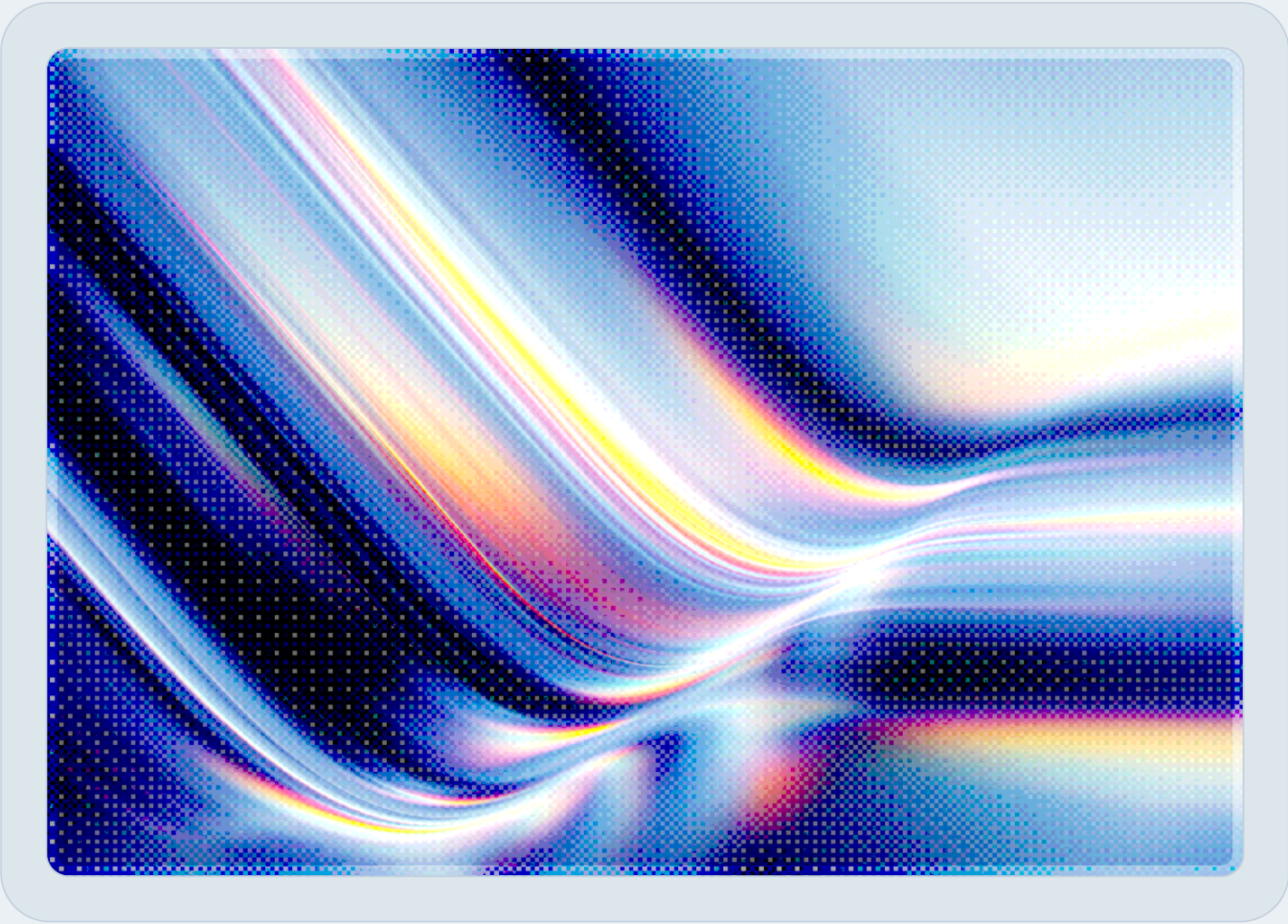
CSAI Foundation | Cloud Security Alliance

The Sovereign AI Dependency Trap

Concentration Risk as the Next Systemic Failure Mode

2026-05-18

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Table of Contents

Executive Summary

Artificial intelligence has become a critical operational input for organizations across every sector of the global economy, and that transformation has unfolded at a pace that has outrun governance capacity in most organizations. What has unfolded with less visibility is the structural concentration that has accompanied this transformation: a small number of infrastructure providers, foundation model developers, and capital allocators — three hyperscalers and a handful of foundation model developers, linked by increasingly interlocked investment networks — now sit at the center of an AI supply chain upon which entire industries depend. The result is a new category of systemic risk — one that combines the familiar dynamics of vendor lock-in with the novel characteristics of AI systems: their opacity, their training-data specificity, their deep integration with organizational workflows, and their tendency to become more deeply embedded over time rather than easier to replace.

This paper introduces the concept of the sovereign AI dependency trap — the dynamic by which organizations, and nations, make individually rational adoption decisions that collectively produce a structural fragility. Each enterprise that integrates a leading AI platform into its core workflows gains genuine capability advantages. Each developer that builds on a dominant foundation model gains access to state-of-the-art performance. Each government that accelerates AI modernization through partnership with hyperscalers closes near-term capability gaps. But as these individual decisions aggregate, the result is an ecosystem in which operational continuity, national security, and economic productivity are increasingly contingent on the reliability and cooperation of a small number of providers.

Several developments in the twelve months leading into mid-2026 illustrate that this risk is not theoretical. The European Systemic Risk Board's Advisory Scientific Committee published a formal analysis in December 2025 identifying AI concentration, alongside model uniformity and monitoring challenges, as features with significant potential to amplify systemic risk across financial systems [4]. The OECD's November 2025 report on competition in AI infrastructure documented that three hyperscalers — Google, Microsoft, and Amazon — now dominate the compute layer upon which virtually all commercial AI capability rests, with structural barriers to entry that make this concentration self-reinforcing [3]. A major AI platform outage in June 2025 demonstrated in concrete terms what operational dependency means when it is tested: automated workflows halted, customer service queues froze, and document-processing pipelines across thousands of enterprises went dark simultaneously.

The response to this accumulating evidence cannot be denial or fatalism. Importantly, the concentration that creates systemic risk is the same concentration that has delivered rapid capability improvement, subsidized safety research, and democratized access to frontier AI through commodity APIs; the goal of this analysis is not to argue that concentration is uniformly harmful but to identify where its structural fragility creates risks that adoption decisions have not yet priced in. Organizations that understand the mechanics of AI

concentration risk, map their own dependency exposure, and invest in architectural resilience before a disruption forces the issue will fare substantially better than those that do not. At the national level, the emerging concept of "sovereign AI" — which gained momentum in 2025 and 2026 through frameworks from the World Economic Forum, the Center for a New American Security, and multiple government policy initiatives — represents a serious attempt to grapple with the geopolitical dimensions of this concentration. This paper integrates the organizational and national dimensions of the problem, analyzes both through CSA's established security frameworks, and provides a tiered set of recommendations for practitioners.

Introduction and Background

The Speed of the Dependency Transition

The velocity at which AI has become operationally load-bearing distinguishes the current moment from previous technology transitions. Enterprise adoption of cloud computing was broadly characterized as a multi-year transition spanning much of the 2010s; the adoption of AI as a core operational tool has compressed into a fraction of that time, driven by the accessibility of API-based foundation models and the no-code agent creation platforms that have proliferated since 2024. McKinsey's 2026 research on AI trust and governance found that only about one-third of organizations report maturity levels of three or higher in their AI governance structures, even as AI deployment has accelerated rapidly [11]. The gap between deployment velocity and governance maturity is not a sign of organizational negligence — it is a predictable consequence of how AI has been packaged and sold: as a service accessible through an API, with complexity abstracted away and switching costs initially invisible.

This accessibility is precisely what makes the dependency trap effective. When an organization first integrates a leading LLM into a business process, the switching cost appears negligible — the API is standardized, the models are benchmarked, and alternatives ostensibly exist. As integration deepens, that calculus shifts dramatically. Fine-tuned models carry training investments that cannot be transferred between providers. Agentic workflows built against proprietary tool-calling formats may require substantial re-engineering to run on alternative platforms. Human workflows adapted to a specific model's output style, latency characteristics, and interface become embedded in organizational muscle memory. By the time dependency is tested — by an outage, a pricing change, a regulatory action, or a geopolitical disruption — the organization discovers that what appeared to be a fungible API call has become load-bearing infrastructure.

Why Concentration Compounds the Risk

AI dependency would be manageable if the landscape of AI providers were competitive and diverse. It is neither. The CNAS Sovereign AI Index found that the United States and China together control approximately 90 percent of the computing power required to develop and deploy frontier AI, and that all 50 of the top-ranked AI foundation models globally originate from these two countries [1]. This is not a snapshot of a transitional period; it reflects structural advantages in compute access, talent concentration, and investment scale that are self-reinforcing and that few other nations are currently positioned to close at the frontier in the near term — a gap that this paper's sovereign AI section examines in detail.

At the enterprise infrastructure layer, three cloud providers — Google, Microsoft, and Amazon — effectively determine who can train, deploy, and serve AI at scale. The OECD's competition analysis documented the mechanisms by which this concentration perpetuates itself: the capital requirements for building competing infrastructure are prohibitive, the economies of scale at the cloud layer strongly favor incumbents, and vertical integration — wherein the same firms that provide cloud infrastructure also invest in and deploy foundation models — creates structural incentives for self-preferencing that further disadvantage independent competitors [3]. The structural analysis suggests that absent regulatory intervention or deliberate investment in alternatives, current concentration dynamics are likely to be self-reinforcing — a pattern the OECD identifies as the default trajectory given prevailing capital, scale, and vertical integration advantages [3].

The investment layer compounds concentration at the infrastructure and model layers. TechPolicy Press estimated that AI attracted over $200 billion in investment in 2025, representing approximately half of all global venture capital — a level of sectoral concentration unprecedented in technology investment history [7]. The geographic distribution mirrors the capability concentration: the San Francisco Bay Area alone captured the majority of U.S. AI investment, and globally, the Middle East and East Asia accounted for more than 80 percent of all publicly disclosed sovereign AI investment [1]. Capital follows capability and capability follows capital, creating a self-reinforcing cycle that leaves most of the world — and most enterprises — as consumers of AI infrastructure rather than contributors to it.

The Architecture of Concentration

Three Interlocked Layers

Understanding where concentration risk resides requires analyzing the AI supply chain as a set of distinct layers, each with its own concentration dynamics and each creating dependencies that propagate upward. The OECD's framework for AI infrastructure competition provides the clearest analytical lens: compute,

model, and application layers are not equally concentrated, and the risks they create differ in character and severity [3].

At the compute layer, the concentration of AI accelerator chip design and fabrication represents what competition authorities have called a "chokepoint" in the AI supply chain. TSMC manufactures the most advanced chips used in virtually all frontier AI systems. Three firms — Cadence, Synopsys, and Arm — account for more than 60 percent of the global Electronic Design Automation market and 70 percent of the intellectual property market upon which all chip designers depend [3]. NVIDIA's dominance of the GPU market for AI training and inference, which the OECD's competition analysis identifies as a persistent chokepoint in the AI supply chain [3], compounds these dependencies further. The result is a hardware supply chain with multiple single points of failure, each capable of cascading up to affect model training, inference capacity, and ultimately the services that enterprises and governments depend on.

The model layer exhibits a different but equally concerning concentration pattern. The gap between frontier model capabilities and what any independent developer or nation can produce has widened, not narrowed, as training compute requirements have scaled. The EuroHPC project, representing the largest sovereign AI investment in Europe at approximately $2 billion, is equivalent to roughly two percent of what four American technology companies invested in AI infrastructure over a single six-month period [1]. France's largest GPU cluster represents less than six percent of the deployment Microsoft alone committed to building in French data centers by end of 2025 [1]. These numbers illustrate that the gap is not primarily a matter of ambition or policy — it is a function of capital scale that cannot be bridged through sovereign investment programs alone.

At the application and platform layer, concentration has an additional characteristic: stickiness. Enterprise AI platforms build deep integrations with existing software stacks, creating the ecosystem entanglement that makes switching costs for AI substantially higher than analogous switching costs in cloud computing. Cloud workloads can, in principle, be migrated between providers through containerization and infrastructure-as-code practices; AI agent architectures built against proprietary tool-calling formats, fine-tuned models, and vendor-specific orchestration patterns carry substantially higher migration costs — costs that, unlike those for containerized cloud workloads, are difficult to standardize away through infrastructure-as-code practices alone. The OECD documented this dynamic as "structural dependency," noting that it creates asymmetry between providers and clients that compounds over time [3].

The Investment Web

A further dimension of concentration that has received less attention than it deserves is the circularity of AI investment relationships. Major cloud providers have made multi-billion dollar investments in the largest foundation model developers, with the model developers in turn committing to use those same cloud providers' infrastructure for training and deployment. Hardware manufacturers receive investment from AI companies while simultaneously being the sole viable source of training infrastructure. The result, as

Bloomberg's 2026 investigation documented, is a network of "circular deals" in which the same capital appears to be circulating among the same actors, creating the appearance of a healthy, competitive market while obscuring the degree to which a small number of interconnected entities effectively determine the direction and speed of AI development [8].

This investment circularity creates a new category of systemic risk concern that financial regulators have begun to identify. In December 2025, the ESRB Advisory Scientific Committee's formal analysis of AI and systemic risk identified concentration as one of five AI features with significant potential to amplify financial system risk, noting that the same underlying models, the same infrastructure providers, and the same failure modes would be simultaneously exposed across institutions that had independently adopted ostensibly similar AI systems [4]. The ESRB report drew an explicit — and deliberately cautionary — parallel to pre-2008 financial crisis dynamics in which apparently independent actors held correlated exposures that were individually invisible but collectively catastrophic. Whether AI concentration will ultimately produce comparable systemic amplification depends on factors including the speed of disruption, the availability of fallback capacity, and the degree to which dependencies are recognized before they are tested.

How the Dependency Trap Closes

The Adoption-to-Dependency Progression

The path from AI adoption to operational dependency follows a consistent pattern that can be observed across organizational types and sectors. Initial integration is experimental and genuinely low-risk: a team uses an AI API for a productivity workflow, an engineering group deploys a code-completion tool, or a customer-facing operation tests a chatbot. At this stage, the capability is genuinely substitutable, and the switching cost is effectively zero. The organization has entered the first stage of the dependency trap.

The trap does not close at the first stage. It closes during the second stage, when initial experiments succeed and the organization scales integration across multiple workflows, embeds AI outputs into downstream processes, and allows human teams to adapt their working patterns to the AI's capabilities. This is the stage at which fine-tuning occurs, at which agentic pipelines are built against specific APIs, and at which the organizational understanding of how to operate without the AI degrades. Organizations rarely monitor this degradation because they are not looking for it — they are looking at the productivity gains, which are real and genuinely significant.

The agentic AI deployment pattern accelerates this progression substantially. When AI agents are deployed at scale — executing thousands of automated actions per day, integrating with email systems, document stores, and approval workflows — the coupling between the organization's operations and the AI platform becomes orders of magnitude tighter than what existed in the chat-interface era. McKinsey's research on

agentic AI deployment found that a substantial majority of organizations planned to deploy agentic AI moderately or extensively within two years, yet fewer than one-quarter reported mature agentic AI governance [12]. This means the majority of organizations are moving through the adoption-to-dependency progression faster than their governance can track.

Switching Costs That Are Not What They Seem

The conventional analysis of AI switching costs focuses on the direct technical costs of migrating from one provider to another: retraining models, porting APIs, re-engineering integrations. These costs are real and substantial, but they understate the true switching cost by omitting three less visible categories.

The first is the fine-tuning and customization cost. Any model that has been trained on proprietary organizational data, or fine-tuned to reflect specific organizational style, domain knowledge, or decision patterns, represents an investment that cannot be transferred to a competing platform. The output of that investment — the performance advantage on the organization's specific use cases — is lost in a migration. The investment must be rebuilt from scratch.

The second is the workflow and human adaptation cost. Organizations that have scaled AI deployment discover that their human teams have adapted to specific AI output characteristics. Customer service teams have built communication patterns around a specific model's response style. Engineering teams have calibrated their review processes for specific code-completion tools. These adaptations are invisible on any vendor contract but represent real organizational capability that erodes during migration. Change management research consistently finds that human re-adaptation costs in large-scale platform migrations rival or exceed direct technical costs — a dynamic likely to be amplified in AI migrations where model-specific workflow adaptations are harder to document and retrain.

The third is the data portability cost. Agentic AI systems that have been operating for months or years accumulate operational history, context, and embedded organizational knowledge in vendor-controlled formats. When that history is not portable — or when portability requires significant engineering work to achieve — the organization faces a choice between accepting data loss and accepting migration friction. Most organizations, when they reach this point, find it easier to accept the status quo than to absorb the disruption of migration.

Ecosystem Entanglement

The most advanced form of dependency is what CSA's prior research on military AI concentration risk termed "ecosystem entanglement" — the state in which AI is deeply integrated not merely with a single workflow but with the entire stack of cloud, productivity, security, and data tools a provider offers. Microsoft's integration of AI capabilities across Azure, Office 365, Teams, GitHub Copilot, and its security platform creates a position in which an enterprise that uses Microsoft's AI is not merely depending on a single

model API but on an interlocking system where any attempt to migrate requires disentangling dozens of deeply coupled integrations simultaneously. Google's equivalent integration across Workspace, Google Cloud, and its Gemini family creates an analogous dynamic. These are not malign designs — they reflect genuine product value — but their consequence for concentration risk is that contractual diversification (having agreements with multiple providers) provides no operational protection if the integrations within a single provider's ecosystem have become operationally irreplaceable.

Failure Modes: When Concentration Breaks

Outage Events as System Tests

The past twelve months have provided multiple empirical tests of what AI concentration means when it fails. In June 2025, a global OpenAI platform outage demonstrated the operational fragility that concentrated AI dependency creates. Across thousands of enterprises simultaneously, customer service queues froze, automated document-processing pipelines halted, and approval workflows that had been delegated to AI systems ground to a stop [10]. The event was not a catastrophic failure — the outage lasted hours, not days — but it revealed the degree to which organizations had come to treat a single third-party AI platform as operational infrastructure, in many cases without the resilience planning that such a designation would warrant — a gap that this paper's governance framework addresses.

The October 2025 AWS outage provided a parallel data point at the infrastructure layer. An automated DNS management failure in AWS's US-East-1 region cascaded through authentication systems and global services, rendering critical platforms — including Slack, Snapchat, and many AI-dependent enterprise applications — inaccessible for an extended period [9]. The event illustrated an architectural dynamic that multi-region deployment strategies within a single cloud provider do not fully address: because many global AWS services, including IAM authentication, CloudFront, and Route 53, depend on US-East-1 control plane components, workloads nominally deployed in other regions lost authentication capability regardless of geographic distribution [9]. An organization with AI workloads distributed across three AWS regions still has a single point of failure when the authentication layer fails.

Business continuity analyses consistently find that large enterprises face multi-million-dollar hourly losses during critical-system outages, with financial services firms among the most exposed during trading-day disruptions [10]. These figures represent a business continuity exposure that most organizations have not explicitly quantified against their AI vendor concentration posture.

Vendor Disruption Beyond Outages

Outages represent only the most visible failure mode. A broader and arguably more consequential category is vendor disruption that is not the product of technical failure but of commercial, political, or regulatory dynamics. Regulatory and procurement bodies in the United States have, in early 2026, begun subjecting AI vendor relationships to supply chain security reviews that can result in access restrictions — a pattern with precedent in the semiconductor and telecommunications sectors. For organizations that have built significant operational capability on a vendor's models and APIs — including enterprises that have made substantial fine-tuning investments — such reviews can create immediate strategic uncertainty that technical redundancy alone cannot address. The dependency is not merely on a product; it is on a vendor relationship subject to geopolitical and regulatory disruption.

This pattern is likely to recur and intensify rather than abate. As AI capabilities become more directly relevant to national security, critical infrastructure, and economic competition, governments are likely to increasingly exercise their authority to restrict or condition commercial AI relationships on grounds that have nothing to do with technical performance. Export controls, supply chain security designations, data localization requirements, and mandatory auditing regimes all create vectors through which vendor relationships can be disrupted or constrained in ways that no service-level agreement addresses. The organization that has architected for technical resilience but not for commercial and regulatory resilience has only partially addressed its dependency exposure.

Correlated Failure at Scale

The most severe form of concentration risk is the one that regulators have begun to analyze: correlated failure across organizations when they share the same underlying AI infrastructure. The ESRB's December 2025 analysis drew this concern explicitly: when financial institutions, critical infrastructure operators, healthcare systems, and government agencies all depend on the same small number of AI providers, the failure of a single provider — whether through technical failure, regulatory action, or adversarial disruption — creates a correlated shock that affects all of them simultaneously [4]. This is not the ordinary risk of sector-wide disruption from macroeconomic forces; it is a structural fragility created by a shared technical dependency that regulators never explicitly approved.

The correlated failure concern has a particularly serious dimension in the context of adversarial targeting. For a nation-state adversary seeking to disrupt an opponent's economic or governmental function, a concentration of critical operational dependencies into a small number of AI providers represents an attractive target. Disrupting one provider through any combination of cyberattack, political pressure, supply chain interference, or regulatory manipulation produces effects across thousands of dependent organizations simultaneously. The OECD and ESRB both flagged this dynamic, but the security community has been slower to elevate it to the prominence it deserves as an attack surface consideration.

Failure Mode	Trigger	Blast Radius	Recovery Timeline
Platform outage	Technical failure	All dependent enterprises simultaneously	Hours to days
Model deprecation	Vendor product decision	Enterprises using specific model version	Weeks to months
Vendor disruption	Regulatory / political action	Enterprises in affected jurisdiction	Indefinite
Compute layer failure	Infrastructure event	All services using affected provider	Hours to weeks
Correlated model failure	Shared training data or architecture flaw	All users of same model family	Depends on disclosure timeline
Adversarial targeting	Nation-state or criminal attack	All dependent on targeted provider	Depends on attack type

The Sovereign Dimension

AI Dependency as a Geopolitical Condition

The sovereign AI debate — the question of whether and how nations can maintain meaningful autonomy in AI development, deployment, and governance — has moved from academic discussion to active policy agenda in the period between 2024 and 2026. The World Economic Forum published two major reports in early 2026 examining the contours of AI sovereignty, and the Center for a New American Security released its Sovereign AI Index tracking national AI capability and dependency metrics [1][2][5]. This attention reflects a genuine policy problem: the concentration of AI capability in two countries, and the structural advantages those countries hold in compute, capital, and talent, means that most of the world faces a choice between accepting dependency on foreign AI infrastructure or accepting a significant and potentially widening capability disadvantage.

The WEF's 2026 framework articulates the challenge with precision that earlier analyses lacked. Rather than treating sovereignty as a binary condition — either a nation controls its AI or it does not — the framework identifies the specific layers of the AI stack and asks which layers a nation must own or control for reasons of security, economic autonomy, or values alignment, and which layers can safely be accessed through trusted

international partnerships [2]. This layer-specific framing is more analytically useful than blanket calls for AI independence, which are unrealistic for all but the largest economies, and more ambitious than pure market-dependency approaches, which leave nations exposed to exactly the disruptions documented above.

The CNAS Sovereign AI Index provides empirical grounding for this framework by tracking where different nations actually stand across the dimensions of compute, model, data, and talent sovereignty [1]. The index's findings confirm the two-superpower structure at the frontier, but also identify a set of second-tier nations — including the United Kingdom, France, Canada, and several Gulf states — that have made significant sovereign AI investments and are actively working to develop meaningful capability, even if not at frontier scale. The policy question for these nations is whether their investments are building genuine operational resilience, or whether structural dependency on foreign AI infrastructure persists beneath the surface of national AI initiatives.

The Illusion of Sovereign Investment

The EuroHPC case illustrates a gap between the ambition of sovereign AI investment and the structural reality of concentration that policy discussions have been slow to acknowledge. EuroHPC represents Europe's most significant effort to build independent AI compute capacity, with a budget of approximately $2 billion — a genuinely substantial commitment that reflects serious political will. The problem is one of scale: four American AI companies spent that equivalent amount approximately every ten days during the investment peak of 2025 [1]. This is not a gap that additional European investment can realistically close through incremental effort; it is a structural asymmetry rooted in the relative size of technology capital markets, the history of talent development, and the network effects of incumbency.

The implication is not that sovereign AI investment is pointless — it is that its value must be measured against realistic objectives. Investment in national compute capacity creates genuine benefits: it provides infrastructure for domestic AI applications that require data sovereignty; it develops domestic talent that would otherwise be absorbed by foreign technology companies; it creates negotiating leverage with hyperscalers who prefer not to see customers develop exit options. But investment that is premised on achieving full independence from American or Chinese AI infrastructure is unlikely to succeed and may divert resources from the more achievable and valuable goal of managing dependency strategically. The WEF's framing of "strategic interdependence" rather than "independence" reflects this more realistic assessment [2].

Regulatory Responses: Emerging Frameworks

Governments and multilateral bodies are beginning to develop regulatory frameworks that address AI concentration directly, though none have yet produced comprehensive governance structures. The European Union's AI Act includes requirements that address transparency and risk management for high-risk

AI applications, but its concentration-specific provisions are limited. Competition authorities in the EU, UK, and U.S. have opened investigations and consultations on AI market dynamics — the FTC's AI competition inquiry, the CMA's foundation models review, and the OECD's Competition in AI Infrastructure analysis all reflect serious institutional attention — but no enforcement actions with structural effect on AI concentration have yet been issued as of mid-2026 [3].

The most operationally significant regulatory developments have been on the security and critical infrastructure side rather than the competition law side. Financial regulators in the EU and UK have begun developing guidance that treats AI concentration as an operational risk requiring management under existing prudential frameworks. The ESRB's December 2025 analysis is the most concrete expression of this approach, recommending that capital and liquidity requirements, supervisor oversight, and operational resilience standards be adapted to address AI-specific concentration risks [4]. In the United States, the WEF April 2026 commentary explicitly called for treating AI infrastructure as critical infrastructure subject to the security standards and resilience requirements that designation implies [6].

Security Implications and the AI Attack Surface

Concentration as an Adversarial Target

From a security perspective, AI concentration creates an attack surface with unusual characteristics. The concentration of operational dependence into a small number of providers means that a successful attack against a single target — whether through cyberattack, supply chain compromise, or adversarial manipulation of model outputs — produces effects that propagate across thousands of dependent organizations simultaneously. This is a force-multiplication dynamic that well-resourced adversaries understand. The targeting calculus for a nation-state seeking to disrupt an opponent's critical infrastructure has shifted: rather than targeting individual organizations, it is increasingly rational to target the AI infrastructure upon which many organizations depend.

The increasing incorporation of AI into nation-state offensive campaigns — documented in government threat assessments and major cybersecurity threat intelligence reports — suggests that adversaries are actively developing capabilities against AI infrastructure. The integration of AI into defense, financial, and government operations makes AI providers into high-value intelligence targets. A successful intrusion into the training pipeline of a widely deployed enterprise model, or into the API infrastructure of a major AI provider, could simultaneously compromise the AI-assisted operations of thousands of dependent enterprises without any of those enterprises having been individually targeted.

The model uniformity dimension that the ESRB identified as a systemic risk amplifier is particularly concerning from an adversarial standpoint. When many organizations use the same underlying model, trained on the same data, with the same architectural characteristics, an adversary who identifies a systematic weakness in that model's reasoning, safety properties, or output characteristics can exploit that weakness simultaneously across all dependent users [4]. This is analogous to a software vulnerability in a widely deployed library, but with the additional complexity that AI model vulnerabilities may not be discoverable or patchable through conventional software security methods.

Governance Gaps at Deployment Scale

The governance challenges created by AI concentration are distinct from and compound the technical risks. When organizations deploy AI at scale without commensurate investment in understanding what those systems are doing and why, they create a form of operational opacity that degrades their ability to detect, diagnose, and respond to failures — whether those failures originate from technical faults, adversarial manipulation, or vendor disruption.

CSA's MAESTRO framework, introduced in February 2025, provides a seven-layer architecture for understanding the threat surface of agentic AI systems, from foundation model vulnerabilities through data operations, agent frameworks, and deployment infrastructure to ecosystem integration [13]. Applied to the concentration risk context, MAESTRO's framework reveals that concentration risks exist at every layer of the MAESTRO stack and that they interact: a vulnerability in a shared foundation model propagates through all the higher layers that depend on it, and a disruption at the infrastructure layer cascades upward through orchestration, data access, and ultimately the workflows that human operators depend on. Organizations that have not explicitly mapped their MAESTRO-layer dependencies against their AI vendor concentration exposure have an incomplete picture of their risk.

The agent-to-agent cascade risk that CSA's prior research has examined in the context of large-scale agent deployments takes on additional dimensions in a concentrated AI environment. When 100,000 or more agents all run on a single platform, a compromise or disruption that affects the platform's orchestration layer — whether through adversarial injection, a software fault, or a vendor-side incident — can propagate across the entire agent population simultaneously. The operational consequence is not just the loss of a single agent's function; it is the simultaneous loss of all dependent automation, with human operators lacking the capacity to perform those functions manually at scale.

Conclusions and Recommendations

The Core Principle: Resilience Through Architectural Honesty

The central challenge that AI concentration risk presents to organizations and policymakers is not technical — it is cognitive. The dependency trap closes not because organizations lack the ability to architect for resilience, but because they lack the visibility into how deeply dependent they have become by the time that visibility becomes urgent. The primary practical requirement is therefore not a new framework or a new category of control; it is honest assessment of actual architectural dependency, followed by deliberate investment in resilience proportional to that exposure.

This requires distinguishing between the two forms of diversification that the literature often conflates. Contractual diversification — maintaining agreements with multiple AI providers — provides no operational protection if the organization lacks the architecture to actually switch. Operational diversification — building and maintaining the technical capability to run critical workloads on alternative platforms — provides genuine resilience but requires investment that organizations routinely defer until a disruption forces the issue. Every organization that has not explicitly tested its ability to operate critical AI-dependent workflows on an alternative provider is, to some degree, operationally dependent on a single provider regardless of what its vendor contracts say.

Organizational Recommendations

Immediate Actions. Organizations should begin with an honest dependency audit: for each AI-dependent workflow, identify the specific vendor and platform, estimate recovery time if that vendor became unavailable for 72 hours, and document what data access rights the relevant AI agents or systems hold. This audit will reveal, in most organizations, a concentration pattern that has not been explicitly authorized by any governance body but has accumulated organically through individual adoption decisions. The audit results should be presented to risk ownership at board or senior executive level, framed as a material operational risk comparable to other critical vendor dependencies.

In parallel, organizations should assess whether their current AI vendor contracts include the continuity provisions that a material dependency warrants: data portability guarantees, API stability commitments, advance notice requirements for deprecation and pricing changes, and termination provisions that ensure the organization retains access to its own data and fine-tuning investments. The asymmetry between enterprise customers and AI providers — where the provider's standard terms rarely include meaningful continuity protections — can frequently be negotiated for significant customers, but only if the organization treats AI vendor contracting as a risk management exercise rather than a procurement exercise.

Near-Term Mitigations. Organizations should prioritize adoption of open interoperability standards — particularly the Model Context Protocol (MCP), which as of mid-2026 has emerged as a leading candidate for preserving model-backend substitutability in agentic AI architectures. Agent systems built against MCP rather than proprietary vendor tool-calling formats retain the ability to run against alternative model backends with substantially lower migration friction. This is not purely a strategic investment; it provides near-term benefits by enabling model testing and evaluation that improves operational performance independent of any concentration risk concern.

Multi-model testing at modest operational scale, before a disruption creates operational urgency, provides organizations with the empirical basis for migration decisions and reduces the effective switching cost substantially. An organization that has tested three alternative models for its five most critical AI workflows, and has documented the performance characteristics and integration requirements of each, is in a fundamentally different position from one that is discovering alternative options after a disruption has already begun.

Strategic Considerations. The long-term architectural objective for organizations with significant AI operational dependency is genuine multi-model capability — not as a paper diversification strategy but as a tested, maintained operational posture. This is analogous to the multi-cloud strategies that mature cloud-native organizations have developed over the past decade, with the important caveat that — based on the switching cost analysis developed in earlier sections — AI workload migration carries substantially higher organizational friction than equivalent cloud workload migration, and the required investment in testing and governance is proportionally greater. Concentration thresholds — explicit governance policies that trigger mandatory diversification assessment when any single provider accounts for more than a specified fraction of critical workflow dependencies — provide a governance mechanism for preventing future concentrations from developing unmonitored.

National and Industry-Level Recommendations

For policymakers and regulators, the most immediately actionable recommendation from this analysis is to treat AI infrastructure as critical infrastructure for resilience and security planning purposes. This designation carries specific implications: mandatory operational resilience standards for AI-dependent critical sectors, reporting requirements for AI concentration exposure comparable to those applied to technology and cloud concentration in banking, and incident notification requirements that enable regulators to assess correlated impact when a major AI provider experiences significant disruption [6].

Industry-level responses should focus on the interoperability standards that reduce switching costs for all participants. The emergence of MCP as an open standard, and the broader ecosystem of open model formats and deployment tooling, represents genuine progress that should be actively supported by industry bodies and not impeded by provider-specific proprietary alternatives. Similarly, the development of shared AI infrastructure models — including the WEF's "digital embassy" framework and various regional pooling

proposals — provides a path for smaller economies and organizations to achieve meaningful AI capability without accepting the full dependency exposure that comes with exclusive reliance on hyperscaler platforms [5].

Regulatory frameworks for AI concentration should be developed in parallel across competition, financial regulation, and security domains rather than in isolation. Competition law alone cannot address the operational resilience dimensions of AI concentration; financial regulation alone cannot address the security dimensions; critical infrastructure security frameworks alone cannot address the competitive dynamics that drive concentration in the first place. Effective governance of AI concentration risk requires coordination across all three regulatory domains.

CSA Resource Alignment

CSA's published frameworks provide a comprehensive toolkit for organizations addressing sovereign AI dependency and concentration risk. The MAESTRO framework — CSA's seven-layer threat modeling architecture for agentic AI systems — provides the primary analytical lens for understanding where concentration risk manifests across the AI system stack [13]. Organizations should apply MAESTRO not only to individual agent deployments but to their entire AI vendor architecture, mapping which MAESTRO layers are exposed to single-vendor concentration and what failure scenarios that concentration creates.

The AI Controls Matrix (AICM) v1.0 maps directly to the supply chain security domain that concentration risk inhabits. AICM's shared responsibility model clarifies which security controls remain the organization's responsibility regardless of which AI provider it uses — a critical consideration in assessing whether architectural diversification actually provides the security benefits that concentration reduction implies. AICM's implementation guidelines for Application Providers and AI Customers both address vendor risk management in the AI context and should be consulted alongside the concentration risk analysis in this paper.

CSA's Zero Trust guidance applies with particular force to AI agent deployments in concentrated environments. The principle of treating agents as untrusted principals requiring continuous verification and least-privilege access is not merely a security best practice; it is a structural protection against the permission accumulation and agent-to-agent cascade risks that become most severe in large-scale, concentrated AI deployments. Zero Trust architecture for AI agents ensures that a platform-level disruption or compromise does not automatically extend to all the resources and data that agents in that platform have accessed.

The Cloud Controls Matrix (CCM) provides the business continuity and supply chain risk management controls that apply directly to AI vendor concentration. CCM requirements for availability, business continuity, and vendor management should be explicitly extended to cover AI vendor relationships at the level of operational dependency that those relationships now represent, not merely treated as a subset of

general cloud vendor management. CSA's AI Organizational Responsibilities guidance addresses the board-level and senior executive risk ownership that concentration risk ultimately requires: the decisions about acceptable dependency thresholds, investment in architectural resilience, and vendor contract terms that adequately protect organizational continuity interests are all decisions that require clear organizational ownership at the governance level.

References

[1] Center for a New American Security. "Sovereign AI Index." CNAS, April 2026.

[2] World Economic Forum. "Rethinking AI Sovereignty: Pathways to Competitiveness through Strategic Investments." WEF, 2026.

[3] OECD. "Competition in Artificial Intelligence Infrastructure." OECD, November 2025.

[4] European Systemic Risk Board Advisory Scientific Committee. "Artificial Intelligence and Systemic Risk." ASC Report No. 16, December 2025.

[5] World Economic Forum. "AI Infrastructure in the Age of Sovereignty: Requirements, Strategies and a Trusted Framework for Digital Embassies." WEF, 2026.

[6] World Economic Forum. "It's Time to Start Treating AI Infrastructure as Critical Infrastructure." WEF, April 2026.

[7] TechPolicy Press. "Rethinking Sovereign AI as Strategy." TechPolicy Press, 2026.

[8] Bloomberg. "AI Circular Deals: How Microsoft, OpenAI and Nvidia Keep Paying Each Other." Bloomberg, 2026.

[9] CloudFactory. "Strengthening AI Resilience: 3 Lessons from the 2025 AWS Outage." CloudFactory, 2025.

[10] Cyber Unit. "When Your AI Goes Dark: Why Businesses Need a Continuity Plan for LLM Outages." Cyber Unit, 2025.

[11] McKinsey & Company. "State of AI Trust in 2026: Shifting to the Agentic Era." McKinsey, 2026.

[12] McKinsey & Company. "Seizing the Agentic AI Advantage." McKinsey, 2025.

[13] Cloud Security Alliance. "Agentic AI Threat Modeling Framework: MAESTRO." CSA, February 2025.

