

# The AI Arms Race and the Dilemma of Global Governance

Alexander Y. J. Sterling

Research Fellow, Lingnan Scientific and Industrial Press, Macao, alexander.yj.sterling@hotmail.com

## Abstract

Global Artificial Intelligence (AI) governance is facing a profound "structural failure." This failure manifests as two mutually reinforcing dilemmas: On the development front, AI's evolution is dominated by a "race to the bottom" driven by "money and power," with tech giants (like OpenAI) disregarding safety and hastily releasing products for market advantage. On the governance front, national governance systems, particularly the Western model led by the United States, exhibit severe "governance myopia." This paper analyzes that the "lawyer-financial" political elite structure in the U.S. makes it difficult for them to comprehend the exponential threat of AI, while the "engineer" political elite structure, despite potentially having a deeper understanding, is also passively drawn into the "AI arms race." By comparing the different failures of capitalist and state governance models on the AI issue, this paper points out that both systems have failed to effectively address the long-term risks posed by AI, ultimately causing global governance to fall into a "prisoner's dilemma."

## Keywords

AI Governance, Structural Failure, Race to the Bottom, Governance Myopia, AI Arms Race

## 1 INTRODUCTION: AI'S "OPPENHEIMER MOMENT"

### 1.1. Problem Formulation: Exponential Development and Linear Lag

Since 2023, breakthroughs in Large Language Models (LLMs) and Generative AI have marked a qualitative leap for artificial intelligence technology. Illustrated by the rapid iteration of models like GPT-5, Gemini 3, Sora, Kimi, DeepSeek, and Claude 3, AI has not only surpassed human benchmarks in specific tasks (like translation and coding) but has also demonstrated complex "emergent abilities," such as zero-shot learning, complex reasoning, and even the construction of rudimentary "world models" (Wei et al., 2022).

This exponential growth in capability is bringing human society to a critical crossroads. On one hand, AI is seen as the core driver of the Fourth Industrial Revolution, promising to solve a range of major challenges from disease discovery to climate change (Hassabis, 2024). On the other hand, the potential risks of AI, especially Artificial General Intelligence (AGI), have triggered deepening anxiety from academia to industry (Bengio et al., 2024; Russell, 2019). This risk includes not only current issues of algorithmic bias, privacy infringement, and disinformation (Noble, 2018) but also points to the fundamental "Alignment Problem" and "Control Problem"—that is, how to ensure that a system far surpassing human intelligence always aligns its goals with human values and intentions (Bostrom, 2014). However, in the face of AI's "exponential" development, the global governance system's response has been "linear" or even "stagnant." Discussions on AI safety and ethics lag far behind capability deployment and commercial application. We are witnessing a repeat of the "Oppenheimer Moment": a powerful technology has been released from the lab, yet we lack effective global mechanisms to control its potentially devastating power (Kissinger et al., 2021).

The core argument of this paper is that the failure of global AI governance is not an accidental technical or temporal lag, but a profound "structural failure." This failure stems from two mutually reinforcing dilemmas: On the development front, AI's evolution is dominated by an "irrational competition" driven by "money and power." For example, Elon Musk's recent lawsuit against OpenAI centrally alleges that OpenAI violated its original "non-profit mission," betrayed humanity, and turned itself into a "de facto for-profit subsidiary" of Microsoft (Kinnard, M et al., 2024); Meta's plan to replace human labor with AI automation to accelerate product launches weakens substantive review of privacy and social risks (Crain, 2025). On the governance front, national governance systems are trapped in "structural myopia," relaxing regulatory requirements (White House Office of Management and Budget, 2020) and failing to effectively respond to long-term exponential threats.

## **1.2. Innovation and Main Contributions**

Current discussions on AI governance, while abundant, often focus disjointedly on three different levels: 1) Micro-level ethical applications (e.g., bias, fairness); 2) Macro-level AGI safety (e.g., existential risk); and 3) Geopolitical tech competition.

The main innovations and contributions of this paper are:

This paper, for the first time, places the two core variables of "race to the bottom" (development side) and "governance myopia" (governance side) into a single analytical framework, demonstrating how they interlock to create the "structural failure" of AI governance, thereby constructing an integrated "Competition-Myopia" framework.

This paper transcends mere institutional analysis to deeply compare the "cognitive structures" of political elites in major powers—namely, the "lawyer-financial" complex in the U.S. and the "engineer" governance tradition in other major powers—and analyzes how this cognitive divergence leads to different understandings of the AI threat and different governance focuses, dissecting the "cognitive gap" in governance.

By analyzing the specific performance of the capitalist (market-driven) governance model on the AI issue, this paper points out that this system is facing a dilemma of "market irrationality" (profit maximization overriding safety).

## **2 LITERATURE REVIEW AND THEORETICAL FRAMEWORK**

### **2.1 Theoretical Framework I: "Race to the Bottom"**

"Race to the bottom" is a classic concept from international political economy, originally used to describe how nations (or regions) compete to attract investment and mobile capital by lowering labor standards, environmental regulations, and taxes in the context of globalization (Cary, 1974). This competitive logic leads to a "Tragedy of the Commons" for the public interest, where individual rationality (lowering standards for short-term advantage) results in collective irrationality (overall welfare is damaged) (Hardin, 1968).

In the field of AI R&D, this "race to the bottom" is particularly extreme. Its core drivers are "money and power" (Acemoglu & Johnson, 2023). AI is seen as a massive, multi-trillion-dollar market and a key force for reshaping the geostrategic landscape (JPMorgan Chase, 2024; Wasi et al., 2025). In this context, the core objective of actors (mainly tech giants and nations) has degenerated from "ensuring safety" to "ensuring a lead." At the corporate level, market logic manifests as "winner-takes-all" and "first-mover advantage." Safety measures, ethical reviews, and long-term "alignment" research are seen as "costs" and "burdens" that slow down time-to-market. At the national level, geostrategic logic manifests as an "AI arms race." This dilemma is profoundly revealed by Graham Allison's (2017) "Thucydides's Trap" theory. The theory posits that a rising power will inevitably challenge the hegemony of an existing established power (like the U.S.), and this structural pressure easily leads to war. AI, as the core of the Fourth Industrial Revolution and even a military revolution, has become the eye of the storm in this "trap." Both sides view AI as a strategic "must-win" high ground. Any "concession" or "pause" on AI safety would be seen by the other side as "strategic weakness" or "unilateral disarmament," thereby intensifying insecurity and forcing both sides (regardless of their original intentions) to adopt an "accelerationist" and irrational stance.

Consequently, AI safety—a desperately needed "Global Public Good" (Kaul et al., 1999)—is severely "externalized" under the dual pressures of the market and geostrategy, meaning its costs are postponed or transferred to society as a whole.

## **2.2 Theoretical Framework II: "Governance Myopia"**

"Governance myopia" (or "political myopia") is a core issue in political science and public administration, referring to the phenomenon where political systems, in their decision-making, systematically favor "short-term gains" while ignoring "long-term risks" (Jacobs, 2011). This myopia is rooted in the structure of modern governance systems. In Western democratic election cycles, the incentive structure for politicians (like America's "lawyer-politicians") is highly tied to the election cycle (2, 4, or 6 years). Their political survival depends on short-term, highly visible achievements, while issues like AI safety—which are "long-term, high-uncertainty, and low-visibility"—are systematically ignored (Downs, 1957). Meanwhile, bureaucratic systems are inherently "reactive" rather than "predictive." They are adept at handling "downstream" problems with existing precedents (e.g., data privacy, copyright disputes), but when faced with "upstream," exponential, and unprecedented "fundamental" threats like AGI, they exhibit cognitive dissonance and instrumental failure. Furthermore, the "cognitive gap" between governance elites and tech elites is widening. As this paper will analyze later, America's "lawyer-financial" elites are more accustomed to linear, precedent-based, and inductive thinking based on laws and rules, making it difficult to comprehend the "exponential" and "emergent" characteristics of AI (Taleb, 2007).

## **2.3 Literature Divergence and This Paper's Positioning**

Existing AI governance literature is mainly divided into three categories:

**AI Ethics and "Downstream" Governance:** This literature (e.g., Noble, 2018; Zuboff, 2019) focuses on the social harms AI has already produced, such as algorithmic bias, surveillance capitalism, data privacy, and labor alienation. This research is necessary, but its focus is on "regulating" the "use" of AI, rarely touching on the "upstream" "existential risk" of AI capabilities themselves.

AI Safety and "Upstream" Governance: This literature (e.g., Bostrom, 2014; Russell, 2019; Ord, 2020) concentrates on AGI's "alignment problem" and "existential risk" (X-risk). This research is highly forward-looking but often remains at the level of technical and philosophical speculation, lacking robust engagement with real-world political and economic structures.

AI and Geostrategy: This literature (e.g., Allison, 2017; JPMorgan Chase, 2024) treats AI as the core variable in the U.S.-China "Thucydides's Trap," focusing on the "arms race" between nations. This research captures the essence of "competition" but often equates "safety" with "national security," ignoring "common human security."

This paper argues that without solving the "upstream" safety problems (Category 2), the "downstream" ethical governance (Category 1) will lose its foundation. And without understanding the "geostrategy" (Category 3) and "capitalist logic" (Zuboff, 2019) revealed in Category 1, the "global cooperation" called for by Category 2 will remain utopian.

### **3 IRRATIONAL COMPETITION DRIVEN BY "MONEY AND POWER"**

#### **3.1 Market Monopolization and Accelerationism**

The failure on the AI development front is rooted in the fundamental forces driving the technology: "money and power" (Acemoglu & Johnson, 2023). Under the current logic of capitalist markets, AI is seen as the next trillion-dollar platform after the internet and mobile internet. Tech giants like Google, Meta, Microsoft/OpenAI, and Anthropic, along with countless startups, are trapped in a "winner-takes-all" "accelerationist" race. The core logic is that the "first-mover advantage" of capturing the market first yields disproportionate returns, including data monopolies, platform hegemony, and the power to set standards. Under this logic, AI safety and "alignment" research, because they consume resources and slow the product launch cycle, are treated as "cost centers" rather than "profit centers" on financial statements. Therefore, safety is systematically "externalized," and its risks are transferred to society at large.

#### **3.2 OpenAI's "Haste" and the Cession of "Safety"**

OpenAI's trajectory is a typical microcosm of "safety" ceding to "speed." The company began with a "non-profit" mission to "ensure AGI benefits all of humanity" but quickly pivoted to a "capped-profit" model and accepted massive investment from Microsoft, deeply binding its fate to the latter's commercial empire (e.g., Bing search, Azure cloud). The "Sam Altman firing and reinstatement saga" in late 2023 was, in essence, a final battle between the company's internal "safety-oriented" faction (represented by the original board and Chief Scientist Ilya Sutskever) and the "accelerationist" faction (represented by CEO Altman and investors). Ultimately, the overwhelming victory of the "accelerationist" faction, and the subsequent collective resignation of the core safety team (including Ilya Sutskever and Jan Leike), showed that under the structural pressure of "money and power," even a mission-driven organization's initial "safety" commitments are fragile.

### **3.3 The "Tragedy of the Commons" in AI Safety**

AI safety, especially AGI "alignment," is a classic "global public good" (Kaul et al., 1999). It is non-excludable and non-rivalrous; if achieved, all of humanity benefits. However, in the current "anarchic state" lacking global mandatory regulation, the pursuit of this public good leads to a profound "tragedy of the commons" (Hardin, 1968). For any single actor (be it a corporation or a nation), proactively "pausing" R&D to ensure safety is tantamount to "unilateral disarmament." It would immediately place them at an absolute disadvantage in market competition and geostrategy. Therefore, the "individual rationality" of all actors (i.e., "accelerate R&D") converges into a "collective irrationality" (i.e., "a global sprint toward an unsafe future").

## **4 THE "MYOPIA" OF CAPITALIST-DOMINATED GOVERNANCE**

### **4.1 Limitations of "Lawyer-Business Politics": Cognitive Gaps and Misplaced Focus**

The failure of the U.S. governance system first manifests as a deep "cognitive gap" between the governance elite and the tech elite. This scene has played out repeatedly in U.S. Congressional hearings on AI: the thinking of the members (the vast majority with legal or financial backgrounds) is linear, precedent-based, and legalistic. Their questions are highly focused on "downstream," attributable issues that fit existing legal frameworks, such as "copyright" disputes, "privacy" violations, "bias" discrimination, and "disinformation." However, when tech CEOs (like Sam Altman) attempt to discuss "upstream," exponential, and unprecedented "existential risks" (like AGI loss of control, compute governance), the members exhibit clear cognitive difficulty. This "dialogue of the deaf" is a typical manifestation of governors being unable to comprehend an "exponential" threat (Taleb, 2007), leading to a severe misplacement of governance focus.

### **4.2 The Contradiction of "Cutting Scientific Research Funding"**

A profound contradiction exists: On one hand, private companies are investing unprecedentedly massive sums (often in the hundreds of billions) into AI capability R&D; on the other hand, the U.S. government (especially Congress) is continuously cutting funding for basic research (especially in public-good areas like AI safety). This has led to a huge gap in talent, compute, and cognition between the government and private giants. The government can neither attract top AI talent to formulate regulatory policy nor does it possess sufficient computing power to "audit" and "evaluate" frontier models. This asymmetry in capability leaves the government completely passive in regulatory negotiations, having thoroughly lost its ability to formulate and execute effective regulation.

### **4.3 Inherent Flaws of the Capitalist Governance Model: Regulatory Capture**

The aforementioned cognitive gaps and capability asymmetries ultimately lead to a "regulatory capture" that is much deeper than traditional lobbying. This is not just tech giants influencing legislation through political donations, but a capture of "agenda-setting" and "expertise." Because the government itself lacks professional knowledge, it is forced to rely on the very tech giants it is supposed to regulate (like OpenAI and Google) to draft executive orders and define "safety standards" and "risk frameworks." The result is that regulatory policies (such as the Biden administration's AI Executive Order) often devolve into a "hodgepodge" of "corporate self-regulation," with vague terms and no mandatory enforcement, in effect protecting the giants' market monopoly rather than public safety.

## **5 THE DILEMMA OF STATE-LED GOVERNANCE**

### **5.1 Strong Control and the Limitations of Competition**

A strong state apparatus provides high execution capabilities in "downstream" AI governance. For example, China was one of the first countries in the world to implement comprehensive, mandatory regulation of "algorithmic recommendations" (the algorithm filing system) and "Deepfakes." This top-down control is highly effective in standardizing data use and maintaining social stability. This "understanding" and "control" of AI primarily aims to serve "national development," "social stability," and "competition with great powers," while more work is still needed to address the global, abstract AGI "alignment problem."

### **5.2 The AI Paradox of the State-Led Governance Model**

In theory, a state-led model can overcome the "myopia" of Western capitalism (Jacobs, 2011). It can "concentrate strength to accomplish major tasks," engage in ultra-long-term strategic planning, and invest enormous resources to solve "choke point" problems like AGI safety. In practice, however, this systemic advantage of "concentrated strength" is "hijacked" by the "Thucydides's Trap" (Allison, 2017). Faced with U.S. technological blockades and the immense temptation of "overtaking on the curve," the national will is directed toward an "AI arms race" with the goal of "surpassing the U.S. as quickly as possible." Thus, the urgency of "development" and "catching up" overrides the long-term nature of "safety" and "alignment." This leads to an intensification of the global "race to the bottom" for the sake of "national irrationality" (geostrategy).

## **6 THE GLOBAL PRISONER'S DILEMMA OF THE "AI ARMS RACE"**

### **6.1 The "Accelerator" of Great Power Competition**

AI technology, especially AGI, is widely regarded as the absolute core of the Fourth Industrial Revolution and future military revolutions. It is not just a new economic sector but the "commanding height" that will determine national power for the next 100 years. Therefore, AI has become the "accelerator" and "main battlefield" of the U.S.-China "Thucydides's Trap." In this competition, the "winner-takes-all" logic extends from the "market" to "geostrategy": the nation that first achieves AGI is believed to gain a "decisive strategic advantage," thereby ending great power competition once and for all.

### **6.2 The "Confluence" of Two Governance Failures: Formation of the Prisoner's Dilemma**

Both the capitalist-dominated and state-led governance models have their shortcomings, and together they create a perfect "prisoner's dilemma." Tech giants race frantically for "money" and "market monopoly," while governments, due to "myopia" and "regulatory capture," are powerless to restrain them. State powers race at full speed for "power" and "geostrategic dominance," placing "safety" second to "catching up." Both great powers are trapped in a "security dilemma": each side views the other's "accelerated R&D" as an "offensive" threat and is thus forced to accelerate its own R&D as "defense." Both sides may know that cooperation (e.g., a global pause on strong AI R&D, establishing global safety standards) is best for "collective humanity" (the Pareto optimum), but the temptation of unilateral defection (continuing R&D) is

too great, and mistrust of the other side is deeply entrenched. Ultimately, both sides choose to "defect" (i.e., "to race"), leading to the worst collective outcome of "mutual insecurity" (the Nash equilibrium).

## **7 CONCLUSION: BREAKING THE VICIOUS CYCLE OF "COMPETITION" AND "MYOPIA"**

### **7.1 Summarizing the Structural Failure**

This paper has systematically argued that global AI governance is facing a profound "structural failure." This failure is not technical, but political and economic. It stems from a vicious cycle of "race to the bottom" in development and "structural myopia" in governance. The capitalist governance model is trapped by "market irrationality" (profit-driven acceleration), while the great power governance model is trapped by "national irrationality" (geostrategy-driven acceleration). These two failures, through different paths, converge on the same outcome, collectively pushing humanity toward a high-stakes "AI arms race."

### **7.2 Path Reflections and Future Outlook**

Breaking this vicious cycle, while extremely difficult, is not impossible. This paper proposes three potential paths for reflection:

A global consensus must be built to treat AGI safety (especially the "alignment problem") as a "common global threat" (Ord, 2020) that transcends sovereignty and ideology, on par with "nuclear war" or "global pandemics." This topic must be "decoupled" from conventional geostrategic and economic competition.

Nations (especially the U.S. and China) must establish "anti-myopia" domestic governance mechanisms—an "AI safety regulatory agency" that is independent of short-term political cycles and granted full authority. This agency must be led by top scientists (not lawyers or bureaucrats) and possess an independent status and professional authority similar to a "nuclear safety agency" or a "central bank" (like the Federal Reserve) to counter "myopic" pressures from the market and politics.

In a "prisoner's dilemma," the only path to building trust is "verifiable oversight." The international community urgently needs a global AI regulatory body, similar to the International Atomic Energy Agency (IAEA). Its core function would be to "monitor" and "verify" ultra-large-scale AI training (e.g., registering and monitoring massive compute clusters, treating them as "AI's enriched uranium") and to conduct mandatory third-party safety audits before "frontier models" are deployed.

### **7.3 Concluding Remarks**

AI technology is advancing on an "exponential" clock, while human political governance, bureaucracy, and international relations remain on a "linear" or even "cyclical" clock. In this replay of the "Oppenheimer Moment," human society must undergo a profound "cognitive revolution" in governance, replacing "short-term impulses" with "long-term rationality," and "zero-sum games" with "collective security." Otherwise, we may not get a second chance to correct our first mistake.

## **REFERENCES**

Acemoglu, D., & Johnson, S. (2023). Power and Progress: Our Thousand-Year Struggle Over Technology and

Prosperity. New York: PublicAffairs.

Allison, G. (2017). *Destined for War: Can America and China Escape Thucydides's Trap?* Boston: Houghton Mifflin Harcourt.

Bengio, Y., Hinton, G., Yao, A., Song, D., Abbeel, P., Darrell, T., Harari, Y. N., Zhang, Y.-Q., Xue, L., Shalev-Shwartz, S., Hadfield, G., Clune, J., Maharaj, T., Hutter, F., Baydin, A. G., McIlraith, S., Gao, Q., Acharya, A., Krueger, D., Dragan, A., Torr, P., Russell, S., Kahneman, D., Brauner, J., & Mindermann, S. (2024). Managing extreme AI risks amid rapid progress. *Science*, 384(6698), 842–845. <https://doi.org/10.1126/science.adn0117>

Boström, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Cary, W. L. (1974). Federalism and Corporate Law: Reflections Upon Delaware. *The Yale Law Journal*, 83(4), 663–705.

Crain, C. (2025, May 31). Meta to replace human risk reviewers with AI, raising safety concerns. *Business & Human Rights Resource Centre*. <https://www.business-humanrights.org/en/latest-news/meta-to-replace-human-risk-reviewers-with-ai-raising-safety-concerns/>

Downs, A. (1957). *An Economic Theory of Democracy*. New York: Harper & Row.

Hardin, G. (1968). The Tragedy of the Commons. *Science*, 162(3859), 1243–1248.

Hassabis, D. (2024). Remarks at the AI Safety Summit.

Jacobs, A. M. (2011). *Governing for the Long Term: Democracy and the Politics of Investment*. Cambridge: Cambridge University Press.

JPMorgan Chase. (2024). The geopolitics of AI: Decoding the new global operating system. <https://www.jpmorganchase.com/content/dam/jpmorganchase/documents/center-for-geopolitics/decoding-the-new-global-operating-system.pdf>

Kaul, I., Grunberg, I., & Stern, M. A. (Eds.). (1999). *Global Public Goods: International Cooperation in the 21st Century*. New York: Oxford University Press.

Kinnard, M., Chan, K., Beaty, T., & O'Brien, M. (2024, March 1). Elon Musk sues OpenAI and CEO Sam Altman, claiming betrayal of its goal to benefit humanity. *Quartz*. <https://qz.com/elon-musk-sues-openai-and-ceo-sam-altman-claiming-betr-1851300019>

Kissinger, H. A., Schmidt, E., & Huttenlocher, D. (2021). *The Age of AI: And Our Human Future*. Boston: Little, Brown and Company.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Ord, T. (2020). *The Precipice: Existential Risk and the Future of Humanity*. Hachette Books.

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking.

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

Wasi, A. T., Eram, E. H., Mitu, S. A., & Ahsan, M. M. (2025). Generative AI as a geopolitical factor in Industry 5.0: Sovereignty, access, and control. *arXiv*. <https://doi.org/10.48550/arXiv.2508.00973>

Wei, J., Tay, Y., Bommasani, R., et al. (2022). Emergent Abilities of Large Language Models. *arXiv*.

White House Office of Management and Budget. (2020). Guidance for regulation of artificial intelligence applications (OMB Memorandum No. M-21-06). Executive Office of the President. Retrieved from <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.