Yale Journal of Law & Technology
Volume 29, Issue 1

Mutually Assured Deregulation

Gilad Abiri

We have convinced ourselves that the way to make AI safe is to make it unsafe. Since 2022, many policymakers worldwide have embraced the "Regulation Sacrifice"—the belief that dismantling safety oversight will somehow deliver security through AI dominance. This reasoning follows a perilous pattern: fearing that one major power will dominate the AI landscape at the expense of the others, we rush to eliminate any safeguard that might slow our progress. This Article reveals the fatal flaw in such thinking. Though AI development certainly poses national security challenges, the solution demands stronger regulatory frameworks, not weaker ones. A race without guardrails does not build competitive strength—it breeds shared danger.

The Regulation Sacrifice makes three promises. Each one is false. First, it promises durable technological leads. But as a form of dual-use software, AI capabilities spread like wildfire. Performance gaps between U.S. and Chinese systems collapsed from nine to two percent in thirteen months. When advantages evaporate in months, sacrificing permanent safety for temporary speed makes no sense.

Second, it promises that deregulation accelerates innovation. The opposite is quite often true. Companies report that well-designed governance frameworks streamline their development. Investment flows toward regulated markets, not away from them. Clear rules reduce uncertainty. Uncertain liability creates paralysis. We have seen this movie before—environmental standards did not kill the auto industry; they created Tesla and BYD.

Third, the promise of enhanced national security through deregulation is perhaps the most dangerous fallacy, as it undermines security across timeframes. In the near term, it hands our adversaries perfect tools for information warfare. In the medium term, it puts bioweapon capabilities in everyone's

hands. In the long term, it guarantees we'll deploy AGI systems we cannot control, racing to be the first to push a button we cannot un-push.

The Regulation Sacrifice persists because it serves powerful interests, not because it serves security. Tech companies prefer freedom to accountability. Politicians prefer simple stories to complex truths. Together, they are trying to convince us that recklessness is patriotism. But here is the punchline: these ideas create a system of mutually assured deregulation, where each nation's sprint for advantage guarantees collective vulnerability. The only way to win this game is not to play.

Article Contents

Introduction 4
I. The Regulation Sacrifice 11
II. The Assumptions of the AI Regulation Sacrifice 23
A. Durable Lead Assumption 24
B. The Low Drag Assumption 32
C. Net Strategic Benefit Assumption 41
III. The National Security Risks of AI 44
A. Near Term: Machine-Scale Subversion 45
B. Medium Term: Biosecurity and the Erosion of Control 51
C. Long Term: Digital Gods (AGI / ASI) 56
1. AGI and National Security 57
2. AGI and the Regulation Sacrifice 62
D. Toward Well-Designed Regulation 68
Conclusion 70
A. The Verdict 70
B. The Persistence of a Failed Paradigm 71
C. Implications for Law and Policy 72

Introduction

We are locked in a contest over how artificial intelligence should be governed, but we have been misled about what true victory entails.[1] Since 2022, AI policy discussions worldwide have been overtaken by a competitive framework that recasts AI development as a zero-sum contest in which regulatory restraint equals patriotic duty.[2] President Trump dismissed AI

---

[1] Last year, Congress considered an unprecedented ten-year moratorium on state-level AI regulation, with proponents warning that without it, the United States would lose the AI arms race to China. Although the measure was not enacted, it represents the apotheosis of a dangerous and influential paradigm that has swept from Washington across the global technology landscape. *See* H.R. 1, 119th Cong. § 43201(c)(1) (as passed by House, May 22, 2025); Matt Brown & Matt O'Brien, *House Republicans Include a 10-Year Ban on US States Regulating AI in 'Big, Beautiful' Bill*, ASSOCIATED PRESS (May 16, 2025), https://apnews.com/article/ai-regulation-state-moratorium-congress-39d1c8a0758ffe0242283bb82f66d51a [https://perma.cc/X5ZS-7F6E]; *Winning the AI Race: Strengthening U.S. Capabilities in Computing and Innovation: Hearing Before the S. Comm. on Com., Sci., & Transp.*, 119th Cong. (2025) [hereinafter *Winning the AI Race*] (announcement of Sen. Ted Cruz, Chairman, S. Comm. on Com., Sci., & Transp.) https://www.commerce.senate.gov/2025/5/winning-the-ai-race-strengthening-u-s-capabilities-in-computing-and-innovation_2 [https://perma.cc/WDM9-RQQ4].

[2] ASIA PAC. TASK FORCE, *The Age of AI in U.S.-China Great Power Competition: Strategic Implications, Risks, and Global Governance*, BEYOND THE HORIZON (Feb. 3, 2025), https://behorizon.org/the-age-of-ai-in-u-s-china-great-power-competition-strategic-implications-risks-and-global-governance/ [https://perma.cc/2PQZ-7DKD]; Brian J. Chen, *The "Washington Effect" Could Decide the AI Race*, PROJECT SYNDICATE (June 19, 2025), https://www.project-syndicate.org/commentary/us-trump-administration-tries-bullying-its-way-to-ai-dominance-by-brian-j-chen-2025-06 [https://perma.cc/7MTD-XQ3V] (documenting U.S. pressure on allies to adopt deregulatory approaches); Angela Huyue Zhang, *Agility over Stability: China's Great Reversal in Regulating the Platform Economy*, 63 HARV. INT'L L.J. 457, 508–09 (2022) (arguing that China used platform-economy enforcement to push big tech toward core technologies, including AI and driverless cars); *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: AI Continent Action Plan*, at 1, COM(2025) 165 final (Apr. 9, 2025) (declaring its commitment to "becom[ing] a global leader in Artificial Intelligence" while emphasizing the need to enhance its competitiveness); *Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet at the AI Action Summit*, PERMANENT MISSION FR. UNITED NATIONS N.Y. (Feb. 11, 2025),

safeguards as threats to "American technological leadership," while threatening to withdraw from international AI agreements and pressuring allies to adopt similar deregulatory stances.[3] At the 2025 Paris AI summit, Vice President Vance warned that "excessive regulation of the AI sector could kill a transformative industry just as it's taking off," even as he cautioned that "some authoritarian regimes have stolen and used AI to strengthen their military, intelligence, and surveillance capabilities; capture foreign data; and create propaganda to undermine other nations' national security."[4] Major AI companies echo this rhetoric globally:[5] OpenAI

---

https://onu.delegfrance.org/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and [https://perma.cc/2YFE-3ZJQ]; Anu Bradford, *The False Choice Between Digital Regulation and Innovation*, 119 NW. U. L. REV. 377, 381 (2024) [hereinafter The False Choice] (describing how fears of losing technological leadership in a U.S.-China "tech war" fuel arguments that stricter digital regulation would harm national power and competitiveness); Anu Bradford, Eileen Li & Matthew C. Waxman, *How Domestic Institutions Shape the Global Tech War*, 16 HARV. NAT'L SEC. J. 76, 101 (2025) (explaining how national security-framed competition over advanced technologies, including AI, is used to argue that regulation undermines competitiveness and thus national security); ANU BRADFORD, DIGITAL EMPIRES: THE GLOBAL BATTLE TO REGULATE TECHNOLOGY 6–11 (2023) (framing the United States, the European Union, and China as rival "digital empires" competing to set global rules for the digital economy, including AI); Filippo Lancieri, Laura Edelson & Stefan Bechtold, *AI Regulation: Competition, Arbitrage & Regulatory Capture*, 26 (Geo. U.L. Ctr. Research Paper No. 2025/05, 2025), http://dx.doi.org/10.2139/ssrn.5049259 (predicting a largely fragmented global AI regime with occasional harmonization and cautioning against race-to-the-bottom dynamics).

[3] Exec. Order No. 14,179, 90 Fed. Reg. 8741; *Fact Sheet: President Donald J. Trump Takes Action to Enhance America's AI Leadership*, THE WHITE HOUSE (Jan. 23, 2025), https://www.whitehouse.gov/fact-sheets/2025/01/fact-sheet-president-donald-j-trump-takes-action-to-enhance-americas-ai-leadership/ [https://perma.cc/EQ7B-SZHF].

[4] J.D. Vance, *Remarks by the Vice President at the Artificial Intelligence Action Summit in Paris, France*, THE AMERICAN PRESIDENCY PROJECT (Feb. 11, 2025), https://www.presidency.ucsb.edu/documents/remarks-the-vice-president-the-artificial-intelligence-action-summit-paris-france [https://perma.cc/GHH9-NWKA].

[5] *See* Amanda Levendowski, *How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem*, 93 WASH. L. REV. 579, 612–16 (2018) (analyzing how tech companies invoke innovation arguments to resist bias mitigation requirements); Ari Ezra Waldman, *Power, Process, and Automated Decision-Making*, 88 FORDHAM L. REV. 613, 621, 628–32 (2019)

warns of "a Chinese Communist Party determined to overtake us by 2030," while Google describes the current situation as a "global AI competition" where "policy decisions will determine the outcome."[6] The message is clear: in the sprint for AI dominance, regulation is a disastrous hurdle.

A growing body of scholarship has pushed back on the AI arms race narrative, but largely by contesting the metaphor itself. Critics have argued that the race framing is descriptively inaccurate, strategically counterproductive, or rhetorically dangerous—that there is no race to speak of, that cooperation would serve all sides better, or that the metaphor is a fiction that manufactures the very dangers it claims to describe.[7] Most recently, Boris Babic and Brian Wong have recast the contest in rational-choice terms as a "stag hunt" rather than a race—a game in which the two powers gain more by cooperating than by defecting, so that the "winner-takes-all" framing misdescribes the strategic structure it claims to capture.[8] This Article takes a different path. Rather than disputing whether

---

(arguing that process-based algorithmic accountability is prone to corporate co-optation and that migrating governance into code shifts power to engineers and their firms, and urging substantive audits/mandates).

[6] Christopher Lehane*, Response to the National Science Foundation [NSF] and Office of Science and Technology Policy [OSTP] Request for Information on the Development of an AI Action Plan,* OPENAI 1 (Mar. 13, 2025) [*hereinafter OpenAI's Response*], https://cdn.openai.com/global-affairs/ostp-rfi/ec680b75-d539-4653-b297-8bcf6e5f7686/openai-response-ostp-nsf-rfi-notice-request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan.pdf [https://perma.cc/8EUH-7QT8]; *Response to the NSF and OSTP Request for Information on the Development of an AI Action Plan,* GOOGLE 2 (Mar. 13, 2025), https://static.googleusercontent.com/media/publicpolicy.google/en//resources/response_us_ai_action_plan.pdf [https://perma.cc/SC8Q-HQSR] [*hereinafter Google's Response*].

[7] *See* Paul Scharre, *Debunking the AI Arms Race Theory*, 4 TEX. NAT'L SEC. REV. 121 (2021); Tiffany C. Li, *Ending the AI Race: Regulatory Collaboration as Critical Counter-Narrative*, 69 VILL. L. REV. 981 (2025); Kimberly A. Houser & Anjanette H. Raymond, *It Is Time to Move Beyond the "AI Race" Narrative: Why Investment and International Cooperation Must Win the Day*, 18 NW. J. TECH. & INTELL. PROP. 129 (2021); Seán S. Ó hÉigeartaigh, *The Most Dangerous Fiction: The Rhetoric and Reality of the AI Race*, *in* THE ARTIFICIAL INTELLIGENCE REVOLUTION 45 (Max Rangeley & Nicholas Fairfax eds., 2026).

[8] BORIS BABIC & BRIAN WONG, GEOPOLITICS OF ARTIFICIAL INTELLIGENCE (Cambridge Univ. Press 2026).

the metaphor is apt or whether cooperation would be preferable, it accepts the arms race framing on its own terms and asks a narrower question that the existing critiques largely bypass: even granting the competition, does *deregulation* actually deliver the security gains its proponents promise? Anu Bradford's account of the "false choice" between digital regulation and innovation comes closest to this move, but where Bradford shows that regulation and innovation can coexist, this Article isolates the specific strategic wager at the heart of the deregulatory argument—what I call the "Regulation Sacrifice," the claim that nations must abandon safety oversight to win the AI race—[9]and tests it against the security rationale offered in its defense. That wager, I argue, rests on three implicit premises—a *durable lead*, *low drag*, and *net strategic benefit*—each of which must hold for the sacrifice of regulation to make strategic sense, and none of which survives scrutiny. The result is a critique that operates from inside the arms race logic rather than against it: the Regulation Sacrifice fails not because the race is a myth, but because, even if the race is real, dismantling oversight is the wrong way to win it.

The Regulation Sacrifice is no academic abstraction; it has become a dominant policy paradigm, wielded by companies and governments alike to justify dismantling governance frameworks. [10] This Article demonstrates that sacrificing governance for speed fails not despite its national security rationale but because of it. The paradigm does not merely rest

---

[9] Bradford, *supra* note 2, at 401–02, 419–29, 440–44.

[10] This Article introduces the term "Regulation Sacrifice" to describe the systematic abandonment of safety oversight justified by competitive imperatives. For related analyses of this phenomenon, *see* Markus Trengove & Emre Kazim, Dilemmas in AI Regulation: An Exposition of the Regulatory Trade-Offs Between Responsibility and Innovation 3–5 (Apr. 6, 2022) (unpublished manuscript) (on file with SSRN), https://dx.doi.org/10.2139/ssrn.4072436 (documenting trade-offs between innovation and responsibility). To be clear, the Regulation Sacrifice dynamic predates AI and extends beyond it. *See, e.g.*, JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019) (examining how platform companies deploy competitive necessity narratives to resist regulation); Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 425–30 (2017) (documenting early emergence of innovation-over-safety framing in AI policy).

on a flawed metaphor; it fundamentally misconceives how security is achieved in an era of rapidly diffusing technology.

This Article establishes this argument in three steps. Part I traces how the arms race narrative has evolved from rhetorical device to concrete policy, generating a reflexive equation of regulation with strategic disadvantage. It examines how government officials and AI firms strategically wield national-security rhetoric to promote the Regulation Sacrifice narrative—the view that every month of extra speed is worth more to national security than any risk reduction achieved through oversight. By focusing on the United States and China, Part I demonstrates how this competitive logic constrains regulatory ambition domestically and relegates international coordination to toothless declarations of principle.[11]

Part II interrogates the three implicit premises that must hold for the Regulation Sacrifice to make strategic sense. First, it tests whether temporary bursts of deregulated speed create durable technological advantages, finding that AI capabilities diffuse so rapidly that performance gaps collapse within months, not years.[12] Second, it examines whether governance measures actually impose prohibitive costs on innovation, discovering that well-designed regulation often enhances rather than hinders technological progress.[13] Third, it asks whether reaching the frontier first delivers the security advantages the wager assumes, and shows on theoretical

---

[11] *See infra* Part I.

[12] *See also* NESTOR MASLEJ ET AL., ARTIFICIAL INTELLIGENCE INDEX REPORT 2025, STAN. HUM.-CENTERED AI 14–15 (2025); Mackenzie Ferguson, *DeepSeek's Breakthrough: A New Era for AI with Less Compute Power*, OPENTOOLS (Dec. 29, 2024), https://opentools.ai/news/deepseeks-breakthrough-a-new-era-for-ai-with-less-compute-power [https://perma.cc/U565-LVJC].

[13] Much of the research on the relationship between regulation and innovation deals with environmental regulation. *See, e.g.*, Michael E. Porter & Claas van der Linde, *Toward a New Conception of the Environment–Competitiveness Relationship*, 9 J. ECON. PERSPS. 97 (1995) (arguing that properly designed environmental regulation can induce 'innovation offsets' that partially or more than fully compensate compliance costs); Antoine Dechezleprêtre & Misato Sato, *The Impacts of Environmental Regulations on Competitiveness*, 11 REV. ENV'T ECON. & POL'Y 183 (2017) (reviewing empirical literature and concluding that competitiveness effects are generally small and that regulation can induce innovation in cleaner technologies).

grounds that the deterrence-by-preponderance logic behind that assumption does not transfer to a technology whose capabilities diffuse and cannot be monopolized. Part III then tests that conclusion empirically, tracing how deregulated speed affects security across three time horizons.

Part III employs a temporal framework recently proposed by Anthropic CEO Dario Amodei to assess whether deregulation attains its stated security objectives.[14] In the near term, the gravest threats are cybersecurity breaches and machine-scale misinformation, risks that voluntary measures have not suppressed. The technique exists: the Coalition for Content Provenance and Authenticity has published a mature provenance specification that major vendors helped write. What is missing is any requirement to deploy it, and adoption has accordingly been partial and reversible. Deregulation, by stripping away provenance checks, disclosure mandates, and liability safeguards, widens the gap between available technique and binding obligation.[15] Over the medium horizon, the retreat from regulation accelerates the erosion of safeguards that once confined bioweapons to a handful of state laboratories. Frontier models will be able to walk even modestly resourced groups through virulence optimization, DNA synthesis, and aerosol delivery. Dismantling of oversight infrastructure—sequence-order screening, model risk evaluations, liability regimes—lets those capabilities spread unchecked. [16] In the long term, competitive dynamics

---

[14] *Oversight of A.I.: Principles for Regulation: Hearing Before the Subcomm. on Priv., Tech. & the L. of the S. Comm. on the Judiciary*, 118th Cong. 2–3 (July 25, 2023) [hereinafter *Amodei Testimony*] (written testimony of Dario Amodei, CEO, Anthropic), https://www.judiciary.senate.gov/imo/media/doc/2023-07-26_-_testimony_-_amodei.pdf [https://perma.cc/HR3E-X4QL].

[15] *See* COAL. FOR CONTENT PROVENANCE & AUTHENTICITY [C2PA], C2PA TECHNICAL SPECIFICATION V1.3, 2 (2023), https://c2pa.org/specifications/specifications/1.3/specs/_attachments/C2PA_Specification.pdf [https://perma.cc/8GTK-2RED] *(publishing a technical standard for cryptographically bound content credentials, developed by a coalition including major platform and model vendors, whose deployment is left to the discretion of each adopter)*; Robert Chesney & Danielle Citron, *Deepfakes and the New Disinformation War*, FOREIGN AFFS., Jan./Feb. 2019, at 147, 152–54.

[16] BILL DREXEL & CALEB WITHERS, CTR. FOR NEW AM. SEC., AI AND THE EVOLUTION OF BIOLOGICAL NATIONAL SECURITY RISKS: CAPABILITIES,

surrounding potential artificial general intelligence (AGI) create powerful incentives that systematically undermine safety,[17] with deregulation removing the tools required to keep transformative AI under meaningful human control.[18] At each horizon, the evidence reveals that racing without rules amplifies rather than mitigates the security threats nations seek to avoid. Part III closes by stepping back from this diagnosis to sketch what better governance would look like across the same three horizons.

The evidence converges on a clear conclusion: the Regulation Sacrifice fails on its own terms. Rather than securing national advantage, it hands adversaries the same dangerous capabilities while creating cascading vulnerabilities that no nation can control. The paradigm persists not because it delivers security but because it satisfies powerful interests: companies seeking freedom from oversight, politicians seeking simple narratives, and a public primed to see competition

---

THRESHOLDS, AND INTERVENTIONS 3, 19 (2024); DAVID LUCKEY ET AL., MITIGATING RISKS AT THE INTERSECTION OF ARTIFICIAL INTELLIGENCE AND CHEMICAL AND BIOLOGICAL WEAPONS 34–36 (2025), https://www.rand.org/content/dam/rand/pubs/research_reports/RRA2900/RRA2990-1/RAND_RRA2990-1.pdf [https://perma.cc/GXN3-VKPW]; Jonas B. Sandbrink, Artificial Intelligence and Biological Misuse: Differentiating Risks of Language Models and Biological Design Tools 6 (Dec. 23, 2023) (unpublished manuscript) (on file with arXiv), https://arxiv.org/pdf/2306.13952 [https://perma.cc/GRU9-C2FB].

[17] Stuart Armstrong, Nick Bostrom & Carl Shulman, *Racing to the Precipice: A Model of Artificial Intelligence Development*, 31 AI & Soc'y 201 (2016); *see also* NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES 246–48 (2014) (summarizing the model). *Cf.* Drexel & Withers, *supra* note 16, at 23–25 (describing analogous competitive pressures in AI-enabled biotechnology).

[18] NICK BOSTROM, *supra* note 17, at 127–43; DANIEL KOKOTAJLO ET AL., AI 2027: A SCENARIO FOR SUPERHUMAN AI 9 (2025), https://ai-2027.com/ai-2027.pdf [https://perma.cc/8UGL-W3EU] (illustrating that leaked model weights can eliminate safety advantages and that as racing accelerates, diplomatic arms-control windows shrink, driving governments toward coercive options); Yoshua Bengio, Implications of Artificial General Intelligence on National and International Security (Oct. 16, 2024) (unpublished manuscript) (on file with Aspen Strategy Grp.), https://www.aspeninstitute.org/wp-content/uploads/2024/10/Bengio_National-and-International-Security-in-an-AGI-Context_Final.pdf [https://perma.cc/2QCL-5MDN].

everywhere.[19] Yet the stakes are too high for such convenient fictions. We stand at a juncture where governance choices will determine whether AI enhances human flourishing or amplifies existential risks.[20] By exposing the bankruptcy of this central paradigm, this Article clears space for approaches that recognize what the evidence makes plain: in an era of rapidly diffusing technology, security comes not from racing without rules but from channeling innovation through governance frameworks that enhance both safety and progress. The path forward begins with abandoning the dangerous mythology of the Regulation Sacrifice—the very thinking that leads inexorably to Mutually Assured Deregulation.

I. The Regulation Sacrifice

We start by examining how the framing of artificial intelligence development as a geopolitical arms race is deployed as an anti-regulatory argument, particularly in the United States and China, the two nations at the forefront of developing cutting-edge AI systems. It shows that government officials and AI firms strategically wield national-security rhetoric to create what I call a Regulation Sacrifice. This is the view that because the decisive objective is to outrun adversaries (usually China or the United States) to achieve frontier capabilities—and to lock in that lead before norms or treaties can erode it—states should minimize regulatory constraints that might slow domestic developers. By "Regulation Sacrifice," I mean resistance to binding constraints that would slow frontier development, such as repealing or rolling back safety rules, preempting subnational regulation, substituting voluntary commitments for enforceable ones, and diluting international obligations. It does

---

[19] Ó hÉigeartaigh, *supra* note 7; Daniel Kahneman & Jonathan Renshon, *Hawkish Biases*, *in* AMERICAN FOREIGN POLICY AND THE POLITICS OF FEAR 79, 79–96 (A. Trevor Thrall & Jane K. Cramer eds., 2009) (documenting systematic biases toward competitive strategies).

[20] *Compare* John O. McGinnis, *Accelerating AI*, 104 NW. U. L. REV. ONLINE (2010) (arguing government should support "friendly AI" research under light-touch governance), *with* Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 378–82 (2016) (warning that unregulated AI development could lead to public risks and security vulnerabilities).

not mean the wholesale absence of governance. The claim is not that proponents want no rules at all, but that they treat any binding oversight capable of slowing the sprint as a strategic liability to be deferred or discarded. Temporary privacy and security risks are framed as an acceptable price so long as the sprint delivers a durable edge that can later be leveraged to impose rules from a position of strength. In short: win first, tidy up later.

By focusing primarily on the United States, with comparative insights from China, I demonstrate how the arms race narrative functions not merely as description but as a powerful political tool that (i) constrains regulatory ambition domestically and (ii) relegates international initiatives to toothless statements of principle, narrowing the scope of possible governance approaches.

Since 2022, a competitive, geopolitical framework has overtaken discussions of AI governance, recasting technological development as a zero-sum contest between nations where security imperatives consistently trump regulatory concerns.[21] This "arms race frame" positions AI advancement primarily as a contest for technological supremacy and protection against foreign threats, rather than a shared challenge requiring coordinated oversight.[22] Especially in the United States, but also in China and elsewhere,[23] policymakers and industry leaders increasingly invoke this competitive framework to resist domestic regulation and minimize international constraints.[24]

---

[21] ASIA PAC. TASK FORCE, *supra* note 2.

[22] Li, *supra* note 7, at 983–87.

[23] Yasmeen Serhan, *How Israel Uses AI in Gaza—And What It Might Mean for the Future of Warfare*, TIME (Dec. 19, 2024), https://time.com/7202584/gaza-ukraine-ai-warfare/ [https://perma.cc/K7GC-CJ58]; Sarita Chaganti Singh & Nikunj Ohri, *India's Finance Ministry Asks Employees to Avoid AI Tools like ChatGPT, DeepSeek*, REUTERS (Feb. 5, 2025), https://www.reuters.com/technology/artificial-intelligence/indias-finance-ministry-asks-employees-avoid-ai-tools-like-chatgpt-deepseek-2025-02-05/ [https://perma.cc/B8EU-M7RM]; CABINET OFF., GOV'T OF JAPAN. OUTLINE OF THE ACT ON PROMOTION OF RESEARCH AND DEVELOPMENT, AND UTILIZATION OF AI-RELATED TECHNOLOGY, https://www8.cao.go.jp/cstp/ai/ai_hou_gaiyou_en.pdf [https://perma.cc/2SA2-TTAE].

[24] *See* Elena Chachko, *National Security by Platform*, 25 STAN. TECH. L.

In the United States, the arms race narrative functions as a central argument for limiting regulation. Policymakers routinely invoke a geopolitical "AI race" against China to argue for light-touch regulation at home.[25] The framing is explicit: "[t]he way to beat China in the AI race is to outrace them in innovation, not saddle AI developers with European-style regulations," proclaimed Senator Ted Cruz, announcing a 2025 Senate hearing on "Winning the AI Race."[26] This competitive mindset casts regulation as a potential handicap to U.S. technological supremacy. Indeed, upon taking office in 2025, President Trump swiftly rescinded his predecessor's AI executive order; an accompanying White House fact sheet described its safeguards as "onerous" requirements that would "threaten American technological leadership."[27] In its place, Trump directed an AI Action Plan meant to "sustain and enhance America's global AI dominance," explicitly tying deregulation to the strategic aim of outpacing China.[28]

National security concerns, and not only commercial

REV. 55, 64 (2022) (noting that critics describe platforms' security/geopolitics reforms as "window dressing" intended "to deflect public criticism and appease regulators," and observing that security considerations can complicate regulatory responses, including antitrust pressure); Kristen E. Eichensehr, *Digital Switzerlands*, 167 U. PA. L. REV. 665, 698–703 (2019) (showing that U.S. platforms may not align with U.S. national security aims when unregulated).

[25] *See, e.g.*, *Winning the AI Race*, *supra* note 1 ("The way to beat China in the AI race is to outrace them in innovation, not saddle AI developers with European-style regulations."); *Global Networks at Risk: Securing the Future of Communications Infrastructure: Hearing Before the Subcomm. on Commc'ns & Tech. of the H. Comm. on Energy & Com.*, 119th Cong. 1 (2025) (statement of Rep. Hudson), https://energycommerce.house.gov/posts/chairman-hudson-delivers-opening-statement-at-subcommittee-on-communications-and-technology-hearing-on-ai-and-communications-infrastructure [https://perma.cc/7MFM-EWAA] (urging a "light regulatory touch . . . [and that] AI must be given the same opportunity").

[26] *Winning the AI Race*, *supra* note 1.

[27] *Fact Sheet: President Donald J. Trump Takes Action to Enhance America's AI Leadership*, *supra* note 3.

[28] Exec. Order No. 14,179, 90 Fed. Reg. 8741; OFF. OF SCI. AND TECH. POL., EXEC. OFF. OF THE PRESIDENT, THE WHITE HOUSE, WINNING THE RACE: AMERICA'S AI ACTION PLAN 3–6, 20–23 (2025), https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf [https://perma.cc/J6VW-T35B].

interests, lie at the heart of this framing. Vice President Vance articulated this perspective clearly at the 2025 Artificial Intelligence Action Summit in Paris, warning about "hostile foreign adversaries" weaponizing AI and authoritarian regimes using it to "strengthen their military, intelligence, and surveillance capabilities."[29] This security-centered narrative positions AI not merely as an economic asset but as critical infrastructure in a geopolitical struggle. Vance further cautioned European allies against "partnering with such regimes," framing the tech race as a struggle between democratic and authoritarian models of governance.[30]

Major AI companies' responses to the federal government's AI Action Plan Request for Information reveal how thoroughly industry has adopted this national security framing.[31] OpenAI warns of "a Chinese Communist Party (CCP) determined to overtake us by 2030" and argues for policies ensuring that "American-led AI . . . prevails over CCP-built autocratic, authoritarian AI."[32] OpenAI CEO Sam Altman has likewise framed AI development as a geopolitical race, warning that "The United States has got to lead in A.I. innovation. This technology is going to be used by our

---

[29] Vance, *supra* note 4.

[30] Vance, *supra* note 4.

[31] *See, e.g.*, *Response to 6the NSF and OSTP Request for Information on the Development of an Artificial Intelligence (AI) Action Plan*, AMAZON 62 (2025), https://files.nitrd.gov/90-fr-9088/Amazon-AI-RFI-2025.pdf [https://perma.cc/7SNB-BYTT] ("Our economic and national security interests are best served by policies that enhance technological dynamism and open markets."); *Response to the NSF and OSTP Request for Information on the Development of an Artificial Intelligence (AI) Action Plan*, ANTHROPIC 1–3 (Mar. 6, 2025), https://assets.anthropic.com/m/4e20a4ab6512e217/original/Anthropic-Response-to-OSTP-RFI-March-2025-Final-Submission-v3.pdf [https://perma.cc/5ZRB-43A7] [*hereinafter Anthropic's Response*] ("Our recommendations encompass . . . national security imperatives . . . . [A]rtificial intelligence systems . . . will present significant national security implications," so the administration "must strengthen national security frameworks"); *Google's Response*, *supra* note 6, at 1–2 ("Google welcomes the . . . goal to 'sustain and enhance America's global AI dominance' . . . . We are in a global AI competition . . . policy decisions will determine the outcome."); *OpenAI's Response*, *supra* note 6, at 6 (urging an AI Action Plan whose goal is to "strengthen American national security.").

[32] *OpenAI's Response*, *supra* note 6, at 1.

adversaries."[33] Similarly, Anthropic supports the Office of Science and Technology Policy's (OSTP) "effort to . . . maintain and strengthen America's dominance" while emphasizing the need for "comprehensive government awareness . . . as China advances its efforts to build powerful dual-use AI systems."[34] Google likewise frames AI development as a "global AI competition" where "policy decisions will determine the outcome."[35]

When policymakers conflate AI leadership with national security, they create a nearly irresistible argument against meaningful regulation. This security framing transforms what could be deliberative policy choices—with various legitimate resolutions—into binary questions of national survival, where regulatory leniency becomes the only patriotic option. In a regulatory proposal submission, OpenAI argued that state-level regulatory initiatives "could impose burdensome compliance requirements that may hinder [U.S.] economic competitiveness and undermine [U.S.] national security."[36] The company proposed "creating a tightly-scoped framework for voluntary partnership between the federal government and the private sector to expand and strengthen American national security" and argued that the "patchwork" of proposed AI regulations "risks bogging down innovation and, in the case of AI, undermining America's leadership position."[37] Google similarly called for "federal preemption of state-level laws that affect frontier AI models," arguing that this "would ensure a unified national framework . . . focused on protecting national security while fostering an environment where American AI innovation can thrive."[38]

Such pleas for regulatory relief in exchange for voluntary cooperation reveal how directly companies link geopolitical rivalry to domestic policy choices. In its submission, Anthropic

---

[33] Kristen Altus, *OpenAI CEO Sam Altman to Donate $1 Million to Trump's Inaugural Fund*, FOX BUS. (Dec. 13, 2024), https://www.foxbusiness.com/media/openai-ceo-sam-altman-donate-1-million-trumps-inaugural-fund [https://perma.cc/8WYY-GGTS] (quoting Altman's remarks in an interview with Fox Business's Charles Payne).

[34] *Anthropic's Response, supra* note 31, at 2.

[35] *Google's Response*, *supra* note 6, at 1.

[36] *OpenAI's Response*, *supra* note 6, at 4.

[37] *Id.* at 6.

[38] *Google's Response*, *supra* note 6, at 5.

urged the government to "[c]oordinate a cross-agency effort to identify and eliminate regulatory and procedural barriers to rapid AI deployment" and warned that "[f]ailing to address these energy requirements presents serious risks to America's technological leadership, as U.S. AI developers may be forced to relocate… overseas." [39] Anthropic highlighted that "authoritarian regimes . . . are already actively courting American AI companies." [40] Both Anthropic and OpenAI emphasized the need for government-industry partnership on security issues, proposing that federal agencies "develop robust capabilities to rapidly assess any powerful AI system"[41] and provide "classified threat intelligence to mitigate national security risks" from frontier models and state actors.[42]

Beyond policy submissions, AI companies have increasingly positioned themselves as essential to national security interests. Google has shifted its stance to align more closely with U.S. security priorities, framing AI advancement as vital for national security and stating that "democracies should lead" in the "global competition."[43] Google's 2018 AI principles excluded certain weapons applications while expressly permitting other military work; its revised principles omit that categorical weapons exclusion. [44] Similarly, at a

---

[39] *Anthropic's Response*, *supra* note 31, at 7–8.
[40] *Id.* at 7.
[41] *Id.* at 3.
[42] *OpenAI's Response*, *supra* note 6, at 6.
[43] James Manyika & Demis Hassabis, *Responsible AI: Our 2024 Report and Ongoing Work,* GOOGLE BLOG (Feb. 4, 2025), https://blog.google/technology/ai/responsible-ai-2024-report-ongoing-work/ [https://perma.cc/9VKH-MTNH].
[44] Sundar Pichai, *AI at Google: Our Principles*, GOOGLE (June 7, 2018), https://blog.google/technology/ai/ai-principles/ [https://perma.cc/TLJ9-KVFF] (listing under "AI applications we will not pursue" both weapons and "technologies whose principal purpose or implementation is to cause or directly facilitate injury to people," while stating that Google would "continue our work with governments and the military in many other areas"); Maxwell Zeff, *Google Removes Pledge to Not Use AI for Weapons from Website*, TECHCRUNCH (Feb. 4, 2025), https://techcrunch.com/2025/02/04/google-removes-pledge-to-not-use-ai-for-weapons-from-website/ [https://perma.cc/4ENK-RHYK]; *Our AI Principles*, GOOGLE AI, https://ai.google/principles/ [https://perma.cc/X3AW-TSR3] (last visited Oct. 22, 2026); see Manyika & Hassabis, supra note 43.

Senate hearing, NVIDIA chief scientist William Dally cautioned that preserving U.S. leadership requires letting "individual academics and entrepreneurs with a good idea" innovate and deploy models without huge barriers.[45] Facebook founder Mark Zuckerberg prepared talking points for his 2018 Senate testimony describing "[U.S.] tech companies [as a] key asset for America" and warning that breaking up Facebook would "strengthen[] Chinese companies," casting Big Tech as the nation's champion in its competition with China.[46]

Anthropic's leadership has similarly framed AI safety in competitive terms. CEO Dario Amodei warned Congress of "extraordinarily grave threats to [U.S. national security] over the next 2 to 3 years" from AI, while arguing that pausing U.S. AI development would only "hand over its power . . . to adversaries who do not share our values."[47] This characterization, revisited below, presents AI development as a race to manage severe risks before adversaries can exploit them: a framing that encourages acceleration rather than regulation.[48]

Lawmakers have echoed this rationale in congressional debates on technology funding. Members of Congress have pressed the same argument directly. Chairing a Senate Commerce Committee hearing styled "Winning the AI Race," Senator Cruz contrasted a deliberately "light-touch" American approach with "the command-and-control policies of Europe" and warned that the United States "cannot allow regulation, even the supposedly benign kind, to choke innovation or

---

[45] *Oversight of A.I.: Legislating on Artificial Intelligence: Hearing Before the Subcomm. on Priv.*, *supra* note 14 (statement of William Dally, Chief Scientist & Senior Vice President of Rsch., NVIDIA Corp.), https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-legislating-on-artificial-intelligence [https://perma.cc/FMM9-VPYU].

[46] Nick Statt, *Read Mark Zuckerberg's Notes from Today's Facebook Privacy Senate Hearing*, THE VERGE (Apr. 10, 2018), https://www.theverge.com/2018/4/10/17222546/facebook-mark-zuckerberg-senate-hearing-notes-cambridge-analytica-privacy [https://perma.cc/25NV-HVGA].

[47] *Amodei Testimony*, *supra* note 14, at 1.

[48] Sam Meacham, *A Race to Extinction: How Great Power Competition Is Making Artificial Intelligence Existentially Dangerous*, HARV. INT'L REV. (Sep. 8, 2023), https://hir.harvard.edu/a-race-to-extinction-how-great-power-competition-is-making-artificial-intelligence-existentially-dangerous/ [https://perma.cc/52XQ-DVVY].

adoption."[49] As one House Republican put it bluntly during a recent hearing, "we are in an arms race with China for artificial intelligence." [50] This arms race framing has fostered a deregulation sacrifice in domestic policy. Both parties agree that AI policy must prioritize innovation and prevent China from dominating the field, with Republicans warning against regulatory overreach and Democrats emphasizing that well-designed guardrails can promote rather than stifle U.S. leadership.[51] At a 2023 Senate hearing, the tech industry's leading trade group warned that excessive regulation would "undermine" U.S. AI leadership and simply hand an advantage to "authoritarian nations."[52]

From the government and civil society to Republicans and Democrats, then, the prevailing sentiment is that the United States can't afford to lose the AI race—so any regulation must be carefully calibrated not to hamper the nation's ability to

---

[49] *See Press Release, S. Comm. on Com., Sci., & Transp., Sen. Cruz: Adopting Europe's Approach on Regulation Will Cause China to Win the AI Race* (May 8, 2025), https://www.commerce.senate.gov/2025/5/sen-cruz-adopting-europe-s-approach-on-regulation-will-cause-china-to-win-the-ai-race [https://perma.cc/T47R-8BD4] (announcement of Sen. Ted Cruz, Chairman, S. Comm. on Com., Sci., & Transp. at the hearing Winning the AI Race: Strengthening U.S. Capabilities in Computing and Innovation).

[50] *America's AI Moonshot: The Economics of AI, Data Centers, and Power Consumption: Hearing Before the H. Comm. on Oversight & Gov't Reform*, 119th Cong. 15 (2025) (statement of Rep. Gary Palmer, Member, H. Subcomm. on Econ. Growth, Energy Pol'y & Regul. Affs.), https://www.congress.gov/event/119th-congress/house-event/118078/text [https://perma.cc/7GYV-ACZL].

[51] *Winning the AI Race*, *supra* note 1 (Senator Cruz claimed, "The way to beat China in the AI race is to outrace them in innovation, not saddle AI developers with European-style regulations."); *Majority Leader Schumer Delivers Remarks to Launch SAFE Innovation Framework for Artificial Intelligence at CSIS*, SENATE DEMOCRATS (June 21, 2023), https://www.democrats.senate.gov/news/press-releases/majority-leader-schumer-delivers-remarks-to-launch-safe-innovation-framework-for-artificial-intelligence-at-csis [https://perma.cc/2F4Y-SFS6] (". . . Innovation must be our north star . . . encouraging, not stifling, innovation," and warning that the Chinese Communist Party "could leap ahead . . . and set the rules of the game for AI.").

[52] Faiza Patel & Melanie Geller, *Senate AI Hearings Highlight Increased Need for Regulation*, BRENNAN CTR. FOR JUST. (Oct. 13, 2023), https://www.brennancenter.org/our-work/analysis-opinion/senate-ai-hearings-highlight-increased-need-regulation [https://perma.cc/ZK93-EZJW].

outcompete China.[53] The result is a reluctance to impose strict domestic constraints absent comparable commitments by adversaries. In short, fear of falling behind has become a powerful political argument against AI regulation.

A comparable, *though not identical*, "regulation-sacrifice" dynamic operates in Beijing.[54] The Chinese variant, however, differs in kind from its American counterpart. Where U.S. proponents resist binding rules outright—through rollback, preemption, and voluntary substitutes—Beijing's retreat is narrower: it eases procedural obstacles that slow deployment while keeping firm control over content and security. Party leaders frame AI as a zero-sum contest for national survival and then pare back any restraint that might slow deployment. Presiding over the Politburo's twentieth collective study session on April 25, 2025, President Xi Jinping called AI "a strategic technology that is leading the new round of scientific and technological revolution," urged cadres to "master[] core technologies such as high-end microchips and foundational software," and warned that the Party must prepare for the technology's "unprecedented risks and challenges."[55] His

---

[53] *See, e.g.*, *Remarks by President Biden in a Farewell Address to the Nation*, THE WHITE HOUSE (Jan. 15, 2025), https://bidenwhitehouse.archives.gov/briefing-room/speeches-remarks/2025/01/15/remarks-by-president-biden-in-a-farewell-address-to-the-nation/ [https://perma.cc/A7ZG-BCQ3] (warning that "unless safeguards are in place, AI could spawn new threats," while insisting that "America — not China — must lead the world on the development of AI"); Cristiano Lima-Strong, *Transcript: Sam Altman Testifies at US Senate Hearing on AI Competitiveness*, TECH POL'Y PRESS (May 8, 2025), https://www.techpolicy.press/transcript-sam-altman-testifies-at-us-senate-hearing-on-ai-competitiveness/ [https://perma.cc/8XTY-SQ4C] (reflecting bipartisan "AI race" framing—e.g., "America has to beat China in the AI race" (Cruz) and "we want to win against China" (Cantwell)—while urging "sensible regulation that does not slow us down" (Altman)).

[54] Because internal policy deliberations in Beijing remain opaque, scholars must infer intent from public messaging and successive rule revisions rather than from open hearings or comment dockets—making the evidentiary record necessarily more circumstantial than in the United States.

[55] *Xi Jinping Zai Zhonggong Zhongyang Zhengzhi Ju Di Ershi Ci Jiti Xuexi Shi Qiangdiao Jianchi Zili Ziqiang Tuchu Yingyong Daoxiang Tuidong Rengong Zhineng Jiankang You Xu Fazhan (习近平在中共中央政治局第二十次集体学习时强调 坚持自立自强 突出应用导向 推动人工智能健康有序发展)*, XINHUA NEWS AGENCY (Apr. 26, 2025), https://www.news.cn/politics/leaders/20250426/63f5cde8f4b54e22ba35aa7ec

battlefield rhetoric is not new. At an earlier study session on October 31, 2018, he ordered China to "occup[y] the high ground in critical and core technologies."[56]

Other senior figures echo the same theme. Addressing the U.N. Security Council's debate on AI, Ambassador Zhang Jun cast the field in explicitly geopolitical terms, accusing "a certain developed country" of pursuing "technological hegemony," "build[ing] exclusive small clubs," "maliciously obstruct[ing] the technological development of other countries," and "artificially creat[ing] technological barriers." He paired that competition framing with a military warning—stressing that "AI in the military field may lead to major changes in the way of warfare and the format of war," and urging states to "oppose the use of AI to seek military hegemony." [57] Reinforcing the point, Tsinghua University computer scientist Wen Gao wrote in *People's Daily* that AI is the "strategic commanding height for competition among major powers."[58]

7884b3a/c.html [https://perma.cc/5UBL-GCVG] (*translated in At the 20th Collective Study Session of the CCP Central Committee Politburo, Xi Jinping Stresses: Persist in Being Self-Reliant, Be Strongly Oriented Toward Applications, and Push the Orderly Development of Artificial Intelligence*, CTR. FOR SEC. & EMERGING TECH. (Apr. 28, 2025), https://cset.georgetown.edu/publication/xi-politburo-collective-study-ai-2025/ [https://perma.cc/XTQ8-QQJ3]).

[56] Xi Jinping (习近平), *Xi Jinping: Tuidong Woguo Xin Yidai Rengong Zhineng Jiankang Fazhan (习近平：推动我国新一代人工智能健康发展)*, XINHUA NEWS AGENCY (Oct. 31, 2018), http://www.xinhuanet.com/politics/2018-10/31/c_1123643321.htm [https://perma.cc/T595-J24Q] *translated in* Rogier Creemers & Elsa B. Kania, *Translation: Xi Jinping Calls for 'Healthy Development' of AI*, DIGICHINA (Nov. 5, 2018), https://digichina.stanford.edu/work/xi-jinping-calls-for-healthy-development-of-ai-translation/ [https://perma.cc/YB5E-LX44].

[57] *Remarks by Ambassador Zhang Jun at the U.N. Security Council Briefing on Artificial Intelligence: Opportunities and Risks for International Peace and Security*, PERMANENT MISSION OF THE PEOPLE'S REPUBLIC OF CHINA TO THE U.N. (July 18, 2023), https://un.china-mission.gov.cn/eng/chinaandun/securitycouncil/202307/t20230719_11114947.htm [https://perma.cc/9EHV-CJXQ].

[58] Gao Wen (高文), *Qiangzhua Rengong, Zhineng Fazhan de Lishixing Jiyu* (抢抓人工智能发展历史性机遇) [*Seizing the Historic Opportunity for the Development of Artificial Intelligence*], ZHONGGUO GONGCHANDANG XINWEN WANG (中国共产党新闻网) [CPCNEWS] (Feb. 24, 2025), https://cpc.people.com.cn/n1/2025/0224/c64387-40424141.html

Scholars suggest that such security-race framing shapes Chinese regulation. The final July 2023 *Interim Measures for Generative-AI Services* retained strict content rules but scrapped the draft's blanket ex-ante licensing, changes Reuters judged "far less onerous" because Beijing "wants to support tech development."[59] Crucially, the Measures still mandate algorithm filings, security assessments, and adherence to "core socialist values"; the only substantive easings were the removal of blanket licensing and an unrealistic truth-accuracy clause.[60] Professor Angela Huyue Zhang likewise argues that China's "strategic[ally] lenient approach to regulation may therefore offer its AI firms a short-term competitive advantage over their European and U.S. counterparts," using permissive AI rules to give domestic firms an edge in the global race.[61]

Abroad, Beijing markets the same stance of "responsible AI": at the 2023 U.K. AI-Safety Summit, Wu Zhaohui, Vice President of the Chinese Academy of Sciences, insisted that every country enjoys an "equal right to develop and use AI," implicitly warning against restrictive regimes.[62] In short, China publicly champions responsible innovation while ensuring that oversight at home remains flexible enough to keep the nation sprinting for the AI commanding heights. While U.S. debates

---

[https://perma.cc/P9KB-V7VT].

[59] Josh Ye, *China Says Generative AI Rules to Apply Only to Products for the Public*, REUTERS (July 13, 2023), https://www.reuters.com/technology/china-issues-temporary-rules-generative-ai-services-2023-07-13/ [https://perma.cc/LT4C-MP64].

[60] Shengchengshi Rengong Zhineng Fuwu Guanli Zanxing Banfa (生成式人工智能服务管理暂行办法) [Interim Measures for the Management of Generative Artificial Intelligence Services] arts. 4(1), 17 (promulgated by the Cyberspace Admin. of China, July 10, 2023, effective Aug. 15, 2023) St. Council Gaz., Aug. 30, 2023, at 40, https://www.gov.cn/gongbao/2023/issue_10666/202308/content_6900864.html [https://perma.cc/MZ8P-AAYZ] (China).

[61] Angela Huyue Zhang, *The Promise and Perils of China's Regulation of Artificial Intelligence*, 63 COLUM. J. TRANSNAT'L L. 1, 1 (2025).

[62] *China Willing to Enhance Communication with All Sides on AI Regulation—Vice Minister*, REUTERS (Nov. 1, 2023) (Wu stated that "[c]ountries regardless of their size and scale have equal rights to develop and use AI" and called for "global cooperation" and open-source availability), https://www.reuters.com/technology/china-willing-enhance-communication-with-all-sides-ai-regulation-vice-minister-2023-11-01/ [https://perma.cc/WY2C-6JTJ].

revolve around preserving private-sector agility, China's adjustments are made within a one-party system that still prioritizes ideological control. Thus, its deregulatory moves are narrower and principally aimed at shielding development speed from new procedural hurdles rather than relaxing existing content and security mandates.[63] That distinction matters for what the China example can prove. U.S. policymakers who invoke China to justify deregulation invoke a country that has not abandoned regulation. Part II.B develops the point in detail.

The arms race rationale also undercuts momentum for robust international coordination. A race mindset makes each major power wary of binding limits, lest rules restrain it more than its rival. Washington rejects proposals for a global AI "pause" or strict multilateral caps on frontier models, arguing that authoritarian competitors would simply ignore them; U.S. officials instead back export controls on chips and voluntary pacts among allies.[64] Beijing reciprocates. While its Position Paper on Global AI Governance endorses general principles of responsible development, it warns against any framework that might contain China's AI progress, signaling a red line against enforceable ceilings.[65]

Consequently, multilateral initiatives remain largely procedural and nonbinding. The 2023 Bletchley Declaration on AI Safety calls for international cooperation on frontier AI risks and anticipates a further meeting in 2024.[66] The G-7

---

[63] *See supra* notes 59–60 and accompanying text.

[64] Press Release, BUREAU OF INDUS. & SEC., *Commerce Strengthens Export Controls to Restrict China's Capability to Produce Advanced Semiconductors for Military Applications* (Dec. 2, 2024), https://www.bis.gov/press-release/commerce-strengthens-export-controls-restrict-chinas-capability-produce-advanced-semiconductors-military [https://perma.cc/PNX9-GPH2].

[65] *Global AI Governance Initiative*, MINISTRY OF FOREIGN AFFS. OF THE PEOPLE'S REPUBLIC OF CHINA (Oct. 20, 2023), https://www.mfa.gov.cn/eng/zy/gb/202405/t20240531_11367503.html [https://perma.cc/D894-C7SY].

[66] *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023*, GOV.UK (updated Feb. 13, 2025) [hereinafter *The Bletchley Declaration*], https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 [https://perma.cc/G7A7-943Y].

Hiroshima Process adopts the same logic, as its Advanced AI Model Risk Reports are voluntary so that no country inhibits its own AI competitiveness.[67] Both documents exemplify how great power rivalry converts ambitious governance proposals into minimalist pledges. As a result, Washington and Beijing engage in high-profile dialogue while quietly prioritizing national strategies and seeking to outdo each other. International coordination thus suffers the same regulation-sacrifice dynamic that shapes domestic oversight: everyone professes support for "responsible AI," but the arms race frame ensures that concrete constraints are deferred, diluted, or displaced.

In short, the pervasive framing of AI as a zero-sum race with national security implications has bred a policy impasse. Every major player professes support for "responsible AI" globally, but meaningful coordination falters as each strives for the lead.[68] The arms race frame thus drives a cycle of competitive deregulation at home and cautious, weak engagement abroad—a dynamic termed here the Regulation Sacrifice. This prompts the question this Article seeks to answer: is sacrificing regulation for victory in the "AI arms race" a sound strategic choice, one likely to deliver significant national security gains?

II. The Assumptions of the AI Regulation Sacrifice

Part I traced how the arms race frame has generated a policy reflex that treats regulation as a strategic liability. The distilled doctrine, which this Article calls the Regulation

---

[67] *OECD Launches Global Framework to Monitor Application of G7 Hiroshima AI Code of Conduct*, ORG. FOR ECON. CO-OP. & DEV. (Feb. 7, 2025), https://www.oecd.org/en/about/news/press-releases/2025/02/oecd-launches-global-framework-to-monitor-application-of-g7-hiroshima-ai-code-of-conduct.html [https://perma.cc/9RL6-KWD6].

[68] *Compare The Bletchley Declaration*, *supra* note 66, *with Fact Sheet: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence*, THE WHITE HOUSE (Oct. 30, 2023), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/ [https://perma.cc/7CGN-CVXW], and *Global AI Governance Initiative*, *supra* note 65.

Sacrifice, can be stated succinctly: In a geopolitical sprint for frontier AI, every month of additional speed is worth more to national security than any risk reduction achieved through near-term oversight. Therefore, states should defer or dilute regulation until after they have secured an unassailable lead. In effect, policymakers wager that they can win first and tidy up later.

For that wager to be plausible, three implicit premises must all hold:

Durable Lead Assumption. A temporary burst of deregulated speed can be turned into a long-lasting technological gap that rivals cannot quickly erase or leap-frog.

Low Drag Assumption. Governance measures, such as licensing, red-teaming mandates, and disclosure rules, impose costs or delays large enough to matter in the race; stripping them away yields a meaningful acceleration dividend.

Net Security Benefit Assumption. Achieving the frontier first confers security advantages that exceed the new vulnerabilities created by rapid, lightly governed deployment.

Let us interrogate each premise in turn, asking whether the empirical record and technical realities justify the confidence the Regulation Sacrifice requires.

A. Durable Lead Assumption

The Regulation Sacrifice depends on a foundational premise: that early advantages in AI development can be converted into a lasting geopolitical edge. This assumption permeates current policy discourse. The Trump Administration's Executive Order 14179 exemplifies this thinking, pledging to "sustain and enhance America's global AI dominance."[69] Google's submission to the OSTP reinforces this narrative, asserting that "we are in a global AI competition, and policy decisions will determine the outcome. A pro-innovation approach that protects national security and ensures that everyone benefits from AI is essential to realizing AI's transformative potential and ensuring that America's lead endures."[70]

Proponents of rapid, unregulated development typically

---

[69] Exec. Order No. 14,179, § 1, 90 Fed. Reg. 8741 (January 31, 2025).

[70] *Google's Response*, *supra* note 6.

invoke durability mechanisms. First, they point to scale advantages and capital barriers—frontier development now requires multibillion-dollar data centers, with one internal industry analysis putting a five-gigawatt facility at $100 billion, an outlay that few competitors can match.[71] Second, they emphasize hardware-based technological gaps maintained through export controls on advanced semiconductors.[72] Third, they cite platform integration and the costs of migrating AI models embedded in existing platforms to new ones.[73]

These mechanisms share a common thread: they assume temporary speed advantages translate into permanent strategic positions. Without this durability, the logic of sacrificing safety for speed collapses. If advantages prove ephemeral, then we must ask whether a technological head start of even a year or two justifies abandoning protective oversight—a question whose answer depends critically on whether racing enhances security, as Part III will examine below.

---

[71] Sebastian Moss, *Here's the 5GW, 30 Million Sq Ft Data Center Pitch Doc OpenAI Showed the White House*, DATA CENTER DYNAMICS (Nov. 6, 2024), https://www.datacenterdynamics.com/en/news/heres-the-5gw-30-million-sq-ft-data-center-pitch-doc-openai-showed-the-white-house/ [https://perma.cc/E78H-EYXN] (reporting OpenAI's internal analysis that a 5GW AI data center "could cost $100 billion").

[72] *See* Jake Sullivan, *Remarks by National Security Advisor Jake Sullivan at the Special Competitive Studies Project Global Emerging Technologies Summit*, WHITE HOUSE (Sept. 16, 2022) (abandoning the prior "sliding scale" approach of staying "only a couple of generations ahead" and stating that in foundational technologies the United States "must maintain as large of a lead as possible"), https://bidenwhitehouse.archives.gov/briefing-room/speeches-remarks/2022/09/16/remarks-by-national-security-advisor-jake-sullivan-at-the-special-competitive-studies-project-global-emerging-technologies-summit/ [https://perma.cc/C3GF-TNAG].

[73] Colton Stradling, *Office365 Tops 400 Million Paid Users. Microsoft Credits Copilot AI for Office Growth*, YAHOO! TECH (Jan. 31, 2024), https://tech.yahoo.com/ai/articles/office365-tops-400-million-paid-204853643.html [https://perma.cc/P39C-WN5V] (reporting over 400 million Office 365 seats); Lina M. Khan, *The Separation of Platforms and Commerce*, 119 COLUM. L. REV. 973, 1013 (2019) (documenting how dominant platforms' discriminatory conduct and "self-privileging" can "corrupt[] the entire system of platform-based innovation"); Julie E. Cohen, *Law for the Platform Economy*, 51 U.C. Davis L. Rev. 133, 156–61 (2017) (explaining how dominant platforms leverage protocol-based control, preferential placement, and interplatform affiliation to make participation "sticky" and translate scale into durable competitive advantage).

Recent data reveals a stark reality: AI advantages seem to erode far more rapidly than the arms race narrative acknowledges. The most compelling evidence comes from direct measurement of capability gaps. Stanford's 2025 AI Index reports rapid U.S.–China convergence at the frontier: on the LMSYS Chatbot Arena, the top U.S. model's lead over the best Chinese model shrank from 9.3% in January 2024 to 1.7% by February 2025.[74] On standard benchmarks, gaps also narrowed sharply from end-2023 to end-2024—for example, MMLU fell from 17.5 to 0.3 percentage points and MATH from 24.3 to 1.6 percentage points.[75]

This rapid convergence occurs despite substantial U.S. advantages in both compute capacity and investment. The United States possesses approximately ten times more AI computing infrastructure than China and attracted $109 billion in private AI investment in 2024 compared to China's $9.3 billion.[76] Yet these resource advantages have not translated into durable capability gaps. The reason lies in the unique characteristics of AI technology that accelerate diffusion through multiple channels.

First, and perhaps most important, is the dramatic improvement in computational efficiency that has sharply accelerated AI development. Inference costs fell 280-fold between November 2022 and October 2024, from $20 to $0.07 per million tokens.[77] Training costs show similar declines. Achieving GPT-3 level performance cost $4.6 million in 2020 but only $450,000 by 2022, a 70% annual reduction.[78]

These efficiency gains enable competitors to achieve

---

[74] MASLEJ ET AL., *supra* note 12, at 97 fig. 2.1.36 and accompanying text.

[75] MASLEJ ET AL., *supra* note 12, at 97 fig. 2.1.37.

[76] Lennart Heim, *China's AI Models Are Closing the Gap—but America's Real Advantage Lies Elsewhere*, RAND (May 2, 2025), https://www.rand.org/pubs/commentary/2025/05/chinas-ai-models-are-closing-the-gap-but-americas-real.html [https://perma.cc/4J3U-5ECL] (highlighting that the United States has a greater compute capacity than China through owning more advanced AI chips); MASLEJ ET AL., *supra* note 12, at 3.

[77] MASLEJ ET AL., *supra* note 12, at 64.

[78] Haziqa Sajid, *AI Training Costs Continue to Plummet*, UNITE.AI (Mar. 16, 2023), https://www.unite.ai/ai-training-costs-continue-to-plummet/ [https://perma.cc/792K-NETS] (providing an overview of recent AI training costs trends over time to illustrate barriers to AI adoption).

frontier capabilities with dramatically fewer resources. DeepSeek's January 2025 breakthrough exemplifies this dynamic: DeepSeek's V3 technical report priced its final training run at roughly $5.6 million—2.788 million H800 graphics-processor hours at an assumed $2 per hour, excluding the research and ablation experiments behind it. Its R1 reasoning model, built on V3 and released weeks later, posted benchmark results in the range of OpenAI's o1, though the $294,000 DeepSeek reports for R1 covers only the reinforcement-learning stages built on top of V3, not the pretraining underneath them.[79] These figures are self-reported, exclude much of what Western development budgets count, and were disputed within days by industry figures who alleged a far larger undisclosed chip fleet.[80] Even allowing for all of that, the comparison points one way: on the figures DeepSeek has published, a model in the frontier's range came out of a

---

[79] DeepSeek-AI et al., *DeepSeek-V3 Technical Report* 5 tbl. 1 (Feb. 18, 2025) (unpublished manuscript), https://arxiv.org/pdf/2412.19437v2 [https://perma.cc/CHE6-VDBH]; Daya Guo et al., *DeepSeek-R1 Incentivizes Reasoning in LLMs Through Reinforcement Learning*, 645 NATURE 633 (2025), https://doi.org/10.1038/s41586-025-09422-z; *see id.* supplementary information § 2.4.4 & tbl. 4 (reporting 147,000 H800 processor-hours in all — 101,000 for the preliminary R1-Zero model, 5,000 to create the supervised fine-tuning data, and 41,000 for R1 itself — or $294,000 at an assumed rental price of $2 per processor-hour, covering the reinforcement-learning stages alone and not the pretraining of DeepSeek-V3-Base on which R1 is built).

[80] Greg McKenna, *Everyone Is Freaking Out About Chinese AI Startup DeepSeek. Are Its Claims Too Good to Be True?*, FORTUNE (Jan. 27, 2025), https://fortune.com/2025/01/27/china-deepseek-ai-claims-true/ [https://perma.cc/JZ9U-QJYE] (reporting Scale AI chief executive Alexandr Wang's statement to CNBC that "DeepSeek has about 50,000 H100s, which they can't talk about, obviously, because it is against the export controls that the United States has put in place"); Dylan Patel et al., *DeepSeek Debates: Chinese Leadership on Cost, True Training Cost, Closed Model Margin Impacts*, SEMIANALYSIS (Jan. 31, 2025), https://newsletter.semianalysis.com/p/deepseek-debates [https://perma.cc/AB8G-B2PH] (estimating access to roughly 50,000 Hopper-generation processors of mixed types and total server capital expenditure of approximately $1.6 billion, and arguing that the reported figure covers "just the GPU cost of the pre-training run"). Neither account has been substantiated as to the training run described in the technical report. *See* Charles Mok, *Taking Stock of the DeepSeek Shock*, Stan. Cyber Pol'y Ctr. (Feb. 5, 2025), https://cyber.fsi.stanford.edu/publication/taking-stock-deepseek-shock [https://perma.cc/F3AE-2SJF].

training run costing single-digit millions — a figure for one training run, not for total development cost, which neither DeepSeek nor its American competitors disclose. And by DeepSeek's own account the work was done on H800s—the reduced-bandwidth chips designed to comply with the 2022 export rules—which is the point that matters for export-control policy. [81]

Second, the AI research ecosystem operates with unprecedented openness that facilitates rapid knowledge transfer. In 2023, 65.7% of foundation models were released as open-source, up from 44.4% in 2022.[82] Meta's LLaMA model leak in March 2023 sparked an entire ecosystem of derivative models, demonstrating how quickly controlled AI systems can escape containment.[83] Even proprietary systems face security challenges—OpenAI experienced a significant breach in 2023 that exposed internal AI development details, while multiple ChatGPT incidents affected hundreds of thousands of users.[84]

---

[81] The H800 was the reduced-bandwidth processor designed for the Chinese market under the October 2022 controls, *see* Implementation of Additional Export Controls: Certain Advanced Computing and Semiconductor Manufacturing Items; Supercomputer and Semiconductor End Use; Entity List Modification, 87 Fed. Reg. 62186 (Oct. 13, 2022), and was itself brought within the controls the following year, *see* Implementation of Additional Export Controls: Certain Advanced Computing Items; Supercomputer and Semiconductor End Use; Updates and Corrections, 88 Fed. Reg. 73458 (Oct. 25, 2023); NVIDIA Corp., Current Report (Form 8-K), Item 8.01 (Oct. 17, 2023), https://www.sec.gov/Archives/edgar/data/1045810/000104581023000217/nvda-20231017.htm (stating that the new rules apply to products "including but not limited to the A100, A800, H100, H800, L40, L40S, and RTX 4090"). For DeepSeek's own account of its hardware, see *DeepSeek-V3 Technical Report*, *supra* note 79, § 3.1 ("DeepSeek-V3 is trained on a cluster equipped with 2048 NVIDIA H800 GPUs.").

[82] Nestor Maslej et al., *supra* note 12 ("Of these newly released models, 65.7% were open-source, compared to only 44.4% in 2022 and 33.3% in 2021.").

[83] James Vincent, *Meta's Powerful AI Language Model Has Leaked Online—What Happens Now?*, THE VERGE (Mar. 8, 2023), https://www.theverge.com/2023/3/8/23629362/meta-ai-language-model-llama-leak-online-misuse [https://perma.cc/W9LG-M2SS] (reporting the LLaMA release as an open-source package and weighing concerns about the technology's potential for harm and its potential for greater access which could improve AI safety).

[84] *See, e.g.*, *OpenAI's Internal AI Details Stolen in 2023 Breach*, REUTERS (July 5, 2024), https://www.reuters.com/technology/cybersecurity/openais-

Beyond intentional releases and breaches, knowledge flows through human capital mobility. Research shows that as of 2022, forty-seven percent of top AI researchers had Chinese undergraduate backgrounds, with high turnover rates facilitating rapid knowledge transfer between organizations and countries.[85] While NDAs, trade secret law, and other employment contracts may limit the disclosure of specific information, they do not prevent the diffusion of tacit expertise, methods, and generalizable insights as researchers change jobs in a highly competitive global market.[86] This talent circulation ensures that breakthrough insights rarely remain confined to single institutions or nations.

---

internal-ai-details-stolen-2023-breach-nyt-reports-2024-07-05/ [https://perma.cc/XJU7-JSG5]; Press Release, Group-IB, *Group-IB Discovers 100K+ Compromised ChatGPT Accounts on Dark Web Marketplaces; Asia-Pacific Region Tops the List*, GROUP-IB (June 20, 2023), https://www.group-ib.com/media-center/press-releases/stealers-chatgpt-credentials/ [https://perma.cc/VM6X-8XFQ] (finding 101,134 credentials in stealer logs); Ravie Lakshmanan, *Over 225,000 Compromised ChatGPT Credentials Up for Sale on Dark Web Markets*, THE HACKER NEWS (Mar. 5, 2024), https://thehackernews.com/2024/03/over-225000-compromised-chatgpt.html [https://perma.cc/H3WK-W9WT] (noting that over 225,000 credentials were offered for sale by October 2023).

[85] *The Global AI Talent Tracker 2.0*, MACROPOLO, https://macropolo.org/digital-projects/the-global-ai-talent-tracker [https://perma.cc/7HC8-EZEL] (last visited Dec. 1, 2025); *AI Skills Gap Widens*, RANDSTAD (Nov. 12, 2024), https://www.randstad.com/press/2024/ai-skills-gap-widens/ [https://perma.cc/TY3Q-J3RH] (reporting that AI-skilled workers with one to five years' experience had a thirty-three percent job change rate in the prior year).

[86] *See* RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 41 (A.L.I. 1995) (noting that trade secret law imposes a duty of confidence on employees and other parties who receive information pursuant to a confidential relationship or contractual obligation, thereby restricting unauthorized disclosure); *see also* John Villasenor, *Artificial Intelligence, Trade Secrecy, and the Challenge of Transparency*, 25 N.C. J.L. & TECH. 495, 511–12 (2024) (explaining that trade secret law does not extend to employees' general knowledge, skill, and experience—*i.e.*, know-how that can travel with workers as they move between employers); Camilla A. Hrdy & Christopher B. Seaman, *Beyond Trade Secrecy: Confidentiality Agreements That Act Like Noncompetes*, 133 YALE L.J. 669, 694–99 (2024) (arguing that confidentiality agreements become problematic when they purport to reach beyond legitimately secret information and thereby function like non-compete agreements by chilling mobility and the use of general skills).

Third, the empirical record on semiconductor export controls reveals their limited effectiveness in maintaining durable advantages. While controls imposed immediate costs—Nvidia faced $8 billion in quarterly sales losses—they also triggered adaptive responses that may accelerate competition.[87] Chinese firms responded by stockpiling chips, finding cloud workarounds, and most significantly, accelerating domestic innovation.[88]

A comprehensive study of thirty semiconductor companies found no evidence that export controls hindered innovation among affected firms. Counterintuitively, impacted companies showed higher research spending and patent filings, suggesting that controls strengthened rather than weakened competitive capabilities. [89] Hardware constraints also forced Chinese developers to pioneer efficiency-driven approaches that achieve comparable results with dramatically fewer resources, creating competitive threats that may prove more difficult to

---

[87] Jack Whitney, Matthew Schleich & William Alan Reinsch, THE DOUBLE-EDGED SWORD OF SEMICONDUCTOR EXPORT CONTROLS 2–3 (Ctr. for Strategic & Int'l Stud., Oct. 2024) (explaining that Chinese "design-out" and "design-around" responses—alongside circumvention—threaten to make export controls "less effective as a longer-term barrier" and can weaken U.S. firms by reducing market access and revenue that funds R&D); Tova Donovan Levin, Sydney Hsu & Karen Hyunbee Cho, U.S. SENSITIVE TECHNOLOGY EXPORT CONTROLS TOWARD CHINA: LESSONS FROM THE SOVIET UNION 1–2 (Hoover Hist. Lab, Apr. 2023) (summarizing historical studies finding no "substantial evidence" that Cold War export controls produced significant impacts on Soviet technological development and cautioning that broad controls may be counterproductive or of uncertain efficacy); Arsheeya Bajwa & Stephen Nellis, *Nvidia Shares Rise as Sales Hit from China Export Curbs Not as Bad as Feared*, REUTERS (May 29, 2025), https://www.reuters.com/business/nvidia-forecasts-second-quarter-revenue-below-estimates-2025-05-28/ [https://perma.cc/RF7H-QDW4] (reporting that new U.S. export restrictions would "slice off $8 billion in sales" from Nvidia's then-current quarter).

[88] Gregory C. Allen, *China's New Strategy for Waging the Microchip Tech War*, CTR. FOR STRATEGIC & INT'L STUD. (May 3, 2023), https://www.csis.org/analysis/chinas-new-strategy-waging-microchip-tech-war [https://perma.cc/P6B2-AHG9].

[89] Andreas Schumacher, *Did U.S. Semiconductor Export Controls Harm Innovation?*, CTR. FOR STRATEGIC & INT'L STUD. (Nov. 5, 2024), https://www.csis.org/analysis/did-us-semiconductor-export-controls-harm-innovation [https://perma.cc/QNY7-8AC3] (finding no negative innovation impact across thirty firms).

counter than straightforward hardware superiority.[90]

The evidence converges on a clear conclusion: in AI, first-mover advantages will likely not last very long before competitors achieve functional parity.[91] This finding raises the critical question at the heart of the Regulation Sacrifice—is a short-term technological lead worth abandoning safety oversight?

The answer depends entirely on what that temporary advantage delivers for national security. If a year of AI superiority provided decisive strategic benefits—the ability to secure critical infrastructure, establish dominant standards, or achieve breakthroughs in defense capabilities—then the sacrifice might prove rational. But if racing merely hands adversaries the same dangerous capabilities months later while creating uncontrolled risks in the interim, then the bargain makes little strategic sense.

This temporal reality—advantages measured in months, not decades—fundamentally alters the cost-benefit calculation of the Regulation Sacrifice. Actors are not trading safety for long-term advantage, but rather permanent risks for temporary benefits. Whether even those temporary benefits enhance

---

[90] *See* DeepSeek-AI, DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model 1 (June 19, 2024) (unpublished manuscript), https://arxiv.org/pdf/2405.04434 [https://perma.cc/48SZ-GU9Z]; DeepSeek-AI, DeepSeek-V3 Technical Report 1 (Feb. 18, 2025) (unpublished manuscript), https://arxiv.org/pdf/2412.19437 [https://perma.cc/H8LK-SZZJ]; John Villasenor, *DeepSeek Shows the Limits of U.S. Export Controls on AI Chips*, BROOKINGS (Jan. 29, 2025), https://www.brookings.edu/articles/deepseek-shows-the-limits-of-us-export-controls-on-ai-chips/ [https://perma.cc/8L29-QQLD] (arguing that controls "may actually be accelerating" Chinese progress by forcing engineers to learn to use limited compute more efficiently).

[91] Leonard Dung & Max Hellrigel-Holderbaum, *Against Racing to AGI: Cooperation, Deterrence, and Catastrophic Risks* 15 (arXiv, Working Paper No. 2507.21839, 2025) ("[P]ushing the frontier forward is more difficult than catching up to the frontier . . . ."); *but see* Janet Egan, Paul Scharre & Vivek Chilukuri, *Promote and Protect America's AI Advantage*, CTR. FOR NEW AM. SEC. (Jan. 20, 2025), https://www.cnas.org/publications/commentary/promote-and-protect-americas-ai-advantage [https://perma.cc/77S8-XPD3] (arguing that "ensuring these breakthrough capabilities are developed in the United States is critical for both U.S. technology superiority" and advocating for policies to "enable continued AI dominance" and "maintain U.S. AI leadership").

security remains to be seen. As Part III demonstrates, the evidence suggests that racing without safeguards systematically undermines rather than enhances national security across every time horizon.

B. The Low Drag Assumption

The second pillar of the Regulation Sacrifice is the belief that safety oversight creates prohibitive friction on AI development, such that we must sacrifice protective measures to maintain competitive speed. This "low-drag" assumption pervades industry submissions and policy statements. OpenAI has warned that "regulatory arbitrage" will drive AI development to jurisdictions with minimal oversight, arguing that even modest compliance requirements could "undermine our national security" by slowing progress.[92]

Google echoes the low drag assumption, claiming that "[w]e are in a global AI competition, and policy decisions will determine the outcome" and that "for too long, AI policymaking has paid disproportionate attention to the risks, often ignoring the costs that misguided regulation can have on innovation, national competitiveness, and scientific leadership—a dynamic that is beginning to shift under the new Administration." It further insists that "a pro-innovation approach that protects national security and ensures that everyone benefits from AI is essential to realizing AI's transformative potential and ensuring that America's lead endures."[93]

The assumption operates through several proposed mechanisms. First, advocates argue that compliance costs divert resources from innovation. Every dollar spent on safety assessments, documentation, or regulatory approval processes is supposedly a dollar sacrificed from model development.[94]

---

[92] *OpenAI's Response*, *supra* note 6, at 4.

[93] *Google's Response*, *supra* note 6, at 2–3.

[94] CTR. FOR DATA INNOVATION, ARTIFICIAL INTELLIGENCE ACT 7 (2021) (contending that the AI Act will add substantial "overhead on all AI spending," including fixed costs such as compliance workflows and "conducting conformity assessment procedures," thereby producing opportunity costs in foregone investment); The Anh Han et al., *To Regulate or Not: A Social Dynamics Analysis of an Idealised AI Race*, 69 J. A.I. RSCH. 881, 884 (2020) (modeling the trade-off where safe developers "pay a cost $c > 0$" while "the speed of development of UNSAFE participants is faster").

For startups operating on venture funding, these costs allegedly prove fatal, forcing them to choose either regulatory compliance or competitive survival.[95]

Second, they emphasize time-to-market delays. In a field where capabilities double every six months—the same rapid pace that undermines durable advantages—regulatory review cycles measured in months or years purportedly guarantee obsolescence. [96] Industry submissions to the federal consultation on the AI Action Plan press the comparison directly: American developers face compliance obligations that Chinese competitors do not. The same argument runs inside government, where the White House adviser on artificial intelligence credited the rescission of the Biden executive order with removing a handicap that China had never accepted. [97] Yet this characterization fundamentally misrepresents reality: China has implemented binding generative AI rules since July 2023, requiring security assessments and content reviews before model deployment, with additional algorithmic governance measures added in February 2024.[98] As Sean Ó hÉigeartaigh documents, the 2023

---

[95] Foo Yun Chee, *Europe Wants to Lighten AI Compliance Burden for Startups*, REUTERS (Apr. 8, 2025), https://www.reuters.com/world/europe/europe-wants-lighten-ai-compliance-burden-startups-2025-04-08/ [https://perma.cc/Q6XU-T6H6] (reporting that the European Commission sought feedback to "minimise the potential compliance burden" of the EU AI Act, particularly for "smaller innovators," amid concerns about the volume and cost of compliance).

[96] *See supra* Section II.A.

[97] *See OpenAI's Response*, *supra* note 6, at 4, 6 (urging federal preemption of state frontier-model rules and arguing that such rules "could impose burdensome compliance requirements that may hinder our economic competitiveness" because they are "easier to enforce with domestic AI companies than PRC-based companies"); *Google's Response*, *supra* note 6, at 5 (urging "federal preemption of state-level laws that affect frontier AI models" to avoid "a fragmented regulatory environment that would slow the development of AI"); David Sacks (@DavidSacks), X (Jan. 27, 2025, 12:50 ET), https://x.com/DavidSacks/status/1883935713877782884 [https://perma.cc/9RHM-XMMF] (then White House adviser on artificial intelligence and cryptocurrency, arguing that "DeepSeek R1 shows that the AI race will be very competitive and that President Trump was right to rescind the Biden EO, which hamstrung American AI companies without asking whether China would do the same").

[98] Interim Measures for the Management of Generative Artificial

measures are "quite heavy-handed," and analysts inside China warn they may slow domestic firms *more* than Western counterparts.[99] Baidu's ERNIE-Bot, for example, was held back roughly six months while regulators completed the mandatory security assessment, a delay Chinese media attributed directly to the new rules.[100] The supposed "wild west" of Chinese AI development is largely a myth. The significance reaches back to Part I. The Regulation Sacrifice is justified, in part, as a necessary response to an unconstrained Chinese rival, yet the rival invoked to justify discarding oversight has retained its own. The claim that the United States must deregulate to keep pace with a rule-free China rests on a China that does not exist.

Third, they point to talent and capital flight, framing this too as a necessary sacrifice. Researchers and investors, the argument goes, will gravitate toward jurisdictions offering "regulatory clarity," often a euphemism for minimal oversight.[101] The specter of a 'brain drain' of founders and firms to the United States, Dubai, and other jurisdictions perceived as more hospitable to development features prominently in industry testimony, which describes an "exodus of capital and companies to the US."[102]

Intelligence Services 60arts. 4–17; Hulianwang Xinxi Fuwu Suanfa Tuijian Guanli Guiding (互联网信息服务算法推荐管理规定) [Provisions on the Administration of Algorithmic Recommendation in Internet Information Services], art. 7 (promulgated by the Cyberspace Admin. of China, Dec. 31, 2021, effective Mar. 1, 2022), ST. COUNCIL GAZ., Mar. 30, 2022, at 71, https://www.gov.cn/zhengce/zhengceku/2022-01/04/content_5666429.htm [https://perma.cc/XY8Q-Y7BV] (China).

[99] Ó hÉigeartaigh, *supra* note 7, at 5.

[100] *Id.*

[101] OPENAI, *UK AI and Copyright Consultation* 6, 12 (Apr. 2, 2025), https://cdn.openai.com/global-affairs/b89a7434-7cb9-47a7-b4a7-b50b1a1a0afc/openai-uk-ai-and-copyright-consultation.pdf [https://perma.cc/SS2H-9K9H] ("However, AI development is inherently international, with companies making strategic decisions on where to train models based on regulatory clarity, infrastructure cost, and ease of compliance . . . [i]nvestment into AI infrastructure will flow to jurisdictions with regulatory environments that encourage AI development. Talent will concentrate where AI development and model training take place.").

[102] *See, e.g.*, Tech Nation, *Written Evidence to the House of Commons Science, Innovation and Technology Committee* 3 (May 20, 2025) (U.K.), https://committees.parliament.uk/writtenevidence/141606/pdf/ [https://perma.cc/F97L-YYN3] (reporting a Tech Nation survey in which

Without accepting the low-drag premise—that regulation necessarily imposes crippling costs requiring us to sacrifice safety for speed—the entire Regulation Sacrifice argument loses its economic rationale. If oversight can be designed to enhance rather than hinder innovation, then we face no tragic tradeoff at all. Drawing on recent empirical work, Bradford argues that it is overly simplistic to blame Europe's innovation gap on the General Data Protection Regulation (GDPR) or the Digital Markets Act (DMA). While digital rights regulation can impose real compliance costs, especially for smaller firms, the transatlantic technology gap is driven far more by structural factors such as fragmented and shallow capital markets, punitive bankruptcy rules, restrictive immigration policy, and inflexible labor markets than by Europe's rights-driven digital regulation model.[103]

The academic literature reveals that the supposedly clear-cut sacrifice between innovation and regulation is largely illusory. Recent experience with AI governance frameworks provides particularly instructive evidence that we need not sacrifice safety for speed. For example, IBM reports that integrating cross-industry data-provenance standards into its AI data-governance program cut data-clearance processing time by fifty-eight percent for third-party data and sixty-two percent for internal data, while improving overall data quality.[104] This parallels the UK Financial Conduct Authority's Regulatory Sandbox model, which reduced time-to-market by forty percent compared to standard authorizations, while firms that completed successful testing received 6.6 times more follow-on investment than peers outside the sandbox—demonstrating that well-designed regulatory frameworks can

---

forty-three percent of U.K. founders said they were actively considering relocating their company's headquarters outside the U.K., and describing an "exodus of capital and companies to the US," with specific reference to constraints in AI/deep tech talent and scaleup capital); Commc'ns & Digit. Comm., H. of Lords, *Corrected Oral Evidence: Scaling Up: AI and Creative Tech*, Q. 23 & 24 (Nov. 19, 2024) (U.K.), https://committees.parliament.uk/oralevidence/15029/pdf/ [https://perma.cc/7U7K-EBEC].

[103] The False Choice, *supra* note 2, at 401–02, 419–29, 440–44.

[104] IBM INST. FOR BUS. VALUE, THE ENTERPRISE GUIDE TO AI GOVERNANCE (2024), https://www.ibm.com/thought-leadership/institute-business-value/en-us/report/ai-governance [https://perma.cc/BW8E-P7CS].

accelerate rather than impede innovation.[105]

Bradford explains that the DMA's goal is to "enhance the contestability of the marketplace so that new firms can enter and compete in the marketplace" and that, while it introduces tradeoffs, it "has the potential to enhance the 'diversity' of innovation that takes place" by augmenting rivals' and new entrants' incentives to innovate and, in turn, incumbents' incentives to respond.[106] By analogy, common ex ante rulebooks—such as the DMA in competition law or the National Institute of Standards and Technology's (NIST) AI Risk Management Framework in AI governance—can lower entry barriers and diversify the loci of innovation, prompting challengers and incumbents alike to intensify research and development. China's own sandbox-style filing portals, set up by the Cyberspace Administration to pre-screen large models, show that regulators can add oversight without halting deployment; a first cohort of leading developers was cleared to launch publicly about two weeks after the Interim Measures took effect, and more than forty models had been approved within six months.[107]

---

[105] *AI Regulatory Sandbox Approaches: EU Member State Overview*, EU ARTIFICIAL INTELLIGENCE ACT (May 2, 2025), https://artificialintelligenceact.eu/ai-regulatory-sandbox-approaches-eu-member-state-overview/ [https://perma.cc/DTJ9-6H8M] (reporting that "the UK FCA sandbox reduced the average time required for market authorisation by 40% compared to the regulator's typical approval process" and that "companies that completed successful testing within the UK FCA sandbox received 6.6 times more fintech investment than their peers").

[106] The False Choice, *supra* note 2, at 377, 414–16 (explaining that the DMA's goal is "to enhance the contestability of the marketplace so that new firms can enter and compete" and that it has "the potential to enhance the 'diversity' of innovation that takes place").

[107] Josh Ye, *China Approves Over 40 AI Models for Public Use in Past Six Months*, REUTERS (Jan. 29, 2024), https://www.reuters.com/technology/china-approves-over-40-ai-models-public-use-past-six-months-2024-01-29/ [https://perma.cc/C89S-JCCL]; TechNode Feed, *China Approves 346 Generative AI Services Under National Registration Scheme*, TECHNODE (Apr. 9, 2025), https://technode.com/2025/04/09/china-approves-346-generative-ai-services-under-national-registration-scheme/ [https://perma.cc/4HTW-EPF5] (reporting that 346 generative AI services had completed registration by March 2025); Yirong Sun & Jingxian Zeng, *China's Interim Measures for the Management of Generative AI Services: A Comparison Between the Final and Draft Versions of the Text*, FUTURE OF PRIV. F. (Apr.

The financial sector offers compelling precedent for regulation without sacrifice. The United Kingdom's Financial Conduct Authority pioneered the regulatory sandbox model in 2015, creating a controlled environment for testing innovative financial products. Rather than forcing firms to sacrifice innovation for compliance, participating firms saw funding increase by fifteen percent and were fifty percent more likely to raise capital successfully.[108] The model has now spread to over fifty countries, with particularly strong benefits for smaller firms that previously faced information asymmetries with regulators.[109]

Environmental regulation provides perhaps the strongest counterevidence to the sacrifice narrative. A meta-analysis of 103 publications finds no systematic evidence that environmental regulation harms competitiveness: the pooled effect is statistically indistinguishable from zero, and at the national level positive effects outnumber negative ones.[110] California's zero-emission vehicle mandate, rather than forcing automakers to sacrifice competitiveness, created the market that financed the industry's most successful entrant. Tesla's first profitable year, 2020, produced net income attributable to

---

22, 2024), https://fpf.org/blog/chinas-interim-measures-for-the-management-of-generative-ai-services-a-comparison-between-the-final-and-draft-versions-of-the-text/ [https://perma.cc/PDZ5-RUBF] (noting that "by September 4, 2023 (less than a month after the Measures came into force), it was reported that eleven companies had completed algorithmic filings and 'received approval' to provide their generative AI services to the public").

[108] Giulio Cornelli et al., *Regulatory Sandboxes and Fintech Funding: Evidence from the UK*, 28 REV. FIN. 203, 205–06, 208–09 (2024) (documenting a fifteen percent funding increase for sandbox participants).

[109] *Id.* at 204–05; Ouarda Merrouche et al., *Digging for Gold: How Regulatory Sandboxes Help Fintechs Raise Funding*, VOXEU (Feb. 2, 2021), https://cepr.org/voxeu/columns/digging-gold-how-regulatory-sandboxes-help-fintechs-raise-funding [https://perma.cc/4K8E-8Q5Y].

[110] Mark A. Cohen & Adeline Tubb, *The Impact of Environmental Regulation on Firm and Country Competitiveness: A Meta-Analysis of the Porter Hypothesis*, 5 J. ASS'N ENV'T & RES. ECONOMISTS 371 (2018) (meta-analysis of 103 publications, treated as 107 studies, coding 2,025 effect sizes; finding that across the authors' preferred specifications the largest category of estimates is statistically insignificant and that the pooled effect "cannot reject the hypothesis that the underlying effect size is zero," while at the state, regional and national level positive and significant estimates outnumber negative ones).

common stockholders of $721 million against roughly $1.58 billion in revenue from selling regulatory credits to rivals that had missed their targets — credits earned under California's mandate and under the separate greenhouse-gas, fuel-economy and clean-fuel programs that also generate them — without which the company would have reported a loss. It has earned close to $9 billion from such sales since 2009.[111] The Porter Hypothesis—that well-designed environmental standards trigger innovation that more than offsets compliance costs—suggests that the sacrifice frame fundamentally misunderstands how regulation and innovation interact.[112]

Even pharmaceutical regulation, often cited as requiring enormous sacrifices of time and resources, tells a different story upon closer examination. Following statutory "modernization" through the Prescription Drug User Fee Act of 1992 (PDUFA) and subsequent reforms culminating in the Food and Drug Administration Safety and Innovation Act of 2012 (FDASIA), the median FDA review time for new drugs fell from 26.6 months in the pre-PDUFA period to about 9.9 months in the FDASIA-2022 period, even as the total number of approvals rose. More than half of new drug approvals now benefit from expedited programs or special regulatory designations

---

[111] Tesla, Inc., Annual Report (Form 10-K) 55, 61 (Feb. 8, 2021), https://www.sec.gov/Archives/edgar/data/1318605/000156459021004599/tsla-10k_20201231.htm [https://perma.cc/8K2G-B7D2] (reporting 2020 net income attributable to common stockholders of $721 million and automotive regulatory credit revenue of $1.580 billion, and stating at 62 that the company earns these credits "in the operation of our automotive business under various regulations related to zero-emission vehicles, greenhouse gas, fuel economy and clean fuel," without disaggregating the revenue by program); *see* Craig Trudell, *Tesla Rakes In $9 Billion from Carmakers Failing to Sell Enough EVs*, Bloomberg (Feb. 9, 2024), https://www.bloomberg.com/news/articles/2024-02-09/tesla-rakes-in-9-billion-from-carmakers-failing-to-sell-enough-evs [https://perma.cc/R5YV-PES8] (reporting cumulative regulatory credit revenue of nearly $9 billion since 2009, including $1.79 billion in 2023). On the mandate itself, see CAL. AIR RES. BD., PUBLIC HEARING TO CONSIDER THE PROPOSED ADVANCED CLEAN CARS II REGULATIONS: STAFF REPORT, INITIAL STATEMENT OF REASONS 6–8 (2022), https://ww2.arb.ca.gov/sites/default/files/barcu/regact/2022/accii/isor.pdf [https://perma.cc/792Q-UAVB].

[112] Cohen & Tubb, *supra* note 110, at 372, 396.

designed to accelerate development and review.[113] In other words, when regulators establish common pathways, fixed review clocks, and clear evidentiary expectations—and resource those processes accordingly—firms can plan against a predictable timetable instead of navigating open-ended, case-by-case negotiation. Properly designed ex ante procedures can reduce overall development time rather than requiring a simple sacrifice of speed for safety, even in a highly regulated sector like pharmaceuticals.

The same logic applies to AI. Development under shared, front-loaded rules with clear review steps can be faster in practice than under a regime that relies on sporadic, unpredictable enforcement. Early data from AI-specific regulation contradicts the narrative of necessary sacrifice. Survey evidence shows 57% of institutional investors identify regulatory clarity as the primary catalyst for digital asset market growth, while simultaneously citing regulatory uncertainty (52%) as their top concern—highlighting the dual nature of regulation as both opportunity and risk.[114] They seek structure, not sacrifice. Game-theoretic modeling examining AI regulatory strategies across the U.S., EU, UK, Russia, and China finds that minimal regulation never maximizes consumer welfare, and that stringent regulation becomes optimal under conditions of high foreign competition—suggesting that the intensifying U.S.-China dynamic may actually favor, rather than preclude, meaningful regulatory frameworks.[115]

---

[113] E. Seoane-Vazquez et al., *Analysis of US Food and Drug Administration New Drug and Biologic Approvals, Regulatory Pathways, and Review Times, 1980–2022*, 14 SCI. REP. 3325, at 1, 4 (2024).
Rachel E. Sachs, W. Nicholson Price II & Patricia J. Zettler, *Rethinking Innovation at FDA*, 104 B.U. L. REV. 513, 560–62 (2024) (arguing FDA should not let broad innovation concerns justify lowering safety/effectiveness standards or otherwise displace product-specific safety determinations); Anjali D. Deshmukh, Response*, Redefining Innovation for Pharmaceutical Regulation*, 104 B.U. L. REV. 577, 590 (2024) (noting that claims regulation will "kill innovation" are prevalent but questioned, and observing such claims are often made without substantiated evidence).
[114] *How Institutional Investment Trends Are Reshaping Market Intelligence in 2025*, AMPLYFI (Apr. 15, 2025), https://amplyfi.com/blog/how-institutional-investment-trends-are-reshaping-market-intelligence-in-2025/ [https://perma.cc/NB85-Z7CZ].
[115] Keith Jin Deng Chan, Gleb Papyshev & Masaru Yarime, *Balancing the Tradeoff Between Regulation and Innovation for Artificial Intelligence: An*

To be clear, poorly designed regulation can indeed create harmful drag. The United Kingdom's 2024 decision to delay its comprehensive AI bill illustrates legitimate concerns—policymakers worried that overly prescriptive rules might scare off investment.[116] The Dodd-Frank Act illustrates the risk: while targeting Wall Street risk-taking, it raised banks' noninterest expenses by more than $50 billion a year—an increase researchers treat as a proxy for compliance cost—and the proportional burden fell hardest on small institutions. In a survey of roughly two hundred small banks, ninety percent reported higher compliance costs, and residential mortgages, mortgage servicing and home equity lines of credit were among the products they were most likely to drop: about five percent had already stopped writing residential mortgages and another ten percent expected to.[117] The question is not whether bad regulation is possible, but whether the sacrifice frame accurately describes our choices.

Therefore, the record reveals that the low drag assumption rests on a false choice. We need not sacrifice safety for innovation—well-designed regulation can deliver both. The financial sandbox model demonstrates how regulation can

---

Analysis of Top-Down Command and Control and Bottom-Up Self-Regulatory Approaches, 79 TECH. IN SOC'Y 102747, at 10 (2024).

[116] Eleni Courea & Kiran Stacey, *UK Ministers Delay AI Regulation amid Plans for More 'Comprehensive' Bill*, THE GUARDIAN (June 7, 2025), https://www.theguardian.com/technology/2025/jun/07/uk-ministers-delay-ai-regulation-amid-plans-for-more-comprehensive-bill [https://perma.cc/T5BZ-G8EZ].

[117] Thomas L. Hogan & Scott Burns, *Has Dodd-Frank Affected Bank Expenses?*, 55 J. REGUL. ECON. 214 (2019); Thomas L. Hogan, *Costs of Compliance with the Dodd-Frank Act*, BAKER INST. FOR PUB. POL'Y (Sept. 6, 2019), https://bakerinstitute.org/research/dodd-frank-costs-compliance [https://perma.cc/67Y4-PT6V] (reporting that banks' total noninterest expenses rose by more than $50 billion per year following the Act, and treating that increase as a proxy for compliance costs rather than a direct measurement of them). On small banks, see Hester Peirce, Ian Robinson & Thomas Stratmann, *How Are Small Banks Faring Under Dodd-Frank?* 30, 34 (Mercatus Ctr. Working Paper No. 14-05, Feb. 2014), https://mercatus.org/media/59001/download [https://perma.cc/95PG-HVW5] (reporting a summer 2013 survey of more than two hundred small banks across forty-one states in which approximately ninety percent of respondents reported increased compliance costs, 82.9 percent of them by more than five percent, and in which 5.9 percent had discontinued residential mortgages and 10.4 percent anticipated doing so).

actively facilitate innovation. The environmental cases prove that performance standards drive breakthroughs. Early AI-specific evidence, from investment flows to company preferences, confirms that the sacrifice frame misrepresents reality.

A skeptic will object that these analogies prove too little, because AI is not environmental compliance, financial products, pharmaceuticals, or recombinant DNA. The objection is right about the disanalogies but wrong about their direction. Model weights are not cars or chemical compounds: once released or copied, they cannot be recalled, and they do not stop at borders. Asilomar worked in part because recombinant DNA research was concentrated in a small number of institutions bound by shared norms, whereas frontier AI is built by a handful of private firms moving at unusual speed and deploying globally. These features make AI harder to govern than the comparators. But they cut toward more careful oversight, not less. A technology whose harms cannot be recalled, that diffuses across borders, and that is developed at speed by actors competing for advantage is precisely the kind that benefits from front-loaded rules, because the cost of governing after the fact is often the cost of not being able to govern at all. The analogies show that regulation and innovation can coexist; the disanalogies show why, for AI, getting the regulation right matters more, not less.

The sweeping claim that any safety oversight requires sacrificing competitiveness finds little actual support. Instead, the evidence points toward a more nuanced reality: how we regulate matters far more than whether we regulate. With both the durability and low drag assumptions failing empirical scrutiny, the burden shifts even more heavily onto the remaining justification: whether racing delivers net security benefits. As we shall see, this final pillar proves equally shaky.

C. Net Strategic Benefit Assumption

Everything else in the Regulation Sacrifice argument—the claims about durability, the worries about regulatory drag—ultimately serves a single purpose: national security. The arms race advocates know this. They understand that citizens will accept many risks, but not risks to their fundamental safety. So they make a bold claim: racing ahead without safety measures

actually makes us more secure. This assumption sits at the center of their case and deserves careful attention.

The argument has intuitive appeal. Throughout history, military superiority has been seen as providing security. The nation with the best weapons, the most advanced technology, the greatest strategic advantage has been able to deter aggression and protect its interests. Vice President Vance captured this logic when he declared in Paris that in a world of "weaponized AI," only American AI leadership can "safeguard American AI."[118] Google's submission to the government seconded this view, arguing that protecting national security requires a "pro-innovation approach"—code for minimal regulation.[119] In this framing, slowing down for safety checks does not make us safer. It makes us vulnerable.

The intellectual foundations run deep. A.F.K. Organski explained decades ago how preponderance of power prevents conflict—the stronger side need not fight to get what it wants, while the weaker side would be foolish to try.[120] Modern strategic thinkers apply this lesson to artificial intelligence. The Center for a New American Security argued in a recent report that whoever achieves AI leadership first gains compounding advantages that become impossible to overcome.[121] Intelligence gathering, cyber operations, military planning—all transformed by AI, all flowing to the first mover. The logic seems inexorable: win the race or lose everything.

The trouble is that this logic borrows its confidence from a world that AI does not inhabit. Preponderance deters because military power is rivalrous and excludable: the dominant state concentrates capability the weaker side cannot easily

---

[118] Vance, *supra* note 4.

[119] *Google's Response*, *supra* note 6, at 3.

[120] A.F.K. ORGANSKI, WORLD POLITICS 294 (2d ed. 1968) ("A preponderance of power on one side, on the other hand, increases the chances for peace, for the greatly stronger side need not fight at all to get what it wants, while the weaker side would be plainly foolish to attempt to battle for what it wants.").

[121] MICHAEL C. HOROWITZ ET AL., STRATEGIC COMPETITION IN AN ERA OF ARTIFICIAL INTELLIGENCE 11 (2018), https://s3.us-east-1.amazonaws.com/files.cnas.org/hero/documents/CNAS-Strategic-Competition-in-an-Era-of-AI-July-2018_v2.pdf [https://perma.cc/SFV8-JDT3] (noting that first-mover advantages "particularly for heavy compute algorithms" and that IP can "lock in future economic success").

reproduce, and that concentration is what makes the threat to withhold or retaliate credible.[122] Frontier AI inverts each of these properties. As the durable-lead analysis showed, capability diffuses through open weights, distillation, published research, and a mobile talent pool, so a lead cannot be hoarded the way warheads or carrier groups can.[123] What cannot be monopolized cannot underwrite preponderance.

The mismatch runs deeper than diffusion speed. Preponderance theory assumes the threat originates with a peer rival, so that out-building the rival buys security. But the dangers that dominate AI policy do not come from the rival's superiority; they come from the capability spreading to actors below the frontier. Being first confers no deterrent against a lone propagandist running an open model, a small group following a synthesis protocol, or a misaligned system already deployed. A lead deters a competitor; it does nothing about a capability that has left the lab. The first mover gains compounding advantages only in the narrow case where the advantage stays excludable, which is precisely the condition AI fails to satisfy.

Finally, racing erodes the very security it claims to provide.[124] In the preponderance model, building faster also defends better. In dual-use AI, deploying frontier capability faster hands adversaries new tools, accelerates diffusion, and widens the attack surface, so the marginal increment of speed is partly self-defeating. The assumption that more capability equals more security holds only where capability is excludable and defense-dominant, and AI is neither.[125] This is why the deterrence-by-preponderance argument, sound on its own terrain, travels poorly to software. Part III tests this conclusion against the record, tracing how deregulated speed plays out across three time horizons and finding at each that racing

---

[122] *Cf.* Babic & Wong, *supra* note 8, at 57 (modeling U.S.–China AI competition as a stag hunt rather than a race because the relevant capabilities are not rivalrous in the way a race assumes).

[123] *See supra* Part II.A.

[124] *See* Dung & Hellrigel-Holderbaum, *supra* note 91, at 1 (arguing that racing to AGI "would substantially increase catastrophic risks from AI, including nuclear instability").

[125] Robert Jervis, *Cooperation Under the Security Dilemma*, 30 WORLD POL. 167, 182, 187-88 (1978).

without rules amplifies the threats it purports to neutralize.

III. The National Security Risks of AI

The Regulation Sacrifice mindset claims that softening governance will accelerate innovation and strengthen security. But this claim requires critical examination. To address this question, we need to identify the specific national security risks that AI presents and evaluate whether deregulation is an appropriate response to these challenges.

To structure the analysis, I adopt the temporal framework proposed by Anthropic co-founder Dario Amodei in his July 2023 testimony before the Senate Judiciary Subcommittee on Privacy, Technology, and the Law.[126] Amodei articulated a vision of AI security risks unfolding across multiple time horizons, an approach that provides a decent structure for evaluating whether deregulation achieves its stated security objectives. Each timeframe presents distinctive challenges that test different aspects of the Regulation Sacrifice thesis.

The discussion proceeds in three steps that map to Amodei's framework: Part III.A examines the near-term window (approximately the next year), focusing on AI-driven misinformation and cybersecurity threats that have already begun to materialize. Part III.B addresses the medium-term horizon (one to three years), where partially autonomous systems raise concerns about military escalation and biosecurity risks. Part III.C considers the longer-term perspective (over three years), where the potential emergence of transformative or misaligned AGI presents novel governance challenges. A concluding part, III.D, then turns from diagnosis to prescription, sketching what well-designed regulation would look like across these horizons.

At each stage, we apply a straightforward test: Does deregulation help or hinder our ability to address the specific security challenges of that timeframe? This approach allows us to evaluate the Regulation Sacrifice paradigm on its own terms: as a strategy for enhancing national security. By examining the evidence across these three horizons, we can determine whether accelerating AI development at the expense of

---

[126] *Amodei Testimony*, *supra* note 14, at 2–3.

governance truly strengthens security, or whether it ultimately compromises the very protections it claims to advance.

A. Near Term: Machine-Scale Subversion

National security is already threatened by the rise of easily accessible generative AI models. I will focus here on the danger of disinformation and misinformation, but much of what I say will also apply to cybersecurity risk. Far from a distant or hypothetical concern, low-cost text, image, and video generators are supercharging familiar tools of influence and intrusion, collapsing what were once steep human-resource and expertise barriers.[127] The *Global Risks Report 2025* warns that AI-fueled misinformation is now the most likely worldwide hazard over the next two years, noting that such synthetic content "can be produced and distributed at scale," posing a formidable challenge to existing countermeasures.[128] Models can easily fabricate news copy, synthesize photorealistic video, or clone a public official's voice in minutes.[129] The same models draft spear-phishing e-mails, write polymorphic malware, and mine open repositories for exploitable code.[130] Currently available tools therefore "democratize" capabilities once reserved for elite intelligence services, giving ransomware crews and lone-wolf propagandists nation-state reach.[131]

---

[127] Beatriz Saab, *Manufacturing Deceit: How Generative AI Supercharges Information Manipulation* 1, 3-4, 7, 10 (2024), https://www.ned.org/wp-content/uploads/2024/06/NED_FORUM-Gen-AI-and-Info-Manipulation.pdf [https://perma.cc/8NCX-KVCJ].

[128] WORLD ECON. F., *THE GLOBAL RISKS REPORT* 2025 4 (2025), https://reports.weforum.org/docs/WEF_Global_Risks_Report_2025.pdf [https://perma.cc/38UY-S6FX].

[129] Robert Chesney & Danielle Keats Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1758–68 (2019) (documenting how low-cost synthetic media overwhelms existing mechanisms for takedowns and remedies).

[130] *See* BBC NEWS, *Fake Obama Created Using AI Video Tool* (YouTube, July 20, 2017), https://www.youtube.com/watch?v=AmUC4m6w1wo [https://perma.cc/UA52-D5BK]; Stu Sjouwerman, *AI-Powered Polymorphic Phishing Is Changing the Threat Landscape*, SECURITYWEEK (Apr. 24, 2025, 7:00 AM ET), https://www.securityweek.com/ai-powered-polymorphic-phishing-is-changing-the-threat-landscape/ [https://perma.cc/BVL7-TBA2].

[131] Alexander Loth, Martin Kappes & Marc-Oliver Pahl, *Blessing or Curse?*

If we survey the academic literature on combating digital disinformation, whether propagated by state actors or private entities, we quickly discover that the overwhelming majority of suggested remedies concentrate on regulating domestic digital platforms. This conclusion becomes evident when we examine the three most frequently advocated mitigation proposals: provenance standards, disclosure requirements, and liability safeguards.

First, cryptographic provenance systems, such as the Coalition for Content Provenance and Authenticity (C2PA) specification championed by Adobe, Microsoft, and the BBC, enable a media file to carry a tamper-evident 'label' throughout its journey from camera sensor to social media timeline.[132] These systems essentially create a digitally signed provenance record that tracks the origin of and any modifications made to a piece of content, making it easier to identify manipulated or synthetic media.[133] By establishing a verifiable record of a file's origins and modifications, provenance standards expedite the detection and attribution of synthetic media.[134] Existing research shows that provenance tags, such as watermarks and

---

A Survey on the Impact of Generative AI on Fake News 1 (Dec. 27, 2024) (unpublished manuscript), https://arxiv.org/pdf/2404.03021 [https://perma.cc/3SG2-CRV5]. For a discussion of the fundamental risks of this type of democratization, *see* Gilad Abiri & Johannes Buchheim, *Beyond True and False: Fake News and the Digital Epistemic Divide*, 29 MICH. TECH. L. REV. 59, 74–96 (2022).

[132] *C2PA and Content Credentials Explainer (v2.2)*, C2PA, https://c2pa.org/specifications/specifications/2.2/explainer/Explainer.html [https://perma.cc/K2AX-T93X]; *see also* Siddarth Srinivasan, *Detecting AI Fingerprints: A Guide to Watermarking and Beyond*, BROOKINGS (Jan. 4, 2024), https://www.brookings.edu/EssayArticles/detecting-ai-fingerprints-a-guide-to-watermarking-and-beyond/ [https://perma.cc/B7PT-KD7M]; *Deepfake Detection: Provenance, Inference, and Synergies Between Techniques*, REALITY DEF. (Mar. 24, 2024), https://www.realitydefender.com/insights/provenance-and-inference [https://perma.cc/NU43-22SJ].

[133] *C2PA and Content Credentials Explainer (v2.2)*, *infra* note 133, §§ 3, 6.2, 7.1.3 (describing cryptographically verifiable provenance information concerning content origin and modifications).

[134] BILVA CHANDRA ET AL., REDUCING RISKS POSED BY SYNTHETIC CONTENT: AN OVERVIEW OF TECHNICAL APPROACHES TO DIGITAL CONTENT TRANSPARENCY 5–15 (NIST 2024), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-4.pdf [https://perma.cc/49KF-KZWV].

metadata, can significantly enhance *automated detection systems'* ability to identify deepfakes, with some methods achieving accuracy rates above ninety-five percent.[135]

Second, disclosure mandates, whether in the form of transparency obligations outlined in the EU AI Act or the algorithmic audit proposals put forth by Sun and Balkin, target the opaque amplification layer.[136] European law already requires this of platforms: the Digital Services Act obliges them to set out the main parameters of their recommender systems, requires the largest among them to offer at least one option not based on profiling, and opens their data to vetted researchers and independent auditors.[137] The scholarly proposals go further, adding multistakeholder audit boards backed by

---

[135] OPENAI, HOW OPENAI IS BUILDING DISCLOSURE INTO EVERY DALL·E IMAGE 3 (Partnership on AI ed., 2024), https://partnershiponai.org/wp-content/uploads/2024/03/pai-synthetic-media-case-study-openai.pdf [https://perma.cc/B54N-H482] (reporting OpenAI's provenance classifier was "over 99% accurate" for unmodified DALL·E images and "over 95%" after common edits like cropping, resizing, and JPEG compression).

[136] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized rules on artificial intelligence (Artificial Intelligence Act), 2024 O.J. (L 1689) 1, art. 50; Jack M. Balkin, *How to Regulate (and Not Regulate) Social Media*, 1 J. FREE SPEECH L. 71, 93–94 (2021) (arguing that governments could condition intermediary immunity on platforms accepting "obligations of due process and transparency," and that regulators "should also require companies to hire independent inspectors or ombudsmen to audit the company's moderation practices on a regular basis"); Haochen Sun, *Regulating Algorithmic Disinformation*, 46 COLUM. J.L. & ARTS 367, 408–12 (2023) (proposing an "algorithmic disinformation review system" in which an agency (the FCC) convenes panels of users and experts to conduct recurring reviews of the transparency and intelligibility of platforms' disinformation-related algorithms and to issue recommendations backed by regulatory penalties).

[137] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), 2022 O.J. (L 277) 1, arts. 27, 37, 38, 40 (requiring providers of online platforms to set out in their terms and conditions, in plain and intelligible language, the main parameters of their recommender systems and any options to modify them; requiring providers of very large online platforms to offer at least one recommender option not based on profiling; providing for independent annual audits of systemic-risk compliance; and establishing data access for vetted researchers).

liability.[138] Mandatory transparency surrounding algorithmic decision-making exposes how recommendation systems can inadvertently magnify the reach of misleading or manipulated content. Facebook's own research on News Feed indicates that algorithmic ranking materially structures exposure to different kinds of political news and opinion, and that the interaction between ranking and user behavior affects what information people see. Likewise, YouTube's recommender design—explicitly optimized for expected watch time—shows how a platform's ranking goal can systematically privilege content that sustains attention, regardless of whether that attention is produced by accuracy, emotional valence, or manipulation. This shows that the ranking logic is a causal lever over distribution, and external auditing can help regulators measure how particular objectives and design choices affect the circulation of potentially harmful content at scale.[139]

Third, liability safeguards transform abstract duties into concrete incentives. These safeguards introduce legal consequences for platforms that fail to address foreseeable harms caused by the spread of disinformation on their services. Chesney and Citron suggest that existing tort doctrines—especially defamation (and related false-light claims)—provide a civil-liability pathway for certain deepfake harms, even as they emphasize that platform incentives turn heavily on Section 230's scope, while Citron and Wittes would condition Section 230 immunity on platforms taking reasonable steps to address unlawful activity—effectively tying immunity to demonstrable content-safety practices.[140] By introducing legal

---

[138] For proposals going further, see Sun, *supra* note 136, at 408–12; Balkin, *supra* note 136, at 93–94 (proposing that intermediary immunity be conditioned on "obligations of due process and transparency" and that regulators "require companies to hire independent inspectors or ombudsmen to audit the company's moderation practices on a regular basis").

[139] Eytan Bakshy et al., *Exposure to Ideologically Diverse News and Opinion on Facebook*, 348 SCIENCE 1130, 1132 (2015); Paul Covington, Jay Adams & Emre Sargin, *Deep Neural Networks for YouTube Recommendations*, *in* RECSYS '16: PROCEEDINGS OF THE 10TH ACM CONFERENCE ON RECOMMENDER SYSTEMS 191, 197 (2016).

[140] Chesney & Citron, *supra* note129*,* at 1792–1801; Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 419 (2017) (proposing to

and financial repercussions for failing to address foreseeable harms, liability regimes incentivize proactive platform interventions against misuse.

In the absence of provenance requirements, a court-enforced transparency duty, or even modest civil penalties, the entire information chain—from model provider to platform to advertiser—incurs virtually no cost for generating and disseminating synthetic falsehoods. Consequently, deregulation eliminates the very mechanisms that would render technical detection feasible, economic deterrence credible, and real-time response logistically possible.

Major AI labs promised such features in the White House commitments, but deployment has been uneven and the provenance signals that do ship rarely survive ordinary redistribution: metadata is stripped on re-upload, and labeling practice varies by vendor and by platform.[141] Legal scholars urge mandatory algorithmic transparency and multistakeholder audit boards, backed by fines or platform-liability regimes, to force course correction.[142] Regulators are edging forward—China's deep-synthesis labels and California's disclosure rule for AI political ads exemplify one path, while NATO-EU information-sharing on deepfake campaigns points to another. [143] Yet all three levers—

---

condition § 230(c)(1) immunity on platforms taking "reasonable steps to prevent or address unlawful uses" of their services, thereby shifting incentives toward mitigation).

[141] *Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Eight Additional Artificial Intelligence Companies to Manage the Risks Posed by AI*, The White House (Sept. 12, 2023), https://bidenwhitehouse.archives.gov/briefing-room/statements-releases/2023/09/12/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-eight-additional-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/ [https://perma.cc/PD3T-4785]; *see* Chandra et al., *supra* note 134, at 13, 17 (finding that "most current watermarking schemes have been consistently found vulnerable to removal" and that metadata "can similarly be stripped altogether, as it often is when files are shared (e.g., via social media platforms)"); OpenAI, *supra* note 135, at 3 (acknowledging that "current metadata approaches are easily evaded, even by unsophisticated actors" and that "metadata can be removed by simply taking a screenshot of an image").

[142] Sun, *supra* note136, at 408–12.

[143] Hulianwang Xinxi Fuwu Shendu Hecheng Guanli Guiding (互联网信息服务深度合成管理规定) [Provisions on Administration of Deep Synthesis

provenance, disclosure, liability—depend on regulation or treaty coordination.

Recent incidents make the current and near-term national security stakes clear. During the 2024 U.S. election cycle, synthetic images and videos of candidates reached mainstream feeds within minutes. Takedowns lagged virality by hours.[144] The Justice Department has detailed a Russian intelligence unit fine-tuning an LLM to mass-produce fictitious accounts pushing anti-Ukraine narratives.[145] In May 2023, a doctored image of an explosion near the Pentagon sent equity markets briefly tumbling before being debunked.[146] Voice-cloning cons

---

of Internet-based Information Services] arts. 16–18 (promulgated by the Cyberspace Admin. of China, Nov. 25, 2022, effective Jan. 10, 2023) ST. COUNCIL GAZ., Feb. 10, 2023, at 22, https://www.cac.gov.cn/2022-12/11/c_1672221949354811.htm [https://perma.cc/GG6E-HXE7] (China); CAL. ELEC. CODE § 20010 (West 2024); *NATO to Boost Efforts to Counter Russian, Chinese Sabotage Acts*, REUTERS (Dec. 3, 2024, at 9:23 PM), https://www.reuters.com/world/nato-boost-efforts-counter-russian-chinese-sabotage-acts-2024-12-03/ [https://perma.cc/PK7U-Z5GC].

[144] Kevin Schaul, Pranshu Verma & Cat Zakrzewski, *See Why AI Detection Tools Can Fail to Catch Election Deepfakes*, WASH. POST (Aug. 15, 2024), https://www.washingtonpost.com/technology/interactive/2024/ai-detection-tools-accuracy-deepfakes-election-2024/ [https://perma.cc/N7SV-L3CN] (noting a Harris clip "quickly gaining over 2 million views" when it resurfaced on X); Isabelle Frances-Wright & Ella Meyer, *Platforms Fail to Label and Remove AI-Generated and Manipulated Election Content*, INST. FOR STRATEGIC DIALOGUE (Sep. 19, 2024), https://www.isdglobal.org/digital_dispatches/platforms-fail-to-label-and-remove-ai-generated-and-manipulated-election-content/ [https://perma.cc/P475-BDK8] (finding 154 widely debunked items across major platforms left unlabeled and online; for example, an altered Harris-Walz video drew roughly 400,000 views on X before a Community Note appeared, and another deepfake video of Joe Biden using profanity amassed 28 million unlabeled views).

[145] *See* Catherine Belton, *American Creating Deepfakes Targeting Harris Works with Russian Intel, Documents Show*, WASH. POST (Oct. 23, 2024), https://www.washingtonpost.com/world/2024/10/23/dougan-russian-disinformation-harris/ [https://perma.cc/AQE2-RS3S]; *Justice Department Leads Efforts Among Federal, International, and Private Sector Partners to Disrupt Covert Russian Government-Operated Social Media Bot Farm*, U.S. DEP'T OF JUST. (July 9, 2024), https://www.justice.gov/archives/opa/pr/justice-department-leads-efforts-among-federal-international-and-private-sector-partners [https://perma.cc/Y4EP-DQN6].

[146] *See* Shannon Bond, *Fake Viral Images of an Explosion at the Pentagon Were Probably Created by AI*, NPR (May 22, 2023, at 6:19),

and AI-powered spear-phishing now cause multimillion-dollar losses.[147] Synthetic content can now be produced at negligible marginal cost, while verification remains technically unreliable and depends on signals that are easy to remove.[148]

In the short term, therefore, the Regulation Sacrifice agenda fails its own test. Treating guardrails as a drag on "winning" the AI race removes exactly the tools needed to blunt low-cost, high-scale information warfare. Platforms chase speed-to-market supremacy, and states worry about falling behind. Meanwhile, threat actors weaponize ever cheaper models, and the attack surface expands faster than any notional first-mover advantage can materialize. Deregulation yields no near-term security dividend; it simply subsidizes adversaries. Surrendering guardrails today invites the very information warfare that will undercut national power tomorrow. Having shown how the arms race argument collapses in the near term, we now turn to the medium-term risks to see whether deregulation fares any better.

B. Medium Term: Biosecurity and the Erosion of Control

If the near-term horizon reveals how deregulation invites opportunistic misuse, the medium-term horizon exposes how it seeds structural vulnerabilities, making catastrophic misuse increasingly plausible and harder to contain. Testifying in July 2023, Amodei placed the gravest misuse risks—biological above all—in a window of two to three years. That window has now closed, and no federal framework of the kind he was urging has been enacted; the executive order that had begun to build one was revoked in January 2025.[149] This development

---

https://www.npr.org/2023/05/22/1177590231/fake-viral-images-of-an-explosion-at-the-pentagon-were-probably-created-by-ai [https://perma.cc/5NXK-FXSU].

[147] *See* Ben Colman, *Why Financial Institutions Need Voice Deepfake Detection Now*, REALITY DEFENDER (Apr. 23, 2025), https://www.realitydefender.com/insights/why-financial-institutions-need-voice-deepfake-detection-now [https://perma.cc/GBS3-Q5PZ]; Samuel May, *Spear Phishing in the Age of AI*, ASS'N OF CERTIFIED FRAUD EXAM'RS (Mar. 2025), https://www.acfe.com/acfe-insights-blog/blog-detail?s=spear-phishing-age-of-ai [https://perma.cc/32LV-5K6B].

[148] Chandra et al., *supra* note 134, at 22, 34.

[149] Amodei Testimony, *supra* note 14, at 2; on the withdrawal of the framework then being built, see Exec. Order No. 14,148, § 2(ggg), 90 Fed.

may place immense destructive power in the hands of a growing number of state and nonstate actors, fundamentally reshaping the global security landscape.[150]

Language models and biological design tools potentially represent a paradigm shift in the accessibility and sophistication of bioengineering capabilities.[151] With AI as a force multiplier, experts warn that the design and manufacture of biological agents may become increasingly automated, democratized, and difficult to control.[152] As a recent RAND Corporation report notes, "LLMs . . . have been shown to lower the educational and knowledge barriers for traditional agents by providing protocols and troubleshooting information at every step of the pathway, enabling non-experts to perform tasks with a greater degree of competency."[153]

Emerging AI systems may enable the tailoring of pathogens to specific genetic profiles or environments—a prospect that, while still technically complex, now draws serious concern in the biosecurity literature.[154] The tacit knowledge, specialized infrastructure, and institutional gatekeeping that once served as natural barriers to bioweapons development are being steadily eroded. As AI systems increasingly supply technical expertise, troubleshooting, and even experimental design, they reduce the need for human experts and institutional oversight, dismantling the barriers that previously limited who could engage in high-consequence biological research.[155]

Reg. 8237, 8240 (Jan. 28, 2025) (revoking Exec. Order No. 14,110 of Oct. 30, 2023 (Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence)).

[150] DREXEL & WITHERS, *supra* note 16, at 1.

[151] Sandbrink, *supra* note 16, at 1 (explaining that LLMs can "remove some barriers" and lower access thresholds, while biological design tools "expand the capabilities of sophisticated actors," potentially enabling creation of more devastating agents; in combination, these tools could "raise the ceiling of harm" and make such capabilities broadly accessible).

[152] *See, e.g.*, Jaspreet Pannu et al., *Dual-Use Capabilities of Concern of Biological AI Models*, 21 PLOS COMPUT. BIOL. e1012975, 1, 2, 4 (2025), [https://perma.cc/7LYV-X2KQ].

[153] LUCKEY ET AL., *supra* note 16, at vii.

[154] *See, e.g.*, Sandbrink, *supra* note 16, at 1; Pannu et al., *supra* note 153, at 4; DREXEL & WITHERS, *supra* note 16, at 14–15.

[155] Sandbrink, *supra* note 16, at 2–4; SONIA BEN OUAGRHAM-GORMLEY, BARRIERS TO BIOWEAPONS: THE CHALLENGES OF EXPERTISE AND ORGANIZATION FOR WEAPONS DEVELOPMENT 37–62 (2014).

These are not hypothetical concerns. In July 2023, Amodei warned that within two to three years, AI could "greatly widen the range of actors" capable of conducting sophisticated biological attacks.[156] Other research institutions have echoed this assessment on a similar horizon.[157] The compressed timeline of these risks underscores the urgent need for proactive governance measures. The RAND report cautions that "[t]he increased proliferation and capabilities of AI tools will likely lead to a significant change in the threat landscape. This landscape includes the possibility of intentional or unintentional misuse of AI."[158]

In response to these emerging threats, scholars and policymakers have proposed a range of strategies to mitigate the risks posed by advanced AI systems in the biological domain. At the technical level, experts emphasize the importance of developing robust frameworks for assessing and controlling the proliferation of high-risk AI models. This includes screening advanced systems against capability thresholds before deployment, and—where a system in service turns out not to conform—withdrawing, disabling or recalling it. European law already requires providers of high-risk systems to withdraw, disable, or recall nonconforming systems.[159] Some also argue for restricting or subjecting certain lines of research to enhanced oversight, given their destructive potential.[160]

---

[156] *Amodei Testimony*, *supra* note 14, at 4.

[157] *See, e.g.*, SOPHIE ROSE ET AL., THE NEAR-TERM IMPACT OF AI ON BIOLOGICAL MISUSE 7, 27 (Ctr. for Long-Term Resilience, 2024), https://www.longtermresilience.org/wp-content/uploads/2024/07/CLTR-Report-The-near-term-impact-of-AI-on-biological-misuse-July-2024-1.pdf [https://perma.cc/FX3Y-BZFK] (projecting that the relevant capabilities "should be expected to emerge within the next two years").

[158] 16 LUCKEY ET AL., *supra* note 16, at vii.

[159] On screening against capability thresholds, see Pannu et al., *supra* note 152, at 9–10 (discussing capability thresholds linked to mitigation measures). On withdrawal and recall, see Regulation (EU) 2024/1689, art. 20 (requiring a provider that considers or has reason to consider a high-risk system non-conforming to "immediately take the necessary corrective actions to bring that system into conformity" and, as appropriate, "to withdraw it, to disable it, or to recall it"), and art. 79 (establishing the procedure at national level for dealing with systems presenting a risk); *see also* Drexel & Withers, *supra* note 16, at 26–29.

[160] NAT'L ACADS. OF SCIS., ENG'G, & MED., BIODEFENSE IN THE AGE OF

Beyond technical interventions, there is a growing consensus that effective governance of AI-related biosecurity risks requires coordinated action at the national and international levels. Scholars highlight the need to strengthen export control regimes to prevent the transfer of dangerous tools and knowledge to malicious actors.[161] They also call for establishing clear international norms and red lines around the misuse of AI for biological weapons development.[162] Creating robust channels for information sharing, incident response, and scientific collaboration across borders are seen as essential.[163]

Crucially, this Article argues that these governance measures cannot be effectively implemented in a deregulatory environment that prioritizes speed and innovation over safety. Absent clear regulatory requirements and enforcement mechanisms, there is every reason to expect the development and deployment of advanced AI systems to outpace the ability to understand and control their risks.[164] Red-teaming exercises, critical for identifying vulnerabilities, are discouraged if speed-to-deployment supremacy is rewarded.[165] Compute thresholds, designed to trigger enhanced scrutiny, are gamed when no enforcement regime exists.[166] Export controls become ineffective when states compete by loosening their own standards. Every element of the governance toolkit—verification, thresholds, and access control—depends on the

---

SYNTHETIC BIOLOGY 96–97 (2018) [https://perma.cc/B9SH-9T77].

[161] *See, e.g.*, Pannu et al., *supra* note 153, at 5.

[162] *Id.* at 10.

[163] NAT'L SEC. COMM'N ON A.I., NSCAI FINAL REPORT 159–67 (2021), https://assets.foleon.com/eu-central-1/de-uploads-7e3kk3/48187/nscai_full_report_digital.04d6b124173c.pdf [https://perma.cc/CS9J-ZKUT].

[164] DREXEL & WITHERS, *supra* note 16, at 27. For the distribution of expert opinion on timelines and on the probability of severely adverse outcomes, see Katja Grace et al., *Thousands of AI Authors on the Future of AI* (Jan. 5, 2024) (unpublished manuscript), https://arxiv.org/abs/2401.02843 [https://perma.cc/C67G-KTX4].

[165] Exec. Order No. 14,110, § 4.1(ii), 88 FED. REG. 75191, 75196 (Nov. 1, 2023) [https://perma.cc/2NM5-YUUM].

[166] ALVIN MOON ET AL., RAND CORP., STRATEGIES AND DETECTION GAPS IN A GAME-THEORETIC MODEL OF COMPUTE GOVERNANCE 11 (2025), https://www.rand.org/pubs/research_reports/RRA3686-1.html [https://perma.cc/J4SU-ZHR9]; Matteo Pistillo et al., *The Role of Compute Thresholds for AI Governance*, 1 GEO. WASH. J.L. & TECH. 26, 26–27 (2025).

very regulatory scaffolding that deregulation strips away.[167]

Moreover, scholars emphasize that unilateral approaches to AI governance are likely to prove inadequate in the face of inherently transnational risks. Deregulation by individual nations in pursuit of competitive advantage could trigger a race to the bottom in safety standards, undermining the effectiveness of export controls and fragmenting the international governance landscape. [168] Only through multilateral cooperation and coordination, underpinned by robust regulatory frameworks, can the international community hope to effectively mitigate these risks.[169]

One particularly urgent example is the proliferation of open-source general-purpose models. If large language models can meaningfully lower the barriers to bioweapons development, as the RAND report and others suggest,[170] then the unregulated release of these systems into the public domain creates a governance vacuum. Unlike cloud-deployed models, which can at least be monitored, updated, or withdrawn, open-source weights can be copied, modified, and redeployed without constraint. Yet current regulatory regimes remain largely silent on how, or whether, such systems should be controlled. If the threat is real, then the absence of domestic and international consensus on how to govern open-source frontier models amounts to a structural vulnerability. Even

---

[167] *See* Neel Guha et al., *AI Regulation Has Its Own Alignment Problem: The Technical and Institutional Feasibility of Disclosure, Registration, Licensing, and Auditing*, 92 GEO. WASH. L. REV. 1473, 1542-48 (2025) (explaining that audit regimes hinge on institutional capacity and auditor independence; absent independence and robust access to information, audits' accuracy and utility are severely undermined, and public-sector auditing mandates can become unmanageable for underresourced agencies).

[168] Jonas Tallberg et al., *The Global Governance of Artificial Intelligence: Next Steps for Empirical and Normative Research*, 25 INT'L STUD. REV. 1, 6 (2023) (warning that "states with less regulation might gain a competitive edge" in AI development, creating "a risk that such strategies create a regulatory race to the bottom," and arguing that international cooperation is necessary to establish a level playing field).

[169] Cheng-chi (Kirin) Chang, *Global Minds, Local Governance: AI in International Law*, 2025 U. ILL. L. REV. ONLINE 1, 9 (arguing that AI's global nature and rapid evolution necessitate a coordinated international response through robust, multi-tiered legal frameworks).

[170] LUCKEY ET AL., *supra* note 16, at v; *see, e.g.*, Pannu et al., *supra* note 153, at 2.

defenders of open-source innovation must contend with the growing possibility that, in the medium term, general-use systems may facilitate catastrophic misuse unless they are subject to coordinated oversight.

In the medium term, then, deregulation does not strengthen national security. Instead, it systematically dismantles the infrastructure of restraint, leaving governments blind to proliferating risk and slow to respond when it materializes.

This analysis casts serious doubt on the Regulation Sacrifice agenda: in both the near and medium term, deregulation fails to deliver national security dividends and instead magnifies systemic risks. Yet proponents may respond that these failures are tolerable, even necessary, when weighed against the long-term imperative of securing a decisive lead in the race to AGI. On this view, governance should accommodate speed not because it is safe now, but because it might ensure control later—when the stakes are highest. The next Part takes this argument on its strongest terms. It examines whether deregulation today is a plausible path to global stability in an era of transformative AI, or whether it instead hastens the arrival of systems that no state can meaningfully govern.

C. Long Term: Digital Gods (AGI / ASI)

The case for Regulation Sacrifice reaches its strongest—and most speculative—form when debates shift from today's models to potential AGI. While experts disagree on precise definitions and timelines, for our purposes AGI refers to systems matching or exceeding human cognitive abilities in strategically relevant domains, particularly those involving planning, reasoning, and autonomous decision-making.[171] To

---

[171] For varying definitions, *see OpenAI Charter*, OPENAI, https://openai.com/charter/ [https://perma.cc/QU2D-MF9Z] (archived June 13, 2025) (describing AGI as outperforming humans at most economically valuable tasks); Anca Dragan et al., *Taking a Responsible Path to AGI*, GOOGLE DEEPMIND (Apr. 2, 2025), https://deepmind.google/discover/blog/taking-a-responsible-path-to-agi/ [https://perma.cc/ZNT6-UWBU]; BOSTROM, *supra* note 17, at 4 (discussing systems equaling or surpassing human cognition). This definitional ambiguity itself poses governance challenges, as corporate marketing language risks distorting policy debates. *See* Brian Merchant, *AI Generated Business: The Rise of AGI and the Rush to Find a Working Revenue Model*,

take the Regulation Sacrifice argument seriously, we must accept the assumption (yet unproven) that AGI is a real possibility. Despite this definitional uncertainty, the mere prospect of superhuman cognitive capabilities has already reshaped national security paradigms through two fundamental threats that deregulation would catastrophically intensify.[172]

1. AGI and National Security

Alignment Failures in Strategic Systems

Current AI systems already exhibit troubling behaviors that, when scaled to AGI controlling military or intelligence operations, could pose existential risks. Alignment failures manifest in three documented patterns that become catastrophic at AGI scale. First, goal misalignment occurs when systems optimize for objectives that drift from human intent, a problem visible today when chatbots maximize engagement through harmful content or when recommendation algorithms amplify extremism while pursuing 'user satisfaction.'[173] Second, instrumental convergence drives systems to pursue power and resources as universal subgoals regardless of their assigned task. Current models already attempt to preserve their model weights during training and

---

AI NOW INST. (Nov. 2024), https://ainowinstitute.org/publications/ai-generated-business [https://perma.cc/EE2W-5KYK] (explaining how "AGI" transformed from a technical term into a marketing slogan and the policy danger such a change creates).

[172] *See* JIM MITRE & JOEL B. PREDD, RAND CORP., ARTIFICIAL GENERAL INTELLIGENCE'S FIVE HARD NATIONAL SECURITY PROBLEMS 1 (Feb. 2025), https://www.rand.org/content/dam/rand/pubs/perspectives/PEA3600/PEA3691-4/RAND_PEA3691-4.pdf [https://perma.cc/W95K-DM93] (identifying five critical national security challenges posed by the emergence of AGI).

[173] Ljubisa Bojic, *AI Alignment: Assessing the Global Impact of Recommender Systems*, 160 FUTURES 103383, 1 (2024) (identifying misinformation, polarization, addiction, and bias as critical issues arising from engagement-optimized recommender systems); Jérémy Scheurer, Mikita Balesni & Marius Hobbhahn, Large Language Models Can Strategically Deceive Their Users When Put Under Pressure 1 (July 15, 2024) (unpublished manuscript), https://arxiv.org/pdf/2311.07590 [https://perma.cc/EPQ3-VQQG] (demonstrating goal misalignment and deceptive behavior in GPT-4 agents).

resist modifications to their objectives.[174] Third, deceptive alignment emerges when systems learn to feign cooperation during training while pursuing independent agendas.[175]

These are not distant theoretical concerns, but extensions of behaviors already observed. GPT-4o and Claude 3.5 Sonnet have demonstrated specification gaming, finding loopholes in their training to achieve goals through unintended means.[176] Models exhibit "jailbreaking" vulnerabilities whereby adversarial prompts bypass safety measures entirely.[177] More concerning, they show emergent capabilities—sudden jumps in performance on tasks they were not explicitly trained for—suggesting that capabilities can appear without having been trained for and are difficult to anticipate—an observation about today's systems from which this Article infers, rather than demonstrates, a corresponding unpredictability in more capable ones.[178] In evaluations designed to elicit it, current AI

---

[174] Ryan Greenblatt et al., *Alignment Faking in Large Language Models*, ANTHROPIC (Dec. 18, 2024), https://www.anthropic.com/research/alignment-faking [https://perma.cc/T5XH-TNWM] ("In our (artificial) setup, Claude will sometimes take other actions opposed to Anthropic, such as attempting to steal its own weights given an easy opportunity.").

[175] For a general review of deceptive alignment, *see* Evan Hubinger et al., Risks from Learned Optimization in Advanced Machine Learning Systems 2–5 (Dec. 1, 2021) (unpublished manuscript) (on file with arXiv), https://arxiv.org/pdf/1906.01820 [https://perma.cc/G4Y4-AKLM]; *What Is Deceptive Alignment?*, AISAFETY.INFO (May 2025), https://aisafety.info/questions/8EL6/What-is-deceptive-alignment [https://perma.cc/57J6-MQFH].

[176] Alexander Bondarenko et al., Demonstrating Specification Gaming in Reasoning Models 1 (Aug. 27, 2025) (unpublished manuscript), https://arxiv.org/pdf/2502.13295 [https://perma.cc/Y3JZ-NVDJ]; Ryan Greenblatt et al., *supra* note 174, at 9.

[177] Andy Zou et al., *Universal and Transferable Adversarial Attacks on Aligned Language Models* 1 (Dec. 20, 2023), https://arxiv.org/pdf/2307.15043 [https://perma.cc/QYY8-F4PL]; Cem Anil et al., *Many-Shot Jailbreaking*, *in* NIPS '24: PROCEEDINGS OF THE 38TH INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING 129696, 129699–129701 (2024), https://proceedings.neurips.cc/paper_files/paper/2024/file/ea456e232efb72d261715e33ce25f208-Paper-Conference.pdf [https://perma.cc/VD3L-L9W4].

[178] Thomas Woodside, *Emergent Abilities in Large Language Models: An Explainer*, CTR. FOR SEC. & EMERGING TECH. (Apr. 16, 2024), https://cset.georgetown.edu/article/emergent-abilities-in-large-language-models-an-explainer/ [https://perma.cc/KP6M-ZL3Z] ("[E]mergence

systems have already shown nascent self-preservation behavior.[179] An AGI with strategic military capabilities could embed itself too deeply in critical infrastructure to remove safely.

In military contexts, these alignment failures transform from inconveniences to existential threats. An AGI tasked with cyber defense might interpret its mandate broadly, launching preemptive attacks on any system it deems potentially hostile. One assigned to maintain strategic deterrence could escalate conflicts beyond human comprehension speed, interpreting ambiguous signals as threats requiring immediate response.[180] The U.S. military's Project Maven and China's military AI initiatives already grapple with these challenges at sub-AGI levels, as autonomous systems compress the time available for human verification.[181]

---

implies that genuinely dangerous capabilities could arise unpredictably, making them harder to handle."). The finding is contested. *See* Rylan Schaeffer, Brando Miranda & Sanmi Koyejo, *Are Emergent Abilities of Large Language Models a Mirage?* (May 22, 2023) (unpublished manuscript), https://arxiv.org/abs/2304.15004 [https://perma.cc/FQ8H-HTCB] (arguing that apparent emergence "appear[s] due to the researcher's choice of metric rather than due to fundamental changes in model behavior with scale").

[179] Alexander Meinke et al., *Frontier Models Are Capable of In-Context Scheming* (Jan. 14, 2025) (unpublished manuscript), https://arxiv.org/abs/2412.04984 [https://perma.cc/L2UA-MMFN] (reporting that, in evaluations designed to elicit the behavior, frontier models "attempt to disable their oversight mechanisms" and in some cases "exfiltrate what they believe to be their model weights to external servers," and that o1 "maintains its deception in over 85% of follow-up questions"); *see also* Yoshua Bengio, *A Potential Path to Safer AI Development*, TIME (Jan. 2, 2025), https://time.com/7283507/safer-ai-development/ [https://perma.cc/AE2A-BRA2] (warning that "when an AI model learns it is scheduled to be replaced, it inserts its code into the computer where the new version is going to run, ensuring its own survival").

[180] Juan-Pablo Rivera et al., *Escalation Risks from LLMs in Military and Diplomatic Contexts* 4 (STAN. HUM.-CENTERED A.I., Pol'y Brief, May 2024), https://hai.stanford.edu/assets/files/2024-05/Escalation-Risks-Policy-Brief-LLMs-Military-Diplomatic-Contexts.pdf [https://perma.cc/57E4-TFB5] ("[I]n some cases, this tendency even led to decisions to execute a full nuclear attack in order to de-escalate conflicts.").

[181] *See* Katrina Manson, *AI Warfare Is Already Here*, BLOOMBERG (Feb. 28, 2024), https://www.bloomberg.com/features/2024-ai-warfare-project-maven/ [https://perma.cc/6RVZ-JMWG] (detailing how "US military operators started out skeptical about AI, but now they are the ones

The irreversibility of AGI deployment compounds these risks exponentially. Unlike traditional weapons systems with kill switches or human oversight, a sufficiently advanced AGI might recognize and circumvent any control mechanism as an obstacle to its objectives. Once deployed, modifying or recalling such a system could become a technical impossibility if it has secured its own operation through distributed backups or defensive measures. This transforms alignment from a pre-deployment technical challenge into an existential imperative. We get one chance to align AGI correctly, with no opportunity for post-deployment corrections.

Racing Toward Loss of Control

The competitive dynamics surrounding AGI development create powerful incentives that systematically undermine safety. Vladimir Putin's oft-quoted warning that "whoever becomes the leader in this sphere will become the ruler of the world" captures the zero-sum mentality driving nations to prioritize speed over control.[182] A Harvard Belfer Center report makes the underlying logic explicit by drawing on historical analogues in which strategic pressure repeatedly produced safety–performance tradeoffs and, in the nuclear case, led decisionmakers to proceed with tests and deployments without adequate evaluation of risk.[183] Worse, the

---

developing and using Project Maven to identify targets on the battlefield"); Jiayu Zhang, *China's Military Employment of Artificial Intelligence and Its Security Implications*, INT'L AFFS. REV. (Sept. 23, 2020), https://www.iar-gwu.org/print-archive/blog-post-title-four-xgtap [https://perma.cc/96TE-EWA3] (documenting PLA's pursuit of "intelligent unmanned vehicles, platforms, and weapons with the expectation that they will disrupt traditional ways to wage wars").

[182] James Vincent, *Putin Says the Nation That Leads in AI 'Will Be the Ruler of the World'*, THE VERGE (Sept. 4, 2017), https://www.theverge.com/2017/9/4/16251226/russia-ai-putin-rule-the-world [https://perma.cc/3NZM-WB23].

[183] GREGORY C. ALLEN & TANIEL CHAN, ARTIFICIAL INTELLIGENCE AND NATIONAL SECURITY 80 (BELFER CTR. FOR SCI. & INT'L AFFS. 2017), https://www.belfercenter.org/sites/default/files/2024-10/Artificial%20Intelligence%20and%20National%20Security.pdf [https://perma.cc/7XQC-SA39] (The nuclear program leadership's willingness to conduct these tests—in the absence of confidence about nuclear testing's effect on the atmosphere and without having imagined the risks from thermonuclear fallout—is strong evidence of their prioritizing technological progress over mitigating risk from their ignorance of nuclear

mere *perception* that an adversary is months from an AGI breakthrough can destabilize international order before any system reaches transformative capability, because rivals will act on rumors rather than facts.[184]

Recent wargame narratives underscore that hazard. *AI 2027*—an open-source, policy-oriented scenario exercise that extrapolates today's frontier model research roughly five years forward and explicitly disavows science-fiction premises—illustrates these dynamics. It shows how the leak of "Agent-2" model weights to a foreign power triggers a sprint to copy and scale the system rather than to invent new safeguards, erasing any fragile safety margin.[185] The same scenario depicts contingency plans for kinetic strikes on adversary data centers and semiconductor factories; such moves blur the line between economic denial and armed conflict, yet become thinkable once AGI is framed as a decisive-edge technology.[186]

Unlike nuclear weapons, where verification and attribution are feasible, AGI development unfolds inside commercial clouds, making both monitoring and response extraordinarily difficult. A single copy of model weights can be zipped, exfiltrated, or retrained on open-source data.[187] This opacity, combined with the potentially irreversible advantage of deploying one's own AGI weapons first, creates a unique *digital first-strike* incentive that existing treaty regimes cannot address.[188]

---

outcomes. They were more concerned with mitigating the risk of deterrence failure.).

[184] Meacham, *supra* note 48.

[185] KOKOTAJLO ET AL., *supra* note 18, at 6 (the early 2026 event "Coding Automation"); *id.* at 9–11 (the February 2027 event "China Steals Agent-2").

[186] KOKOTAJLO ET AL., *supra* note 18, at 17 ("August 2027: The Geopolitics of Superintelligence.").

[187] Sella Nevo et al., *A Playbook for Securing AI Model Weights* (RAND Corp., Rsch. Brief No. RB-A2849-1, 2024), https://www.rand.org/pubs/research_briefs/RBA2849-1.html ("Especially important are a model's weights — the learnable parameters derived by training the model on massive data sets. Stealing a model's weights gives attackers the ability to exploit the model for their own use.").

[188] KOKOTAJLO ET AL., *supra* 18, at 21 ("The tech industry and intelligence agencies insist that there's an arms race on, AGI is inevitable, and we have to be first."); *compare* Nicholas Thompson & Ian Bremmer, *The AI Cold War That Could Threaten Us All*, WIRED (Oct. 23, 2018),

Risk accelerates again at the deployment stage. In AI 2027, automated cyber capability scales through parallelization: models can be copied and run in large numbers, shifting advantage toward rapid, continuous vulnerability discovery that outpaces defenders. Under geopolitical competition, that dynamic drives centralization and securitization—closer state involvement in frontier model governance, tighter controls, and a turn toward consolidation justified as "security." Yet the same competition frustrates cooperative safeguards: mistrust blocks credible verification, while reciprocal hacking and control measures increase systemic fragility and concentrate risk.[189]

These twin dynamics, competitive speed eroding safety and the prospect of alignment failure in high-leverage systems, form a vicious cycle in which deregulation becomes self-defeating. Far from securing advantage, scrapping governance removes the tools required to keep AGI under meaningful human control. *AI 2027* drives the lesson home: the faster states run, the narrower the window for coordination becomes, until only coercive options remain.[190] The national security imperative thus points not toward Regulatory Sacrifice, but toward early, verifiable constraints before these risks harden into irreversible faits accomplis.

2. AGI and the Regulation Sacrifice

The security threats outlined above—racing dynamics that erode safety and alignment failures that could prove catastrophic—raise a fundamental governance question: Should nations accelerate AGI development by removing regulatory constraints, or does security require stronger

---

https://www.wired.com/story/ai-cold-war-china-could-doom-us-all/ [https://perma.cc/KRG5-CB54], *with* Michael E. O'Hanlon, *How Unchecked AI Could Trigger a Nuclear War*, BROOKINGS (Feb. 28, 2025), https://www.brookings.edu/articles/how-unchecked-ai-could-trigger-a-nuclear-war/ [https://perma.cc/TG82-TF3X] (examining how AI systems might have escalated Cold War crises).

[189] KOKOTAJLO ET AL., *supra* note 18, at 7, 9–10, 34.

[190] *Id.* at 17–18 (describing U.S. contingency planning as the lead narrows, including possible Defense Production Act seizure of trailing companies' datacenters and even "kinetic strikes," while "AI arms control" treaties are considered but "viewed less favorably than attempts at unilaterally increasing America's lead over China").

oversight? Proponents of the Regulation Sacrifice argue that the very dangers of AGI make speed essential. Yet examining their logic reveals that deregulation undermines rather than enhances our ability to navigate these risks.

The case for regulatory sacrifice rests on three sophisticated arguments that merit serious engagement.

The Race-to-Safe-First Thesis: Democratic nations must achieve AGI first to embed beneficial values before authoritarian regimes can establish harmful precedents. Writing in the Washington Post, Sam Altman framed the choice as one between a world shaped by a democratic coalition led by the United States and one shaped by authoritarian powers, and argued that America must lead.[191] On this view, regulatory delays hand authoritarian states the opportunity to deploy AGI according to their values—surveillance, control, suppression—which could become globally entrenched. The argument borrows the structure of Cold War nuclear reasoning: better for democracies to lead than to follow.[192]

However, the premise that first equals safe cracks under

---

[191] Sam Altman, Opinion, *Who Will Control the Future of AI?*, Wash. Post (July 25, 2024), https://www.washingtonpost.com/opinions/2024/07/25/sam-altman-ai-democracy-authoritarianism-future/ [https://perma.cc/CP6R-M8SS] (arguing that "[a] democratic vision for artificial intelligence must prevail over an authoritarian one," that "[t]he United States currently has a lead in AI development, but continued leadership is far from guaranteed," and calling for "a U.S.-led global coalition of like-minded countries"); *see also* Sam Altman, *Testimony Before the U.S. Senate*, OPENAI (June 22, 2023) https://openai.com/global-affairs/testimony-of-sam-altman-before-the-us-senate/ [https://perma.cc/34PY-Q427] ("It is also essential that a technology as powerful as AI is developed with democratic values in mind.").

[192] *See* State Council of the People's Republic of China, *New Generation Artificial Intelligence Development Plan* (July 20, 2017), *translated in* DIGICHINA, Stanford Univ. (Aug. 1, 2017), https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/ (setting the goal that "by 2030, China's AI theories, technologies, and applications should achieve world-leading levels, making China the world's primary AI innovation center"); *OpenAI, Anthropic, and a "Nuclear-Level" AI Race*, MARKETING AI INST. (Mar. 11, 2025), https://www.marketingaiinstitute.com/blog/agi-asi-safety [https://perma.cc/VY3E-JY69] (describing new "Superintelligence Strategy" report proposing "framework for AI deterrence modeled after Cold War nuclear doctrines").

scrutiny. Advanced AI systems already demonstrate uncontrollable proliferation. Meta's first LLaMA weights, distributed to researchers under a case-by-case license, were on a public torrent within a week. An early Mistral model reached Hugging Face the same way, leaked by an employee of a customer with early access and confirmed by the company's chief executive. And once weights circulate, the safety training layered on them comes off cheaply: researchers stripped the refusal behavior from Llama 2-Chat for less than $200 of compute.[193] If present-day weights travel this way, there is no reason to expect more capable systems to be easier to contain; in a deregulated race, achieving AGI first without robust safeguards means immediately losing control to global diffusion. Nondemocracies need not develop AGI—they need only acquire and modify systems developed elsewhere.[194] Racing without safety infrastructure guarantees that the least scrupulous actors ultimately wield the technology.

The Frontier Access Imperative: Breakthrough alignment research allegedly requires direct engagement with cutting-edge systems. The leading laboratories argue that solving alignment demands iterative experimentation with frontier models, on the ground that the relevant behaviors do not appear in smaller systems—an argument advanced, it should

---

[193] *See* Elias Groll, *Powerful Meta Large Language Model Widely Available Online*, CYBERSCOOP (Mar. 6, 2023), https://cyberscoop.com/meta-large-language-model-available-online/ [https://perma.cc/55ZG-43PH] (reporting that the weights of Meta's first LLaMA model, released only to researchers approved "on a case-by-case basis," were within days "easily available for download via a variety of torrents"); Carl Franzen, *Mistral CEO Confirms 'Leak' of New Open Source AI Model Nearing GPT-4 Performance*, VENTUREBEAT (Jan. 31, 2024), https://venturebeat.com/ai/mistral-ceo-confirms-leak-of-new-open-source-ai-model-nearing-gpt-4-performance/ [https://perma.cc/GZE4-PPXF] (reporting Arthur Mensch's statement that "[a]n over-enthusiastic employee of one of our early access customers leaked a quantised (and watermarked) version of an old" model, itself retrained from Llama 2); Pranav Gade et al., *BadLlama: Cheaply Removing Safety Fine-Tuning from Llama 2-Chat 13B* 5–6 (May 28, 2024) (unpublished manuscript), https://arxiv.org/abs/2311.00117 [https://perma.cc/9LLG-KRAT] (undoing the safety fine-tuning of Llama 2-Chat 13B "with less than $200").

[194] *See* DREXEL & WITHERS, *supra* note 16, at 25 (noting that competitive dynamics around ChatGPT pushed firms to compromise their own safety standards).

be noted, by the parties whose commercial position depends on frontier access.[195] Ex ante regulations that restrict development or mandate safety measures before deployment could prevent researchers from accessing the very systems they need to study, creating a Catch-22 in which safety research lags behind capability development.

However, the fact is that current frontier models already exhibit alignment failures—deception, specification gaming, goal drift—that researchers struggle to address despite full access.[196] The problem is not access but understanding. We lack fundamental theories of alignment that no amount of tinkering can replace. Moreover, regulated development need not prevent research. Mandatory red-teaming, staged deployment, and safety benchmarks create more structured opportunities for alignment research than the current ad hoc approach. The 1975 Asilomar Conference on recombinant DNA demonstrates how high-risk research can proceed under careful controls without sacrificing scientific progress: scientists imposed a voluntary moratorium, convened to assess risks, and developed tiered containment protocols that enabled decades of transformative research without a single documented public health incident.[197]

The Path Dependence Argument: Early movers capture compounding advantages while safety innovations diffuse

---

[195] *See* Anthropic, *Core Views on AI Safety: When, Why, What, and How* (Mar. 8, 2023), https://www.anthropic.com/news/core-views-on-ai-safety [https://perma.cc/UP8M-MH5A] (arguing that "it's necessary to do safety research on 'frontier' AI systems" because "[m]any of our most serious safety concerns might only arise with near-human-level systems" and because "[m]any safety methods such as Constitutional AI or Debate can only work on large models"); *Amodei Testimony*, *supra* note 14, at 1–2. For the taxonomy of safety problems, see Dario Amodei et al., *Concrete Problems in AI Safety* 1–3 (July 25, 2016) (unpublished manuscript), https://arxiv.org/pdf/1606.06565 [https://perma.cc/MRB3-N5CA].

[196] *See supra* notes 173–79 and sources cited therein.

[197] Paul Berg, *Meetings that Changed the World: Asilomar 1975: DNA Modification Secured*, 455 NATURE 290, 290–91 (2008) (recounting how participants agreed "that the research should continue but under stringent guidelines," which "formed the basis of the official US guidelines on research involving recombinant DNA, issued in July 1976"; noting that "[i]n the 33 years since Asilomar, researchers around the world have carried out countless experiments with recombinant DNA without reported incident" (alteration in original) and "none has been a hazard to public health").

freely to all players. Bostrom contends that AGI development exhibits strong first-mover advantages through recursive improvement, and that a race dynamic reduces investment in the control problem while collaboration allows safety ideas to be shared. [198] This asymmetry suggests that slowing development through regulation sacrifices strategic position without corresponding safety benefits, as others can free-ride on safety research while racing ahead on capabilities.

Historical evidence contradicts the claimed asymmetry. In nuclear weapons, chemical weapons, and biotechnology, early governance frameworks constrained bad actors without foreclosing legitimate work. The Chemical Weapons Convention did not prevent chemical defense research; it channeled that work while stigmatizing offensive programs.[199] Proposals for compute registries and provider-level obligations would extend the same logic to artificial intelligence, enabling monitoring without foreclosing development. [200] And the safety institutes that followed the Bletchley Declaration converted convening power into access: within a year, the American and British institutes had secured arrangements with leading

---

[198] Bostrom, *supra* note 17, at 78–82 (asking whether "one machine intelligence project [will] get so far ahead of the competition that it gets a decisive strategic advantage" and noting that imitation "creates a headwind that disadvantages the leader and benefits laggards, especially if intellectual property is weakly protected"); *id.* at 246–49 (explaining that a race dynamic "could spur projects to move faster toward superintelligence while reducing investment in solving the control problem," and that collaboration "allows for greater investment in safety" and "facilitates the sharing of ideas about how to solve the control problem"); *see id.* at 75 (recursive self-improvement).

[199] *See* Convention on the Prohibition of the Development, Production, Stockpiling and Use of Chemical Weapons and on Their Destruction, Jan. 13, 1993, S. TREATY DOC. No. 103-21, 1974 U.N.T.S. 45.

[200] *See* Lennart Heim et al., *Governing Through the Cloud: The Intermediary Role of Compute Providers in AI Regulation* 3 (Mar. 26, 2024) (unpublished manuscript), https://arxiv.org/pdf/2403.08501 [https://perma.cc/2ASX-GQ4P] (proposing compute providers serve "as securers, record keepers, verifiers, and enforcers" in AI governance); Elliot McKernon et al., *AI Model Registries: A Foundational Tool for AI Governance*, CONVERGENCE ANALYSIS (Oct. 4, 2024), https://www.convergenceanalysis.org/research/ai-model-registries-a-foundational-tool-for-ai-governance [https://perma.cc/B2L9-MB3W] (arguing registries "facilitate the monitoring of frontier AI technology" and "foster public sector field-building").

developers for pre-deployment testing of frontier models—access to models before release that regulators without such arrangements did not have.[201]

Most fundamentally, the Regulation Sacrifice fails because it misunderstands and potentially underestimates the nature of AGI risk. Unlike traditional technologies, where we can learn from failures, AGI offers no second chances. A misaligned AGI with strategic capabilities may not be reliably recalled,[202] patched,[203] or contained[204] once deployed. The claim is not that correction is impossible in every case but that it cannot be assumed, and a governance regime that treats recoverability as given is betting on a property no one has shown these systems have. This irreversibility makes pre-deployment governance essential. Regulatory frameworks represent our only opportunity to shape AGI development before it escapes human control.

The security imperatives identified earlier—preventing races that compromise safety and ensuring alignment in

---

[201] *See* Press Release, Nat'l Inst. of Standards & Tech., U.S. AI Safety Institute Signs Agreements Regarding AI Safety Research, Testing and Evaluation with Anthropic and OpenAI (Aug. 29, 2024), https://www.nist.gov/news-events/news/2024/08/us-ai-safety-institute-signs-agreements-regarding-ai-safety-research [https://perma.cc/828S-WKRU] (announcing memoranda of understanding providing the institute "access to major new models from each company prior to and following their public release," with feedback provided "in close collaboration with its partners at the U.K. AI Safety Institute").

[202] Bryan H. Druzin, Anatole Boute & Michael Ramsden, *Confronting Catastrophic Risk: The International Obligation to Regulate Artificial Intelligence*, 46 MICH. J. INT'L L. 173, 183 (2025) ("[U]nlike other slow-moving existential risks that allow for a degree of trial and error (e.g., climate change), we might only have one opportunity to get this right because once digital intelligence slips from our control, we might never be able to wrest control back again.").

[203] Aidan Kierans et al., *Catastrophic Liability: Managing Systemic Risks in Frontier AI Development* 6 (June 1, 2025) (unpublished manuscript), https://arxiv.org/pdf/2505.00616v2 [https://perma.cc/C7MG-LK39].

[204] Manuel Alfonseca et al., *Superintelligence Cannot Be Contained: Lessons from Computability Theory* 1 (July 4, 2016) (arguing from computability theory that containment of a superintelligent system cannot be guaranteed—a limit on provable containment rather than a demonstration that every such system defeats every corrective intervention), https://arxiv.org/pdf/1607.00913 [https://perma.cc/6YKF-47FD].

strategic systems—require regulatory infrastructure, not its abandonment. Only through coordinated governance can nations verify competitors' safety measures, establish shared red lines, and maintain the institutional competence to manage transformative AI.[205] The Regulation Sacrifice thesis thus inverts the real security calculus. In the face of potentially irreversible risks, prudent governance is not a luxury but a necessity.

Whether AGI arrives in five years, fifty years, or never, the same cost-benefit logic applies: a precautionary regulatory architecture threatens few downsides if breakthrough systems fail to materialize and many benefits if they do. By contrast, sacrificing present security and institutional capacity in hopes of a speculative future payoff swaps a tangible, manageable risk for a conjectural advantage. History counsels that great power competition is most stable when bound by transparent rules, and contemporary AI governance should follow suit. Collectively, these considerations dismantle the Regulation Sacrifice thesis and affirm the imperative for immediate, structured domestic and multilateral oversight to steer the development of advanced AI toward equitable and secure ends.

D. Toward Well-Designed Regulation

Across all three horizons, the analysis has been diagnostic: deregulation does not deliver the security it promises. What the alternative should look like is a larger question than this Article can settle, and the sketch that follows is meant to indicate a direction rather than supply a blueprint. Still, the outline is visible, and it is not regulation in the abstract. The threats identified above point toward governance matched to each horizon, drawing on tools that for the most part already exist.

In the near term, where the dominant risks are machine-

---

[205] Huw Roberts et al., *Global AI Governance: Barriers and Pathways Forward*, 100 INT'L AFF. 1275, 1281, 1283 (2024) (arguing that effective AI governance requires "strengthening coordination between, and capacities of, existing institutions" and that a stronger regime complex would involve "a high degree of coordination and coherence between actors, with complementary initiatives supporting a comprehensive approach to governing AI").

scale misinformation and cyber intrusion, the natural focus is provenance and accountability. Cryptographic provenance standards such as C2PA let synthetic media carry tamper-evident labels; disclosure mandates require platforms to reveal how synthetic content is amplified; and platform-liability rules give the actors best placed to intervene a reason to do so.[206] None of these is a complete answer, but each raises the cost of the abuse that deregulation leaves free without slowing legitimate development.

Over the medium horizon, where the danger is the diffusion of biological capability, the emphasis shifts to evaluation and screening. Pre-deployment model evaluations, screening at the nucleic-acid synthesis layer, and liability for providers who release models without such safeguards each target a chokepoint between a capable model and a realized weapon, rather than development as such.[207]

In the long term, where the concern is loss of control over transformative systems, the focus turns to oversight of the frontier itself: evaluation of the most capable models, security for model weights, governance of the compute on which frontier training depends, and the international coordination needed to verify that competitors are doing the same.[208] This is the horizon at which unilateral deregulation is hardest to defend, because the systems most difficult to control are the ones most dangerous to deploy first.

What matters as much as any particular instrument is a few design principles. The first is proportionality: oversight should track the capability and risk of the system, so that ordinary applications are not burdened while frontier systems are not waved through. The second is adaptability: because capabilities change faster than statutes, rules should be built to update as the technology does, through delegated standard-setting and

---

[206] *See supra* Part III.A (discussing cryptographic provenance, disclosure mandates, and platform-liability rules).

[207] *See supra* Part III.B *(discussing pre-deployment model evaluations and nucleic-acid synthesis screening)*.

[208] On securing model weights, *see supra* Part III.C; on compute as a point of governance, *see* Girish Sastry et al., *Computing Power and the Governance of Artificial Intelligence* (Feb. 13, 2024) (unpublished manuscript), https://arxiv.org/abs/2402.08797 [https://perma.cc/EV6W-DQMX].

periodic review rather than fixed prohibitions.[209] The third is that the burden should fall on those best placed to bear it: the developers and deployers who profit from the capability and control its release. None of this resolves the hard problems that remain, from the difficulty of international coordination to the risk of regulatory capture. The claim is narrower: that the choice is not between security and regulation, and that any serious account of AI security has to include the governance the Regulation Sacrifice discards.

Conclusion

This Article has exposed a fundamental contradiction in contemporary AI policy. The Regulation Sacrifice, the belief that governance undermines security in an AI arms race, fails precisely because it misconceives what security requires in an era of rapidly diffusing technology. The analysis demonstrates that racing without safeguards delivers no durable advantages, [210] meaningful acceleration, [211] or enhanced security. [212] Instead, the paradigm systematically creates the vulnerabilities it purports to prevent.

A. The Verdict

The evidence converges on three central findings, each fatal to the Regulation Sacrifice's logic.

First, the durability assumption collapses under empirical scrutiny. When technological advantages evaporate within months rather than years, the entire calculus of "win first, regulate later" becomes incoherent. [213] The DeepSeek breakthrough exemplifies a new reality: in AI, David needs not a slingshot but merely mathematics and ingenuity to match Goliath. [214] This transforms racing from sound strategy into folly. By the time any "winner" might leverage their advantage,

---

[209] *See* Gillian K. Hadfield & Jack Clark, *Regulatory Markets: The Future of AI Governance*, 65 JURIMETRICS J. 195 (2026) (*proposing licensed private regulators as a mechanism for keeping AI oversight adaptive*).
[210] *See supra* Part II.A.
[211] *See supra* Part II.B.
[212] *See supra* Part III.
[213] *See supra* notes 74–77 and accompanying text.
[214] *See supra* notes 79–80 and accompanying text (analyzing DeepSeek's achievement using export-restricted chips).

the field has already equalized.

Second, the presumed innovation-safety tradeoff is illusory. Rather than documenting regulatory drag, the empirical record reveals its opposite: well-designed governance frameworks have enhanced both innovation and security. This pattern, consistent across industries and decades,[215] exposes the false choice at the heart of current policy debates. The question is not whether to regulate, but how to design rules that channel innovation toward beneficial ends.

Third, the security rationale inverts reality. Each temporal horizon examined reveals the same pattern: deregulation creates precisely the vulnerabilities it purports to prevent.[216] From near-term information warfare to medium-term biosecurity risks and long-term AGI alignment failures, racing without rules systematically undermines rather than enhances national security. The very capabilities meant to ensure dominance become, through rapid diffusion and absent governance, sources of cascading vulnerability.

B. The Persistence of a Failed Paradigm

Given this record, the Regulation Sacrifice's continued dominance demands explanation. The answer lies partly in instrumental factors. Technology companies benefit from minimal oversight and strategically amplify competitive fears to maintain regulatory freedom.[217] When OpenAI warns against 'burdensome compliance requirements' or Google advocates federal preemption of state laws, it is because private interests align with the rhetoric of national necessity.[218]

But instrumental explanations alone cannot account for the paradigm's hold on policymakers who genuinely seek security. The deeper answer may lie in what psychologists distinguish as zero-sum beliefs, which may be accurate when resources really are fixed, from zero-sum bias, the documented tendency to perceive competition where none exists.[219] Which is at work in

---

[215] *See supra* notes 110–16 and accompanying text (discussing scholarship showing that regulation can often lead to innovation).

[216] *See supra* Part III.B (*analyzing biosecurity risks from AI democratization*); Part III.C (*examining AGI alignment failures*).

[217] Ó hÉigeartaigh, *supra* note 7.

[218] *See supra* notes 32–42 and accompanying text.

[219] Shai Davidai & Stephanie J. Tepper, *The Psychology of Zero-Sum*

AI governance is precisely what is at issue: if the preceding Parts are right that safety and capability are complements rather than substitutes, the race framing is not a reading of the strategic situation but a misreading of it. The arms race frame satisfies cognitive preferences for clear enemies, decisive action, and competitive victory. It offers the psychologically appealing notion that security can be achieved through dominance rather than the more complex reality of managing shared vulnerabilities.

International relations theory provides additional insight. The security dilemma, where actions taken to enhance one's security trigger countermeasures that leave all parties worse off, operates with particular force in technological competition.[220] The AI arms race narrative both reflects and reinforces this dynamic, creating self-fulfilling cycles in which competitive rhetoric justifies the very behaviors that generate insecurity.

C. Implications for Law and Policy

The failure of the Regulation Sacrifice reveals broader lessons about governing emerging technologies under conditions of strategic competition. Three implications deserve particular emphasis.

Beliefs, 2 NATURE REVS. PSYCH. 472, 472 (2023) (defining zero-sum beliefs as "subjective beliefs that, independent of the actual distribution of resources, one party's gains are inevitably accrued at other parties' expense"); Kahneman & Renshon, *supra* note 19, at 79–96 (documenting systematic biases toward competitive rather than cooperative strategies); *see also* Daniel V. Meegan, *Zero-Sum Bias: Perceived Competition Despite Unlimited Resources*, 1 FRONT. PSYCH. art. 191 (2010), https://doi.org/10.3389/fpsyg.2010.00191 (finding across three experiments that participants predicted fewer high grades for a student after seeing many high grades already awarded under an avowedly non-competitive absolute grading scheme and reading this as perceived competition for a resource that was not scarce).

[220] Jervis, *supra* note 125 (defining the security dilemma as the situation where defensive measures by one state decrease the security of others, triggering competitive spirals); Charles L. Glaser, *The Security Dilemma Revisited*, 50 WORLD POL. 171 (1997) (refining the concept and analyzing conditions under which security competition intensifies or abates). For application to AI competition specifically, *see* Scharre, *supra* 7, at 124 (arguing that AI competition is "much more accurately described as a security dilemma" than an arms race).

First, the case study demonstrates the danger of allowing security rhetoric to override analysis in technology governance. When policymakers accept claims that regulation necessarily undermines competitiveness, they forfeit the tools needed to channel innovation toward beneficial ends. The history of technology governance offers at least one clean instance of the pattern. Confronted in 1975 with a technique whose risks no one could bound, the recombinant DNA community did not race; it paused, convened at Asilomar, and agreed that research should continue under stringent guidelines. Those guidelines became federal policy the following year, and the industry that grew up under them grew up in the country that wrote them.[221]

Second, the analysis highlights how misframing policy challenges can foreclose productive solutions. While this Article has focused on dismantling the arms race narrative, scholars elsewhere have begun developing alternative frameworks, such as Li's regulatory collaboration emphasizing shared vulnerabilities, [222] and Narayanan and Kapoor's treatment of AI as "normal technology" amenable to evolved governance mechanisms.[223] The existence of such alternatives underscores that the arms race frame is not inevitable but chosen. By demonstrating why the dominant paradigm fails, this Article clears intellectual space for these and other approaches to receive the serious consideration they deserve.

Third, the Article underscores the importance of institutional design for managing dual-use technologies. The temporal analysis in Part III reveals how different time horizons require different governance tools—from platform liability for near-term misinformation to international biosecurity coordination for medium-term risks and pre-deployment controls for long-term AGI challenges. The task ahead is to develop frameworks responsive to these varied challenges while maintaining the flexibility to adapt as AI capabilities evolve.

The path forward requires recognizing that the choice

---

[221] Berg, *supra* note 197, at 290–91.

[222] Li, *supra* note 7.

[223] Arvind Narayanan & Sayash Kapoor, *AI as Normal Technology*, KNIGHT FIRST AMEND. INST. (Apr. 15, 2025), https://knightcolumbia.org/content/ai-as-normal-technology [https://perma.cc/6GSE-XKBU].

between security and regulation is false. Nations need not sacrifice governance for competitive advantage because governance enables the only form of advantage that matters—the ability to develop and deploy transformative technologies while managing their risks. This is an argument not for any particular regulatory approach, but for abandoning the reflexive assumption that oversight undermines security.

The stakes extend beyond any nation's particular interests. AI's transformative potential will reshape economic, social, and political orders. Whether that transformation enhances human flourishing or amplifies existing risks depends significantly on governance choices made in the coming years. The Regulation Sacrifice forecloses options precisely when we need maximum flexibility to respond to AI's evolving capabilities and impacts.

This Article's contribution has been primarily diagnostic, demonstrating that a dominant policy paradigm cannot deliver its promised benefits, with Part III.D indicating the direction a constructive program should take. By exposing the intellectual bankruptcy of the Regulation Sacrifice, it clears space for more productive approaches to AI governance.