies at AI.gov, "The Government Is Using AI to Better Serve the Public" (September 1, 2023), https://ai.gov/ai-use-cases/.
27. 27. Government Accountability Office, *Improper Payments: Information on Agencies' Fiscal Year 2023 Estimates*, GAO-24-106927 (March 26, 2024), pp. 1–2, www.gao.gov/products/gao-24-106927.
28. 28. Government Accountability Office, *Fraud Risk Management: 2018–2022 Data Show Federal Government Loses an Estimated $233 Billion to $521 Billion Annually to Fraud, Based on Various Risk Environments*, GAO-24-105833 (April 16, 2024), www.gao.gov/products/gao-24-105833.
29. 29. Kevin De Liban, *Inescapable AI: The Ways AI Decides How Low-Income People Work, Live, Learn, and Survive* (TechTonic Justice, November 2024), www.techtonicjustice.org/reports/inescapable-ai.
30. 30. Government Accountability Office, *Fraud Risk Management: 2018–2022 Data Show Federal Government Loses an Estimated $233 Billion to $521 Billion Annually to Fraud, Based on Various Risk Environments*, GAO-24-105833 (April 16, 2024), p. 35, www.gao.gov/products/gao-24-105833.
31. 31. Department of Defense, *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to Advance Our Security and Prosperity* (February 12, 2019), https://tinyurl.com/mrxpv6tm.
32. 32. Office of the Under Secretary of Defense, Comptroller/Chief Financial Officer, *United States Department of Defense Fiscal Year 2024 Budget Request* (March 2023), p. 16, https://tinyurl.com/mr26n622.

33. 33. See AI.gov, “The Government Is Using AI to Better Serve the Public” (September 1, 2023), https://ai.gov/ai-use-cases/. That database lists specific instances of the use of AI by the federal government. Also, agencies implementing AI typically have a portion of their website dedicated to it.
34. 34. Government Accountability Office, *Artificial Intelligence in Health Care: Benefits and Challenges of Technologies to Augment Patient Care*, GAO-21-7SP (November 30, 2020), www.gao.gov/products/gao-21-7sp.