HOOVER INSTITUTION
HISTORY WORKING PAPERS

Artificial Intelligence and Global Security

Niall Ferguson
Milbank Family Senior Fellow
Chairman, Hoover Applied History Working Group
Hoover Institution, Stanford University

History Working Paper 202601

HOOVER INSTITUTION
434 GALVEZ MALL
STANFORD UNIVERSITY
STANFORD, CA 94305-6010

June 2, 2026

ABSTRACT: The rapid development of artificial intelligence has produced two barely controlled AI races, one between around a handful of companies, and the other between two superpowers. Neither race is in any meaningful sense regulated, so that the only constraints are financial and physical. The U.S. government has resisted state initiatives and punted responsibility to Congress. There is little sign of an AI arms control agreement between the United States and China. Growing public disquiet would appear to anticipate adverse consequences, such as unemployment, that have yet to materialize and may not. Yet there are good reasons to fear the unintended consequences of the two unfettered races. Drawing parallels with events in the late 1950s and early 1960s, when roughly comparable races occurred in the domains of prohibited narcotics and nuclear weapons, this paper argues for a rapid transition to a new détente based on AI arms control. A reduction of breakneck competition between the superpowers would reduce the need for mafia-like behavior by the leading U.S. companies.

I

In 1957, Henry Kissinger published *Nuclear Weapons and Foreign Policy*. Though flawed in the eyes of its many critics, the book nevertheless clearly identified the central problem that nuclear arms posed for the United States: that any strategy of brinkmanship based on the threat to use strategic nuclear weapons created such a high risk of Armageddon that it lacked credibility.

Five years later, the Cuban Missile Crisis brought the world to the brink of just such a catastrophic Third World War. Even before then, the United States had begun a shift to "flexible response," relying on an array of intermediate-range and tactical nuclear weapons to give itself options other than all-out nuclear war. Later, it sought to limit the growth of strategic arsenals through arms limitation agreements with the Soviet Union and treaties banning nuclear weapons tests and restricting nuclear proliferation. Although Kissinger's theory of limited nuclear war was never accepted by the academic theorists, in practice it became integral to NATO's plans for the defense of Europe against Soviet forces that were manifestly much larger in their conventional capabilities.

Six decades later, we desperately need someone of Kissinger's intellectual breadth and depth to write *Artificial Intelligence and Global Security*.

The unfolding history of artificial intelligence has now arrived at the worst-case scenario from the point of view of Demis Hassabis and other leading computer scientists. There are two barely controlled AI races, one between around five (perhaps just two—at most a dozen) companies, and the other between the two superpowers. The leadership of the competitors in this race is, to say the least, of mixed quality. The chief executives of the most important companies include at least one with a record of duplicity and at least two egomaniacs, one of whom has commissioned an AI version of himself from a team of engineers, the other of whom also builds rockets, satellites, electric cars, and neural implants, and in his limited spare time takes drugs.

The president of the United States is a former real estate developer and reality TV star, roughly half of whose public utterances are mere bluffs. The leader of the People's Republic of China is a Marxist-Leninist who aspires to eclipse Mao Zedong as a dictator. The most gifted of the AI pioneer, a polymath who might have been the J. Robert Oppenheimer of a modern Manhattan Project, sold out cheaply to Google and from that moment lost his autonomy.

II

Assuming that the AI version of Moore's Law* continues to apply to the advanced semiconductors used by AI models, the constraints on the AI race today are:

1. The loss of plasticity that may be inherent in large language models;
2. the amount of data available for training models, which ought to increase with the size of the models, something now only achievable with synthetic data;

---

* Gordon Moore's "law" was that the number of transistors in an integrated circuit doubled about every two years, with little increase in cost. The AI equivalent might be that computing hardware doubles in price-performance every 12 to 16 months. I am grateful to John-Clarl Levin for this formulation. I would also like to thank Sophie Coste, Eyck Freymann, Harry Halem, Nick Kumleben, Chris Miller and Manuel Rincon-Cruz for comments on earlier drafts. All remaining errors are my own.

3. the amount of computational power available for inference;
4. the number and capacity of frontier fabs;
5. the number of advanced lithography machines;
6. the supply of electricity (U.S.) and the externalities of generating so much of it by burning natural gas (U.S.) and coal (China);
7. the number of elite engineers and data scientists, who can be numbered in the hundreds, as well as the much larger but still limited number of electricians and other skilled laborers who can build data centers;
8. the capacity of capital markets to finance capital expenditure of up to $9 trillion in the next five years; and
9. the growing public backlash at the consequences of an unregulated AI race, including the proliferation of resource-guzzling datacenters.

In the United States today, regulation does not feature on that list.

In the past three years, the debate on AI regulation has been noteworthy for the extreme positions it has generated. At one end of the spectrum are those who argue that AI should not be developed any further because the advent of "superintelligence" would pose a threat of human extinction. At the other are those who oppose any kind of regulation whatsoever. In theory, AI could be regulated as other new technologies have been in the past. In the 1890s, Underwriters Laboratories turned private safety certification into a de facto market requirement by embedding its standards into building codes, insurance requirements, and retailer policies. The Federal Trade Commission, which dates back to 1914, oversees all kinds of self-regulatory bodies in different sectors of the economy, with the power to take enforcement actions against violators. In the 1930s, after the Wall Street Crash, the securities industry accepted a regulatory system, with the Financial Industry Regulatory Authority supervising broker-dealers under the oversight of the Securities Exchange Commission. Aviation safety is ensured by non-government agencies such as Aviation Safety Information Analysis and Sharing and the Mitre Corporation, with the Federal Aviation Administration providing the "teeth" of enforcement. The electricity power grid is regulated by the North American Electric Reliability Corporation under the auspices of the Department of Energy. After the Three Mile Island accident in 1979, every nuclear utility in the United States joined the Institute of Nuclear Power Operations. As Andy Hall has written, effective regulation in the American tradition has "four crucial ingredients: independent assessment by people with genuine expertise, incentives that make nonparticipation costly, broad enough participation to prevent free-riding, and an external backstop from regulators, insurers, or courts that gives the system's judgments real weight."

However, no comparable regulatory apparatus as yet exists for AI. The Frontier Model Forum was set up in 2023 with the backing of six of the key AI companies. But it is not a regulator of the industry. The agency formerly known as the AI Safety Institute has been rebranded as the Center for AI Standards and Innovation (CAISI), with a budget of just $10 million, something less than the compensation of a senior AI researcher at a top lab. There remains an AI Safety

Institute, but it was established in the United Kingdom. The European Union has AI regulation, of course, but it is not clear that it can gain traction when there is really only one European AI company, France's Mistral, whose most advanced model currently ranks 114th in the LLM leader board. Little heed has been paid to the argument, made by Mustafa Suleyman in 2023, that we urgently need an international "technoprudential" regime for AI analogous to those already in place for climate change and financial instability.

The key point is that the United States government has taken a conscious decision not to regulate AI. This is perfectly clear from the National AI Legislative Framework, published on March 20, 2026, which punts responsibility for AI regulation to Congress. The document states: "Congress should establish commercially reasonable, privacy protective, age-assurance requirements (such as parental attestation) for AI platforms and services likely to be accessed by minors." Congress should also "ensure that residential ratepayers do not experience increased electricity costs as a result of new AI data center construction and operation." However, Congress should also "streamline federal permitting for AI infrastructure construction and operation so AI developers can develop or procure on-site and behind-the-meter power generation to accelerate AI infrastructure buildout and enhance grid reliability." And Congress should prevent anyone—including the federal government itself—"from coercing technology providers, including AI providers, to ban, compel, or alter content based on partisan or ideological agendas."

Most explicitly, the White House enjoins Congress "not [to] create any new federal rulemaking body to regulate AI." Rather, it should "support development and deployment of sector-specific AI applications through existing regulatory bodies with subject matter expertise and through industry-led standards." And Congress should also "ensure that State laws do not … act contrary to the United States' national strategy to achieve global AI dominance. … States should not be permitted to regulate AI development, because it is an inherently interstate phenomenon with key foreign policy and national security implications." The White House also leaves it to the courts to resolve whether "training of AI models on copyrighted material" does or does not violate copyright laws, while making it clear that the government thinks it does not.

The extent to which the administration of President Donald Trump has embraced the principles of the "techno-optimists" is impressive. Opponents of AI regulation in the private sector are meanwhile making every effort to ensure that the next Congress follows the direction of the executive branch. Leading the Future—a political action committee supported by OpenAI co-founder Greg Brockman and the venture capitalist firm Andreessen Horowitz—argues that AI rules would "stifle innovation [and] enable China to gain global AI superiority." It has raised more than $125 million in the past year, according to the *Financial Times*. An executive order creating a voluntary oversight system, under which developers of advanced AI models could submit their products for review by federal agencies before releasing them, was supposed to be issued on May 21. It was postponed at the last moment, according to *Politico*, because of opposition from the White House's former AI "czar" David Sacks. "I didn't like certain aspects of it," President Trump told reporters. "I think it gets in the way of—we're leading China. We're

leading everybody, and I don't want to do anything that's going to get in the way of that." Last month, the Trump administration came close to publishing an executive order envisioning the lightest-touch regulation imaginable. It got pulled at the last minute. On June 2 the president signed a watered-down version. According to *Politico*, it politely "asks some AI companies to submit their powerful new models to a voluntary government review 30 days before releasing the products to the public."

Yet the political backlash seems unstoppable. A recent poll found that two-thirds of Americans are concerned about the rapid development of AI, while 76 percent believe it needs to be regulated. AI, like crypto before it, was poised to become a partisan issue. If Republicans favor unrestrained acceleration, then Democrats logically must support tight regulation. Yet a striking feature of anti-AI sentiment is that it has become bipartisan. 60 percent of Trump voters say they are worried about AI's rapid development and almost 80 percent think the technology needs more regulation. Older voters are more likely than younger voters to think AI is moving too fast, according to an *Economist* / YouGov poll, but even two-thirds of 18- to 29-year-olds would prefer a slower pace. A recent poll by the Institute for Family Studies found that nearly four-fifths of voters in Republican-leaning states want tech companies to be liable if AI harms children. Last year, local opposition blocked or delayed at least 48 projects valued at $156 billion, according to Data Center Watch. The number will be higher this year. In Kentucky, Georgia, and Texas dozens of projected datacenters are being blocked. Last December, Building America's Future ran an advertising campaign exhorting Congress to pass federal legislation banning state-level AI regulation. It was unsuccessful, not least because of opposition by Steve Bannon, formerly Trump's campaign manager and strategist. Around 370 AI-related measures have been introduced in state legislatures this year, according to the *Financial Times*, of which nearly a third came from Republican-led chambers.

This backlash seems likely to grow. Missouri Senator Josh Hawley has said that AI is "working against the working man, his liberty and his worth." Republican opponents of AI also fear a further concentration of power in the hands of big technology companies that previously showed themselves willing to engage in ideologically motivated censorship, sometimes in concert with the federal government. Roman Catholics must take heed of Pope Leo XIV's encyclical "Magnifica Humanitas," with its powerful warning against the construction of a new Tower of Babel. Yet opposition to AI comes more naturally to the "progressive" left, who have a history of being against any new technology that threatens established jobs. Anti-AI activists have thrown Molotov cocktails and fired bullets at the home of Sam Altman, the chief executive of OpenAI. The more AI is associated with inflation and / or unemployment, the more Democrats will want to accuse Republicans of having done the bidding of the Big Tech overlords, prioritizing the needs of data centers over those of working people. John Burn-Murdoch's evidence suggests that large language models are innately "converging"—that is, tend to narrow the range of popular opinion—in contrast to social media's tendency to polarize or fragment it. It will be ironic if public sentiment converges on hostility to AI itself.

III

The striking thing is that American society is reacting against AI before its effects have really been felt. This may be paranoia. It may be prescience.

As Matt Shumer has argued, "AI isn't replacing one specific skill. It's a general substitute for cognitive work … Whatever you retrain for, it's improving at that too." The most plausible losers from corporate AI adoption are entry-level graduate recruits. Research by Erik Brynjolfsson suggested that employment of 22- to 25-year-olds in the most AI-exposed occupations such as software development and customer-service agents fell 6 percent—16 percent when controlling for relevant variables—in the three years after the introduction of ChatGPT, while that of older workers and workers in unexposed occupations rose. But this might just be the beginning. An entertaining report by Citrini Research imagined that by June 2028, because of AI, the unemployment rate would have risen to 10.2 percent, and the S&P 500 index would have fallen 38 percent from its "October 2026 highs."

Writing ironically, in the style of a future commentator, Citrini went on:

> It should have been clear all along that a single GPU cluster in North Dakota generating the output previously attributed to 10,000 white-collar workers in Midtown Manhattan is more economic pandemic than economic panacea … The human-centric consumer economy … withered.
>
> AI capabilities improved, companies needed fewer workers, white collar layoffs increased, displaced workers spent less, margin pressure pushed firms to invest more in AI.
>
> Agentic AI created frictionless transactions that did not need human intermediation. For every new role AI created, though, it rendered dozens obsolete. The new roles paid a fraction of what the old ones did.

Unemployment led to recession, which led to financial crisis, which led to a housing market crisis. In effect, 2008-9 happened all over again, but in reverse—thanks to AI.

When posed by a British comedian dressed as a Waffen-SS officer, the question, "Are we the baddies?" is funny. A better question today might be: Are we the horses? This was the question posed by the historian Matthew Lowenstein in a 2023 paper. "In a world with Artificial Superintelligence," he wrote "humans will flourish only at the pleasure of more intelligent machines … the replacement of horses by industrial technology took centuries, whereas the replacement of humans after the advent of powerful AI is likely to take less than a decade."

Following the horse's domestication on the Eurasian steppe nearly 6,000 years ago, the world horse population grew from less than half a million to a peak of over 100 million. The UK horse population grew from roughly 1.3 million in 1811 to 3.3 million in 1911. But it then slumped to

1.9 million in 1924. By 1965, there were only 21,000 horses still being used in British agriculture, a decline from peak to trough of 98 percent. The U.S. farm horse population peaked between 1910 and 1920 at roughly 20 million, but then plummeted to 1.6 million in 1974, a decline of 92 percent. The steam engine and then the internal combustion engine had rendered the horse obsolete. This did not lead to unemployment. It led to population collapse. The fact that there remain roughly 7 million horses in the United States reflects our sentimental or atavistic fondness for riding as a leisure-time activity.

As Lowenstein argues, the rise and fall of the horse illustrates "how a more intelligent agent can bestow enormous benefits on a less intelligent agent that it finds useful; but once the dominant no longer depends on the less intelligent agent, ruin for the latter follows in short order."

> The horses that survive today do so thanks to human largesse, either cared for by humans or living on lands protected by our laws. An ASI [artificial super intelligence] that valued humanity deeply enough could be relied on to seek our flourishing in perpetuity; but the unfortunate reality is we do not know how to program an "aligned" or "friendly" ASI. … [Moreover], an ASI will be able to exploit resources and optimize its environment more thoroughly than humans. The opportunity cost that living humans impose on AI will therefore be higher than the costs that horses exact on human civilization.*

The standard economist's response is that all technological change is expected to destroy jobs but then turns out to create new ones to replace the obsolescent ones—exposing the "lump of labor" fallacy. In the words of Greg Ip of the *Wall Street Journal*:

> Technological advancements always cost some people their jobs—those whose skills can be easily substituted by tech. But their loss is more than offset through three other channels. The new technology enhances the skills of some survivors, who become more productive and better paid; it helps create new businesses and new jobs; and it makes some stuff cheaper, increasing consumers' incomes, adjusted for inflation, which can be spent on other stuff, generating yet more jobs. These offsets explain why, through the sweep of U.S. history, technological advance hasn't, by itself, raised unemployment for the country as a whole.

In other words, ATMs did not cause bank tellers to go extinct. (Unfortunately for this argument, the smartphone probably succeeded where ATMs failed.) If there is to be an "AI Shock," comparable with the China Shock after 2001, Steven Davis argues that it will be more geographically diffuse and therefore less socially painful. The net result should be job creation, not job destruction. Brynjolfsson's critics argue that he is conflating the impact of ChatGPT with other variables, such as interest rates. The less reassuring view is Pablo Duran Steinman that "cognitive displacement (multi-week autonomy, novel problem solving) only arrives ~2031." Seb Krier's argument is that, "as long as the combination of Human + AGI yields even a marginal gain over AGI alone, the human retains a comparative advantage." He does not specify

how long that will be. But he argues that humans have skills—especially when they are organized into larger groups—that give us better prospects than the horses a hundred years ago.

From my vantage point, large language models are rapidly getting better and better at precisely the things I do well. Gemini can write sophisticated and well-researched historical articles very fast indeed, with only occasional "hallucinations" (i.e., fabrications). ChatGPT and other LLMs already generate a very large proportion of student assignments and probably grade a rising share of them, too. Claude is clearly better at coding than almost everyone. It has also developed a rather endearing nutty-professor persona. I devote some of my time to trying to anticipate economic and political developments on the basis of historical data and human intelligence. But Mantic's prediction engine now seems to be as good as the best human superforecasters. Another contender is Cassi. The Metaculus median forecast suggests that LLMs will out-perform human forecasters in mid-June 2027. Its makers say that AIA Forecaster already achieves "performance equal to human superforecasters."

Given the available evidence, I find the idea that I am essentially a mid-twentieth-century horse quite plausible, though the employment prospects for, say, plumbers may be brighter so long as world models remain in their infancy and humanoid robots struggle to perform simple tasks such as running. This is one reason I find collapsing human fertility rates unsurprising. Some attribute this trend to distracting smartphones. Perhaps, more than the horses, we sense our obsolescence and adjust our reproduction accordingly. Meanwhile, the overwhelmingly negative effects of LLMs on educational standards and juvenile cognitive development—by largely freeing students from the need to read, think, or write—only bring nearer the era of human redundancy.

The less speculative adverse scenario is just that the AI bubble bursts because capex exceeds revenues by too vast a sum. Let us assume that Alphabet, Microsoft, Amazon and Oracle do indeed deploy between $4 and $9 trillion of capital expenditure over the next five years. Let us also assume that a rising share of this will be financed on the bond market rather than just from free cashflow. As has been noted, a $9 trillion investment would require a 10 percent return, or $900 billion of profit per year after the costs of energy and depreciation. That implies a need for $2.7 trillion of revenue—which seems, to say the least, ambitious. A society which such a large share of its collective net worth in technology-related equities will plainly suffer a significant economic shock, potentially a recession, in the event of a major stock market correction.

IV

The perils of a completely unconstrained race between companies and superpowers are therefore clear. Artificial intelligence has the potential to render the human race obsolete in many of its key economic functions, much as steam engines and internal combustion engines rendered horses obsolete. At the very least, there is a significant risk that much larger sums are being spent on datacenters than the likely revenues from AI can justify.

Yet unemployment and financial instability are not the biggest problems that AI may give rise to. AI also has the potential to empower rogue regimes, criminals, and psychopaths to manufacture new kinds of lethal weapons, such as the "respiratory AIDS" one leading technologist in the field recently described to me (imagining how a frontier model without guardrails could quite easily devise a lethal biological weapon), and new scales of cybercrime such as an all-out attack on the national payments system, which would be one of the quickest ways to plunge the United States into chaos. Even the techno-optimist Marc Andreessen admits that AI "will make it easier for bad people to do bad things." The U.S. Treasury Secretary takes this threat seriously. "If we don't win in AI," Scott Bessent recently told the journalist Mark Halperin, "then it's game over." The "ultimate threat," Bessent went on, is that "somebody can back into something that's 10 times worse than Covid, like [by] just using biological data."

Domestically, the situation is akin to that described in the novel *The Godfather*. There are five companies, like the five families, in a completely unregulated competition with one another for dominance in a new and potentially enormous market. A notional government regulator could no more restrain them than the FBI could restrain the Mafia in the 1950s and 1960s. (The analogy is not as far-fetched as it may first appear. The production and sale of drugs such as cocaine and heroin was one of most dynamic sectors of the global economy in the later 20th century. Whereas AI offers to replace the human mind, narcotics offered to stimulate and sedate it. Whereas the federal government has opted to unleash AI, it opted to prohibit narcotics.)

But the most serious danger of artificial intelligence is geopolitical. As we have seen, it is the stated "national strategy" of the Trump administration "to achieve global AI dominance." China lags behind the United States in advanced chips and frontier-model performance, though not in electricity generation, the speed at which it can build datacenters, and the near monopoly it has on key rare-earth elements such as gallium and germanium. Because the United States is likely to get to AGI and ASI before China, Beijing must feel considerable insecurity. Chen Yixin, the head of State Security, recent argued in a *Qiushi* article on "Enhancing National Security Capabilities" for a Chinese policy of "strengthening research and prevention of potential risks from frontier technologies [and] conducting special protection of national strategic assets." This is not an unwarranted fear. It is a rational one.

As with nuclear fission in the later 1940s, it is starkly obvious that AI has the potential to be a weapon of mass destruction. Headlines in recent months have been revealing. "How AI Is Turbocharging the War in Iran" was the *Wall Street Journal*'s headline early in the U.S.-Israeli war against Iran. "The AI-driven 'kill chain' transforming how the US wages war" was the *Financial Times* equivalent:

> AI is reshaping how the US military makes decisions in war—a shift clear in Iran, where the Pentagon says it struck more than 2,000 targets in just four days.

> The primary operating system for the Pentagon's data is Palantir's Maven Smart System, which alongside Anthropic's Claude model forms a real-time data analysis dashboard for operations in Iran.
>
> During a live military operation such as Operation Epic Fury in Iran, Palantir's Maven platform acts as the software "brain." It supports the entire so-called chain—finding and hitting a target during active conflict.
>
> As of May 2025, the Maven system was used by more than 20,000 users across 35 military entities in the field, according to public comments by Vice Admiral Frank Whitworth, director of the National Geospatial-Intelligence Agency. That number may be closer to 50,000 users in the US today.
>
> The bombing of a girls' primary school in Minab, in southern Iran … illustrates the lethal risks.

The significance of reports such as these is twofold. First, they almost certainly exaggerate the importance of AI in Operation Epic Fury. Secondly, they nevertheless almost certainly arouse fear in America's adversaries.

It never seemed plausible, as the techno-optimists claimed, that AI was "going to improve warfare, when it has to happen, by reducing wartime death rates dramatically [and giving] military commanders and political leaders … AI advisors that will help them make much better strategic and tactical decisions, minimizing risk, error, and unnecessary bloodshed." There is already compelling evidence from the war in Ukraine that AI makes weapons systems such as drones more lethal. The mortality rate on the front line in Ukraine is not low by historical standards; it is exceptionally high. Yet AI adoption has not reduced, at least on the Russian side, the proclivity for attacks on civilian populations. In the coming AI-enabled wars, mortality rates in armed forces will be elevated, not reduced, precisely because AI will make the missiles and other weapons so much more accurate. Any half-decent AI that has read Clausewitz will want to achieve the annihilation of the enemy as soon as possible. AI-enabled commanders may also be more willing to sacrifice their own men to secure victory, in the same way that AI chess programs sacrifice their own pieces more ruthlessly than human grandmasters.

There was a time when American spoke softly and carried a big stick. No longer. On January 12, at Starbase, Texas, Secretary of War Pete Hegseth announced an "AI acceleration strategy" to create an "AI-first warfighting force across all domains, from the back offices of the Pentagon to the tactical edge on the front lines." No one has been a more enthusiastic proponent of this strategy than Shyam Sankar of Palantir, whose new book, *Mobilize: How to Reboot the American Industrial Base and Stop World War III*, envisions an AI acceleration strategy. New models, he argues, should be deployed "within 30 days of public release, standardizing the acceptable use policy for AI models used by the military. Among the innovations Sankar envisions are:

> *Swarm Forge:* [Creating a] competitive mechanism to iteratively discover, test, and scale novel ways of fighting with and against AI-enabled capabilities;

> *Agent Network:* Unleashing Al agent development and experimentation for Al enabled battle management and decision support, from campaign planning to kill chain execution.
>
> *Ender's Foundry:* Accelerating AI-enabled simulation capabilities—and sim-dev and sim-ops feedback loops—to ensure we stay ahead of AI-enabled adversaries.
>
> *GenAI.mil*: Democratizing AI experimentation and transformation across the Department by putting America's world-leading AI models directly in the hands of our three million civilian and military personnel, at all classification levels … Giving them armies of robot agents.

This is an impressive agenda, no doubt. But it is worth pondering what impression Sankar's plan for AI mobilization will leave on its most important audience.

V

As Sam Altman said to Lex Friedman in March 2024, "The road to AGI should be a giant power struggle. … Well, not should. I expect that to be the case." He was talking about OpenAI's turbulent corporate governance. He might equally well have been talking about geopolitics.

China is still at a relatively early stage of a race to catch up with the United States in every military domain. As Lt. Gen. Stephen Sklenka, the U.S. Marine Corps deputy commandant for installations and logistics has argued, the People's Liberation Army may now be a "peer," not a "near peer," in terms of the number of surface ships and conventional forces such as ballistic missiles and drones. However, in numbers of nuclear missiles, and in qualitative terms across all capabilities, it clearly lags behind. Crucially, it also lags in AI, mainly because the U.S. invests much more in computation and has unlimited access to the most advanced semiconductors, whereas China does not have access to ASML's extreme ultraviolet (EUV) lithography machines, among other chipmaking tools, and thus cannot manufacture advanced chips domestically at scale and cost-efficiently.

It is a supposition that Xi Jinping might reasonably entertain that the United States is likely to act preemptively before China has completed its vast arms buildup. The news that Anthropic had developed Mythos, a model so powerful that it poses a threat to most cybersecurity systems, can hardly have reassured him. Mythos Preview found a 27-year-old vulnerability in OpenBSD—a long-established and respected operating system used to run firewalls and other critical infrastructure. The vulnerability enabled an attacker to crash any computer running OpenBSD. Mythos also found a 16-year-old vulnerability in a line of code in FFmpeg—which is used many software programs to encode and decode video. Mythos found and linked several vulnerabilities in the Linux kernel—the software that runs most of the world's servers—so that an attacker

could upgrade from ordinary user access to total control. Even the founder of Mozilla has acknowledged that Mythos successfully hacked his company's Firefox browser.

Most commentary on Mythos has focused on Project Glasswing,[*] whereby Anthropic gave access to Mythos—and $100 million in credits to use it—to more than 50 of the world's largest organizations, including Amazon, Apple, CrowdStrike, the Linux Foundation, Microsoft, Google and JPMorgan Chase. Might this be the beginning of a serious system of AI self-regulation? Or the creation of a cartel that stifles innovation? The fact that Anthropic's principal rival, OpenAI, is not part of Glasswing—and has immediately proposed an alternative approach—suggests to at least one critic that it is "an Anthropic-led club rather than an industry standard." An improved version of Glasswing, it has been argued, "would need to include the whole industry and not be led by one lab, credibly constrain the labs with legal force, and protect not just against novel cybersecurity threats but against a broader array of threats in ways that build political trust."

I would go further. Glasswing is disguised as an altruistic, precautionary venture. It is in reality a threat to all Anthropic's rivals, showing just enough of the major users of AI the immense power of its weapon. It is a little like the moment in *The Godfather* when it becomes clear that the Tattaglias— one of the five families—are about to take control of the new and rapidly growing narcotics business and will kill anyone, including Vito Corleone, who stands in their way.

But the significance of Mythos is much greater for the race between the superpowers than for the race between the U.S. companies. Is Mythos a "cybernuke"? Does it give offense a lethal edge over defense? Naveen Krishnan has described it as "a recipe for chaos and asymmetry in the wielding of cyber power," which, "like the atomic bomb … marks a step change that calls into question all prior cyber deterrence logic." As Krishnan argues, "an offensive model needs to find one exploitable vulnerability in a target system." But "a defensive one needs to find and patch every vulnerability across every system, continuously, before adversaries find any of them. Attackers operate at the speed of a prompt, while defenders operate at the speed of a patch cycle." Over time, the advantage may swing back to defense.[**] In the short run, however, advances at the frontier favor aggressors.

By Anthropic's own estimate, Mythos will be matched by open-source models, foreign programs, and uncontrolled actors within six to 18 months. OpenAI's GPT 5.5 may have matched it already—hence the company's decision to make it available first to "critical cyber defenders." Given that Chinese state-sponsored hackers were already caught using Claude for cyberattacks, there is every reason to expect open-source versions of Mythos to be used for the same purpose as soon as they are available—though without the scale of computational power

---

[*] According to Anthropic's press release, Glasswing is named for the *Greta oto* butterfly, whose "transparent wings let it hide in plain sight, much like the vulnerabilities [exposed by Mythos]; they also allow it to evade harm—like the transparency we're advocating for in our approach."

[**] See on this issue Ben Garfinkel and Allan Dafoe, "How does the offense-defense balance scale?" in *Emerging Technologies and International Stability* (Routledge, 2021), 245-274.

U.S. models can use for inference. Yet the crucial point is that for at least six months the United States has Mythos (and GPT 5.5) and China does not.

In its most recent public pronouncement, released on May 14, Anthropic leaves no doubt as to its intentions. "It's essential," the company states, "that the US and its allies stay ahead of authoritarian governments like the Chinese Communist Party." That means maintaining the "incredibly successful" export controls that have limited Chinese access to the most advanced semiconductors and therefore quantities of "compute" comparable with those available to U.S. AI labs. It also means clamping down on the method known as "distillation" whereby Chinese AI companies "illicitly extract the innovations of American companies." The goal for U.S. policymakers should be "to lock in a 12-24 month lead in frontier capabilities," as such a lead "by 2028 would be enormously advantageous." In sum: "If the US strengthens its restrictions on the CCP's ability to access US compute … America will have access to roughly 11 times more compute than China's AI sector—a potentially once-in-a-generation opportunity to secure our lead."

To extend the admittedly imperfect analogy with nuclear fission, we are entering a period like that between August 1945 and August 1949, when the United States enjoyed a monopoly over the atomic bomb. Unless President Xi has great faith in the self-restraint of the United States—which was undoubtedly striking a feature of the late 1940s—China may now be contemplating "getting its retaliation in first," to preempt a potentially disastrous cyber-attack on its inferior systems.

The most obvious preemptive move for China under these circumstances would be to establish maritime control of the Taiwan Strait and political control of TSMC. As Eyck Freymann has argued, this would not necessarily require the People's Liberation Army to invade or blockade Taiwan. It may be sufficient for China simply to assert its right to collect customs duties on Taiwanese imports, and to challenge the United States to resist such an assertion of sovereignty, pitting warfare against lawfare. Control over TSMC would secure for Beijing the one thing that it crucially lacks, namely, the most sophisticated semiconductors, nearly all of which are manufactured on the island, which it claims to own. Given the latest U.S. entanglement in the Persian Gulf, and ongoing western commitments to Ukraine, there is now a serious munitions crisis that is clearly weakening U.S. deterrence. AI enabled the U.S. to hit a great many targets in Iran very fast. It also enabled the U.S. to run down, equally rapidly, an alarmingly large proportion of its stocks of certain categories of weapon that will take many months to relace. The Chinese have also seen the limits of AI-enabled military power. It may be very good at identifying targets for precision missiles. It does not seem to have good answers to the Iranian closure of the Strait of Hormuz, which has been achieved mostly with threats and very limited numbers of drone strikes and mines.

Chinese control over the Taiwan Strait would be even more potent than Iranian control over the Strait of Hormuz. And the old U.S.-Taiwanese threats that TSMC's fabs would be destroyed in the event of a Chinese takeover, or that ASML would no longer service its equipment, surely ring

hollow if it is precisely those fabs that provide the United States with an eleven-fold advantage in computational power—and if a Dutch company is expected to enforce U.S. sanctions at a time when transatlantic relations are in disarray.

VI

In the late 1950s and early 1960s, two dangerous races reached their climaxes. In May 1957 Vito Genovese tried and failed to kill Frank Costello, part of an extended struggle for power between rival mafia families that inspired *The Godfather*. Five years later the United States and the Soviet Union came to the brink of nuclear war over Cuba, an island which both the mafia and the Soviets had aspired to control.

The analogy implies that the world in the 2020s is simultaneously on the brink of a mafia-like war for dominance between the leading American AI companies and of the Taiwan Semiconductor Crisis, potentially the most dangerous moment of the new Cold War between the United States and China. I believe we are much closer to this second crisis than most experts assume. Those who assume that Xi will wait until January 2028, the next Taiwanese election—in the expectation that he can achieve control over Taiwan by political methods—fail to understand, first, the paranoid psychology of any Marxist-Leninist leader and, second, his quite rational calculation that the longer China waits to act, the more risk it runs of an American cyberattack.

The mafia wars continued from the mid-1960s plots of Joe Bonnano until the mid-1980s trials that broke up the so-called mafia "Commission." The nuclear arms race stabilized in the arms control negotiations of the 1970s and 1980s. Thus, just as either Michael Corleone or "the Feds" need to bring order to the cutthroat competition between the hyperscalers and frontier-model-builders, so the United States and China must accelerate the history of Cold War II and move beyond brinkmanship (or blinkmanship) to détente.

The basis for peace between the companies is clear. There must be a binding agreement between all of them to stop releasing ever more advanced models to the public when they are not just likely but—as Secretary Bessent says—certain to be used by criminals, lunatics, and adversaries. We will look back with incredulity on this period when such a dangerous technology was made available at subsidized prices to an unsuspecting public. The models have their hugely powerful uses, but they are not a consumer product. They should never have been presented to the public as chatbots with subscriptions absurdly low in relation to the computational power that was being made available.

Between the United States and China, meanwhile, détente must be based on an arms control agreement that extends into the realm of artificial intelligence—just as Kissinger recommended in the last year (and last foreign trip) of his life. In the first Cold War, such agreements were not reached until the 1970s and even then were fragile. That was because the Soviets did not feel ready to discuss arms limitation until they had caught up with the United States in terms of strategic forces. But China needs to discuss arms limitation now precisely because it is behind

the United States and is therefore vulnerable to preemptive American action. U.S. leadership today shows far less sign of the self-restraint that characterized Harry Truman's administration during the period of the U.S. atomic monopoly. It seems doubtful that last month's summit in Beijing did much to reassure President Xi on this score. While President Trump may issue conciliatory statements, the posture of his administration remains antagonistic, especially on the critical issue of AI.

The United States must therefore restrain itself lest its penchant for brinksmanship tips the world into a conflagration. It currently acts as if it really were a hyperpower when it is, in fact, anything but. We have seen this clearly in the war in the Persian Gulf. The United States had the capacity to assassinate the supreme leader of Iran and to lay waste to the military capabilities of that country. But when the rump regime controlled by a faction of the IRGC succeeded in closing the Strait of Hormuz with mere threats and a few drone attacks, the United States seemed paralyzed. At the time of writing, the U.S. government tacitly admits that its armed forces cannot reopen the Strait by force at an acceptable cost. The U.S. Navy apparently lacks the capacity to create an effective minesweeping and convoy escort system. Despite being decapitated and all but disarmed, the Islamic Republic has shown the limits of American power simply by taking as hostages the Strait and the Gulf kingdoms around it.

The United States is also constrained by its fiscal position. Ferguson's Law states that a great power that spends more in interest payments than on defense will not be great for much longer. The United States has been in that position since 2024 and the excess of interest payments over defense spending is projected to rise until, by 2036, the former is double the latter. The defense budget is bound to be constrained by the nasty arithmetic of public finance and congressional opposition. It is not in the interests of the United States, therefore, to be in an uncontrolled arms race with anybody, and certainly not with China. It makes much more sense for the U.S. and China to use their duopoly on advanced AI to create a non-proliferation regime for Mythos-level models, similar to the nuclear non-proliferation regime that originated in the late 1960s, but which threatens in our time to unravel.

VII

Nuclear fission was discovered in Berlin by two German chemists, Otto Hahn and Fritz Strassmann, in 1938. It was explained theoretically (and named) by the Austrian-born physicists Lise Meitner and her nephew Otto Robert Frisch in 1939. The possibility of a nuclear chain reaction leading to "large-scale production of energy and radioactive elements, unfortunately also perhaps to atomic bombs" was the insight of the Hungarian physicist Leó Szilárd. The possibility that such a chain reaction might also be harnessed in a nuclear reactor to generate heat was also recognized at that time. Yet it took little more than five years to build the first atomic bomb, whereas it was not until 1951 that the first nuclear power station was opened.

Ask yourself: Which did human beings build more of in the past eighty years: nuclear warheads or nuclear power stations? Today there are approximately 12,300 nuclear warheads in the world, and the number is currently rising as China adds rapidly to its nuclear arsenal. By contrast, there are 436 nuclear reactors in operation. In absolute terms, nuclear electricity generation peaked in 2006, with the share of total world electricity production that is nuclear declining from 15.5 percent in 1996 to 8.6 percent in 2022, partly as a result of political overreactions to a small number of nuclear accidents whose impacts on human health and the environment were negligible compared to the effects of rising carbon dioxide emissions from the burning of fossil fuels. The share of global primary energy consumption that comes from nuclear sources was 3.92 percent in 2024—roughly the same share as in 1983.

It is not yet clear if Dario Amodei is Michael Corleone: the seemingly clean-cut young man who turns out to be ruthless enough to end the war between the five families by winning it. Perhaps Project Glasswing was a strategic error. If he were running Anthropic, Corleone would have used Mythos without warning to expose the inferiority of the other frontier-model companies. Nor is it clear if President Trump or his successor is a sufficiently skilled negotiator to achieve a meaningful détente with China. There is a world of difference between *The Art of the Deal* and the intricate game theory of superpower arms control.

But unless such steps are taken—unless the unconstrained races between the companies and the superpowers are slowed down—then humanity is in grave danger of destruction at the hands of an alien intelligence that did not have to invade Earth, because we created it. This was the great preoccupation of Henry Kissinger's final years, inspiring him to write two books and numerous essays. In his first essay on the subject, published in 2018, Kissinger postulated a reversal of Max Weber's "demystification of the world," prophesying that, in the face of AI, the achievements of the Enlightenment might "go the way of the Incas, faced with a Spanish culture incomprehensible and even awe-inspiring to them." In *The Age of AI* (2021), co-authored with Eric Schmidt and Daniel Huttenlocher, Kissinger asked: "When a human-designed software program learns and applies a model that no human recognizes or could understand, are we advancing towards knowledge? Or is knowledge receding from us?"

Nearly three years after his death, we badly miss Kissinger's distinctive kind of human intelligence, which could absorb the implications of a new and potentially very destructive technology and foresee its far-reaching implications for the traditional fields of politics and strategy.