Convergence Analysis

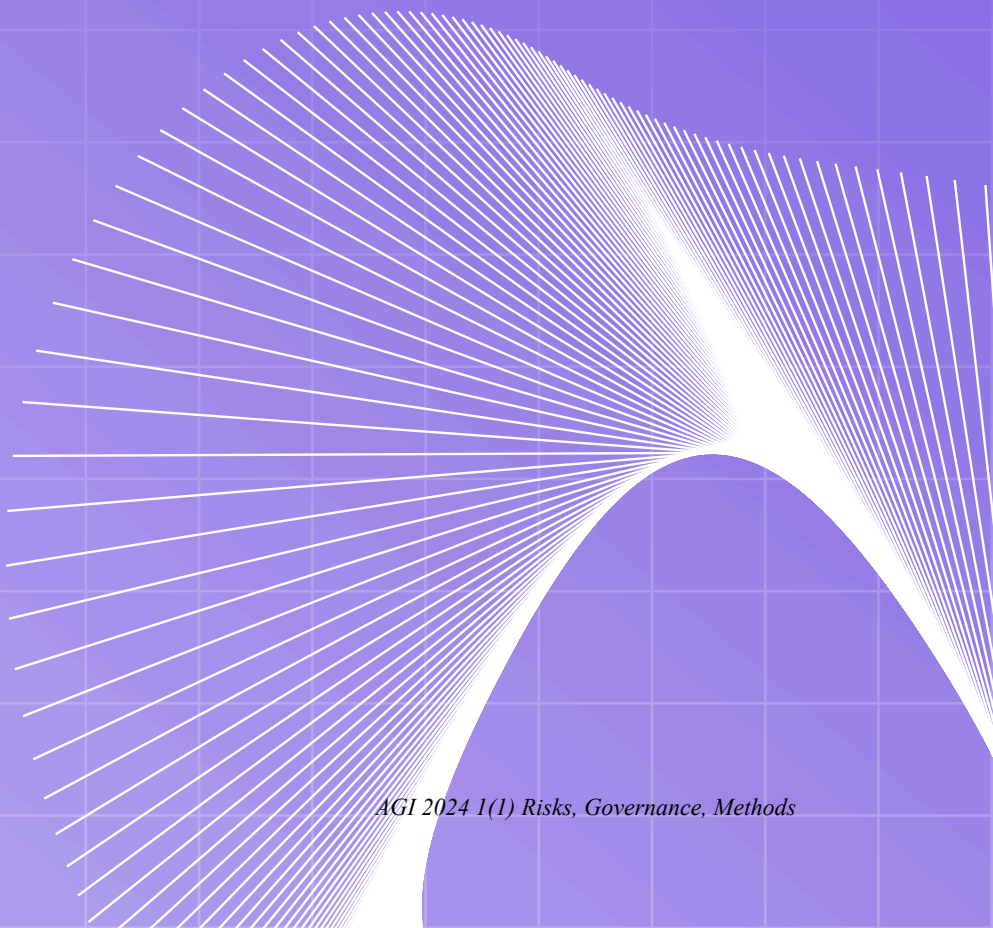
GOVERNANCE RECOMMENDATIONS REPORT

Soft Nationalization

How the US Government Will Control AI Labs

BY DERIC CHENG, CORIN KATZKE

AUGUST 2024



AGI 2024 1(1) Risks, Governance, Methods

Table of Contents

Introduction 3

Part 1: What is Soft Nationalization? 4

- Why Total Nationalization Is Not The Most Likely Model 6
- Upcoming Projects on Soft Nationalization 8

Part 2: Policy Levers for Soft Nationalization 10

- Management & Governance Mechanisms 11
- Operational Control 13
- Security & Containment Measures 16
- Security & Containment Measures 16
- Financial Ownership & Control 18

Part 3: Scenarios Illustrating Soft Nationalization 20

- US “Brain Drain” 20
- Escalation of an AI Arms Race 21
- Nationalization of Bioweapon Technologies 22

Conclusion 24

References 25

Introduction

We have yet to see anyone describe a critical element of effective AI safety planning: **a realistic model of the upcoming role the US government will play in controlling frontier AI.**

The rapid development of AI will lead to increasing national security concerns, which will in turn pressure the US to progressively take action to control frontier AI development. This process has already begun[1], and it will only escalate as frontier capabilities advance.

However, we argue that existing descriptions of nationalization[2] along the lines of a new Manhattan Project[3] are unrealistic and reductive. The state of the frontier AI industry – with more than $1 trillion[4] in private funding, tens of thousands of participants, and pervasive economic impacts – is unlike nuclear research or any previously nationalized industry. The traditional interpretation of nationalization, which entails bringing private assets under the ownership of a state government[5], is not the only option available. Government consolidation of frontier AI development is legally, politically, and practically unlikely.

We expect that AI nationalization won't look like a consolidated government-led "Project", but rather like an evolving application of US government control over frontier AI labs. The US government can select from many different policy levers to gain influence over these labs, and will **progressively pull these levers as geopolitical circumstances, particularly around national security, seem to demand it.**

Government control of AI labs will likely escalate as concerns over national security grow. The boundary between "regulation" and "nationalization" will become hazy. In particular, we believe the US government can and will satisfy its national security concerns in nearly all scenarios by combining sets of these policy levers, and would only turn to total nationalization as a last resort.

We're calling the process of progressively increasing government control over frontier AI labs via iterative policy levers soft nationalization.

AUTHOR'S NOTE

It's important to clarify that we are not advocating for a national security approach to AI governance, nor yet supporting any individual policy actions. Instead, we are describing a model of US behavior that we believe is likely to be accurate to improve the effectiveness of AI safety agendas.

Part 1: What is Soft Nationalization?

We'd like to define a couple terms used in this article:

- *Total nationalization:* The traditional meaning of "nationalization", where a government transforms private industry or organizations into a public asset, taking over full ownership and control.
- *Soft nationalization:* In contrast to total nationalization, soft nationalization encompasses a wide-ranging set of policy levers governments can use to increase control over the direction, impact, and applications of a private industry or organization. These levers may allow governments to achieve their high-level goals without taking full ownership of said entity.

We argue that *soft nationalization* is a useful model to characterize the upcoming involvement of the US government in frontier AI labs, based on our following observations:

1. Private US AI labs are currently the leading organizations pushing the frontier of AI development, and will be among the first to develop AI with transformative capabilities
2. Advanced AI will have significant impacts on national security and the balance of global power.
3. A key priority for the US government is to ensure global military and technological superiority – in particular, relative to geopolitical rivals such as China.
4. Hence, the US government will begin to exert greater control and influence over the shape, ownership, and direction of frontier AI labs in national security use-cases.

1. Private US labs are currently the leading organizations pushing the frontier of AI development, and will be among the first to develop AI with transformative capabilities.[6]

Substantial evidence points towards the current and continued dominance of US AI labs such as OpenAI, Anthropic, Google, and Meta in developing frontier AI.[7]

The strongest competitors to private US AI labs are Chinese AI labs, which have strong government support but are limited by Chinese politics[8], as well as US export controls[9] stymying access to cutting-edge AI chips.

Metrics predicting the gap between US and Chinese AI technological development vary:

- Paul Scharre estimates that Chinese AI models are 18 months behind US AI models.[10]

- Chinese AI chip development is estimated to be between 5 - 10 years behind US-driven chip development.[11] This lag will become a critical factor if the US effectively enforces export controls on AI chips.[12]

2. Advanced AI will have significant impacts on national security and the balance of global power.[13]

Upcoming Capabilities: Experts forecast that advanced AI will enable a number of capabilities that have significant implications for national security[14], such as:

- **Lethal Autonomous Weapons:** LAWs may enable vastly superior military capabilities, leading to automated warfare scenarios that may distribute decision-making beyond the direct control of humans.
- **Cyberwarfare:** AI will increase the scale, accessibility, and success of cyberattacks, which have the ability to destroy critical infrastructure, among many other consequences.

National Security Outcomes: Transformative capabilities such as these may lead to outcomes that the US would view as critically detrimental for national security[15], such as:

- **Malicious Uses:** Capabilities like these could be used against the US populace, which the US government is highly incentivized to prevent.
- **AI Arms Race:** It's likely that nation-states will race to develop military AI technologies to gain geopolitical advantages, which may increase the likelihood of international destabilization and conflict.
- **Loss of Control:** Advanced AI systems or LAWs may become extremely dangerous if they behave in unexpected ways, such as making incorrect decisions in automated warfare scenarios or developing agency.

Economic Outcomes: Additionally, advanced AI systems could also result in significant negative outcomes for the US and global economies, including:

- **Mass Unemployment:** Strong financial incentives to automate human labor may lead to rapid unemployment and dependence on AI systems.
- **Wealth Inequality:** An AI-driven economy may drastically increase wealth inequality, amplifying social instability and discontent.[16]
- **Economic Instability:** AI-driven financial trading systems may amplify flash crashes or financial instability[17], which is a major concern for the US government.

3. A key priority for the US government is to ensure global military and technological superiority.

The US government has for decades operated on the assumption that the existing world order depends on its military and technological dominance, and that it is a top national priority to maintain that order[18]. As a result, it views any challenge to this dominance as an unacceptable threat to its national security.

As AI system capabilities are demonstrated to matter for national security, the

US government will likely continue to escalate its involvement in AI technologies to maintain this superiority, even at the cost of exacerbating its AI arms race with China.[19]

A key takeaway from this observation is that the US government will *not choose to slow the pace of frontier AI development* absent international agreement that includes geopolitical adversaries like China. The US may choose to moderate certain aspects of AI that demonstrate substantial risk with little advantage, but by default it will avoid actions that inhibit American R&D in AI. Today, unilaterally pausing AI[20] development would be in opposition to the US government's current goals.

Finally, a relevant priority of the US government is maintaining social and economic stability. As has been demonstrated in numerous economic crises[21], the US is willing to take drastic action to ensure the stability of the US economy, including the takeover and bailout of multi-billion dollar private corporations[22]. Though it seems to us this priority is of less relevance to the policy levers for soft nationalization, there are plausible scenarios where the US may choose to enact these levers to preserve social and economic stability.

4. Hence, the US government will begin to exert greater control and influence over the shape, ownership, and direction of frontier AI labs in national security use-cases.

The US has already demonstrated that it is pursuing greater control over AI chip distribution – nearly a year before passing the Executive Order on AI, in 2022 the Biden administration began enforcing export controls limiting Chinese access to cutting-edge semiconductors.

We believe that this process of exerting greater control can take a wide range of possible paths, where the US progressively utilizes a wide range of *policy levers*. These levers will likely be applied to satisfy national security concerns in response to technological and geopolitical developments. Though the total nationalization of frontier AI labs is one possible outcome, we don't think it is the most likely one.

Why Total Nationalization Is Not The Most Likely Model

In a recent example of AI scenario modeling, Leopold Aschenbrenner's "Situational Awareness" describes a plausible scenario involving an extremely rapid timeline to superintelligence. He describes superintelligence's likely impact on the geopolitical landscape, concluding with the prediction that a "Manhattan Project for AI" will be soon organized by the US government. He argues that this project will consolidate and nationalize all existing frontier AI research due to the national security implications of superintelligence.

We argue that "The Project"[23] and other similar descriptions of nationalization[24] represent only a narrow subset of possible scenarios

modeling US involvement, and are not the most likely scenarios.

Total nationalization is not the most likely scenario for a few reasons:

1. American policymakers would likely believe that total nationalization would **undermine the US' technological lead in AI and broader economic interests.**
    a. Nationalizing frontier AI development could be seen as jeopardizing the pace of innovation and R&D currently driven by the private sector. It would remove competitors, incentives, and a diversity of approaches from the US AI landscape.
        i. The American model of innovation is built on free-market private competition, and is arguably one of the reasons the US is leading the AI race today.[25]
        ii. Since the 1980s, the United States has seen a significant trend towards increased private sector involvement in various industries[26], driven by factors such as:
            1. A perception among policymakers that market-based solutions can be more efficient than direct government management.
            2. The belief that private sector competition could foster greater innovation and cost reduction.
        iii. US policymakers generally endorse free-market competition on innovation and are reluctant to regulate the AI industry[27]. It would require a massive ideological shift for the US government to nationalize an industry that has critical consequences for the US economy.
2. The total nationalization of frontier AI labs would face unprecedented **practical, legal, and political challenges.**
    a. Organizations in control of frontier AI labs such as Microsoft, Google, and Meta are among the largest corporations in the world today, with market capitalizations over $1 trillion each.[28]
        i. Practically, total nationalization of these corporations is financially and logistically implausible.
        ii. Nationalization of only their frontier AI labs is more plausible. However, these corporations are developing their long-term strategies around frontier AI models, and their frontier AI labs are tightly integrated with the rest of their business.
        iii. Any form of nationalization would undermine their long-term business models, plummet shareholder value, and upend the global tech industry. It would result in massive legal and political resistance.
    b. The leading chip manufacturer Nvidia, which is a primary driver of frontier AI research by controlling 80% of the AI chip market[29], has a current market capitalization of $3 trillion[30].

   i. Many total nationalization scenarios would involve government ownership of Nvidia. However, it's challenging to imagine a legally and financially feasible pathway for the US government to gain full ownership of a public corporation of this size.

3. The US may be able to **achieve its national security goals with substantially less overhead than total nationalization** via effective policy levers and regulation.
   a. We argue that various combinations of the policy levers listed below will likely be sufficient to meet US national security concerns, while allowing for more minimal governmental intrusion into private frontier AI development.
   b. We expect that such an approach would likely be more appealing for the US government, due to the challenges of total nationalization described above.

Despite these arguments, it's still possible that the US government may eventually choose total nationalization given the right set of circumstances. We don't believe that it is possible yet to confidently predict a future set of outcomes, and that over-indexing on any scenario is a mistake.

Rather than committing to a specific model of the future, **we believe the most effective analysis today will consider a wide range of scenarios** that describe actions the US government will take in response to global circumstances. By enumerating many of the plausible scenarios regarding soft nationalization, we believe AI governance researchers can better ground our research in likely futures and design better interventions.

Upcoming Projects on *Soft Nationalization*

We are conducting scenario modeling and governance research to describe how **upcoming national security concerns will lead to greater US governmental control over frontier AI development.** We expect this research will ground AI governance discourse in a realistic understanding of plausible scenarios involving US control of frontier AI.

To execute, we're spearheading a collaborative research project with the following three parts:

1. **Describing Soft Nationalization:** Describe the policy levers and scenarios that encompass soft nationalization
2. **Conducting Further Scenario Research:** Evaluate the implications of this research on further scenario modeling topics
3. **Aligning AI Safety with Soft Nationalization:** Research how this process can be shaped to achieve the broader goals of AI safety organizations

If you're interested in collaborating or receiving updates on any of this work, shoot us a message at research@convergenceanalysis.org.

1. Describing Soft Nationalization

In the upcoming quarter, we will publish a report exploring the following:

- What types of *policy categories* (e.g. oversight, security requirements, use limitations) will governments use in order to increase control of AI labs and achieve its national security goals?
  - For each of these categories, what is the spectrum of possible policy levers that the government can pull, from least to most invasive?
  - What is the legal and practical *feasibility* of these levers?
  - What are the *externalities* of these levers?
- What *societal circumstances* will lead governments to pull on certain policy levers?
- What are *plausible scenarios* of comprehensive responses by governments to specific circumstances, combining sets of policy levers as described?
  - What levers are most plausible to be used in key situations?

Conducting Further Scenario Research

The results of our soft nationalization report will inform further scenario modeling that builds on our research, on questions such as:

- What forms of **international cooperation** are viable when national security is a primary concern of AI governance? Will we see a NATO-like alliance[31] of Western countries led by the US?
- How will soft nationalization shape **society & governments** beyond AI policy and US national security? What are plausible secondary impacts (e.g. AI race dynamics, AI safety outcomes)?
- How will soft nationalization **impact economic scenarios**? How will this impact job automation, resource allocation, and the distribution of GDP?

3. Aligning AI Safety with Soft Nationalization

A clear set of scenarios implied by soft nationalization will enable further research into how these outcomes can be shaped to achieve the broader goals of AI safety organizations, such as:

- How does soft nationalization affect the **reduction of extreme, large-scale risks?** What new strategies should be pursued? How can AI safety projects be aligned with national security concerns?
- How can we **mitigate AI race dynamics?** What policy levers slow competitive incentives, rather than accelerating them?
- What actions can we take to avoid **AI power concentration in the hands of the military-industrial complex?** What checks and balances should exist to protect society from this new hierarchy of power?
- What **economic interventions** should governments take to improve outcomes for the average person?

Part 2: Policy Levers for Soft Nationalization

We describe thirteen preliminary sets of policy levers the US government might pull to exert control over frontier AI. Each set of levers offers a series of options that afford the government increasingly more influence, on a spectrum ranging from standard regulations to more comprehensive government control.

We envision that certain policy levers will be combined and deployed by the US government given a particular societal environment. That is, we believe that given a certain scenario, the US will choose a strategy involving policy levers that exert enough control to sufficiently protect its national security, and that is also legally, politically, and practically feasible.

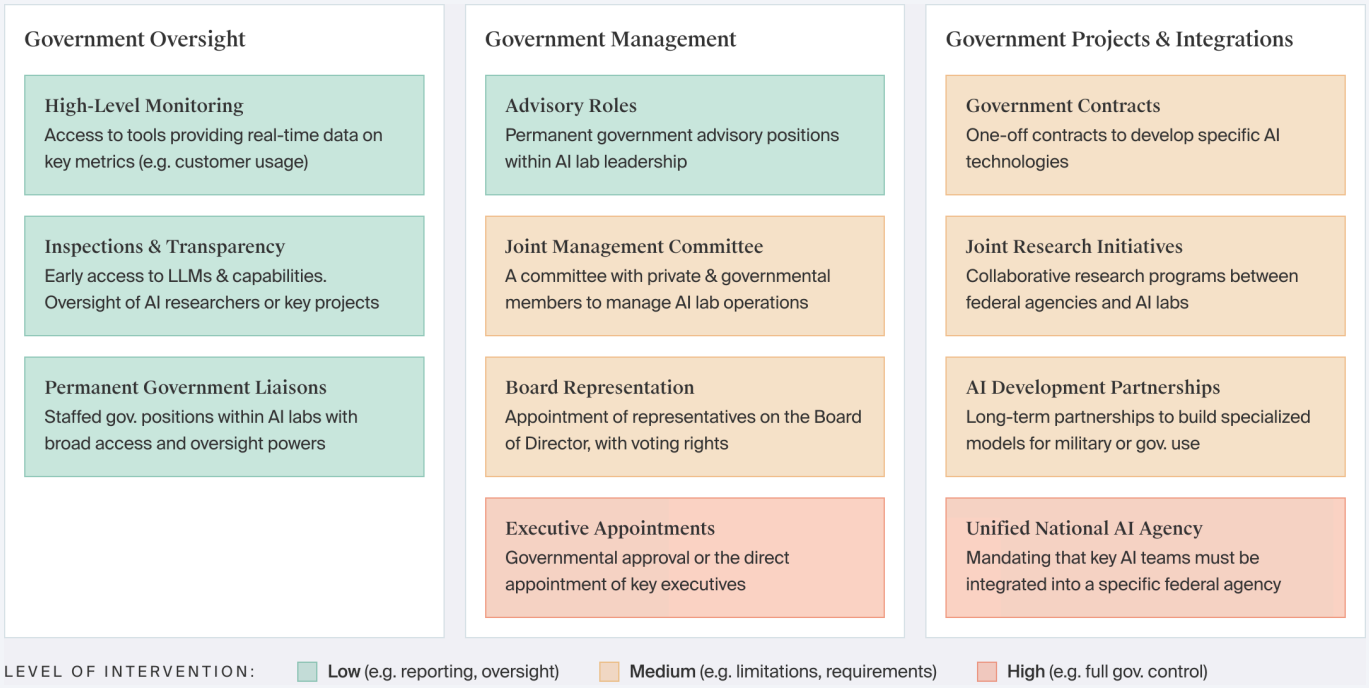
This list of policy levers is an active work in progress and will be explored in detail in a report we'll publish in the upcoming quarter, considering aspects such as:

- The legal precedent and feasibility of each policy lever
- The potential effectiveness and externalities of each lever
- The likelihood of its use by the US government in certain scenarios

AUTHOR'S NOTE

We do not advocate or recommend for the application of any of these policy levers. This section is informative in nature – it is intended solely to describe the space of plausible policy levers that may occur. In the future, we may recommend certain levers after conducting further research.

Management & Governance Mechanisms



Government Oversight

The US may seek to implement better tools to monitor the day-to-day operations of key AI labs, including policy levers such as:

- **High-Level Monitoring:** The US government may require access to comprehensive monitoring tools providing real-time data on key metrics, customer usage, or incident reporting (see: SEC Enforcement Division). It may require the creation of novel monitoring tools to meet national security goals.
- **Inspections & Transparency:** The US government may require early access to LLMs, capabilities, and results to ensure compliance with national security. It may require access to and recurring oversight of AI researchers, engineers, or key projects (see: NRC Inspections).
- **Permanent Government Liaisons:** The US may establish permanent government liaisons within AI labs, with broad access and oversight ability (see: NRC Resident Inspectors). This would create a direct touchpoint for government oversight and accountability.

Government Management

The US may seek to have direct control over the day-to-day operations of key AI labs, including policy levers such as:

- **Advisory Roles:** The US may establish permanent government advisory positions within AI lab leadership. It may require regular consultation with governmental safety or national security panels (see: Defense Science Board).
- **Joint Management Committee:** The US may require the formation of a

joint public-private management committee to control and manage AI lab operations (see: War Industries Board).

- **Board Representation:** The US may force the appointment of key representatives on the boards of directors for AI labs, with voting rights on key issues (see: Fannie Mae & Freddie Mac, 2008). This could range from a single seat to full board control.
- **Executive Appointments:** The US may require governmental approval or the direct appointment of key executives in AI labs (see: Fannie Mae & Freddie Mac, 2008).

Government Projects & Integrations

The US may seek to integrate the R&D and output of AI labs with its national security goals. This could look like any of these policy levers (in order of increasing interventionism):

- **Government Contracts:** The US could give AI labs one-off contracts to develop specific AI technologies, similar to its current relationship with Palantir.
- **Joint Research Initiatives:** The US could establish collaborative research programs between government agencies (e.g., DARPA, NSF) and AI labs on specific AI challenges (see: Human Genome Project). This would allow for shared resources and expertise while maintaining separate organizational structures.
- **AI Development Partnerships:** The government could work in partnership with AI labs to form projects building private AI models specifically for military or governmental purposes (see: Lockheed Martin's Skunk Works).
- **Unified National AI Agency:** The US could mandate that key AI labs or teams must be integrated into a specific federal agency. This would effectively merge key AI programs into the federal government.

Operational Control

Development Limitations

Reporting Requirements
Required disclosure of key AI development milestones

Pre-Approval for Large Training Runs
Government approval before initiating frontier-level training runs

Development Licensing
Licensing system to obtain permission for frontier AI development

Ban on Research & Development
Banning of large training runs or development exceeding certain thresholds

Customer Limitations

Transparency Requirements
Required disclosure of major customers and their use cases

Know Your Customer (KYC)
Standardized reporting from customers accessing advanced AI models

Government Pre-Approval
Licensing for significant customer contracts or partnerships

Customer Blacklists / Whitelists
Prohibiting frontier model access or use to certain customers or foreign entities

Government-Controlled Distribution
The US becomes the sole distributor of advanced AI technologies

Deployment / Use Limitations

Mandatory Use Case Reporting
Required reporting on the uses of advanced models and any misuse

Use Case Licensing
Licensing for the use of specific AI applications

Use Case Restrictions
Prohibiting specific high-risk applications such as autonomous weapons

Government-Controlled Distribution
The US becomes the sole distributor of advanced AI technologies

Compute Usage Limitations

Mandatory Compute Reporting
Required reporting of compute, compute clusters, or cloud compute usage

Compute Export Controls
Banning the use of US-based compute for foreign AI development

Compute Usage Caps
Setting limits on compute resources for a single project or organization

Compute Use / Ownership Permits
Permitting for the purchase / use of large-scale compute resources

Centralized Compute Allocation
A government agency that oversees compute resource allocation

Nationalizing Compute Resources
Centralized management & ownership of some / all major compute clusters

LEVEL OF INTERVENTION:

Low (e.g. reporting, oversight) **Medium** (e.g. limitations, requirements) **High** (e.g. full gov. control)

Development Limitations

The US may decide to set limitations on large-scale AI R&D for frontier AI labs:

- **Reporting Requirements:** The US may mandate the disclosure of AI development milestones such as frontier-level training runs or capability breakthroughs.
- **Pre-Approval for Large Training Runs:** The US may eventually require government approval before initiating training runs that exceed certain compute or data thresholds.
- **Development Licensing:** The US may require a licensing system for AI development, requiring labs to obtain and regularly renew government permission to work on advanced AI systems (see: FDA Development & Approval).
- **Ban on Research & Development:** In extreme scenarios, the US may unilaterally ban US AI labs from conducting training runs or development exceeding certain thresholds (see: US moratorium on gain-of-function research).

Customer Limitations

The US may require that AI labs report, vet, or restrict its customers to prevent usage of frontier AI by adversaries:

- **Transparency Requirements:** The US may require AI labs to disclose a list of major customers and their use cases to federal agencies.
- **Know Your Customer (KYC) Protocols:** The US may require strict KYC procedures for customers accessing advanced AI models, similar to financial industry standards.
- **Government Pre-Approval:** The US may require governmental pre-approval (e.g. licensing) for significant customer contracts or partnerships (see: ITAR Export Licensing).
- **Customer Blacklists / Whitelists:** The US may prohibit access or commercial use of frontier AI models by categories of foreign entities (see: Entity List).
- **Government-Controlled Distribution:** The US may establish a government agency as the sole distributor of advanced AI technologies, determining all customer relationships (see: DoE uranium management)

Deployment / Use Limitations

The US may limit the availability of specific use cases of frontier AI models:

- **Mandatory Use Case Reporting:** The US may require AI labs to report on the uses of their advanced models and any potential misuse detected (see: Suspicious Activity Reports).
- **Use Case Licensing:** The US may institute a licensing system for specific AI applications, requiring government approval for deployment in certain

use cases (see: FDA Development & Approval).

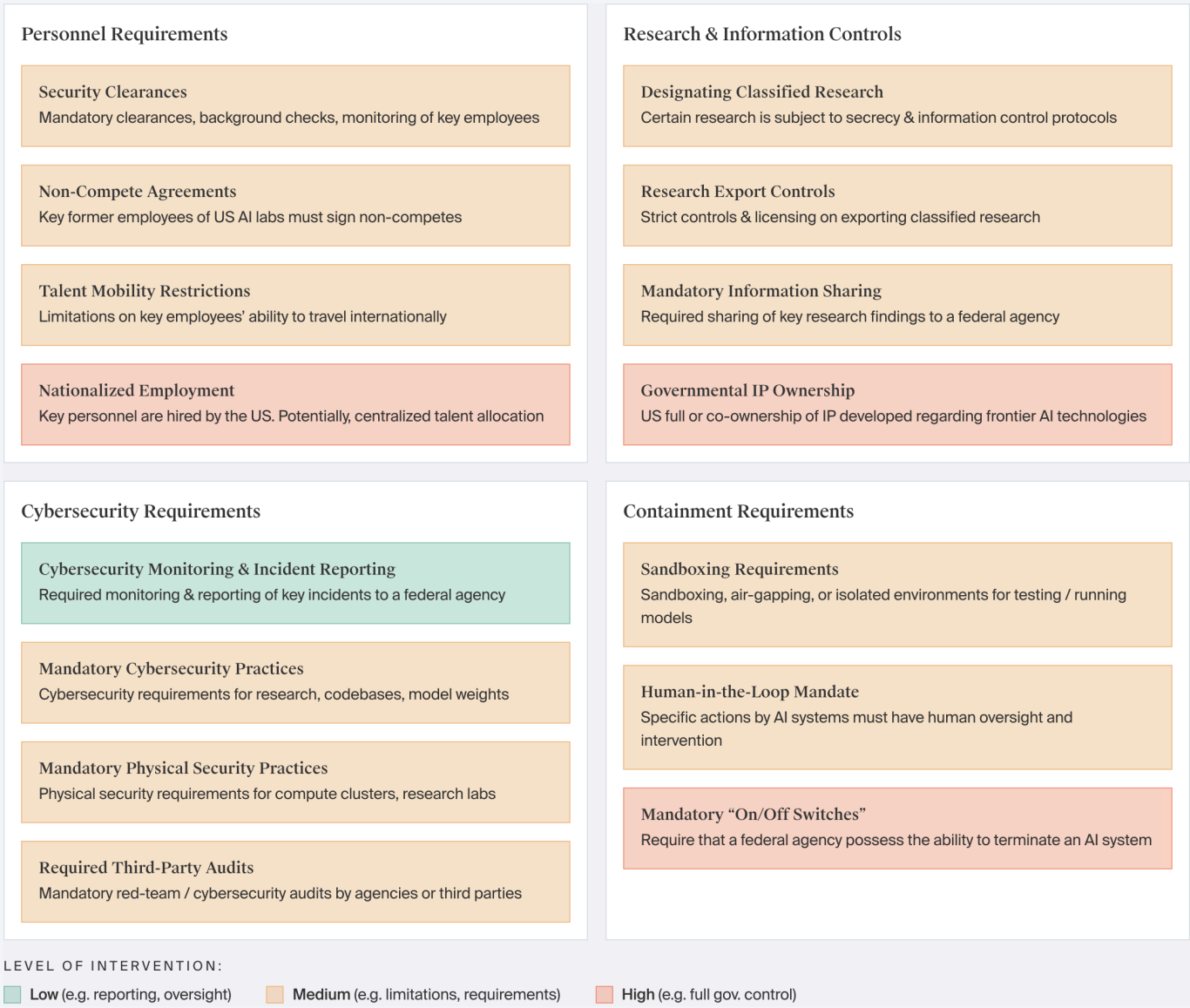
- **Use Case Restrictions:** The US may prohibit specific high-risk applications of AI, such as autonomous weapons systems or certain surveillance technologies (see: EU AI Act's Prohibited AI Practices).
- **Government-Controlled Distribution:** The US may establish a government agency as the sole distributor of advanced AI technologies, determining all use-cases and deployment methods.

Compute Usage Limitations

The US may decide to influence AI development via control over the allocation and availability of compute resources:

- **Mandatory Compute Reporting:** The US may require semiconductor manufacturers and AI labs to report their compute resources (e.g. a chip registry), cloud compute usage (e.g. KYC), or computing clusters beyond a certain size (see: Sastry et al, 2024).
- **Compute Export Controls:** The US may ban the use of US-based compute resources for foreign AI development. (see: existing US export controls).
- **Compute Usage Caps:** The US may set limits on the total compute resources that can be allocated to single projects or organizations within a given timeframe (see: Sastry et al, 2024).
- **Compute Use / Ownership Permits:** The US may implement a permit system for the purchase of cutting-edge compute resources, or for the use of large-scale AI compute resources.
- **Centralized Compute Allocation:** The US may establish a centralized government body that oversees and approves significant AI compute resource allocations.
- **Nationalizing Compute Resources:** The US may centralize ownership and management of some or all major compute clusters. AI labs may need to work with the government to obtain compute resources.

Security & Containment Measures



Security & Containment Measures

The US may seek to control key personnel within AI labs, by limiting their ability to disseminate sensitive information, to work for geopolitical rivals, or in extreme cases by requiring that they work for the US government:

- **Security Clearances:** The US may mandate security clearances for key AI researchers and engineers, similar to defense contractors. This would permit background checks, restricted information dissemination, and the ongoing monitoring of critical personnel.
- **Non-Compete Agreements:** The US may require that key former employees of US AI labs sign non-compete agreements preventing them from working for organizations affiliated with geopolitical rivals.
- **Talent Mobility Restrictions:** The US may put limitations on AI researchers' ability to travel internationally, similar to restrictions on nuclear scientists (see: Manhattan Project travel restrictions).

- **Nationalized Employment:** Key personnel from AI labs may be hired / managed by the government. In the extreme case, the US may require a centralized, government-run allocation of key AI talent. Such an unprecedented policy would involve conscription.

Research & Information Controls

The US may seek to control the classification or distribution of AI research developed by private AI labs:

- **Designating Classified Research:** The US may designate certain AI research to be classified, subject to government secrecy and information control protocols (see: US stealth technology classification).
- **Research Export Controls:** The US may place strict controls on exporting classified research, with associated licensing requirements (see: US cryptography export controls, 1950s).
- **Mandatory Information Sharing:** The US may require sharing of key research findings, including proprietary information, to a federal agency. This would allow the government to monitor and determine ownership of critical IP. This may be accomplished via licensing (see: FDA Development & Approval).
- **Governmental IP Ownership:** The US could require co-ownership or full ownership of intellectual property developed by AI labs, or developed under contract with the US government. This would enable federal legal control over key innovations, and may come with classification requirements.

Cybersecurity Requirements

The US may require specific digital or physical cybersecurity practices for highly capable AI models to protect against malicious exploitation:

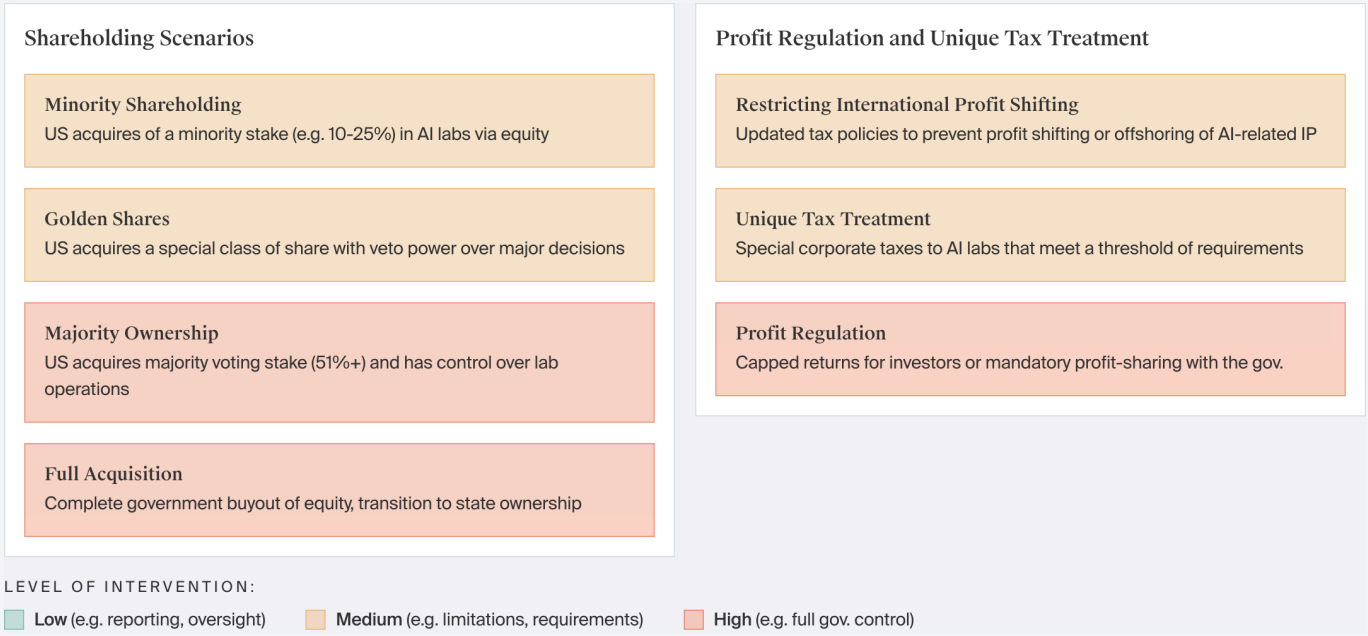
- **Cybersecurity Monitoring & Incident Reporting:** AI labs may be required to report any incidents detected to a federal agency, or to have specific monitoring requirements (see: DFARS Clause 252.204-7012).
- **Mandatory Cybersecurity Practices:** The US may require that AI labs must comply with specified cybersecurity practices to secure AI research, codebases, or model weights (see: DFARS Clause 252.204-7012)[32].
- **Mandatory Physical Security Practices:** Similar requirements may be required for physical access to key AI labs or compute clusters (see: NRC Physical Protections).
- **Required Third-Party Audits:** AI labs may be required to undergo red-team cybersecurity audits by governmental agencies (see: FISMA annual audits).

Containment Requirements

The US may require certain practices that allow AI labs or federal agencies to protect, contain, or restrict deployed AI models:

- **Sandboxing Requirements:** The US may mandate the use of sandboxing, air-gapping, or isolated environments during testing or runtime, to prevent risks such as autonomous replication or hacking (see: cybersecurity for nuclear power plants).
- **Human-in-the-Loop Mandate:** The US may require that specific actions taken by AI systems must have human oversight and intervention capabilities. The human-in-the-loop may need to be certified or work for the government.
- **Mandatory "On/Off Switches":** The US may require that a federal agency possess the ability to terminate an active advanced AI system (see: SEC-mandated "circuit-breakers").

Financial Ownership & Control



Shareholding Scenarios

The US government may consider acquiring stakes of private AI labs, achieving control through market-based mechanisms.

- **Minority Shareholding:** The US may acquire a minority stake (e.g. 10-25%) in key AI labs through stock purchases or capital injections. This gives the government some influence over the direction of the company.
- **Golden Shares:** The government may require the creation of a special class of share with veto power over major decisions, similar to "golden shares" used in privatizations. This may allow for the blocking of actions deemed against national interests.
- **Majority Ownership:** If the US were to acquire a majority voting stake (51%+) in AI labs, it would have effective control over operations and strategy while maintaining some private investment (see: General Motors, 2009).

- **Full Acquisition:** A complete government buyout of a company's equity would repay investors and reduce pushback during a transition to a fully state-owned enterprise (see: Conrail, 1976).

Profit Regulation and Unique Tax Treatment

It's plausible that leading AI labs may eventually control a sizable percentage of the revenue and valuation of private companies in the US. If this were the case, the US may seek to treat these leading AI labs uniquely from traditional corporations in pursuit of more equitable or economically beneficial outcomes, using levers such as:

- **Restricting International Profit Shifting:** The US may update its tax policies to prevent AI labs from engaging in traditional multinational corporation techniques, such as profit shifting or offshoring of AI-related IP.
- **Unique Tax Treatment:** The government could apply a certain set of corporate taxes specifically to AI labs that meet its threshold of requirements, such as an "AI Windfall Tax".
- **Profit Regulation:** The government could cap returns for private investors or mandate profit-sharing with the government, outside of traditional tax structures via custom regulation.

Part 3: Scenarios Illustrating Soft Nationalization

In this section, we describe a few preliminary scenarios in which the US exerts control over frontier AI development in response to national security concerns. For each scenario, we illustrate broad strokes of the circumstances that may occur. Then, we describe a plausible package of "soft nationalization" policy levers that the US would be likely to deploy as a comprehensive strategic response.

We present three scenarios with three different "levels" of relative governmental control: low, medium, and high. We will be exploring scenarios such as these in more detail via a report we'll publish in the upcoming quarter.

AUTHOR'S NOTE

It's important to note that these are hypothetical, illustrative scenarios to demonstrate that our model of soft nationalization may be an effective tool for describing US national security concerns. We do not propose that any of these scenarios are likely to happen, nor do we advocate for any of the suggested policy levers. We don't necessarily believe securitization is the ideal outcome, and that there are still possible scenarios involving international cooperation.

US "Brain Drain"

In early 2027, China and Saudi Arabia launch motivated, well-funded governmental initiatives to compete in AI technological superiority. In particular, one key branch of their initiative focuses on financial compensation - they offer hugely lucrative compensation packages for top AI researchers, with yearly salaries in the tens of millions, paid upfront. US AI labs are unable to compete with these offers, as most of the value of their compensation packages is in equity and illiquid. The US government does not offer similarly competitive packages.

These initiatives create a wave of talent migration, with hundreds of top AI researchers leaving for well-paid opportunities in countries the US considers to be geopolitical rivals. The exodus raises alarm in both Silicon Valley and Washington about maintaining US technological leadership in AI. In particular, the US government is concerned that top researchers are moving from capitalist, private AI applications to state-organized AI initiatives, which may

conflict with US geopolitical goals.

US Governmental Response:

- The US implements limited Talent Mobility Restrictions for key AI researchers, mandating that they work for US-based organizations and do not travel to certain countries (such as China and Saudi Arabia).
- The US sets up Permanent Government Liaisons with key AI labs. Initially, these government liaisons are tasked with identifying key AI researchers with exceptional talent or cutting-edge knowledge of AI development, to enforce the new talent mobility restrictions.
- The US increases funding for Joint Research Initiatives conducted in collaboration with top AI labs. These projects funnel millions of dollars in upfront compensation to key AI researchers, and redirect focus from free-market AI applications to projects aligned with US governmental interests.

Escalation of an AI Arms Race

In late 2029, US intelligence agencies obtain credible information that China has made significant breakthroughs in AI-enabled autonomous weapons systems. Satellite imagery and intercepted communications suggest that China is developing swarms of AI-controlled drones capable of coordinated combat operations without human intervention. These developments threaten to upset the global military balance, allowing the Chinese military to break through missile & air defense systems and undermining US & Taiwanese defensive capabilities. The news leaks to the press, causing public alarm and intensifying the ongoing debate about lethal autonomous weapons. The US is pressured to respond, fearing that China's advancement could embolden it to take more aggressive actions against Taiwan.

These developments occurred because China has been pursuing a tight-knit integration of its AI research labs and the Chinese defense industry, pouring tens of billions into military AI technologies. In comparison, the US government has been relatively hands-off on AI, preferring to fund exploratory research initiatives with AI labs rather than directly overseeing the development of cutting-edge AI technologies. As a result, the US is now behind in developing similar lethal autonomous weapons.

The US government recognizes that its approach to AI technologies has left it flat-footed relative to its geopolitical rivals, risking its position as the leading superpower. It commits to integrating frontier AI labs and technologies more directly into governmental initiatives and the defense industry.

US Governmental Response:

- The US invests heavily into scaling an **AI Development Partnership** developed in close collaboration with private AI labs. It requires that labs dedicate substantial resources to military AI development.
- The US mandates **Security Clearance Requirements** for key AI

researchers and engineers working on frontier AI model development and projects related to defense.

- The US establishes strict **Research Export Controls**, limiting the distribution of key research developments with actors from non-allied nation-states. It restricts specific forms of collaboration and communication related to AI research.
- The US establishes governmental **Board Representation** on the boards of directors for key AI labs. These individuals are tasked with ensuring that the output of AI labs accelerates US defense projects, and that key AI developments are secured in service of US national security interests.
- The US begins enforcing a system of **Use / Ownership Permits** for cutting-edge compute resources (e.g. AI chips). It finds that the existing **Compute Usage Controls** (as initiated by Biden's Oct 7 Export Controls) have been ineffective at reducing chip smuggling, and decides to strengthen its limitations on who can use next-gen AI chips to further reduce China's military AI research capabilities.

Nationalization of Bioweapon Technologies

In 2035, significant and disturbing developments at a new biotech startup occur. A novel AI virus modeling technique for vaccine development has the side effect of allowing lab researchers to easily develop bioweapons of unprecedented lethality and specificity. The AI system, trained on vast datasets of genetic and epidemiological information, can design viruses tailored to target specific ethnic groups or even individuals based on their genetic makeup. These viruses are relatively feasible to produce, and knowledge of the design of these viruses would permit any of 100+ research labs worldwide to easily create such a pathogen.

The US government determines that the capabilities of this biotech startup are too risky to permit for a private corporation. Furthermore, it believes that any further research into this novel virus modeling technique is too dangerous to permit, as it could easily lead to targeted pandemics. It moves to nationalize this biotech startup fully to prevent any further consequences, and passes legislation prohibiting private research and development into similar virus modeling techniques.

US Governmental Response:

The US government performs what we might consider a **Full Acquisition** of the specific biotech startup described above.

- Financially, the US **Purchases All Existing Equity** and pays out the current valuation to existing shareholders.
- The US **Nationalizes Employment** of all personnel currently within the biotech startup. It forces key AI researchers in this startup to have mandatory **Security Clearances** and requires **Talent Mobility**

Restrictions, similar to key government employees today.

- The biotech startup and its employees are brought into a **Unified National AI Agency**, intended to securely conduct R&D on defense and national security AI topics.
- It restricts usage of the key technologies produced by this biotech startup to have **Government-Only Access**.

Outside of this biotech startup, the US government moves quickly to create stringent national (and international) restrictions on research regarding this set of AI virus modeling techniques:

- It implements a nationwide **Ban on Research and Development** related to this AI virus modeling technique.
- It simultaneously implements a **Licensing System** for specific categories of biochemical research that are similar or related to this set of techniques. Large-scale AI model training runs by a licensed biotechnology company in this domain of research must have a **Use Permit** before proceeding.
- It claims **Ownership of all IP** related to this specific AI virus modeling technique, as well as related biochemical research.
- It **Restricts the Use Case** of private AI biotechnology models, preventing the commercial usage of models that may allow parties to progress in developing such virus modeling techniques.

These two sets of drastic actions significantly deter US private companies from undertaking any further R&D in this area of virus and pathogen modeling. The full nationalization of a private company signals that the US is likely to take similar actions in the future.

Conclusion

National security concerns suggest the US will exert more control over frontier AI development. However, predictions of a "Manhattan Project for AI" are reductive and misleading. The US isn't likely to "nationalize" frontier AI development, at least in the sense of all at once bringing it under full public ownership and control. Doing so would be legally, politically, and practically challenging, and it could ultimately undermine the US' technological lead in AI.

Instead, we propose that the US government's control over frontier AI is likely best modeled by our framework of "soft nationalization." According to this framework, the US will exert progressively greater power over frontier AI development as national security concerns arise by employing several different policy levers. The options described by these levers constitute a spectrum from "soft touch" regulation to de facto government ownership.

This model assumes that the US will act to preserve its national security. However, exactly which combinations of options across policy levers the US will choose depends on the contingencies of global and domestic technopolitics, as well as balancing goals other than national security.

We hope our model will enable the evaluation of AI safety agendas across realistic scenarios of US involvement, and encourage further related research. In upcoming work, we intend to more rigorously describe the policy levers the US will choose to exercise such control, and the scenarios that will cause the US to deploy them.

References

1 US Semiconductor Export Controls

2 Let's nationalize AI. Seriously. - POLITICO

3 IV. The Project - Situational Awareness by Leopold Aschenbrenner

4 Will the $1 trillion of generative AI investment pay off? | Goldman Sachs

5 Nationalization - Wikipedia

6 The transformative potential of artificial intelligence - ScienceDirect

7 AI Index Report 2024 – Artificial Intelligence Index

8 China Puts Power of State Behind AI—and Risks Strangling It - WSJ

9 Newly Updated US Export Rules to China Target AI Chips | Altium

10 AI: How far is China behind the West? – DW – 07/24/2023

11 China is falling behind in race to become AI superpower | Semafor

12 Newly Updated US Export Rules to China Target AI Chips | Altium

13 How Artificial Intelligence Is Transforming National Security | U.S. GAO

14 An Overview of Catastrophic AI Risks

15 Ibid.

16 AI's economic peril to democracy | Brookings

17 Artificial intelligence and financial crises

18 For example: Bush's The National Security Strategy of the United States of America. Or: Biden-Harris Administration's National Security Strategy

19 The Battle for Technological Supremacy: The US–China Tech War. Or: Global Strategy 2023: Winning the tech race with China.

20 We need to Pause AI, Pause Giant AI Experiments: An Open Letter

21 2007–2008 financial crisis - Wikipedia

22 Emergency Economic Stabilization Act of 2008 - Wikipedia

23 IV. The Project. Note that Leopold does allude to implementations that do not involve total nationalization, such as defense contracting or voluntary agreements. However, the majority of his argument is built around the idea of a fully centralized government-led research project.

24 AI and Geopolitics: How might AI affect the rise and fall of nations? | RAND

25 Competing Values Will Shape US-China AI Race – Third Way

26 Does Privatization Serve the Public Interest?

27 SAFE Innovation Framework

28 Companies ranked by Market Cap - CompaniesMarketCap.com

29 What you need to know about Nvidia and the AI chip arms race - Marketplace

30 Companies ranked by Market Cap - CompaniesMarketCap.com

31 See: Chips for Peace: How the U.S. and Its Allies Can Lead on Safe and Beneficial AI | Lawfare

32 A Typology of China's Intellectual Property Theft Techniques — 2430 Group