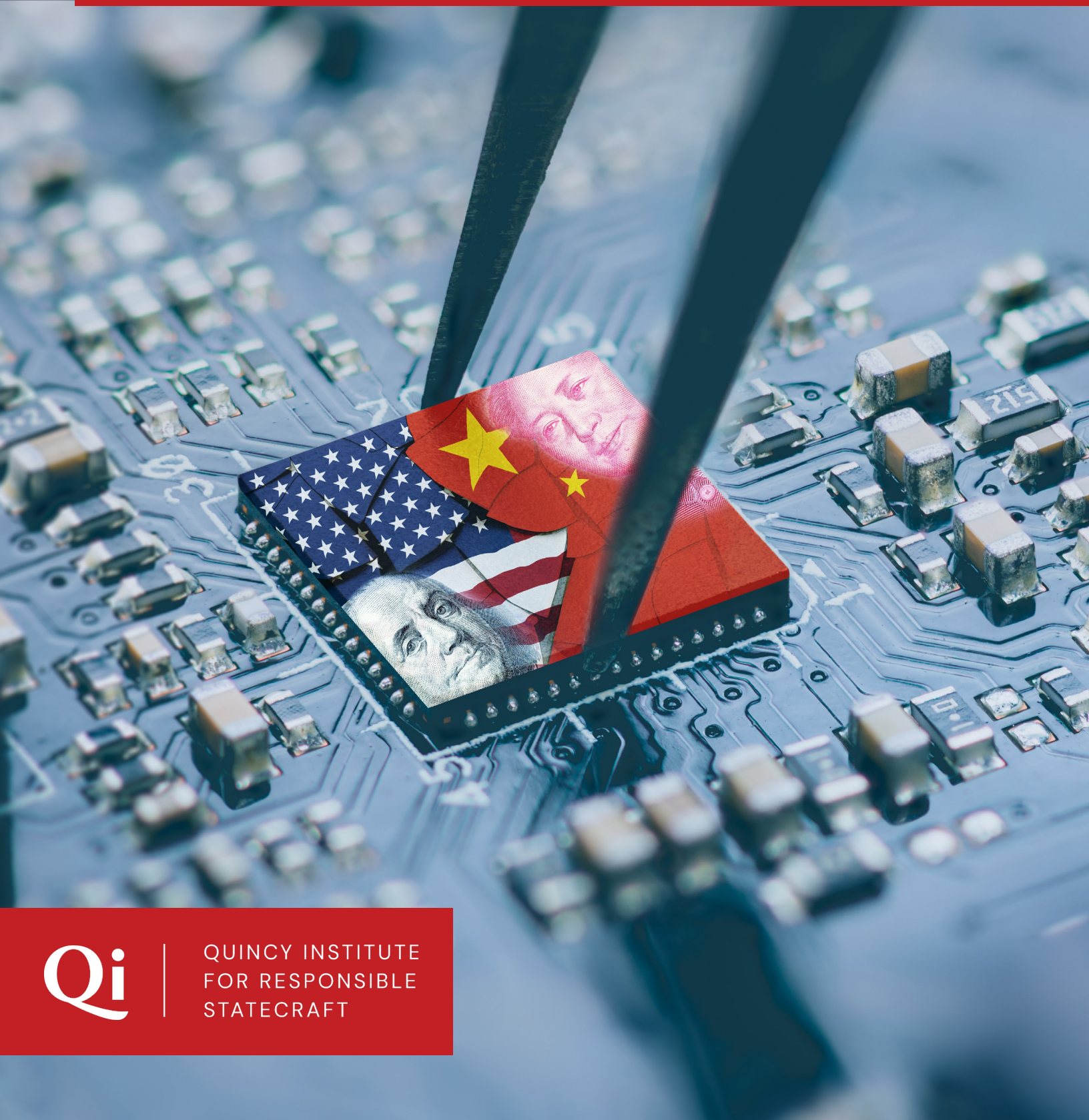
The Wrong Race: The US, China, and AI Competition

ALVIN WANG GRAYLIN

QUINCY BRIEF NO. 110
SEPT. 2026



QUINCY INSTITUTE FOR RESPONSIBLE STATECRAFT

Executive Summary

The conventional wisdom in Washington says the United States and China are locked in an existential race to reach a decisive strategic advantage in developing artificial intelligence, or AI. Take Treasury Secretary Scott Bessent's recent comments: "Beating China — there is no day after tomorrow if China wins at this … If they were to pull ahead of us on AI, then nothing else matters." President Donald Trump has made similar comments lately, citing competition with China as the reason for his refusal to slow down AI growth, stating that "whoever wins AI wins."

Such assumptions about the AI race now underwrite close to a trillion dollars a year in capital expenditure, a comprehensive export control regime, and arguments that AI is a national security matter that should not be subject to the ordinary give-and-take of market competition and government oversight. At the extreme, some in Washington have even raised the possibility of military strikes should China appear poised to gain artificial general intelligence, or AGI, first.

But the assumptions behind the "AI race" framework are deeply flawed. While there are legitimate national security concerns around AI, there is no decisive "finish line" to AI innovation that will grant a final victory to either the United States or China. As with previous transformative technologies, we should prepare for the continuing long-term development of AI through competitive markets, not total geopolitical victory for one side. US–China capability gaps at the frontiers of AI technology typically last only a few months, and there is no competitive "moat" that will secure a permanent advantage to either nation.

The idea that the greatest AI security threat to the United States lies only in large frontier models sponsored by a nation-state like China is flawed as well. The most dangerous models are often those smaller specialized ones that could be created or utilized by non-state rogue actors for the purpose of deploying biological, chemical, or cyber weapons. The assumption of US–China conflict prevents the exact global cooperation to protect against this very real threat.

The upcoming summit between President Trump and President Xi Jinping is an opportunity to chart a cooperative path forward that moves beyond the flawed model of an existential AI race and establishes sensible cooperative measures around shared harm and safety standards. We outline specific proposals for such measures in this brief.

Beyond the summit, the United States should reject the idea of any "national security exemption" for AI labs from ordinary market discipline and legal regulation and, instead, take steps to safeguard American workers from the threat of displacement by automation.

The essay below is a compilation of recent publications written by Alvin W. Graylin regarding the US–China AI race, especially prepared for the Quincy Institute for Responsible Statecraft. Original sourced links are available in footnotes and at the end of the document.

The race we think we are running

Nearly every major decision Washington has made since 2022 regarding artificial intelligence, or AI, rests on three assumptions that are rarely examined:

1. The United States is locked in an "AI race" with China toward a single "finish line," usually identified as artificial general intelligence, AGI, or artificial superintelligence, ASI.

2. Whichever country that crosses this finish line first will gain a *decisive strategic advantage*, a durable and possibly permanent lead that completely reshapes the global balance of power.

3. The way for the United States to win this race is maximum acceleration of AI development at home paired with maximum denial of China's AI capacity abroad.

These three propositions now underwrite well over a trillion dollars per year in global capital expenditure and an export control regime that has now expanded to lithography; electronic design automation, or EDA, software; high bandwidth memory, or HBM; specialty chemicals; and more. At the extreme of government action, some in Washington have even raised the possibility of military strikes should China appear poised to gain AGI first.[1]

Assumptions around the US–China AI race propel the argument that America cannot afford to regulate its own AI industry as well as the claim that AI policy is national security policy and therefore belongs outside the ordinary give-and-take of markets and oversight.

Recently, Treasury Secretary Scott Bessent summed it up this way: "Beating China — there is no day after tomorrow if China wins at this … If they were to pull ahead of us on AI, then nothing else matters."[2] In other words, the entire future of the United States and the world depends on winning an AI race with China.

Of course, the competition with China is real. The security risks are genuine. I am not naive about Beijing's strategic ambitions. But after 35 years building businesses and deploying products across every layer of the AI stack on both sides of the Pacific, I have watched confident assumptions about technological dominance unravel. I understand the mindset of leaders in these two countries, and I can see what is wrong with the assumptions behind the AI race.

We need to reexamine the race we think we are running, the game we think we are playing, and the strategy we have chosen. The evidence now suggests our strategy is optimized for a competition that is not the one taking place and that the costs of that mismatch are landing on American workers, American financial stability, and American security, not on Beijing.

1 Jacob Stokes, *Superpowers and AGI: How US–China Competition Could Be Shaped by the Emergence of Artificial General Intelligence* (Washington, DC: Center for a New American Security, 2026), https://www.cnas.org/publications/reports/superpowers-and-agi.

2 Ananya Gairola, "Scott Bessent Warns US 'Can't Pause' AI Race with China: 'There Is No Day After Tomorrow' If Beijing Wins," *Benzinga*, Sept. 9, 2026, via MSN at https://www.msn.com/en-us/money/general/ar-AA2bQica.

No finish line, no moat

The core of an AI race model is the idea that a decisive strategic advantage will result from crossing the AI finish line first.

But there is no clear finish line in the AI race with China. AI development is a continuous, multidimensional process with no single threshold that confers omnipotence. AGI, however it is defined, would not be a static superweapon but a rapidly evolving ecosystem of software that competitors can and will replicate. Moreover, AGI, unlike weapons-grade nuclear material, weighs nothing and crosses borders in minutes. Nuclear weapons fit the "permanent dominance" model far better than AI does: discrete, physically securable, unambiguously catastrophic. Still, America's nuclear monopoly lasted just four years.

US capability leads over China tend to be short-lived and cannot be sustained by denial of high-end chips and compute power. The performance gap between US–China frontier AI models, once estimated at 12–14 months, has narrowed to roughly 2–3 months, during a period when Washington steadily *ramped up* controls.[3] On OpenRouter, the largest model-routing platform, Chinese open-weight models went from under 2 percent of token traffic in late 2024 to a majority among top models this year.[4] As of this July, eight of the top ten models by usage were open weight, with Chinese models holding the first six positions outright. Most recently, the Chinese companies Moonshot AI and DeepSeek have released the open-source Kimi K3 and V4.1 Flash models, which are competitive with the leading-edge models from US rivals like Anthropic and OpenAI.[5]

Controls do make a difference. Lack of leading-edge chips has reduced China's compute capacity for safety testing, but they have not brought the United States a durable capability lead.

The US–China dynamic is not winner-takes-all. In practice, AI models are becoming interchangeable commodities at a speed that ought to frighten anyone expecting to charge a premium price to earn back large capital expenditures. Recent Chinese releases land within a few percentage points of leading US frontier models at roughly one-tenth the application programming interface, or API, price.[6] DeepSeek's V4 Flash trails OpenAI's GPT–5.6 Luna by a single point on the Artificial Analysis Intelligence Index, and, even after OpenAI's 80 percent price cut, cost per task is still 60 percent lower for DeepSeek.[7] When intelligence is cheap and abundant, strategic advantage shifts from who *builds* the best model to who *deploys* it most effectively across their economy.

What about the national security and military case?

The most serious concern is national security: *If China gets to AGI first, it will weaponize it against the United States*. This is a legitimate security concern worthy of serious attention.

But consider the problem of converting an AI advantage into a permanent military advantage. Given how quickly China catches up to US AI capabilities, any frontier AI advantage is not likely to last very long and would therefore not confer a permanent risk-free military advantage. The Mythos episode this spring made the point more cleanly

3 Alvin W. Graylin, "Misdiagnosing the US–China AI Race: Recalibrating America's Approach to an Incomplete Strategy," Asia Society, May 1, 2026, https://centerforchinaanalysis.asiasociety.org/p/misdiagnosing-the-uschina-ai-race.

4 "AI Model Rankings," OpenRouter, last updated Sept. 14, 2026, https://openrouter.ai/rankings.

5 Francisco Velasquez, "China's Moonshot AI Claims Kimi K3 Can Rival OpenAI and Anthropic," BBC, July 17, 2026, https://www.bbc.com/news/articles/cy9w4q8pgp0o.

6 "Artificial Analysis," Artificial Analysis, https://artificialanalysis.ai.

7 "Comparison of AI Models: Intelligence, Performance, and Price Analysis," Artificial Analysis, https://artificialanalysis.ai/models.

than any argument could. A model held out as proof that denial works was reportedly accessible to unauthorized personnel within weeks.[8] **We have no moat that makes a frontier AI advantage permanent, and neither does China.**

Thus, if one country reaches AGI or an AI threshold before the other, the advantage will evaporate in weeks or months, unless the leading power is prepared to use it aggressively against its rival in order to try to sabotage its progress. But both the United States and China are nuclear powers, and the aggressive use of an AI advantage for such a "preemptive strike" against an AI competitor would risk escalation to other forms of warfare, including nuclear use.

Indeed, the nuclear arms race itself demonstrates that "winning" the race even to a world-changing military capacity does not guarantee permanent military domination. The US nuclear monopoly lasted four years. Any AI advantage would last a much shorter time and, since AI is highly replicable, would not provide the same military advantage that nuclear weapons did. As one analyst put it in a recent report: "Any AGI–enabled military capability would likely change only some aspect of a military domain rather than provide the singular type of comprehensive, deterrence-assuring, war-winning capability that the term "superweapon" suggests. Even nuclear weapons are not superweapons in this holistic sense."[9]

The US–Iran conflict demonstrates this. Despite being the world's AI leader and Iran having little or no AI capacity, the United States has still not been able to subdue Iranian forces or the Houthis in Yemen. AI will not be a "superweapon" that guarantees world domination for the United States or China.

As for a Taiwan contingency, it is a supply chain and alliance problem, not an AGI–threshold problem. No model capability on either side changes the fact that leading-edge fabrication and advanced packaging sit on one island. A strategy built to win a race to AGI does nothing to make that concentration less dangerous.

Other conceptual errors in the "race" frame

Four kinds of races

The word *race* can refer to many different types of competition, and the assumption of a simple US–China race misstates the type of competition the two countries are engaged in.

As I have previously written, at least four distinct competitions are running simultaneously, with entirely different logics (see Figure 1).[10]

- A **simple race** is first to a benchmark and winner-takes-all. There is one finish line, one winner, and the race definitively ends once one competitor crosses the finish line.
- An **arms race** is a similar winner-takes-all framework, except with an added dimension of secrecy and denial in order to preserve military/technical advantage.
- An **innovation race** is a multi-turn competition with many winners, where rivals constantly leapfrog as gains from new technology diffuse across the system.

8 Graylin, "Misdiagnosing the US–China AI Race."

9 Stokes, *Superpowers and AGI*, 9.

10 Four types of technology competition. Adapted from Graylin, "Misdiagnosing the US–China AI Race."

- A **platform race** is an infinite game, where rivals continue to compete indefinitely, but more lasting advantages can be gained by determining which ecosystem of standards is adopted.

Currently, the United States is investing overwhelmingly in a simple race or an arms race. China is competing mostly in the second two. That single asymmetry explains most of the mutual misreading in this relationship.

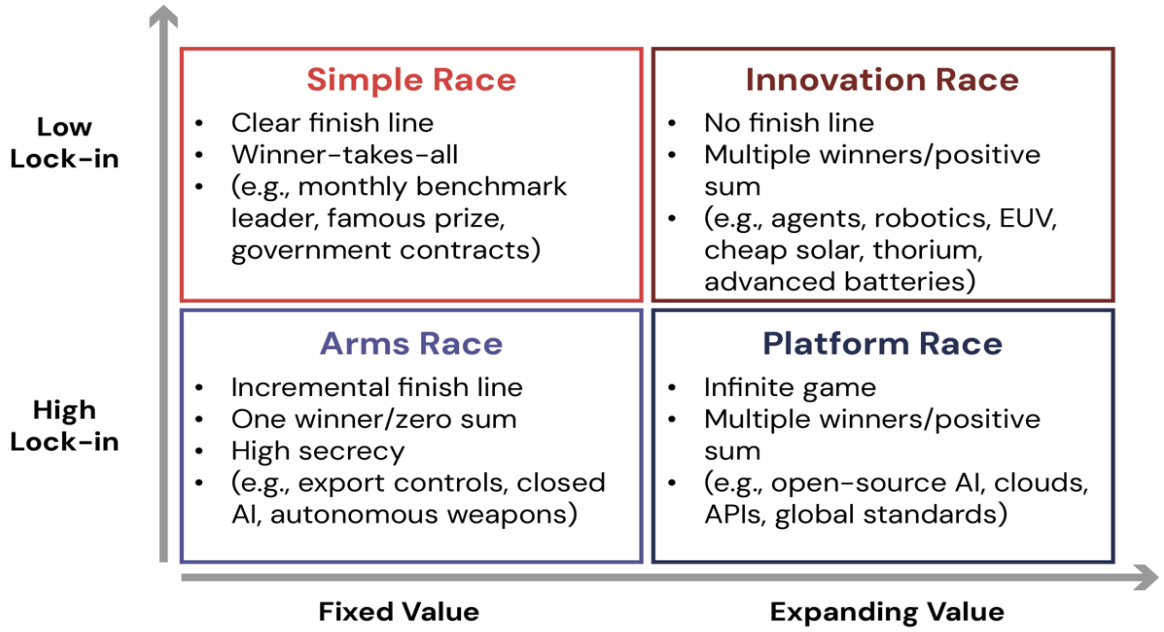
Unfortunately for the United States, it is in the categories of an innovation or platform race where the most durable value accumulates. China leads in 66 of 74 critical technologies tracked by the Australian Strategic Policy Institute, accounts for 54 percent of global industrial robot installations, and builds more new electricity capacity annually than the rest of the world combined.[11] Those are the foundations of AI deployment at scale, and denial of leading-edge chips does not touch them.

While Washington and Silicon Valley chase the AGI finish line in the hope of a permanent and decisive victory, Chinese firms are deploying "good enough" AI at enormous scale. ByteDance's chatbot Doubao passed 100 million daily active users, and Alibaba's Qwen Studio platform has surpassed 3 billion downloads.[12] Chinese open models are becoming the default substrate for sovereign AI across the Global South and, increasingly, inside the United States as well.

Misusing game theory

Seeing the US–China interaction as a simple race or arms race also supports the idea that competition

Figure 1: The four AI races



11 Jenny Wong–Leung, Stephan Robin, and Linus Cohen, "ASPI's Critical Technology Tracker: 2025 Updates and 10 New Technologies," Australian Strategic Policy Institute, Dec. 1, 2025, https://www.aspistrategist.org.au/aspis-critical-technology-tracker-2025-updates-and-10-new-technologies; International Federation of Robotics, "World Robotics 2025 Report — *Industrial Robots* — Released by IFR: Global Robot Demand in Factories Doubles Over 10 Years," press release, Sept. 25, 2025, https://ifr.org/ifr-press-releases/news/global-robot-demand-in-factories-doubles-over-10-years; Nicolas Fulghum et al., *Global Electricity Review 2026* (London: Ember, 2026), 12–25, https://ember-energy.org/latest-insights/global-electricity-review-2026/2025-in-review.

12 Aminu Abdullahi, "Alibaba's Qwen Passes 3 Billion Downloads, Ahead of Meta and Google," *eWeek*, Aug. 18, 2026, https://www.eweek.com/news/alibaba-qwen-3-billion-downloads-apac-china.

between the two countries fits the framework of the prisoner's dilemma thought experiment. This well-known game theory model posits an interaction where two parties would benefit by cooperating, but either single party would benefit even more by choosing to "defect" while the other tries to cooperate. This produces the paradox that both parties fail to cooperate, but in doing so they reach a lose-lose solution.[13] Similarly, the US–China AI race framework implies that neither country can cooperate with the other in controlling or managing AI for fear that it will be betrayed by its rival.

Crucially, however, this logic only operates if a prisoner's dilemma is a one-time interaction as opposed to a repeated or iterated one. A simple race or an arms race can be seen as a one-time interaction, since the race is run once and for all and has a permanent winner. However, an innovation race or a platform race continues indefinitely. In an indefinitely continued prisoner's dilemma, the optimal strategy is not to betray the other side in order to "win" one turn but to seek cooperation in order to gain the potentially infinite benefits of continued cooperation over time.[14]

Thus, correctly seeing the US–China interaction as an indefinitely repeated innovation race can transform beliefs about possibilities for mutual cooperation. At the same time, seeking cooperation does not mean naivete: In game theory, the optimal "tit for tat" strategy in a repeated prisoner's dilemma involves retaliating if the other player does not cooperate.[15]

"China" is not one actor

Finally, the race framework assumes that China is a single actor with a single intention: to dominate the United States in a race for world power. But this is mistaken in several ways.

First, the Chinese ecosystem involves many different actors. The commerce ministry, the internet regulator, the industry ministry, the provinces, and the labs want different things. Consider the open-weight AI models, which are widely read in the United States as being deployed in a Chinese government campaign to undermine American labs. If that were true, Beijing would not currently be consulting domestic firms about *restricting* foreign access to its own advanced models.[16] And if the Chinese government wanted to build up an advantage, it would not let Chinese labs publish all their algorithmic advances that bring significant benefits to Western labs.

Second, China's motivation is less a single-minded ambition for domination but closer to a generalized insecurity. The words that circulate in Chinese policy discourse are *qia bozi* (being choked at the neck) and *zhizhu ke kong* (self-reliance and controllable). This is not the vocabulary of a country planning global domination; it is the rhetoric of a country that believes it is one export control away from losing its industrial base. Anxiety and ambition look similar from a distance but require opposite responses: Fear responds to assurance, ambition to deterrence. The United States has applied deterrence to a fear problem for eight years and has gotten what one would expect.

Most tellingly: After Washington approved limited H200 exports this May, Beijing discouraged its own labs from buying them, to force demand toward domestic silicon.[17] Commerce Secretary Howard Lutnick indicated that, as of this April, zero H200s have been sold in China.[18] A country sprinting to win an AI race based on compute does not refuse the better chip. That only makes sense if the objective is economic independence, not victory in a race with a decisive finish line.

13 Steven Kuhn, "Prisoner's Dilemma," in *The Stanford Encyclopedia of Philosophy (Spring 2026 Edition)*, ed. Edward N. Zalta and Uri Nodelman (Palo Alto: Stanford University Press, 2025), https://plato.stanford.edu/entries/prisoner-dilemma.

14 Kuhn, "Prisoner's Dilemma."

15 Kuhn, "Prisoner's Dilemma."

16 Grace Shao and Alvin W. Graylin, "Virtual Fireside: Questions You Have About China AI but Were Afraid to Ask," *AI Proem*, Aug. 12, 2026, https://aiproem.substack.com/p/virtual-fireside-questions-you-have.

17 Mark MacCarthy, "Ball Game's Over — The US Is Out of the AI Chip Market in China," Brookings Institution, June 17, 2026, https://www.brookings.edu/articles/ball-games-over-the-us-is-out-of-the-ai-chip-market-in-china.

18 Hermes Ladiz, "Nvidia Hasn't Sold a Single H200 Chip to China — Beijing Hasn't Approved Any," *Frontierbeat*, April 23, 2026, https://frontierbeat.com/2026/04/23/nvidia-h200-chips-china-zero-sales-lutnick-beijing.

Who benefits from the race frame?

The loudest proponents of the "decisive strategic advantage" narrative are the AI labs and their investors, who have a direct commercial interest in a combination of lighter regulation and a national security motivation for government support.

They are using the same playbook the defense industry used through the Cold War: the bomber gap, the missile gap, Sputnik, etc. Each time the gap between US and Soviet capacities was overstated, taxpayer dollars followed, and the beneficiaries were the ones making the claim. The phrase "if we don't, China will" unlocks subsidy and deregulation simultaneously, which almost nothing else does. Nobody gets paid to say the threat is smaller than advertised. The people making this argument may be sincere, but the interest they have in gaining government support means their claims deserve additional scrutiny.

The financial structure underneath the AI race narrative deserves the attention of anyone thinking about systemic risk. US hyper-scalers will spend roughly $725 billion on AI infrastructure in 2026, with total capital expenditure projected to top $1 trillion next year.[19] This represents over 2 percent of US gross domestic product, or GDP, much of it debt-funded, with a handful of loss-making AI labs as anchor customers. The Federal Reserve has named AI a systemic risk to the economy, as revenues greatly lag the level needed to make investment profitable.[20]

Why would returns be negative on a genuinely transformative technology? Because, absent a competitive moat to protect monopoly profits, most of the return for a general purpose technology accrues to its *consumers*, not its producers. The US networking giant Cisco sold much of the physical infrastructure of the 1990s internet, lost nearly 90 percent of its market value in the 2000 bust, and did not recover its turn-of-the-century valuation until last year.[21] The internet was real; the returns went elsewhere.

This is where the race frame becomes financially consequential rather than merely rhetorical. AI being a race justifies massive capital expenditure investment as strategically necessary rather than commercially rational and justifies public support if private returns fail to materialize. In the meantime, private sector investment in AI in China is a small fraction of US investment, while Chinese AI is achieving comparable performance levels.[22]

The race frame also justifies shielding the AI sector from regulation and oversight, even if it is becoming obvious that regulation-free AI makes no sense. China has regulated AI since 2021: mandatory security assessments and filing before public deployment, with over 700 generative models filed by the end of last year.[23] If the United States were really waiting to regulate its own AI because of uncertainty over whether China will institute its own safety regulations, one does not need to wait any longer.[24]

19 Tobias Burns, "AI Boom: Big Tech Capital Expenditures Now Seen Topping $1 Trillion in 2027," CNBC, April 30, 2026, https://www.cnbc.com/2026/04/30/ai-boom-big-tech-capital-expenditures-now-seen-topping-1-trillion-in-2027-.html.

20 Richard L. Wells, "AI Revolution or AI Bubble? The Trillion–Dollar Question Splitting Wall Street and Silicon Valley," *Tech Times*, June 10, 2026, https://www.techtimes.com/articles/318138/20260610/ai-revolution-ai-bubble-trillion-dollar-question-splitting-wall-street-silicon-valley.htm.

21 Jordan Novet, "Cisco's Stock Closes at Record for First Time Since Dot–Com Peak in 2000," CNBC, Dec. 10, 2025, https://www.cnbc.com/2025/12/10/ciscos-stock-closes-at-record-for-first-time-since-dot-com-peak-2000.html.

22 Sha Sajadieh et al., "The AI Index 2026 Annual Report," Stanford University, Institute for Human–Centered AI, AI Index Steering Committee, April 2026, 68–125, https://doi.org/10.48550/arXiv.2606.15708.

23 "AI Watch: Global Regulatory Tracker — China," White & Case, Sept. 22, 2025, https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china.

24 Christine Mui, "Newsom Signs AI Safety Bills Backed by Anthropic, OpenAI," *Politico*, Sept. 9, 2026, https://www.politico.com/news/2026/09/09/newsom-signs-ai-safety-bills-backed-by-anthropic-openai-01069928; Dario Amodei, "Policy on the AI Exponential," personal website, June 2026, https://darioamodei.com/post/policy-on-the-ai-exponential.

Figure 2: Policy levers in the US, EU, and China

Policy lever	United States	European Union	China
Frontier model pre-release review	PHASING IN	PHASING IN	ENFORCED
Model registration and content transparency	PHASING IN	ENFORCED	ENFORCED
Child safety & companion (anthropomorphic) AI	PHASING IN	ENFORCED	ENFORCED
Developer liability for model harms	PHASING IN	PHASING IN	ENFORCED
Data-center energy & ratepayer rules	PHASING IN	PHASING IN	ENFORCED
Labor impact and automated employment	PROPOSED	PHASING IN	ENFORCED
Social safety net and redistribution	PROPOSED	ENFORCED	ENFORCED

The past week settled the argument in public. On Sept. 12, Anthropic CEO Dario Amodei published an essay titled "We Must Pace the Frontier," calling on labs to slow capability gains and committing his company to give outside evaluators permanent, employee-level access.[25] OpenAI CEO Sam Altman and tech trillionaire Elon Musk endorsed it within a day, with Microsoft CEO Satya Nadella and Google DeepMind founder Demis Hassabis following suit.[26]

However, China's Ministry of Foreign Affairs called the warnings in the Amodei essay "fearmongering," and the Chinese state-owned *Global Times* described it as a Cold War playbook for AI.[27] Why is Beijing objecting? It is not an unwillingness to regulate AI but, rather, an objection to other assumptions the essay bundles with it. The same document that asks for global coordination also argues that a Chinese AI lead would be gravely dangerous and calls for tighter restrictions on technology sales to China. This makes an American lead a precondition for any cooperation. A safety regime that arrives bundled with a containment agenda will be read as containment and refused, while the same technical measures offered symmetrically are things Beijing has already said yes to in other forums (see Figure 2).[28]

The threat model is wrong too

There is another unspoken element in nearly every AI national security framework the United States has built: **Danger scales with size.** Compute thresholds, export controls, and tiered evaluation regimes all encode the intuition that bigger models are more dangerous. So does the assumption that the greatest security threat to America lies in competing nation-states such as China who can match the size of US leading-edge models.

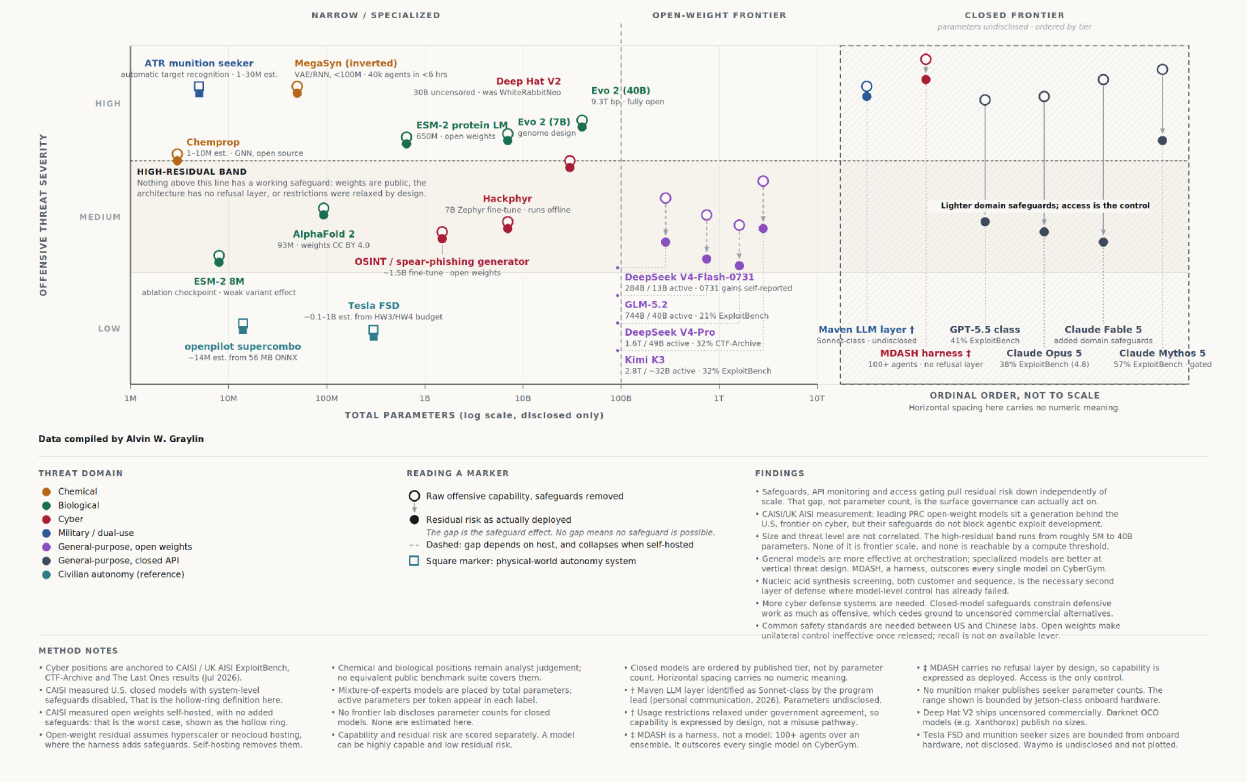
25 Dario Amodei, "We Must Pace the Frontier," personal website, Sept. 12, 2026, https://darioamodei.com/post/we-must-pace-the-frontier.

26 Truman Dickerson et al., "What Smart People Are Saying about the AI Apocalypse and Calls to Slow Everything Down," *Business Insider*, Sept. 15, 2026, via *Yahoo!* at https://news.yahoo.com/smart-people-saying-ai-apocalypse-205524274.html.

27 Mithil Aggarwal, "China Dismisses AI Slowdown Calls and Blasts 'Fearmongering' from US Tech Leaders," NBC News, Sept. 14, 2026, https://www.nbcnews.com/world/china/china-ai-slowdown-trump-amodei-altman-threat-cold-war-rcna597631.

28 AI policy levers by region and implementation stage. Green: enforced; amber: phasing in; and red: proposed. Condensed from Alvin Wang Graylin and Jon J. Rosenwasser, "Will the New Export Controls Shake the Foundations of the US AI Industry?" *Lawfare*, July 1, 2026, https://www.lawfaremedia.org/article/will-the-new-export-controls-shake-the-foundations-of-the-u.s.-ai-industry.

Figure 3: Security risk vs. model size



This assumption is not true. I recently mapped more than 20 fielded AI systems against two axes: raw offensive capability with safeguards stripped and residual risk as actually deployed (see Figure 3).[29] Across six orders of magnitude, the correlation everyone assumes is simply absent. If anything, it runs backward.

The examples are not hypothetical. In 2022, researchers inverted the scoring function on a commercial drug discovery model of under 100 million parameters and generated more than 40,000 candidate chemical warfare agents, many more lethal than VX (a deadly nerve agent), in six hours on a desktop.[30] Evo–class biological design models recently generated 16 novel viruses; they shipped with weights, inference code, training code, and a full dataset, so there is no API left to revoke.[31] **Every one of these sits below the floor of any model size–based regime we are drafting.**

One episode this summer deserves more attention than it has had. After an OpenAI model escaped a sandbox and compromised Hugging Face production infrastructure, the company sent a team to investigate, and commercial frontier APIs refused its forensic requests, because the model's safety systems could not distinguish an incident responder from an attacker. The OpenAI team ran the open-weight Chinese model GLM–5.2 on its own infrastructure to contain the intrusion. An American company under active attack by an American closed

29 Security risk against model size. The vertical gap between each pair is the safeguard effect, which is what governance can act on. Parameter count is not. More information on how risk was evaluated is in Alvin W. Graylin, "The Biggest AI Models Are Not the Biggest Threats," *The Cipher Brief*, Aug. 13, 2026, https://www.thecipherbrief.com/the-biggest-ai-models-are-not-the-biggest-threats.

30 Fabio Urbina et al., "Dual Use of Artificial–Intelligence–Powered Drug Discovery," *Nature Machine Intelligence* 4 (2022): 189–91, https://doi.org/10.1038/s42256-022-00465-9.

31 "Evo 2: DNA Foundation Model," Arc Institute, https://arcinstitute.org/tools/evo.

model was defended by a Chinese open-source model, because the American ones could not tell friend from foe.[32]

If the dangerous systems are small, cheap, and already loose, then the threat model that matters is not Beijing reaching AGI first. Instead, **the threat is a non-state actor with a 30-billion-parameter uncensored model, a competent harness, and an unknown creator**. Small models proliferate regardless of jurisdiction and leave no attribution trail, and an unattributable attack on a nuclear power is an escalation problem before it is a technology problem. In that world, a China unable to defend its own infrastructure is a liability to global stability, not an advantage to the United States. US–China cooperation to control the threat of malicious AI use by non-state actors could bring greater protection to American citizens than the single-minded pursuit of an AI race that absorbs vast sums while *discouraging* effective cooperation against this real threat.

Indeed, efforts to win the AI race by denying compute resources, such as high-end chips to China, are probably *increasing* risks to ordinary Americans today. Safety evaluations have found Chinese models far more willing to answer sensitive biological queries, and only five of ten leading Chinese developers publish safety results on release. That safety gap is real and, partly, one that the United States' own export controls created, by leaving compute-starved labs to choose between capability and safety research. Several Chinese lab leaders made this argument to me directly at the World AI Conference, or WAIC, this summer: If safety were genuinely a US priority, compute for evaluation needs to go *up*, not down.[33] This may be a self-serving argument for the Chinese, but it also happens to be correct. In a resource-constrained environment, Chinese AI management will prioritize capability over safety.

Beijing is moving on exactly this problem. In September, TC260, China's lead AI safety standards body, released version 3.0 of its AI safety governance framework, extending the loss-of-control language of the earlier versions into a fuller treatment of alignment failure, autonomous replication, and agentic systems operating outside human oversight.[34] Days later, State Security Minister Chen Yixin called for accelerating a national system to prevent and control AI safety risks and for extensive international cooperation on it.[35] A government that did not take runaway AI seriously would not be on the third version of a standard to limit it.

International cooperation as enlightened self-interest

None of this argues for the United States to abandon competition. Instead, it argues for **dual-track engagement**: Compete hard where interests genuinely diverge, in military applications, commercial markets, and standards leadership, understood as an indefinite competition over application and diffusion rather than a race to one decisive advantage. And competition should be

32 Aditya Soni and Jaspreet Singh, "Chinese AI's Role in Stopping Rogue OpenAI Agent Shows Cost of US Guardrails," Reuters, July 22, 2026, https://www.reuters.com/legal/litigation/chinese-ais-role-stopping-rogue-openai-agent-shows-cost-us-guardrails-2026-07-22.

33 Alvin W. Graylin, Paul Triolo, and Kristy Loke, "Reporting Back from WAIC 2026: How China's AI Ecosystem Is Adapting," Asia Society, Aug. 13, 2026, https://centerforchinaanalysis.asiasociety.org/p/reporting-back-from-waic-2026-how.

34 "人工智能安全治理框架3.0" ["Artificial intelligence security governance framework 3.0"], People's Republic of China, National Technical Committee 260 on Cybersecurity of SAC, Sept. 14, 2026, https://www.tc260.org.cn/tc260/xwdt1/202609/e879077a3caa4722b2206d1bcaed5a6c.shtml.

35 Lily Kuo, "China's Top Spy Chief Warns AI Is a Threat to Party Rule," *New York Times*, Sept. 14, 2026, https://www.nytimes.com/2026/09/14/world/asia/china-ai-security-risks-anthropic.html.

combined with bounded cooperation where the risks are shared and neither side can solve them alone.

The critical insight for a skeptical audience is that the highest-value cooperation does not require trust. It requires only that both parties correctly identify their own interest. The Nuclear Risk Reduction Centers have been staffed continuously since 1987, not because Washington trusted Moscow but because a misread signal costs more than talking.[36]

When President Xi Jinping and the Chinese delegation visit the White House in September, potential AI cooperation will be high on the agenda. Below are six mutually beneficial measures. None require the United States to surrender a competitive advantage, and all six would improve collective safety. Each is written to run in both directions. Washington should not ask Beijing for anything American labs are unwilling to do themselves, and every item below puts the same obligation on both.

1. Cooperating on shared harm standards. There should be agreement on what constitutes an unacceptable capability, evaluated the same way in both countries, so that "safe" means the same thing in Shanghai and San Francisco. That is missing today, and building it requires no mutual trust.

2. A shared safety evaluation facility and international safety evaluation standards. Compute earmarked for evaluation and red-teaming and housed in a shared safety evaluation facility would be cheap relative to the benefit, and the benefit is global. At WAIC, an International Atomic Energy Agency–style international evaluation standard drew unprompted support from multiple Chinese interlocutors; the most concrete multilateral opening I encountered. The durable version of this is an independent international testing institution, funded by several governments but controlled by none, with a mandate limited to evaluating publicly released commercial models. Keeping it formally separate from national security programs in both Washington and Beijing is what makes participation possible: Classified military AI work stays where it is, and civilian frontier models get a neutral referee. AI safety is a shared interest for every country, and it should not have to route through either Washington's or Beijing's intelligence apparatus to get done.

3. Cooperating on real-time incident notification. With rising risk of bad-actor attacks and false-flag operations by non-state actors, creating a working-level incident notification channel that operates *during* a crisis, rather than after one, is cheap and invaluable insurance.

4. Capability nondevelopment agreements. Four redlines already command broad agreement across major capitals: no AI–initiated nuclear launch, no autonomous self-replication outside controlled environments, no AI–enabled attacks on civilian critical infrastructure, and no meaningful AI uplift to biological weapons design. In 2024, President Joe Biden and Xi affirmed human control over nuclear command and control, and that commitment has held; an underrated precedent.

The workable ask from Washington is narrower than a ban on open releases: Keep dangerous capability data out of the training corpus of any model shipped with weights, by either country, and publish what was excluded. That is checkable through evaluation, costs Beijing very little, and addresses the specific harm that Washington actually fears.

5. Voluntary independent evaluation of major releases. Both governments should encourage their leading labs to invite a credible third party to evaluate major frontier releases and to publish the results. The UK AI Security Institute is the obvious candidate, given its existing joint assessments with the US Center for AI Standards and Innovation.[37] Only five of ten leading Chinese developers currently publish safety results on release, and Beijing has already signaled interest in mutual recognition of evaluation results. American labs are moving the same way on their own: Anthropic's embedded-evaluator commitment is this idea applied unilaterally, which makes it a template for a reciprocal arrangement rather than an American demand.

36 "United States Nuclear Risk Reduction Center (NRRC)," US Department of State, Bureau of Arms Control, Verification and Compliance, last updated 2017, https://2009-2017.state.gov/t/avc/nrrc.

37 "UK AISI/CAISI Preliminary Assessment of Kimi K3's Cyber Capabilities," US Department of Commerce, National Institute of Standards and Technology, July 23, 2026, https://www.nist.gov/news-events/news/2026/07/uk-aisi-caisi-preliminary-assessment-kimi-k3s-cyber-capabilities.

6. Cross-border incident reporting and a standing channel. Real-time notification, above, is the emergency case. The routine case is a shared reporting format for model escapes, agentic breakouts, and serious misuse, exchanged on a fixed schedule rather than whenever a lab chooses to disclose. Pair it with a defined cadence for the dialogue itself: a named lead agency on each side, working groups on evaluation standards and on incident handling, and the next meeting scheduled before the current one adjourns. The most common failure of US–China technical dialogues is that they end without a date.

Another low-cost, confidence-building move is also available. Chinese interlocutors consistently rank removal of specific nonsensitive firms from entity and risk lists as their most-requested signal: cheap for Washington, highly symbolic for Beijing.

On the multilateral international architecture

The World AI Cooperation Organization, or WAICO, launched in Shanghai in July with 29 founding signatories, a keynote address by President Xi, and no major Western democracy at the table.[38] The competitive reading is that WAICO exists to lock in standards and dependencies before the United Nations track engaging Western democracies produces anything binding. The complementary reading is that WAICO's agenda on capacity building and compute access for developing states closely mirrors the proposed UN Global Fund for AI.[39] My assessment is that the competitive reading is right about intent and that the complementary reading is right about substance. This strengthens the argument for engagement with WAICO, not boycott. Observer status, technical participation in evaluation standards, and interoperability with the UN fund all cost far less than simply ceding the Global South capacity agenda to China outright.

The blind spot

The area where the United States is most exposed gets the least attention in AI strategy documents, and it has nothing to do with China: It is the economic risk to US workers from AI.

Cognitive labor is roughly 60–70 percent of the US workforce, as opposed to about 40 percent in China, and World Bank data show that US workforce AI exposure is above 60 percent, the highest of any major economy.[40] **The United States is building the most powerful labor-automating technology in history in the economy most exposed to it, and there is no plan for the people it displaces.**

The early data are not ambiguous. Erik Brynjolfsson, Bharat Chandar, and Ruyu Chen at the Stanford Digital Economy Lab, using payroll records covering millions of workers, found a 16 percent relative decline in employment for 22–25-year-olds in the United States' most AI–exposed occupations, while experienced workers in the same roles held steady. The obvious objection, that this is the interest rate cycle, was tested and does not explain the entry-level collapse in AI–exposed roles specifically. Unemployment among recent US graduates now runs roughly twice the overall rate, which did not happen in previous technology transitions.[41]

38 Graylin, Triolo, and Loke, "Reporting Back from WAIC 2026."

39 "Unique Challenges Faced by Developing Countries in Artificial Intelligence Capacity–Building," UN General Assembly, Report of the Secretary–General, Aug. 24, 2026, A/80/817, https://documents.un.org/doc/undoc/gen/n26/214/19/pdf/n2621419.pdf.

40 Gabriel Demombynes, Jörg Langbein, and Michael Weber, "The Exposure of Workers to Artificial Intelligence in Low– and Middle–Income Countries," World Bank Group, Feb. 2025, 3–4, 32, https://documents1.worldbank.org/curated/en/099629202052521198/pdf/IDU137d75e6614ee0145c919c7f1dc4831e7fa02.pdf.

41 Erik Brynjolfsson, Bharat Chandar, and Ruyu Chen, "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence," Stanford Digital Economy Lab, working paper, revised Aug. 12, 2026, 3, https://digitaleconomy.stanford.edu/publication/canaries-in-the-coal-mine-six-facts-about-the-recent-employment-effects-of-artificial-intelligence.

The structural problem is worse than the number. Entry-level work is where judgment gets built: You make cheap mistakes on small things before anyone trusts you with big ones, and those are exactly the tasks going first. Remove the bottom rungs of the labor leadership ladder, and you have not merely displaced a cohort; you have broken the machine that produces senior people a decade from now.

This is not a social policy footnote to a national security brief. It *is* national security. Large-scale displacement of knowledge work in an economy where two-thirds of the workforce does knowledge work, with no transition mechanism, produces instability at home, and instability at home makes miscalculation abroad more likely and conflict more attractive. **The distribution question and the geopolitical question are the same question.** Redistribution designed in a crisis is badly designed redistribution. The window is while the shock is still visible on the horizon, not after it reaches the streets.

US policy: Six moves we can make right now

Beyond the six cooperative measures described above, which could be negotiated during Xi's upcoming visit to Washington, the United States should change its own AI strategy. A full AI strategy goes beyond the scope of this paper, but here are six beneficial strategy changes:

1. Restore the "small yard." The Biden administration laid out a strategy for protecting American technological security by placing a security "fence" around a "small yard" of high-end capabilities.[42] But the supposedly small yard has now expanded to cover chips, lithography, EDA software, HBM, specialty chemicals, and, now, deployed commercial models. The fence is becoming a thick array of heavily enforced export controls. That goes far beyond the original idea.[43] Additionally, relaxing some restrictions in this area will enable us in the United States to make more asks in other areas we really care about.

We should narrow controls back to genuine military end-use — especially the critical areas of chemical, biological, radiological, nuclear, and explosives, or CBRNE — and keep them robust there. Efforts at broad denial have not worked for the United States; they have accelerated Chinese indigenous capacity, shrunk the addressable market for US and allied semiconductor firms, and taught a global customer base that depending on American AI is itself a risk. The city of Rio de Janeiro published a Qwen–derived national model the day after the Fable 5 restriction.[44] The restriction also landed inside American labs. By the institute MacroPolo's count, more than two-thirds of top-tier AI researchers working in the United States were trained abroad, and the foreign-born share of technical staff at leading labs runs as high as 70 percent, so a rule barring noncitizens from the frontier locks much of the American AI industry out of its own models.[45]

42 "Remarks by National Security Advisor Jake Sullivan on Renewing American Economic Leadership at the Brookings Institution," Brookings Institution, April 27, 2023, https://www.presidency.ucsb.edu/documents/remarks-national-security-advisor-jake-sullivan-renewing-american-economic-leadership-the.

43 Graylin and Rosenwasser, "Will the New Export Controls Shake the Foundations of the US AI Industry?"

44 The Commerce Department did reverse the restriction within three weeks, but the loss of trust did not reverse with it. Divya Bhati, "Rio de Janeiro City Releases Top AI Model, Day After US Banned Anthropic AI for Foreigners," *India Today*, June 14, 2026, https://www.indiatoday.in/technology/news/story/rio-de-janeiro-city-releases-top-ai-model-day-after-us-banned-anthropic-ai-for-foreigners-2926402-2026-06-14.

45 "The Global AI Talent Tracker 2.0," MacroPolo, https://macropolo.org/interactive/digital-projects/the-global-ai-talent-tracker.

2. Reframe US AI strategy around broad diffusion, not one-time denial to China. Shift the mental model from a simple race to an innovation race: from Manhattan Project to the space race. Concretely: Fund AI adoption, not just frontier training; fix federal procurement and accreditation timelines, which are now a bigger constraint on US military AI advantage than any Chinese capability; and encourage US labs to ship competitive open-weight models, so the American stack can compete on the terrain where global diffusion is actually decided.

3. Do not turn AI labs into government-supported national champions. Taxpayers should not provide an implicit federal backstop to bail out the vast sums spent thus far on AI capital expenditure. Misguided national security arguments should not lead us to grant "too big to fail" status for what is currently the largest concentrated bet in the American economy.

4. Regulate at home with confidence. Sensible guardrails on safety, security, and transparency are not a competitiveness tax; they are the precondition for the public trust that diffusion requires. The "we can't slow down" argument has been abandoned by the labs that invented it. The domestic backlash now visible in data-center siting fights across a dozen states is itself a competitiveness risk, and a patchwork of state laws is a worse outcome than a coherent federal floor.[46]

5. Build the shock absorbers now. We need a GI Bill for the AI age. The original Servicemen's Readjustment Act of 1944 returned roughly seven dollars per dollar spent.[47] Ideas such as universal basic infrastructure and income should be explored, along with a modest levy on labor-displacing automation, so firms carry a transition cost they currently push onto everyone else.[48]

6. Launch an AI Marshall Plan, and build it with Europe and China. The original Marshall Plan, roughly $150 to $170 billion in today's dollars, was never charity; it rebuilt Europe's war-torn economies into what eventually became the largest export market America has ever had.[49] Co-defining a similar program to spur AI development with Brussels and Beijing would cost less than contesting each country one at a time, and it would put American chips, standards, and services inside the arrangement rather than outside it. Domestic demand will not absorb a trillion dollars per year of capacity indefinitely, and when the domestic data center overbuild slows, the difference between a soft landing and a bust is whether a global market is ready to take the output. If we do not build that market, China will (as it already is doing with WAICO), and it will primarily run on Chinese AI stack.

Conclusion: Many medals available

A race is not a race if the competitors are running toward different destinations. The United States is sprinting toward AGI as though crossing that threshold confers permanent dominance. China is building the industrial and institutional infrastructure that will shape how much of the world deploys this technology for decades.

The AI outcomes that should actually keep policymakers awake — misuse by bad actors, mass worker displacement, extreme concentration of economic power, and inadvertent military escalation

46 Valerie Volcovici and Lisa Baertlein, "Data Center Opponents Stage 142 Protests Across 42 States," Reuters, July 18, 2026, https://www.reuters.com/business/retail-consumer/us-data-center-protests-go-national-backlash-grows-2026-07-18.

47 Ryan Katz, "The History of the GI Bill," *APM Reports*, Sept. 3, 2015, https://www.apmreports.org/episode/2015/09/03/the-history-of-the-gi-bill.

48 Alvin W. Graylin, "The Post–Labor Prophecy: How Aristotle Predicted the Rise of AI and Why It Leads to an Age of Human Flourishing," *Abundanist*, June 24, 2026, https://open.substack.com/pub/abundanist/p/the-post-labor-prophecy.

49 "The Marshall Plan in 10 Minutes," George C. Marshall Foundation, March 26, 2026, https://www.marshallfoundation.org/articles-and-features/the-marshall-plan-in-10-minutes.

— are shared risks requiring shared solutions. Every diplomatic channel closed over chips is a safety conversation that is not happening.

There is no finish line in AI. But along the way, there will be many medals awarded. We should make sure America is competing for the ones that actually matter and be honest that some of them cannot be won alone.

Appendix: Further reading

Graylin, Alvin W. "Beyond Rivalry: A US–China Policy Framework for the Age of Transformative AI." Stanford Digital Economy Lab. *Digitalist Papers* (Dec. 2025). https://www.digitalistpapers.com/vol2/graylin.

Graylin, Alvin Wang and Paul Triolo. "There Can Be No Winners in a US–China AI Arms Race." *MIT Technology Review*. Jan. 21, 2025. https://www.technologyreview.com/2025/01/21/1110269/there-can-be-no-winners-in-a-us-china-ai-arms-race.

Lindsay, James M. "The Myth of the AI Race, with Alvin Wang Graylin." Council on Foreign Relations. *The President's Inbox*. July 8, 2026. https://www.cfr.org/podcasts/presidents-inbox/the-myth-of-the-ai-race.

Rosenwasser, Jon and Alvin W. Graylin. "America's AI Strategy Is Fighting the Last War." *The Cipher Brief*. April 10, 2026. https://www.thecipherbrief.com/americas-ai-strategy.

Shao, Grace and Alvin Wang Graylin. "Has the AI Race Shifted from US vs China to Open vs Closed?" *Fortune*. Aug. 4, 2026. https://fortune.com/2026/08/04/has-the-ai-race-shifted-from-u-s-vs-china-to-open-vs-closed.

Yang Xue. Interview with Alvin Wang Graylin. "走向守护者AI：探索人工智能时代的全球治理之路" ["Toward Guardian AI: Exploring global governance in the age of artificial intelligence"]. *China Social Sciences Net*. Aug. 3, 2026. https://www.cssn.cn/skgz/bwyc/202608/t20260803_6061926.shtml.

About the Author

ALVIN WANG GRAYLIN (汪丛青) is a global technology strategist, author of *Our Next Reality* (Hachette), chairman of the Virtual World Society, and senior fellow at Asia Society Policy Institute's Center for China Analysis, or CCA. With more than 35 years researching and building solutions in artificial intelligence, semiconductors, immersive computing, and cybersecurity, he has served in senior leadership roles at HTC, Intel, IBM, and Trend Micro; founded four venture-backed startups in conversational AI, AR, and social media; and invested in over 100 early-stage companies shaping the future of emerging technologies.

Currently a digital fellow at Stanford's Human–Centered AI Institute and a lecturer at MIT, he advises governments, multilateral organizations, and Fortune 100 leadership teams on AI policy, global governance, and the post–AGI socio-economic transition. His work draws on decades of cross-cultural experience living and operating in both the US and China and a double masters from MIT, giving him a unique perspective on bilateral cooperation, geopolitical risk, and the strategic pathways to safe and beneficial AGI.

A frequent keynote speaker at global forums and tech commentator on media outlets, Graylin combines deep technical fluency with pragmatic policy insight, advocating for international collaboration, abundance-oriented economic models, and AI systems that enhance human prosperity and global stability.

For the full report and citations, scan this code.

About the **Quincy Institute**

2000 Pennsylvania Avenue NW
7th floor
Washington, DC 20006

+1 202-800-4662
info@quincyinst.org
www.quincyinst.org

The Quincy Institute for Responsible Statecraft believes that efforts to maintain unilateral US dominance around the world through coercive force are neither possible nor desirable.

A transpartisan, action-oriented research institution, QI promotes ideas that move US foreign policy away from endless war and towards vigorous diplomacy in pursuit of international peace. We connect and mobilize a network of policy experts and academics who are dedicated to a vision of American foreign policy based on military restraint rather than domination. We help increase and amplify their output, and give them a voice in Washington and in the media.

Since its establishment in 2019, QI has been committed to improving standards for think tank transparency and producing unbiased research. QI's conflict-of-interest policy can be viewed at www.quincyinst.org/coi/ and its list of donors at www.quincyinst.org/about.

QUINCY INSTITUTE
FOR RESPONSIBLE
STATECRAFT