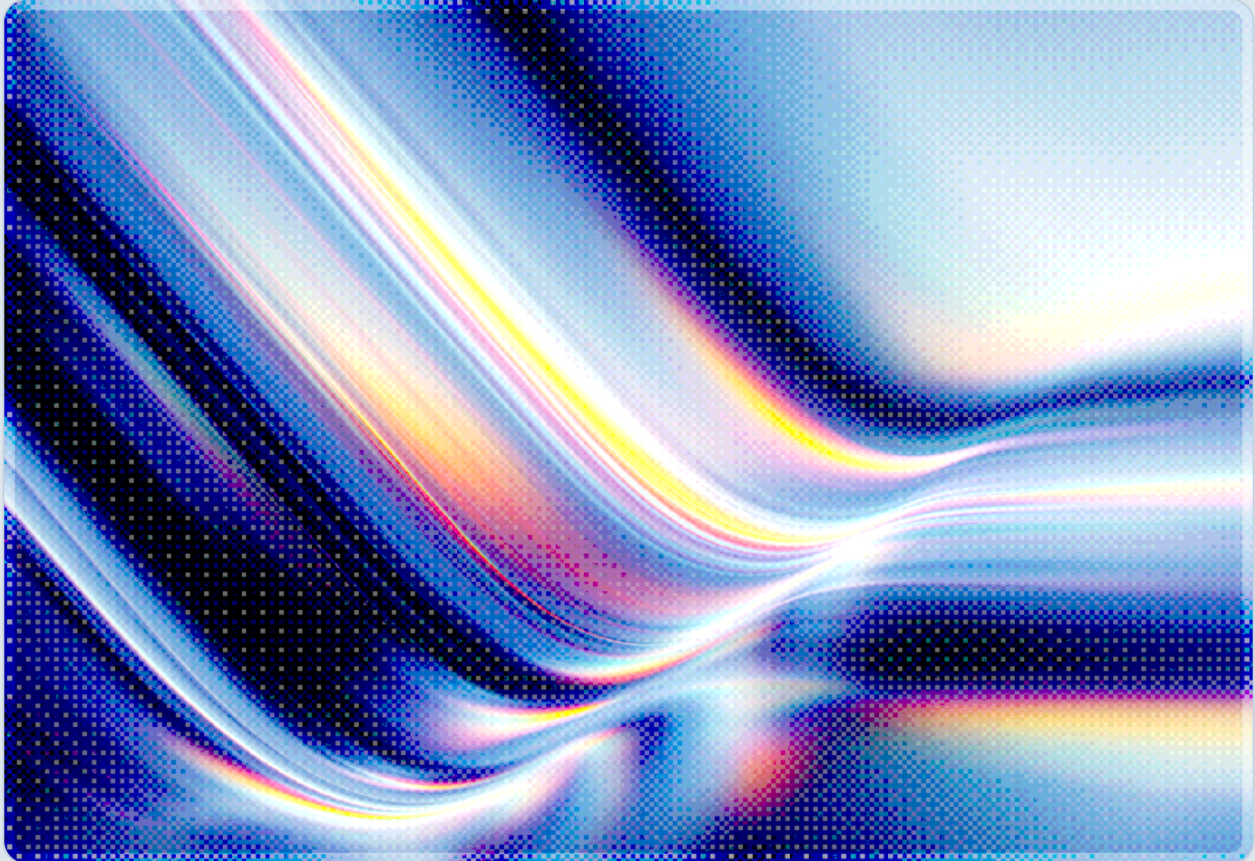


The Access Layer Is No Longer Neutral: AI Gateway Gatekeeping

Provider Approval of Gateways, Sovereign Exclusions, and Enterprise Concentration Risk

2026-10-24

 AI-assisted Rapid Research



CSAI

cloud
security
alliance®

© 2026 Cloud Security Alliance. Some rights reserved.

You may download, store, display, view, print, redistribute, and link to this document in its original, unmodified form, provided that attribution to the Cloud Security Alliance is maintained and all trademark and copyright notices remain intact.

This document may not be modified or altered. You may quote portions of the document as permitted by the Fair Use provisions of the United States Copyright Act, provided that attribution is given to the Cloud Security Alliance.

This document may be shared on professional and social media platforms in its original form with attribution.

This document was generated with AI assistance and has not undergone official CSA review and approval processes.

Key Takeaways

On 21 October 2026, OpenAI revised its usage policies so that, effective 20 November 2026, operators of model-access intermediary platforms must obtain OpenAI's prior review and approval before offering third-party access to its models [1][2][3]. Secondary coverage, quoting OpenAI's accompanying announcement, describes this category as including model routers and AI gateways [1][2][17]. Applications that only use OpenAI models to power their own features are reported to be carved out [1][2]. To CSA's knowledge, this is the first time a frontier provider has publicly made the routing layer itself, rather than individual customers or use cases, an explicitly permissioned tier of its distribution model.

The change lands on top of a sequence of access withdrawals that were not driven by technical failure. One provider cut a coding platform's bundled model access after reports that a competitor would acquire it [5], and another has given notice that it will end model supply to a coding platform on 12 November 2026 following that platform's acquisition [4]. A third pattern emerged when a provider excluded customers based on corporate ownership by entities in specified jurisdictions [6]. Governments have acted as well: a U.S. directive suspended foreign nationals' access to specific frontier models [7], and a second government has reportedly discussed limits on overseas access to its own domestic models [8]. Taken together, these events indicate that access to frontier models is increasingly conditioned on who the customer is, who owns the intermediary, and which jurisdiction both sit in.

For enterprises, the practical consequence is that the AI gateway, which CSA's prior concentration-risk guidance positioned as the main resilience control, is becoming an explicitly permissioned point where access can be revoked. A multi-provider strategy that routes through a single hosted gateway may therefore concentrate risk rather than spread it. Security and risk leaders should determine before 20 November whether any gateway they use or operate falls within the new clause, add provider-policy and ownership-change triggers to third-party risk monitoring, and test failover paths that do not depend on one intermediary's standing with one provider.

Background

The gateway as the enterprise control plane

Over the past two years, AI gateways and model routers have become a common way for enterprises to reach large language models. A gateway sits between applications or agents and one or more model providers. It holds provider credentials, applies authentication and rate limits, logs prompts and responses, enforces data-handling policy, and routes requests by cost, latency, or capability. Hosted services aggregate many providers behind one API and one bill, while self-hosted proxies such as LiteLLM let an organization run the same function inside its own environment. CSA's 2026 work on provider concentration recommended exactly this abstraction layer as the mechanism for multi-provider failover and as the enforcement point for enterprise AI access controls [12][14].

That design assumed the gateway was a neutral conduit: a piece of plumbing that the enterprise chose, configured, and could replace. The assumption has been eroding from two directions. On the security side, gateways have become high-value targets because they concentrate provider keys, cloud roles, and logged model traffic. The LiteLLM 1.107.0 release compromise of 21 October 2026 shipped a credential stealer and a remote loader through official PyPI wheels and Docker images [9][10], and CSA analysis of that event noted that gateway request logs can effectively function as a credential store [15]. On the commercial and policy side, which is the subject of this note, the gateway is becoming a point that providers and governments regulate directly.

OpenAI's intermediary approval requirement

According to secondary coverage of the 21 October revision, the updated policy states that operators of a model-access intermediary platform, or another service that enables third parties to directly access OpenAI models, must obtain OpenAI's prior review and approval [1][2]. Partners that host or resell OpenAI models are expected to maintain safeguards consistent with OpenAI's own, investigate potential violations, and take appropriate action when concerns are identified [2][17]. The same revision introduces "SafetyIDs," which are described as identifiers that allow abuse investigations to distinguish individual end users behind a customer without OpenAI learning their identities [2]. OpenAI had not, as of this writing, published a list of approved intermediaries, a review timeline, or a detailed application procedure [2]. CSA was unable to retrieve the primary policy text directly at the time of writing, so the exact clause wording should be confirmed against OpenAI's published Usage Policies [3].

OpenAI frames the change as a matter of consistent safeguards wherever its models are accessed [2] [17]. Commentators have noted that a gateway aggregates many end users behind one account, which leaves the provider one step removed from the end user and limits its ability to attribute misuse [1]. That rationale is reasonable on its face, and the requirement does not, by itself, cut anyone off. What it changes is the default: from 20 November, an intermediary's ability to serve OpenAI models to third parties depends on an approval that the provider grants and can presumably withdraw.

A pattern of non-technical access withdrawal

The intermediary clause is best read alongside earlier events in which access to frontier models ended, or is scheduled to end, for reasons unrelated to outages or vulnerabilities. The table below summarizes the cases most relevant to enterprise planning.

Date	Event	Trigger	Who lost or may lose access
June 2025	Anthropic cut Windsurf's first-party access to Claude 3.x models with short notice [5]	Reported (not completed) acquisition by a competing provider	Users of the platform's bundled access; bring-your-own-key access continued [5]
September 2025	Anthropic barred entities more than 50% owned, directly or indirectly, by companies headquartered in unsupported regions [6]	Corporate ownership and jurisdiction	Subsidiaries worldwide of firms in restricted regions [6]
June 2026	U.S. export control directive suspended access to two Anthropic frontier models by foreign nationals inside or outside the United States [7] [18]	Government national security determination	Foreign national users and staff, including at U.S. firms [7]
July 2026	China's commerce ministry reportedly discussed limiting overseas access to advanced domestic models [8]	Prospective government policy; nothing decided	Overseas users of Chinese hosted models and future open-weight releases, if adopted [8]

Date	Event	Trigger	Who lost or may lose access
August 2026	OpenAI gave notice ending model supply to Cursor effective 12 November 2026 after its acquisition by SpaceX [4]	Change of ownership and provider trust assessment	Platform-provided OpenAI access; whether users' own API keys will remain an option has not been reported [4]
October 2026	OpenAI requires prior approval for model-access intermediaries, effective 20 November 2026 [1][2]	Provider policy on distribution and enforcement	Unapproved routers and gateways serving third parties

Two features recur across these cases. First, in most of them the decisive variable was identity, ownership, or jurisdiction rather than a security failure: who acquired or was reported to be acquiring the intermediary, who owns the customer, or what nationality the user holds. The intermediary clause is the partial exception, since it is framed around enforcement capability, and the China measure remains prospective. Second, in the one completed platform case, end users who held their own direct provider relationship retained access while users who relied on the intermediary's bundled access did not [5]. Whether the same will hold for Cursor users after 12 November is not yet clear [4].

Security Analysis

The resilience control has become a single point of revocation

CSA's earlier concentration-risk guidance identified provider outages, model deprecation, regulatory enforcement, and strategic pivots as distinct failure modes, and recommended a gateway or abstraction layer to fail over between providers [14]. The intermediary approval requirement adds a failure mode that this architecture does not address on its own: loss of the gateway's standing with a provider. If an enterprise routes all model traffic through one hosted aggregator, a decision about that aggregator, whether a denied application, a lapsed approval, an acquisition, or a sovereign exclusion triggered by the aggregator's ownership, can remove a provider for every customer at once. The enterprise's own relationship with the model provider need not enter the decision.

This suggests that hosted aggregation can convert provider diversity into intermediary concentration. An organization may hold access to four model families on paper while depending on one company's contractual position with each of them. The Cursor case illustrates the dynamic: the provider's stated concern related to the acquirer rather than the platform's own conduct, and developers will have to switch models or change how they access OpenAI once the contract ends [4]. Enterprises whose fallback plan is "the router will switch providers" should ask whether the router itself could be the thing that is switched off.

Sovereign exclusions propagate through ownership chains

Sovereign exclusions are now defined by ownership and nationality as well as geography. Anthropic's 2025 policy applies to entities controlled from restricted jurisdictions wherever they operate [6], and the June 2026 directive applied to foreign nationals regardless of location [7]. For an AI gateway, these rules raise questions that AI vendor due diligence has not typically emphasized. Who ultimately owns the intermediary, and could a future funding round or acquisition change that? Does the gateway pass through user nationality or entity ownership attributes that a provider might need to enforce a nationality-based restriction? Can the gateway segment traffic so that a restriction affecting one population of users does not force the operator to disable a model for everyone?

The June 2026 suspension shows why the last question matters. Anthropic disabled the affected models for all customers to ensure compliance with a directive scoped to foreign nationals [7], because its systems could not verify user nationality at the per-session API level within the compliance window [13] [18]. Gateways that cannot attest to the attributes a regulator cares about may face the same choice. The reported discussions in China point in a different direction: enterprises that adopted Chinese models for cost reasons, often through aggregators, may face a supply-side restriction from the model's home jurisdiction rather than their own, although officials have reportedly decided nothing and any curbs might apply only to future models [8].

Identity propagation creates new data-handling obligations

The SafetyID mechanism signals that OpenAI wants intermediaries to pass pseudonymous per-user identifiers with requests [2], and other providers may follow. That has security consequences on both sides. Positively, it lets abuse investigations target an individual end user instead of suspending an entire gateway account, which would reduce the blast radius of enforcement actions. On the other hand, a gateway that generates and logs per-user identifiers becomes a store of linkable user activity. Given that gateway logs have already proved to be a credential-exposure path [9][15], operators should treat identifier generation, salting, rotation, and retention as part of gateway security design rather than as a compliance afterthought.

Scope ambiguity is itself a risk

The available reporting does not settle whether an internal enterprise gateway that serves employees, subsidiaries, contractors, or customers counts as enabling "third parties" to access OpenAI models directly [1][2]. A gateway serving one legal entity's staff appears closer to the carve-out for applications powering their own features. A shared platform team serving several subsidiaries, a managed service provider operating gateways for clients, or a software vendor that exposes raw model access to its customers sits closer to the intermediary definition. Until the provider publishes criteria, organizations in these grey zones carry an undetermined continuity risk with a fixed date attached. The absence of a public approval list also limits customers' ability to verify a supplier's status independently [2].

Regulatory framing: concentration risk now includes intermediaries

Financial-sector regulation already treats concentration on ICT third-party providers as a risk to be assessed and reported. The EU Digital Operational Resilience Act requires financial entities to assess ICT concentration risk and to maintain exit strategies for critical third-party arrangements [11]. Applying that lens, an AI gateway that supports a critical function is itself an ICT third-party arrangement, and its dependence on approvals from upstream model providers is a form of subcontracting or fourth-party risk. CSA has argued previously that frontier model access should be treated as a regulated, controlled commodity [13]. The developments described here extend that argument from the model to the access layer that delivers it.

Recommendations

Immediate Actions

Organizations should first inventory every path by which their applications and agents reach frontier models, distinguishing direct provider contracts, cloud-marketplace access, hosted routers or aggregators, and self-hosted gateways. For each hosted router or gateway supplier that serves OpenAI models, request written confirmation of the supplier's approval status or application timeline under the policy taking effect on 20 November 2026, and record the answer in the supplier's risk file. Teams that operate their own gateway on behalf of subsidiaries, customers, or partners should obtain a legal and contractual reading of whether that service enables third-party access within the meaning of the clause, and engage the provider directly if the answer is uncertain.

At the same time, confirm that critical workflows have a tested fallback that does not depend on the same intermediary. Where the provider permits it, maintain a direct API relationship, with keys held in the enterprise's own secrets manager, for each model family that supports a critical function. In the Windsurf case, users with their own provider keys kept access to the affected models when bundled access ended [5]; whether the same will hold for Cursor after 12 November is not yet clear [4].

Short-Term Mitigations

Over the next quarter, extend third-party risk management for AI suppliers to cover the triggers observed in this note. Contracts and due diligence for gateways and AI-enabled platforms should address notice obligations for changes in upstream provider approval, changes of control, ownership by entities in jurisdictions subject to provider or government restrictions, and the supplier's plan for continuity if an upstream provider withdraws. Monitoring should include provider usage-policy revisions and export control announcements alongside conventional security advisories, since each has now caused access loss.

Gateway architecture should also be reviewed. A self-hosted or enterprise-controlled gateway with direct provider credentials reduces exposure to an intermediary's standing, but it moves supply-chain and credential risk in-house, as the LiteLLM 1.107.0 compromise demonstrated [9][15]. Operators should pin gateway images by digest, avoid logging full response bodies, and design any per-user identifier scheme so that identifiers are pseudonymous, rotated, and retained only as long as abuse investigations require. Where a gateway must support nationality-based or entity-based restrictions, build the ability to segment and restrict traffic by user attribute so that a restriction can be enforced narrowly rather than by disabling a model for all users.

Strategic Considerations

Over the longer term, boards and risk committees should treat the AI access layer as a distinct concentration domain, separate from both cloud hosting and model providers. Resilience testing should include a scenario in which a gateway loses one provider for policy reasons, a scenario in which a provider restricts a population of users by nationality or ownership, and a scenario in which a model's home jurisdiction limits export. Each scenario should have a documented degraded mode, whether that means a secondary provider, a self-hosted open-weight model, or a manual process.

Enterprises should also weigh the trade-off between diversity and governability. Routing through many aggregators to many providers increases the number of parties whose decisions can affect availability, while routing everything through one direct provider maximizes exposure to that provider's policy changes. CSA's view, consistent with its earlier sovereign access guidance [12], is that the enterprise

should own its control plane, hold its own provider relationships for critical functions, and use intermediaries for convenience and cost optimization rather than as the sole path to a capability. Finally, organizations should participate in industry and standards efforts that push providers toward published intermediary criteria, transparent approval lists, and predictable notice periods, because the absence of those elements is itself a source of unmanaged risk.

CSA Resource Alignment

CSA's *Sovereign AI Access Controls and Frontier Model Dependency Risk* [12] is the most directly relevant prior work. It proposed an enterprise-side governance layer built around AI gateways, identity and access management for AI, data residency architecture, and multi-provider resilience. This note refines that framework rather than replacing it: the gateway remains the right place to enforce enterprise policy, but the paper's recommendation should now be read as "an enterprise-controlled gateway with direct provider relationships," because a hosted intermediary's own access can be conditioned by providers.

Sovereign AI Access Controls and Enterprise Concentration Risk [13] analyzed the June 2026 suspension of two Anthropic frontier models and framed frontier AI access as a regulated commodity. CSA's white paper *Sovereign AI Risk: When Your AI Vendor Gets Export-Controlled* [18] examined the same event from the enterprise's perspective. The ownership-based and nationality-based exclusions discussed here show that the same commodity logic applies to intermediaries, and the scenario planning these papers recommend should be extended to include the gateway's ownership and the provider's approval of it.

AI Provider Concentration Risk: Enterprise Resilience [14] identified strategic pivots, including acquisitions and terms-of-service changes, as a distinct failure mode and recommended an abstraction layer with failover routing. The Cursor wind-down and the intermediary approval requirement are concrete instances of that failure mode, and they indicate that the abstraction layer must itself be included in the dependency inventory that paper recommends.

The security of the gateway as an asset is addressed in CSA's research note *LiteLLM 1.107.0: When Model-Gateway Logs Become the Credential Store* [15], which is relevant to the identifier-handling and in-house gateway recommendations above. For control mapping, the AI Controls Matrix (AICM) v1.1 [16] provides the supply chain management, identity and access management, and business continuity control objectives against which gateway suppliers and internal gateway operations can be assessed.

References

- [1] Crypto Briefing. "[OpenAI updates usage policies to require approval for model routers and AI gateways.](#)" Crypto Briefing, October 21, 2026.
- [2] CellCog. "[OpenAI Usage Policy 2026: Gateways Need Approval Nov 20.](#)" CellCog Blog, October 21, 2026.
- [3] OpenAI. "[Usage Policies.](#)" OpenAI, revised October 2026, reported effective November 20, 2026.
- [4] Channel Insider. "[OpenAI to Cut Cursor Off From Its Models After SpaceX Acquisition.](#)" Channel Insider, September 3, 2026.
- [5] The Decoder. "[Anthropic cuts Claude access for Windsurf after OpenAI's \\$3B takeover news.](#)" The Decoder, June 2025.
- [6] Anthropic. "[Updating restrictions of sales to unsupported regions.](#)" Anthropic, September 4, 2025.
- [7] Al Jazeera and Reuters. "[US asks Anthropic to block global access to top AI models: Why it matters.](#)" Al Jazeera, June 14, 2026.
- [8] The Next Web. "[China weighs curbing overseas access to its top AI models.](#)" The Next Web, July 7, 2026.
- [9] Datadog Security Labs. "[LiteLLM supply chain compromise: What you need to know.](#)" Datadog, October 21, 2026.
- [10] Socket. "[LiteLLM PyPI Package and Docker Images Compromised to Steal Cloud and AI Credentials.](#)" Socket, October 21, 2026.
- [11] European Union. "[Regulation \(EU\) 2022/2554 on digital operational resilience for the financial sector \(DORA\).](#)" Official Journal of the European Union, December 2022.
- [12] Cloud Security Alliance. "[Sovereign AI Access Controls and Frontier Model Dependency Risk.](#)" CSAI Foundation, June 27, 2026.
- [13] Cloud Security Alliance. "[Sovereign AI Access Controls and Enterprise Concentration Risk.](#)" CSAI Foundation, June 16, 2026.

- [14] Cloud Security Alliance. "[AI Provider Concentration Risk: Enterprise Resilience](#)." CSAI Foundation, June 19, 2026.
- [15] Cloud Security Alliance. "[LiteLLM 1.107.0: When Model-Gateway Logs Become the Credential Store](#)." CSAI Foundation, October 23, 2026.
- [16] Cloud Security Alliance. "[AI Controls Matrix \(AICM\) v1.1](#)." Cloud Security Alliance, 2026.
- [17] OpenAI. "[Broad access through consistent standards](#)." OpenAI, October 2026.
- [18] Cloud Security Alliance. "[Sovereign AI Risk: When Your AI Vendor Gets Export-Controlled](#)." CSAI Foundation, July 2, 2026.