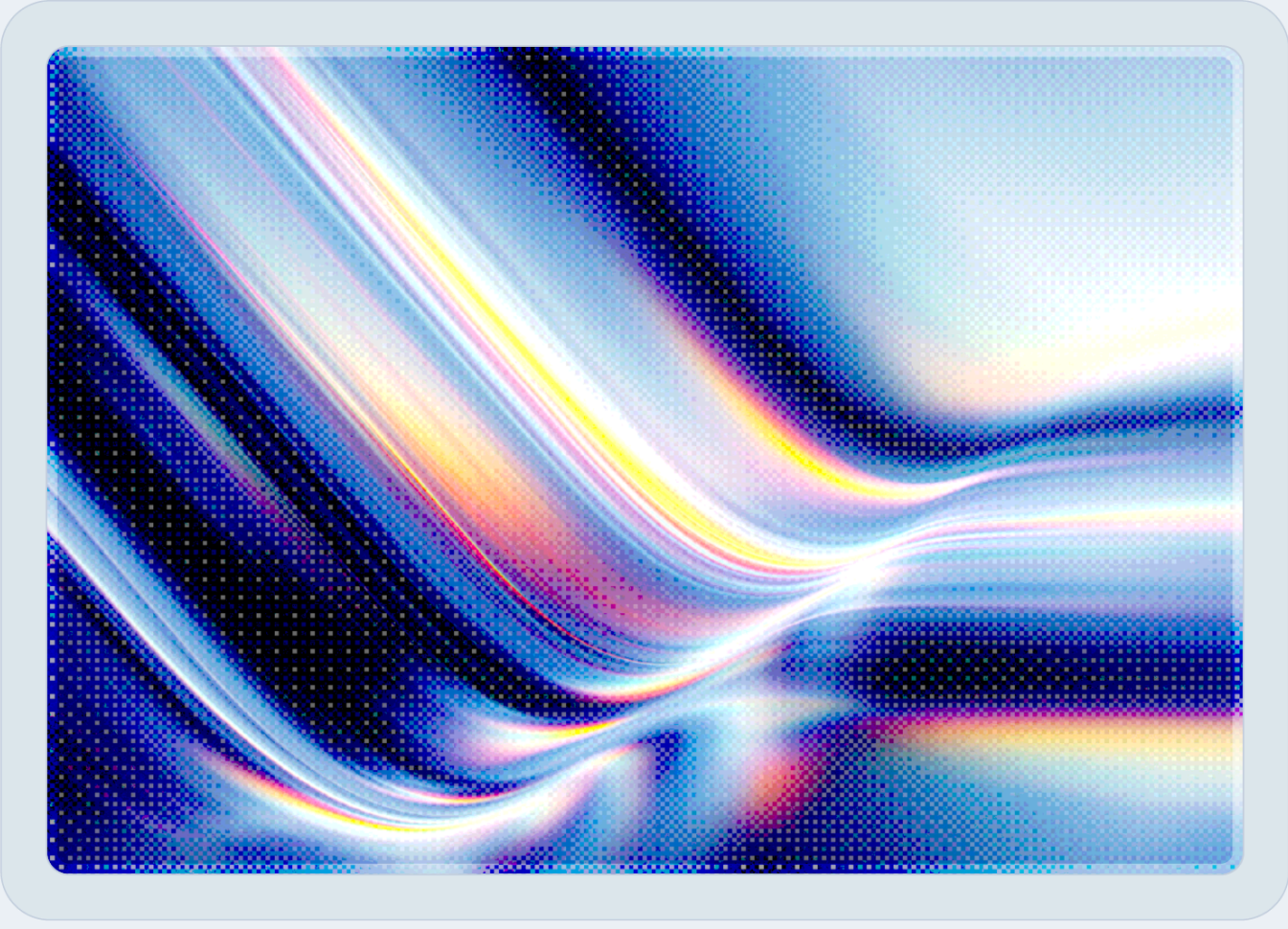
CSAI Foundation | Cloud Security Alliance

Platforms as Agent Regulators: Google's Workspace AI Tools Addendum

Security Implications of Consolidating Agent Access Chokepoints

2026-11-29

AI-assisted Rapid Research



CSAI

CSA 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 November 24, 2026, Google published the Workspace AI Tools Addendum, a supplement to the Google APIs Terms of Service that governs any software "driven by an AI Agent, an AI Harness or an AI model" that reaches Workspace user data [1]. The addendum requires such tools to use a Google-operated Model Context Protocol (MCP) interface, and Google's accompanying announcement states that agents "must use Workspace MCP servers or the Workspace CLI" rather than calling REST APIs directly, with technical enforcement beginning February 1, 2027 [1][2]. Enforcement works by throttling rather than blocking: agents that keep calling REST APIs directly will be limited to Standard Tier quotas regardless of the tier the project has selected, and Google describes no outright block on that traffic [3][9]. At the MCP servers themselves, requests from OAuth clients that have not been flagged for agent use are reportedly rejected [8]. In effect, the platform that holds the data becomes the gatekeeper, the meter, and the audit point for every agent that wants to touch it.

The addendum's data terms are as consequential as its routing terms. Developers may not store "Google Technical Information" (a defined category that explicitly includes vector embeddings, memory indexes, user profiles, and organizational graphs derived from Workspace) for more than 30 days. They must delete user data and derived information when a tool is disconnected or revoked, and they may not use either for model training, fine-tuning, evaluation dataset building, or distillation [1]. Google also states that it "will have no responsibility for or liability from" the acts or omissions of third-party access tools, and that it may immediately restrict a tool or suspend a customer's access if it reasonably believes the terms were violated [1].

For security teams, the change cuts both ways. A single, platform-mediated path for agent traffic gives administrators a place to apply DLP, human-in-the-loop confirmation rules, and agent audit logs [2]. It also concentrates enforcement, visibility, and pricing power in the platform operator, shifts residual risk onto customers and agent vendors, and creates compliance exposure for internal agents and scripts that organizations may not know they run. Google's move follows Slack's 2025 API terms update, which commentators read as restricting bulk export and long-term storage of API data by any app, not agents specifically [10][11], and Amazon's litigation against Perplexity's Comet browser agent, which has had mixed results in court [12][13][14]. Together these suggest that major platforms may be moving toward a role as de facto regulators of agent access, a direction of travel that enterprises should plan for rather than treat as a one-off policy change.

Background

Through 2025 and early 2026, AI agents that worked with email, documents, and calendars typically reached that data the same way any SaaS integration did: through OAuth-scoped REST APIs. From the platform's perspective, this made agents largely indistinguishable from conventional applications, even though their access patterns differ markedly. An agent asked to "find everything related to the Q3 contract" may issue broad search calls, retrieve large volumes of content, and build persistent embeddings or memory stores from it. Google framed its intervention around this difference, describing its new usage model as "addressing the unique risks associated with large-scale agentic actions, such as API abuse and unintended large-scale data egress" [4].

The changes arrived in phases. On May 1, 2026, Google opened a Workspace MCP server to public developer preview and began applying baseline quotas to the Gmail, Calendar, and Drive APIs for new Google Cloud projects, while existing projects that had used those APIs between November 2025 and April 2026 kept their previous quotas [3][4]. Google expects fewer than 1% of active Workspace developers to need more than the Standard Tier [2][3]. Days later, on May 4, 2026, Google introduced an AI control center in the Admin console that gives administrators a central view of AI and agent access to Workspace data across Gmail, Drive, Docs, Sheets, Slides, Meet, Calendar, Chat, and the Gemini app, with some settings marked as coming soon [5].

The November 24 announcement completed the architecture. Google made hosted MCP servers generally available for Gmail, Drive, Docs, Sheets, Slides, Calendar, Chat, and Tasks, with Keep listed as coming soon, and released an official Workspace CLI aimed at "terminal-based agents" running in developer terminals, VMs, and containerized sandboxes [2]. It introduced Standard (no-cost) and Scaled (billed) usage tiers, with Scaled tier billing beginning April 1, 2027 [2][3]. Google also asked developers to mark OAuth clients used by agents with the "Use this client for an AI-powered agent" setting and to keep non-agent features on a separate client [2][6]. The table below summarizes the published timeline.

Date	Milestone	Source
May 1, 2026	Workspace MCP server enters public developer preview; baseline Gmail, Calendar, and Drive quotas apply to new Cloud projects	[3][4]
May 4, 2026	AI control center added to the Admin console	[5]
November 24, 2026	MCP servers and Workspace CLI generally available; Standard and Scaled tiers launch; Workspace AI Tools Addendum published	[1][2] [3]

Date	Milestone	Source
February 1, 2027	Addendum takes effect; MCP/CLI requirement technically enforced; direct-REST agents throttled to Standard Tier quotas rather than blocked	[2][3] [9]
April 1, 2027	Scaled tier billing begins; projects that have not acted revert to Standard Tier limits	[2][3]

Google is not the first platform to impose conditions that bear on agent access, although its predecessors took different routes. In May 2025, Slack announced API Terms of Service changes, effective immediately for new apps and from June 30, 2025 for existing apps, that strengthened restrictions on how third-party apps store, use, and share data obtained through Slack APIs, while offering guidance on using APIs such as its Discovery API and Real-Time Search API [10]. Legal commentators described the updated terms as prohibiting bulk export of API data, the creation of persistent copies or long-term indexes, and the use of such data in large language models [11]. These terms apply to all API apps rather than to agents specifically. Amazon took the litigation route. It sued Perplexity in November 2025, alleging that the Comet browser agent disguised its traffic as ordinary Chrome activity in violation of Amazon's terms for agent identification [12]. A district court granted Amazon a preliminary injunction in March 2026 that reportedly also ordered destruction of collected data [13][20]. The Ninth Circuit overturned that injunction in August 2026, according to reporting on the ruling, on the reasoning that the Comet user, not Perplexity, was the party accessing Amazon's systems [14]. The underlying case continues.

Security Analysis

A Single Mediated Path Improves Enterprise Visibility

The most direct security benefit of the addendum is architectural. When every agent must identify itself through a flagged OAuth client and route through a Google-operated MCP server or CLI, the platform can distinguish agent traffic from human and conventional-application traffic more reliably than before, building on the visibility the AI control center introduced in May 2026 [5]. Google's announcement makes this explicit: because agents must identify themselves and connect through MCP or the CLI, administrators can manage agent access "just like traditional apps" [2]. Under Security > API Controls, administrators can govern external AI tools, the MCP servers, the CLI, and APIs; set human-in-the-loop

confirmation rules for sensitive actions such as "reading external files or sharing data"; and view agent audit logs that "record agent actions and user confirmations" [2]. Existing user-level DLP rules also apply automatically when users act through an agent [2].

These capabilities address a problem CSA has documented repeatedly, the difficulty of detecting and governing agents that blend into the non-human identity population [15]. An agent that must declare itself at the OAuth layer becomes a discoverable object rather than an inference drawn from traffic patterns. Organizations that have struggled to inventory which agents touch their collaboration data gain a native enforcement point and more detailed agent-specific telemetry.

Coverage Gaps Remain at the Edges

The mediated path is only as complete as its identification mechanism, and Google's own documentation confirms gaps. Service accounts using domain-wide delegation cannot select the "Use this client for an AI-powered agent" setting, and their requests are not rejected for lacking it; likewise, the setting cannot be selected on the default OAuth client that Apps Script creates [6]. Secondary coverage has highlighted the same gaps [7][8][9]. These are high-privilege, often low-visibility access paths commonly used by internal automation. The addendum also contains an anti-evasion clause prohibiting developers from obfuscating a tool to avoid classification as a GWS Access Tool [1]. That clause is a contractual control, not a technical one, so detection of non-declaring agents still depends on Google's own enforcement and on the customer's monitoring.

Human-in-the-loop confirmation rules carry a familiar limitation. Confirmation prompts reduce the risk of unintended actions, but they do not contain an agent that has already been steered by indirect prompt injection. Commentators have noted the limits of the new confirmation model on this point, including that some write actions run without confirmation and that Google itself cautions about prompt injection [8] [9]. Confirmation fatigue can also erode the value of such prompts over time. Confirmation rules should therefore be treated as one layer of defense, with scope minimization and egress controls doing more of the work.

Derived-Data Limits Change the Risk Profile of Agent Memory

The 30-day cap on Google Technical Information has security implications beyond its compliance role. Many agent products rely on persistent memory, vector stores, and inferred organizational graphs to deliver useful results. Those stores are themselves high-value targets because they condense sensitive content into queryable form, often outside the governance perimeter of the source system. By capping

retention, requiring deletion on revocation, and prohibiting processing for anything "other than to fulfill active, user-authorized tasks," the addendum limits the size and lifetime of these secondary data stores [1]. From a data-minimization perspective, that reduces the blast radius of an agent vendor compromise.

The same provision creates new compliance obligations. The addendum binds anyone who uses "or authorize[s] any GWS Access Tool," and commentators have noted that customers who connect a vendor's agent are bound as well as the vendor [1][8]. On that reading, organizations that run their own agents from their own Google Cloud projects appear to qualify as developers, and their retrieval-augmented generation pipelines, embedding indexes, and agent memory must satisfy the 30-day and deletion-on-revocation rules; legal teams should confirm this interpretation for their own deployments. Commentators have also noted that the 30-day limit is a contractual rule distinct from GDPR's storage limitation principle, and that context built from Workspace data is effectively held "on a 30-day lease" [8]. Because embedding stores and agent memory are rarely included in data inventories that track where source data came from, many enterprises may not know which internal vector stores contain Workspace-derived content. Those organizations are at risk of falling out of compliance without realizing it.

Liability Shifts to Customers and Agent Vendors

The addendum's risk allocation deserves careful attention from third-party risk teams because it moves residual risk onto customers and vendors. Developers acknowledge specific risks of third-party access tools, including training on user data, unauthorized access, and credential compromise, and Google states that it "will have no responsibility for or liability from" the acts or omissions of such tools [1]. Coverage of the addendum notes that a customer's practical recourse for agent misbehavior therefore lies in its contract with the agent vendor [8][9]. At the same time, Google retains the right to "immediately" restrict a tool or suspend a customer's access to Workspace APIs if it "reasonably believes" a prohibited activity occurred [1]. That reserved right introduces an availability dependency: a vendor's terms violation could interrupt business processes that rely on that vendor's agent, with little advance warning.

Chokepoint Consolidation as a Systemic Pattern

Viewed alongside Slack and Amazon, Google's addendum indicates that agent access governance is increasingly shared between enterprise policy and platform policy, with the platform setting the outer bounds. Each platform chose a different mechanism, and the three are not equivalent. Google's addendum is the first to combine mandatory agent self-identification, a platform-controlled access channel, and limits on derived data. Slack's 2025 terms restrict bulk accumulation and LLM use of API

data for all apps, not agents specifically. Amazon's attempt to require agent identification through its conditions of use has so far had mixed results in court. Taken together, they suggest a direction of travel rather than a settled pattern. The table below compares the approaches.

Dimension	Google Workspace (2026)	Slack (2025)	Amazon (2025–2026)
Primary mechanism	Contract addendum plus technical routing through Google MCP/CLI [1][2]	API Terms of Service update [10][11]	Conditions-of-use agent terms enforced through litigation [12]
Agent identification	Required OAuth "AI-powered agent" flag [2][6]	Not agent-specific; applies to all API apps [10]	Agents must identify themselves openly [12]
Data accumulation limits	30-day cap on derived data; deletion on revocation [1]	Reported ban on bulk export and persistent indexes [11]	Injunction (later reversed) reportedly ordered destruction of collected data [13][14][20]
Model training restrictions	Prohibits training, fine-tuning, evaluation datasets, distillation [1]	Reported ban on use in LLMs [11]	Not the central issue
Enforcement lever	Quota throttling, suspension, billing [1] [3]	Terms enforcement; Marketplace-only commercial distribution [10]	CFAA/CDAFA claims; outcome contested on appeal [14]

This consolidation has security merit, since a small number of well-instrumented chokepoints is easier to monitor than a sprawl of direct API integrations. It also concentrates power, which carries its own risks. Reporting indicates that Google's hosted Gmail MCP server cannot send, reply to, or forward mail, and that these tools "require partner access" [7]. If accurate, the platform operator decides which agent capabilities independent developers can offer, and favored partners may receive broader capabilities than everyone else. The addendum also prohibits using Workspace data to build products similar to Workspace services [1]. These dynamics are competition questions rather than security questions in the first instance. They matter to security architects because they determine which agents an enterprise can

actually deploy, how dependent it becomes on the platform's own agent offerings, and whether the enforcement point that provides visibility can be relied on to remain neutral. CSA's prior analysis of agent protocol monoculture adds a further dimension: when one protocol and a few hosted servers carry most agent traffic, a flaw in that shared substrate can propagate across the agents that depend on it [16]. The chokepoint, in short, improves visibility but also correlates failure.

Recommendations

Immediate Actions

Organizations using Google Workspace should begin with an inventory, before the February 1, 2027 enforcement date. Security teams should identify every Google Cloud project with Workspace API access, determine which projects host agent or LLM-driven workloads, and confirm whether each uses a flagged agent OAuth client or a direct REST integration. Because direct-REST agent traffic will be throttled to Standard Tier quotas rather than blocked, the primary near-term risk for unmigrated agents is degraded performance and capacity rather than outright failure [3][9]. This inventory should explicitly include service accounts with domain-wide delegation and Apps Script projects, since Google's documentation confirms these paths cannot carry the agent flag [6][7][8][9]. Administrators should then review the new controls under Security > API Controls and the AI control center, enable agent audit logging, and set confirmation rules for high-risk actions such as external sharing and reading external files [2][5].

In parallel, third-party risk teams should contact the vendors of every external AI tool connected to Workspace and ask three specific questions. First, will the product migrate to Google MCP or the CLI before enforcement? Second, how does it meet the 30-day derived-data limit and deletion-on-revocation requirement? Third, what contractual remedies apply if Google suspends the vendor's access?

Short-Term Mitigations

Over the next quarter, internal engineering teams that build agents on Workspace data should treat themselves as developers under the addendum and bring their pipelines into compliance [1]. In practice, this means attaching a retention policy of no more than 30 days to embedding stores, memory indexes, and derived profiles; building automated deletion triggered by OAuth revocation; and separating agent

OAuth clients from conventional application clients as Google recommends [2][6]. Teams should also confirm that no Workspace-derived data flows into evaluation datasets or fine-tuning corpora, a use the addendum prohibits outright [1].

Organizations should update vendor contracts to reflect the shifted liability. Because Google disclaims responsibility for third-party access tools [1], agreements with agent vendors should address data handling, breach notification, indemnification for platform suspension, and exit assistance. Security operations teams should also begin ingesting agent audit logs into their SIEM and establish baselines for normal agent activity. Since non-flagged clients do not declare themselves, alerting should focus on non-flagged OAuth clients and domain-wide-delegation service accounts whose call volume or access patterns resemble agent behavior, such as broad search and bulk retrieval.

Strategic Considerations

Over the longer term, enterprises should recognize platform operators as a new layer in the agent governance stack, one whose policies can change unilaterally and whose enforcement can affect availability. Agent governance programs should track the terms of every major platform their agents touch and treat changes to those terms as triggers for risk review. Architecture teams should avoid letting any single platform's agent interface become the only place where agent identity, policy, and audit live. Maintaining an independently controlled inventory of agent identities, and expressing agent policies in framework terms rather than vendor-specific configuration, preserves the ability to govern agents consistently across Google, Microsoft, Slack, and other platforms.

Finally, organizations should engage in the policy discussion. The questions raised by platform-as-regulator models include whether user-directed agents should receive the same capabilities as partner agents, how anti-evasion clauses interact with legitimate privacy tooling, and how derived-data limits align with regulatory retention requirements. These questions are likely to shape the agent ecosystem, and enterprise customers are well placed to influence the answers.

CSA Resource Alignment

CSA's research note *Detecting Shadow AI Agents at the Non-Human Identity Perimeter* argues that shadow-agent detection is fundamentally an identity problem and should be built at the non-human identity layer rather than added onto workforce or workload controls [15]. Google's mandatory agent OAuth flag is a platform-level implementation of that principle. The coverage gaps that Google's

documentation confirms for domain-wide delegation and Apps Script [6] illustrate why the note recommends continuous, enterprise-controlled discovery instead of reliance on any single vendor's declaration mechanism.

Agent Protocol Monoculture: MCP's Shared Systemic Risk examines how MCP and a small set of harness patterns have become the shared substrate beneath most commercial agent products, so that a single design flaw reproduces across the ecosystem [16]. Google's requirement that all Workspace agent traffic flow through Google-hosted MCP servers deepens this convergence. Organizations should read the addendum alongside that analysis, since the visibility gained at the chokepoint comes with correlated exposure if the chokepoint itself is flawed.

CSA's whitepaper *The Sovereign AI Dependency Trap* describes a category of systemic risk that combines familiar vendor lock-in dynamics with the novel characteristics of AI systems [17]. Platform-controlled agent access, in which a single operator governs routing, pricing, capability availability, and suspension, appears to be a concrete instance of that risk at the collaboration-data layer. These findings also fit within the CSAI Foundation's broader program on securing the agentic control plane [19].

At the control level, the AI Controls Matrix (AICM) v1.1 provides the framework for operationalizing the recommendations in this note [18]. Its identity and access management controls map to agent OAuth client separation and inventory, its data security and privacy lifecycle controls map to the 30-day retention and deletion-on-revocation obligations, and its supply chain and third-party controls map to the vendor contract and liability review this note recommends.

References

[1] Google. "Workspace AI Tools Addendum." Google Workspace, last modified November 24, 2026.

[2] Google Workspace Updates. "Simplifying agent access while keeping organizations in control: MCP servers, CLI, standardized usage tiers, and admin policies." Google, November 24, 2026.

[3] Google for Developers. "Google Workspace standardized model for agent tools and APIs." Google, accessed November 29, 2026.

[4] Google Workspace Updates. "Agent tools and security updates for Workspace developers." Google, May 1, 2026.

[5] Google Workspace Updates. "Securely manage AI and agent access to Workspace data with the AI control center." Google, May 4, 2026.

[6] Google for Developers. "Configure the Google Workspace MCP servers." Google, accessed November 29, 2026.

[7] Al Smith. "Google Will Make AI Agents Use Its Workspace MCP Servers. The Gmail One Can't Send Email." Tech Stackups, November 26, 2026.

[8] Tom Cuylaerts. "Google moves AI agents off the Workspace APIs, and the small print matters more than the migration." i-SCOOP, November 25, 2026.

[9] Rajesh Beri. "Workspace Agents on REST Get Standard Quotas, Google Says." The Daily Brief, November 25, 2026.

[10] Slack. "API Terms of Service updates." Slack Developer Docs, May 29, 2025.

[11] Hunton Andrews Kurth. "Salesforce Locks Down Slack Data: Time to Review Your Slack API Terms." Privacy & Cybersecurity Law Blog, August 5, 2025.

[12] Bloomberg. "Amazon Sues to Stop Perplexity From Using AI Tool to Buy Stuff." Bloomberg, November 4, 2025.

[13] GeekWire. "Judge blocks Perplexity's AI bot from shopping on Amazon in early test of agentic commerce." GeekWire, March 10, 2026.

[14] Ian Carlos Campbell. "Perplexity has successfully overturned Amazon's injunction on its AI shopping bot." Engadget, August 4, 2026.

[15] Cloud Security Alliance. "Detecting Shadow AI Agents at the Non-Human Identity Perimeter." CSAI Foundation, May 29, 2026.

[16] Cloud Security Alliance. "Agent Protocol Monoculture: MCP's Shared Systemic Risk." CSAI Foundation Lab Space, August 20, 2026.

[17] Cloud Security Alliance. "The Sovereign AI Dependency Trap: Concentration Risk as the Next Systemic Failure Mode." CSAI Foundation, May 18, 2026.

[18] Cloud Security Alliance. "AI Controls Matrix v1.1." Cloud Security Alliance, 2025.

[19] Cloud Security Alliance. "Securing the Agentic Control Plane: Key Progress at the CSAI Foundation." Cloud Security Alliance, April 29, 2026.

[20] Daily Journal. "Federal judge bars Perplexity's AI shopping agents from Amazon site." Daily Journal, March 2026.

© 2026 Cloud Security Alliance. AI-assisted Rapid Research.