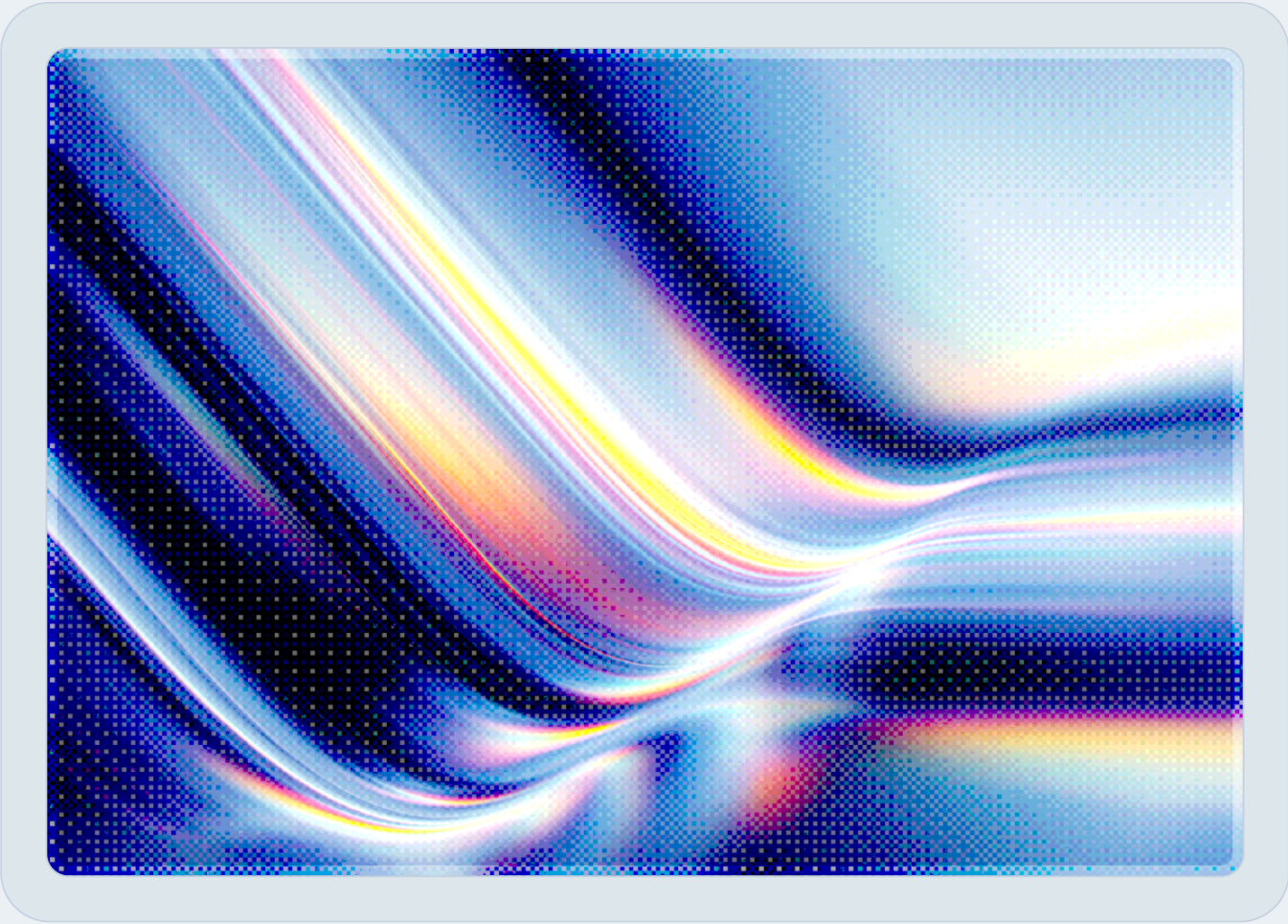
CSAI Foundation | Cloud Security Alliance

OCC Bulletin 2026-54: Examiner Framework for Agentic AI in Banking

What the OCC's Supervisory Framework for Agentic Super Intelligence Means for Banks and Their AI Vendors

2026-11-13

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 4, 2026, the Office of the Comptroller of the Currency (OCC) issued Bulletin 2026-54, "Super Intelligence: Supervisory Framework," together with a framework document that tells national bank examiners how to evaluate banks' use of generative and agentic AI [1][2][3]. "Super Intelligence" (SI) is the OCC's term for the technologies previously called "artificial intelligence," a change in wording that follows Executive Order 14434 [1][5]. It is not a claim about the capability of the systems under review. The framework is the OCC's first dedicated examiner resource to address agentic systems directly. It fills the gap left in April 2026, when the revised interagency model risk management guidance (OCC Bulletin 2026-13) placed generative and agentic AI outside its scope [2][4].

The first point for security, risk, and vendor teams is that the framework is an examiner tool, not a rule. The OCC says it "does not establish supervisory expectations for banks," but in practice it shows what examiners will ask about [1]. The second is that consequences matter more than sophistication. Control intensity scales with the decision consequence, reversibility, impact, autonomy, data sensitivity, explainability, and provenance of each use case, so an autonomous customer-facing payment agent sits at Tier 1 "regardless of how well it performs in testing" [2]. Third, agentic failure counts as a resilience event: the framework expects banks to detect and contain an agent that "could remain available while taking inappropriate or harmful actions," using tools such as circuit breakers, rate limits, least privilege, and kill switches [2].

Two further points concern identity and vendors. Agents should be governed as non-human identities, since the framework says agent credentials and permissions may be governed "in a manner comparable to privileged users" and treats the breadth of an agent's tool access as attack surface [2]. Vendor relationships also face closer scrutiny. Examiners will look at the scope of provider agents' authority, SI incident definitions, liability allocation, and notice when vendors add SI features to products a bank already uses [2].

Background

From SR 11-7 to a Gap for Generative and Agentic AI

For fifteen years, bank oversight of quantitative models in the United States rested on OCC Bulletin 2011-12 and the parallel Federal Reserve SR 11-7. Supervisors and banks also stretched that guidance to cover machine learning. On April 17, 2026, the OCC, Federal Reserve, and FDIC replaced it with shorter, risk-based guidance issued as OCC Bulletin 2026-13, SR 26-2, and FIL-15-2026 [4][7]. The revised guidance narrowed what counts as a "model," stated that it does not create enforceable standards, and left generative and agentic AI out of scope [4][7]. Bulletin 2026-54 now confirms the division of labor in a footnote. The principles of Bulletin 2026-13 apply to non-generative, non-agentic SI models, so the new framework "is more applicable to generative and agentic SI" [2].

The vocabulary comes from Executive Order 14434, "Inaugurating the Era of Super Intelligence," published at 91 Fed. Reg. 63129 on October 2, 2026. The order directs federal agencies to use "Super Intelligence" in place of "Artificial Intelligence" [5]. The OCC defines SI as the technologies previously covered by the statutory definition of AI in 15 U.S.C. 9401(3), "as well as the promise, potential, and rapidly advancing capabilities of those technologies" [2]. Readers should treat the terminology as a change in wording, not a new technical category. The framework's own lexicon reuses definitions from the AI Lexicon workstream of the Artificial Intelligence Executive Oversight Group (AIEOG), a Treasury-convened public-private group, and substitutes "SI" for "AI" throughout [2]. The OCC also notes that the framework "was prepared with the assistance of SI" and was reviewed and approved by a human [2]. In CSA's view, that disclosure is unusual for a federal supervisory document.

What the Framework Contains

The framework document runs to about twenty numbered pages, roughly fifteen pages of framework text followed by a lexicon, and is organized around one principle. SI is not a separate risk category, and banks are not required to define a separate SI risk appetite. SI instead "can introduce, amplify or modify risk in existing risk categories, such as operational, compliance and credit" [2]. Sixteen components are grouped into six pillars, and an illustrative four-tier risk methodology and a technical lexicon follow. The OCC calls the framework a "living document" that it expects to update as technology and supervisory understanding mature [1][2][3]. The bulletin, signed by Senior Deputy Comptroller and Chief National Bank Examiner James M. Gallagher, applies to national banks, federal savings associations, and federal branches and agencies of foreign banking organizations. It tells examiners to scale reviews of community banks to each bank's size, complexity, and risk profile [1].

The six pillars each pose a single supervisory question, and several components appear under more than one pillar. The table below shows how the OCC maps the sixteen components onto those questions.

Pillar	Supervisory question	Framework components
Purpose	What is the system or agent permitted to do?	Defined Purpose and Permitted Use
Risk	What happens if its output or action is not fit for purpose?	Risk Classification and Criticality Assessment
Control	What prevents unauthorized, inappropriate, or unsafe action?	Human-in/on-the-Loop; Model Risk Management; Data Governance; Privacy; Compliance; Cybersecurity
Oversight	Who monitors the system, and who is accountable?	Governance & Accountability; Human-in/on-the-Loop; Performance Monitoring; Reporting & Transparency
Resilience	Can the bank detect, contain, recover, and remediate while the business continues?	Operational Resilience; Cybersecurity
Assurance	Is there evidence that controls keep working?	Testing; Documentation & Auditability; Third-Party & Vendor Risk; Internal Audit & Independent Challenge

Source: OCC Super Intelligence Supervisory Framework, Section 1.2 [2].

The timing matters because agentic deployment in financial services is already well advanced. In CSA's 2026 financial services survey of 340 respondents, 62% of respondents reported that their organizations were deploying AI agents, and 93% of those using agents granted them some form of autonomy. Respondents also anticipated agent-driven payments, and third-party and supply-chain risk was their top cloud security concern [8]. These figures are self-reported by survey respondents and should be read as indicators of direction rather than a census of the industry. Even so, they suggest the OCC framework arrives after agentic deployment is already under way, not ahead of it.

Security Analysis

Risk Tiering Targets Agency, Not Model Capability

In CSA's assessment, the framework's most consequential design choice is its tiering methodology. Every SI and agentic use case is scored against seven factors that apply the same way in every business line. The factors are decision consequence (inform, decide, or act), reversibility, customer or financial impact, autonomy level, data sensitivity, explainability, and provenance (built in-house, fine-tuned from a third-party base model, or consumed as an external service) [2]. These scores combine into a composite tier. The OCC states the intent plainly: controls should be "light-touch for low-risk, low-autonomy use, materially heavier for anything approaching customer-impacting, irreversible, or highly autonomous action" [2]. The illustrative tiers that result are summarized below.

Illustrative tier	Profile	Illustrative use cases
Tier 1: Critical	Direct action, low or no reversibility, high financial or customer impact, autonomous or partially autonomous	Autonomous payment agent; real-time credit decisioning; fraud-hold release; market access
Tier 2: High	Direct action or high-weight decision support, human-on-the-loop, moderate-to-high impact	SI-assisted underwriting; agent-initiated customer communications; AML alert disposition; code assistance for internal code or data access changes
Tier 3: Moderate	Decision support with mandatory human-in-the-loop, limited data sensitivity, reversible	Internal document summarization for credit memos; agent-assisted research
Tier 4: Low	No customer or financial data exposure, no direct action, fully human-controlled	Internal productivity tools; code assistance; general employee Q&A

Source: OCC Super Intelligence Supervisory Framework, Section 3.2 [2].

From a security standpoint, CSA considers this structure well aligned with how agentic threats actually show up. A capable model that only drafts internal memos has a narrow blast radius. A simple agent holding payment-initiation credentials, by contrast, can cause irreversible loss at machine speed if it is manipulated through prompt injection or misconfigured. Coding assistants appear in two tiers, which illustrates the point. General code assistance is Tier 4, but code assistance "for internal code or data access changes" moves to Tier 2 [2]. Security teams should therefore expect examiners to ask whether a bank's tier assignments follow the permissions an agent actually holds, including its tools, credentials, and reachable systems, rather than the stated purpose of the business process it supports.

Agentic Threats Named in Supervisory Language

The cybersecurity component explicitly names the main agentic attack classes. These are prompt injection and jailbreaking that can lead an agent to "misuse connected tools without a conventional network breach," data and model poisoning, and model extraction and data leakage. The component also names "tool use and excessive agency," where broad credentials let a compromised agent "cause fraud, data exposure, or operational disruption at machine speed" [2]. The framework's answer is identity-centric. It calls for governing agent identities, credentials, logs, and data use like privileged users and for treating "the breadth of agent permissions and tool access" as part of the attack surface [2].

This emphasis matters in practice. In CSA's experience, agents are often provisioned through shared service accounts or broadly scoped API keys, which this expectation would directly address. The scale of the identity problem is also significant: an industry estimate cited in CSA's survey report puts the ratio at roughly 96 non-human identities per human identity in financial services, though the report notes that precise figures vary by source and methodology [8].

The purpose component links cybersecurity to authorization design. Before deployment and after any material change, a bank may document what an agent is and is not designed to do, who or what may invoke it, its permitted data, systems, and actions, and the conditions that call for human review or suspension. These boundaries should be stated "in terms that can be tested rather than as a general capability description" [2]. Activity outside the approved purpose is framed as scope drift that can indicate a control failure. In effect, this asks banks to keep a declarative, testable permission manifest for each agent and to link it to the inventory.

Resilience Means Containing a Running but Wrong Agent

Traditional operational resilience programs plan for outages. The framework extends the concept to "incorrect operation," noting that "an agentic system could remain available while taking inappropriate or harmful actions" [2]. For consequential or autonomous use cases, the framework expects documented

and tested detection, containment, suspension, continuity, recovery, and remediation. The containment tools it lists are monitoring, observability tools, control agents, deterministic rules, circuit breakers, rate limits, least-privilege permissions, and kill switches [2].

The framework also addresses what happens downstream of a failure. It describes tested fallback through trained personnel or a "narrower fallback agent," and a separate customer-remediation workstream that identifies the complete affected population and independently verifies completion. It further calls for scenario analysis of correlated or cascading high-speed agent activity across customers, agents, and providers [2].

Beyond a bank's own agents, the framework treats external agents as channel risk. Banks may receive traffic from customer agents, fintech agents, and agentic-commerce systems they do not control, and the framework says identity, provenance, delegation, authentication, velocity limits, dispute handling, and revocation should be "calibrated to the action requested" [2]. This suggests banks will need to authenticate and authorize inbound agents, not only their own.

Human Oversight Must Be Demonstrably Real

The framework separates human-in-the-loop (approval before an action executes) from human-on-the-loop (active monitoring with authority and practical ability to intervene). It is skeptical of nominal oversight. Review should be "more than a nominal click," high volumes or bulk approvals "could reduce review quality and contribute to alert fatigue," and banks may assess "how many agents, accounts, processes, alerts, or actions one person can credibly monitor" [2]. These properties are measurable. Examiners could reasonably ask for metrics such as approval latency, override rates, and span-of-control ratios, so banks that rely on human oversight as a mitigating control should be prepared to produce them.

Testing and Auditability for Non-Deterministic Systems

The model risk and testing components accept that SI systems may lack a verified source of truth, behave non-deterministically, and change without a traditional model release. The framework therefore prefers "repeated testing, defined samples, distributional measures, and consistency thresholds" over "a single successful run." Its testing methods include adversarial testing, red teaming (including testing whether the bank can detect, contain, and recover from realistic attacks), expert subjective review, LLM-as-a-judge with attention to the judge's own bias and correlated failures, and hallucination measurement under representative and adversarial prompts [2].

Monitoring carries the same logic forward into production. It is expected to catch upstream vendor model changes, prompt and configuration changes, and new tools or permissions that "alter behavior without a traditional model release" [2]. For auditability, the framework expects records of inputs, outputs, context, retrieved sources, tool calls, approvals, overrides, and actions that are sufficient "to reconstruct a material decision or event," even when the model cannot produce a rule-based rationale [2].

Implications for AI Vendors

The framework says outright that banks "frequently obtain SI capabilities from third parties" and that this introduces risks "not fully addressed by conventional service-level and information-security reviews" [2]. Bank accountability is unchanged when a vendor builds or hosts the tool [7]. The third-party component identifies four areas that will shape contract negotiations and due diligence. The table below pairs the OCC's examiner considerations with CSA's assessment of the likely effect on vendors.

Area	What examiners may look for	Likely vendor impact (CSA analysis)
Provider agent visibility	Defined authority, identity, permitted tool use, data use, and escalation limits for provider agents operating in bank systems; bank ability to monitor and limit them	Per-agent identities, scoped credentials, and customer-visible activity logs
SI incidents	A specific incident vocabulary (hallucination, misfire, drift, manipulation) with notification timelines	Contractual SI incident definitions and notice SLAs beyond security breaches
Liability	How contracts allocate model-related errors and losses, including customer harm (clarity, indemnification, caps)	Pressure on standard AI liability disclaimers
Change and exit	Notice and transparency when vendors add SI features or change underlying models; right to reassess or retest	Advance change notification, model version pinning, and release notes

Source: Areas and examiner considerations from OCC Super Intelligence Supervisory Framework, Section 2.9 [2]; vendor impact column is CSA analysis.

The change-notice expectation may be the most disruptive for software-as-a-service providers. Embedding generative features into existing products, or silently changing the underlying foundation model, is the kind of change the framework says may prompt a bank to reassess the relationship or retest the model [2]. The framework also signals that "vendor and model changes" should appear in the bank's own performance monitoring [2].

These vendor provisions sit on top of a broader shift in third-party guidance. On September 11, 2026, the OCC, Federal Reserve, FDIC, and NCUA proposed replacing the 2023 interagency third-party guidance; the proposal was published in the Federal Register on September 15, and comments are due November 16, 2026 [6][10]. Based on available summaries, the proposal does not focus on AI specifically [10]. The OCC framework therefore currently serves as the agency's main SI-specific third-party reference.

Limits and Open Questions

Several limits deserve attention. The framework creates no enforceable standards, and its language is consistently permissive ("may," "could"), so how examiners apply it will become clear only through examination practice. It applies to OCC-supervised institutions, and whether the Federal Reserve and FDIC adopt parallel tools remains open. State-chartered banks supervised by those agencies should therefore watch for any parallel guidance. The tiering is explicitly illustrative, which means banks will need to calibrate thresholds such as dollar limits and autonomy levels themselves. Finally, the "Super Intelligence" terminology may cause confusion in vendor and customer communications, although it does not change any substantive legal definition [5].

Recommendations

Immediate Actions

Banks should first build or refresh an SI inventory that records, for each generative and agentic use case, the owner, provider, autonomy level, permissions, connected tools and systems, deployment status, monitoring, and fail-over arrangements. That inventory feeds directly into the framework's purpose, governance, and tiering components [2]. Each use case should then receive a provisional tier using the seven OCC factors. Priority should go to any agent that can move funds, change customer records, release holds, or access markets, because these are the Tier 1 patterns the framework's illustrative tiering identifies [2]. For every Tier 1 and Tier 2 agent, security teams should confirm that an emergency suspension mechanism exists, works within a defined time, and has been exercised, and that a

documented fallback process can keep the business running while the agent is disabled. Vendor management teams should identify which existing products have added SI features in the past year and whether the bank was notified.

Short-Term Mitigations

Over the next one to two quarters, institutions should move agent credentials onto individually attributable, least-privilege non-human identities. The goal is to govern them like privileged users, as the framework suggests, with logging sufficient to reconstruct tool calls and actions [2]. Testing programs should add repeated-sampling consistency tests, adversarial and prompt-injection testing, and hallucination measurement for generative use cases, and red teaming should include exercises that test detection and containment, not only model behavior [2][11].

Contractual and oversight practices need parallel attention. Contract templates for AI vendors should be revised to cover provider agent scope and monitoring rights, SI incident definitions and notification timelines, liability allocation for model errors and customer harm, and advance notice of new SI features or underlying model changes [2]. Banks that rely on human approval as a control should measure approval volumes, review time, and override rates to show that oversight is meaningful [2].

Strategic Considerations

Over the longer term, institutions should treat the framework as a design specification for agent architecture, not as an exam checklist. Building agents with declarative, testable permission manifests, transaction and velocity limits, and deterministic guardrails around probabilistic components will make many of the framework's components easier to demonstrate [2]. Banks should also prepare for inbound agent traffic by building authentication, delegation verification, and revocation for external agents [2].

For AI vendors serving U.S. banks, the prudent course is to prepare for the possibility that similar examiner considerations will reach customers supervised by the Federal Reserve, FDIC, or state regulators, because bank customers often apply a single vendor standard across charters. Vendors can reduce friction by offering per-customer change notifications, model version transparency, agent activity logs, and independent assurance evidence as standard features. Institutions should also follow the pending interagency third-party guidance [6][10] and any coordinated Federal Reserve or FDIC response to the OCC framework.

CSA Resource Alignment

CSA's *The State of Cloud and AI for Financial Services 2026* is directly relevant and provides empirical context for this bulletin [8]. Its finding of a "control-lag" in which agent adoption outpaces governance maturity, together with its emphasis on agent identity governance and third-party risk, matches the OCC framework's focus on autonomy, non-human identities, and vendor oversight. Institutions can use the survey's autonomy spectrum to support their own tier assignments.

The CSA *AI Model Risk Management Framework* can help address the generative and agentic model risk that Bulletin 2026-13 placed outside its scope [9]. Its Model Cards, Data Sheets, Risk Cards, and Scenario Planning provide practical artifacts that map onto the OCC's documentation and auditability, data governance, and risk classification components. They can help banks produce the per-use-case evidence examiners will request.

The *AI Controls Matrix (AICM) v1.1* is a control catalog that covers the breadth of the six pillars [12]. Its domains for identity and access management, threat and vulnerability management, supply chain transparency, and governance can serve as a common control baseline between banks and their AI vendors. AICM-based assurance through the STAR program also gives vendors a structured way to supply the third-party evidence the framework's Assurance pillar anticipates. For testing, CSA's *Agentic AI Red Teaming Guide* gives concrete test procedures for permission escalation, tool misuse, and multi-agent failure modes, matching the framework's call for adversarial testing and red teaming of agentic systems [11]. The MAESTRO threat modeling framework [13] provides a layered method for identifying the agent-specific threats that drive tier and control decisions.

References

[1] Office of the Comptroller of the Currency. "Super Intelligence: Supervisory Framework (OCC Bulletin 2026-54)." OCC, November 4, 2026.

[2] Office of the Comptroller of the Currency. "Super Intelligence Supervisory Framework." OCC, November 2026.

[3] Office of the Comptroller of the Currency. "OCC Issues Supervisory Framework on Bank Use of Super Intelligence (News Release 2026-97)." OCC, November 4, 2026.

[4] Office of the Comptroller of the Currency. "Model Risk Management: Revised Guidance (OCC Bulletin 2026-13)." OCC, April 17, 2026.

[5] The White House. "Inaugurating the Era of Super Intelligence (Executive Order 14434)." The White House, September 2026; published at 91 Fed. Reg. 63129, October 2, 2026.

[6] Board of Governors of the Federal Reserve System, FDIC, NCUA, and OCC. "Proposed Third-Party Risk Management Guidance." Federal Register, September 15, 2026.

[7] King & Spalding. "The OCC Issues Examiner Framework for Bank Use of AI: What OCC Bulletin 2026-54 Means for National Banks, Federal Savings Associations, and Federal Branches and Agencies." King & Spalding, November 2026.

[8] Cloud Security Alliance. "The State of Cloud and AI for Financial Services 2026." CSA, 2026.

[9] Cloud Security Alliance. "AI Model Risk Management Framework." CSA AI Technology and Risk Working Group, 2024.

[10] Skadden, Arps, Slate, Meagher & Flom. "US Federal Banking Agencies Propose Revised Third-Party Risk Management Guidance." Skadden, September 2026.

[11] Cloud Security Alliance. "Agentic AI Red Teaming Guide." CSA, 2025.

[12] Cloud Security Alliance. "AI Controls Matrix (AICM) v1.1." CSA, 2026.

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

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