

The Problem of Unknowable Power: Why Understanding AI May Matter More Than Owning It



At the Group of Seven summit in Évian-les-Bains this month, the most consequential gathering took place away from the main table. While presidents and prime ministers worked through the formal agenda, a working lunch brought together the leaders of the world's frontier artificial intelligence laboratories alongside heads of government. Sam Altman sat beside Donald Trump, while Dario Amodei and Demis Hassabis occupied seats once reserved for national leaders. Ministers and senior officials crowded the room, and the discussion drew as much attention as the summit itself. The symbolism mattered. A small group of private executives now commands influence once associated with sovereign states because the systems they build increasingly shape economic competitiveness, military capability, scientific progress and public administration. Much of the commentary focused on sovereignty. Should nations rely on models developed elsewhere? What risks arise when critical capabilities sit



within a handful of firms? What happens if access is restricted, degraded or withdrawn?

These questions matter.

They also point towards a deeper shift that is beginning to redefine the structure of power itself.

For four centuries, power was broadly calculable. States estimated the size of armies, the output of factories, the strength of economies and the yield of weapons with increasing precision. Intelligence was never perfect, yet competition unfolded within shared assumptions about capability. Governments disagreed about intentions while maintaining a broadly common understanding of means.

The modern international system rests on that foundation.

Artificial intelligence is beginning to challenge it.

Days before the Évian summit, that challenge came into sharp focus.

Following concerns regarding the security implications of Anthropic's most advanced models, the United States Commerce Department reportedly issued an export-control directive requiring the suspension of access for foreign nationals. Because the company could not reliably distinguish users by nationality, the effect extended globally. Within hours, the models went offline across more than one hundred countries.

The details matter less than the broader lesson.

Senior government officials, intelligence agencies, company executives and outside experts all appeared to hold different views about the nature and significance of the capability in question. Government assessments pointed to a serious concern. The company offered a narrower interpretation.

Independent analysts disagreed among themselves.

A decision with global consequences ultimately rested on competing technical judgements regarding what a single system could actually do.

This episode highlights a defining characteristic of frontier AI.

Strategic decisions increasingly depend upon systems whose capabilities remain only partially understood, including by those closest to them. The challenge is no longer simply the power of these systems. It is the difficulty of characterising that power with confidence.

Earlier generations of technology followed a different logic. Engineers worked with known tolerances. Military planners operated within defined performance envelopes. Nuclear strategy relied upon calculable yields and verifiable stockpiles.



Frontier AI evolves through large-scale training rather than explicit specification. Researchers observe behaviour, test performance and refine outputs, yet they do not produce a complete inventory of capability. New behaviours emerge after deployment. Performance shifts across contexts. Outcomes vary according to prompts, environments and use cases. Capability unfolds over time.

The strategic challenge is not merely that these systems are powerful. It is that consequential capability may exist before institutions can confidently define its limits.

Previously, uncertainty centred on intention. States assessed known capabilities and debated how adversaries might use them. Artificial intelligence introduces uncertainty into capability itself. Decision-makers must increasingly evaluate systems whose full range of behaviour remains only partially mapped.

That distinction carries profound consequences.

Strategy depends upon reliable assessments of relative power. Governments allocate resources, build alliances and develop doctrine based on expectations of what competitors can achieve. When capability becomes fluid, evolving and difficult to bound, the risk of misjudgement increases. The implications extend well beyond national security.

Economic performance increasingly depends upon systems whose behaviour evolves over time. Public administration, critical infrastructure and corporate decision-making are becoming reliant upon technologies that continue to develop after deployment. Organisations are integrating capabilities that remain only partially characterised and whose future performance cannot always be predicted with confidence.

Reliance expands even as understanding struggles to keep pace.

This reality exposes a limitation in many current policy debates.

Governments around the world are investing heavily in sovereign compute, sovereign infrastructure and sovereign models. These investments address legitimate concerns regarding access, resilience and dependence.

Yet ownership alone does not solve the deeper problem.

A government may own a frontier model while continuing to discover what it can do. Conversely, an organisation with sophisticated evaluative capacity may understand both its own systems and those of its competitors more thoroughly than the owners themselves.

The emerging contest therefore carries an increasingly important epistemic dimension.



Advantage will belong not simply to those who possess the most capable systems, but to those who can understand, evaluate and verify those systems most effectively.

At present, this capability remains scarce.

Many governments lack the technical expertise, computational resources and institutional frameworks required for independent evaluation. Regulators rely heavily on information provided by developers. Political leaders often make decisions based on competing claims advanced by companies, researchers and international partners.

This creates a form of strategic vulnerability.

A government that depends on external interpretations to understand critical technologies has effectively outsourced a core element of strategic judgement. Decisions increasingly rely upon perspectives shaped by different incentives, priorities and commercial interests.

A government that develops strong evaluative capacity occupies a fundamentally different position.

It can test systems directly. It can compare claims against observed behaviour. It can develop an evidence-based understanding of technologies that increasingly influence national outcomes.

Most importantly, it can exercise sovereign judgement in a domain defined by uncertainty.

Strong evaluative capacity also improves resilience, enabling institutions to respond more effectively when systems behave unexpectedly or access is disrupted.

The lesson from Évian extends beyond the influence of technology companies. It points to a deeper transformation in strategic competition. The central challenge of the coming decade will not simply be acquiring advanced AI systems but understanding them.

States will continue to invest in compute, talent, infrastructure and model development. These investments remain essential and will define the baseline of capability.

Increasingly, however, competitive advantage will depend on something else: the ability to identify emerging capabilities, evaluate performance independently and generate trusted judgements about rapidly evolving systems.

A new layer of strategic capacity is emerging: independent testing regimes, specialised expertise and institutions capable of producing trusted assessments under conditions of uncertainty. These capabilities will underpin



more disciplined decision-making, more credible risk assessment and more stable strategic interaction.

The emerging competition therefore extends beyond a race to build more powerful systems. It is becoming a race to understand them.

The governments that recognise this shift will invest not only in capability, but in the institutions that generate insight. They will develop the means to understand both their own systems and those of their competitors with speed and independence.

Those that do so will help shape the next phase of technological competition.

Those that do not will increasingly act on the interpretations of others.

In an era where capability does not always present itself in measurable form, understanding becomes the decisive form of power.