AI Sovereignty, Diffusion, and Risk: Strategy in a Contested Landscape

Conor Chapman, Jon Bateman, Andrew Grotto, Colin Kahl, Arthur Nelson, Lakshmee Sharma

Conor Chapman, Andrew Grotto, and Colin Kahl are affiliated with the Program on Geopolitics, Technology, and Governance at Stanford University. Jon Bateman, Arthur Nelson, and Lakshmee Sharma are affiliated with the Technology and International Affairs Program at the Carnegie Endowment for International Peace.

OVERVIEW

The global demand for "AI sovereignty" is increasingly shaping AI policy and safety discussions. What was once a niche concern has become a widely shared priority, with more countries seeking control over how AI is built, deployed, and governed within their borders. This memo synthesizes key insights on the drivers of AI sovereignty, how countries are operationalizing it in practice, and the implications for U.S. diffusion strategy and managing global risks.

Three premises frame the analysis. First, emerging economies and middle powers are becoming crucial partners, consumers, and suppliers in the global AI ecosystem. Second, U.S. leadership in frontier AI creates a narrow—and likely diminishing—window to shape how this technology diffuses. Third, understanding what countries actually want from AI is critical to designing a sustainable U.S. diffusion strategy and navigating shared risks.

These insights draw on ongoing research from the Carnegie Endowment for International Peace (CEIP) and Stanford's Program on Geopolitics, Technology, and Governance (GTG), and were further developed at a workshop, "AI Sovereignty, Diffusion, and Risk: Strategy in a Contested Landscape," convened by CEIP and GTG on June 17, 2026 with experts from civil society, industry, and academia.

KEY INSIGHTS

"AI sovereignty" is a politically potent but analytically ambiguous concept, driven primarily by a desire to manage dependence and extractivism, and to obtain AI suited to local contexts.

While the term "AI sovereignty" is ambiguous, it has become an increasingly influential part of international AI discussions. Sovereignty goals seem less about achieving true technological autarky (widely seen as unfeasible), and more about a desire to secure national interests (economic, security, cultural) within a tech landscape dominated by the U.S. and China. In this way, AI sovereignty might be practically understood better as "AI agency:" a country's ability to execute choices in line with national interests. In search of this agency, countries will likely pursue a strategy that combines assured access arrangements for foreign AI models and hardware with efforts to diversify and build domestic capacity (indigenous or modified foreign open-source) if such access is disrupted. These domestic capacity efforts tend to focus on cheaper, multilingual, and multimodal models contextually attuned to underserved markets.

Sovereignty can serve as a source of resilience, and, increasingly, may serve as a deterrent against being cut off from frontier AI capabilities. To achieve this deterrent effect, countries are likely to seek sources of leverage along the AI value chain. These may include:

- Control over scarce resources in the AI supply chain, such as critical minerals or low-cost energy for powering data centers.
- Significant market power that makes a country an indispensable consumer of AI services.
- Refining high-value local data for domestic value capture.

These sources of leverage, however, could be a depreciating asset, as a nation may only be able to threaten or use its leverage once before providers diversify away from it.

Perceptions of AI risk within sovereignty debates rarely focus on concerns over shared catastrophic and large-scale threats.

Discussions of AI sovereignty in many middle powers and emerging economies are primarily driven by concerns over dependency, extractivism, market concentration, and geopolitical subordination. Cross-border, large-scale AI risks like cyberattacks, biosecurity, or rogue AI agents have not been central to these debates to date, as they are often perceived as remote or lower priority than immediate economic and political concerns and assured access to AI technology.

Recent events, such as the U.S. government blocking Anthropic's Fable model access, have reinforced this focus on dependency. While the incident highlighted the potential for dangerous capabilities, the primary international reaction has been concerned with the U.S. wielding a “kill switch” over model deployment, reinforcing fears of unilateral control and strengthening the case for AI sovereignty.

This dynamic could shift as AI capabilities diffuse. The widespread availability of powerful models capable of causing significant cross-border harm, such as cybersecurity failures in critical infrastructure, may force a re-evaluation. In such a scenario, the salience of shared safety and security could rise, potentially aligning national interests more closely with global risk mitigation efforts.

Countries are actively pursuing sovereign AI projects, but a significant gap persists between ambitious strategies and operational capacity.

A clear trend of sovereign-related AI projects and announcements is underway globally, especially in the EU, Indo-Pacific, and Gulf States. These initiatives range from building national compute clusters and developing domestic models to pursuing legal arrangements to ensure data localization and provide assured access to foreign AI models.

However, a significant gap often exists between high-level ambitions and the operational capacity to implement them, as many efforts lack clear funding and technical expertise.

The viability of these national ambitions may hinge on a critical, and often unresolved, distinction: identifying which use cases can use “good enough” AI, relying on less advanced models and infrastructure, versus those that require access to the frontier.

The future trajectory of AI—whether dominated by a few frontier labs or a broader open-source ecosystem—is a central uncertainty shaping national strategies.

A fundamental tension exists between two potential AI futures: one where a handful of frontier labs create a runaway capability gap, and another where open-source models remain competitive and useful for most purposes.

In a world where frontier models pull away, a U.S.-led stack could become central to the global economy and international security, giving the U.S. unprecedented insight and international leverage.

In contrast, in a world where open-source and fast follower models remain competitive, the strategic calculus shifts, potentially lowering the stakes of frontier competition for many countries and enabling more diversified technology ecosystems.

The concept of an organized "third stack" as an alternative to U.S. and Chinese ecosystems built on open-source models and diverse hardware is a potential pathway for middle powers seeking to avoid dependency. This alternative is only viable if a coalition of middle powers align on standards for procuring and deploying AI to pool their collective purchasing power; the European Union's regulatory and tech sovereignty efforts are informed by this logic. However, multi-state organization around a third stack faces significant collective action problems, and any effort to create an independent stack could be viewed as a challenge to U.S. national security, incentivizing Washington to pursue bilateral engagements that thwart an emergent alternative.

Even with open models, countries may still want assured access to frontier AI for limited, exquisite capabilities.

Even if many countries' economic, development, and security goals can be achieved through "good enough" AI, countries may want access to high-end capabilities for specific purposes. This could give the United States an opportunity to offer assured access to frontier AI in exchange for safety and security commitments.

However, an assured access framework faces major challenges:

- Trust in U.S. assurances: The U.S. is often not currently seen as a credible, long-term partner. Without trust, assurances are secondary to mutual leverage.
- Third country leverage: A security–frontier bargain appears most viable with countries that possess something the U.S. wants (e.g., India's data, Brazil's energy resources). Leverage is unclear for states that lack such bargaining chips.
- Risk perception gap: Safety commitments, especially those framed around catastrophic risks, do not resonate strongly in many parts of the world, where developmental and economic priorities are paramount. For Global Majority economies, legible near-term AI risks include displacement of labor within domestic informal sectors and the devaluation of labor within global value chains.

Significant doubts about the U.S. government's ability to execute a nuanced AI partnership strategy highlight the roles of market forces and non-state actors.

Many believe the U.S. government currently lacks the capacity and planning to execute a complex global AI diffusion strategy. The U.S. government's existing toolkit for promoting the U.S. tech stack, including bodies like the Development Finance Corporation (DFC) and EXIM Bank, is difficult to deploy effectively. The government's most effective role may be to de-risk investment and lend credibility in geopolitically critical areas where a natural market does not exist.

In the absence of a robust government strategy, market forces are the primary driver. The quality of U.S. technology creates a natural pull, but this may be counteracted by U.S. policy unpredictability.

In this vacuum, non-governmental actors are stepping in. Philanthropic organizations are brokering agreements between U.S. AI labs and Global Majority countries for specific use cases. However, it remains unclear if these ad-hoc efforts can scale into a coherent ecosystem that benefits U.S. interests.

PRIORITIES FOR FURTHER RESEARCH

This analysis surfaces several unresolved questions that warrant further research and debate. Priorities for future inquiry include:

- **Clarifying the competing futures of AI:** Under what conditions might frontier AI models achieve a decisive, compounding advantage over open-source alternatives? What are the key technical and economic indicators that policymakers should monitor to assess which future is becoming more likely? What are the implications for AI risk management?
- **Mapping points of national leverage:** Beyond theoretical control over chokepoints, what forms of economic, political, or geographic leverage have proven most effective for middle powers in securing favorable terms for AI access? How durable is this leverage, and can it be pooled regionally to overcome collective action problems?
- **Designing a viable U.S. partnership model:** What specific, actionable policy tools are required for the U.S. to offer a compelling "assured access" bargain? How can such a policy be designed to be credible across administrations, especially for countries that lack significant intrinsic market or resource leverage? How should AI risks factor in?
- **Integrating safety into diffusion frameworks:** How do perceptions of AI risk influence national sovereignty strategies? What practical mechanisms can embed safety and security commitments into technology partnerships?