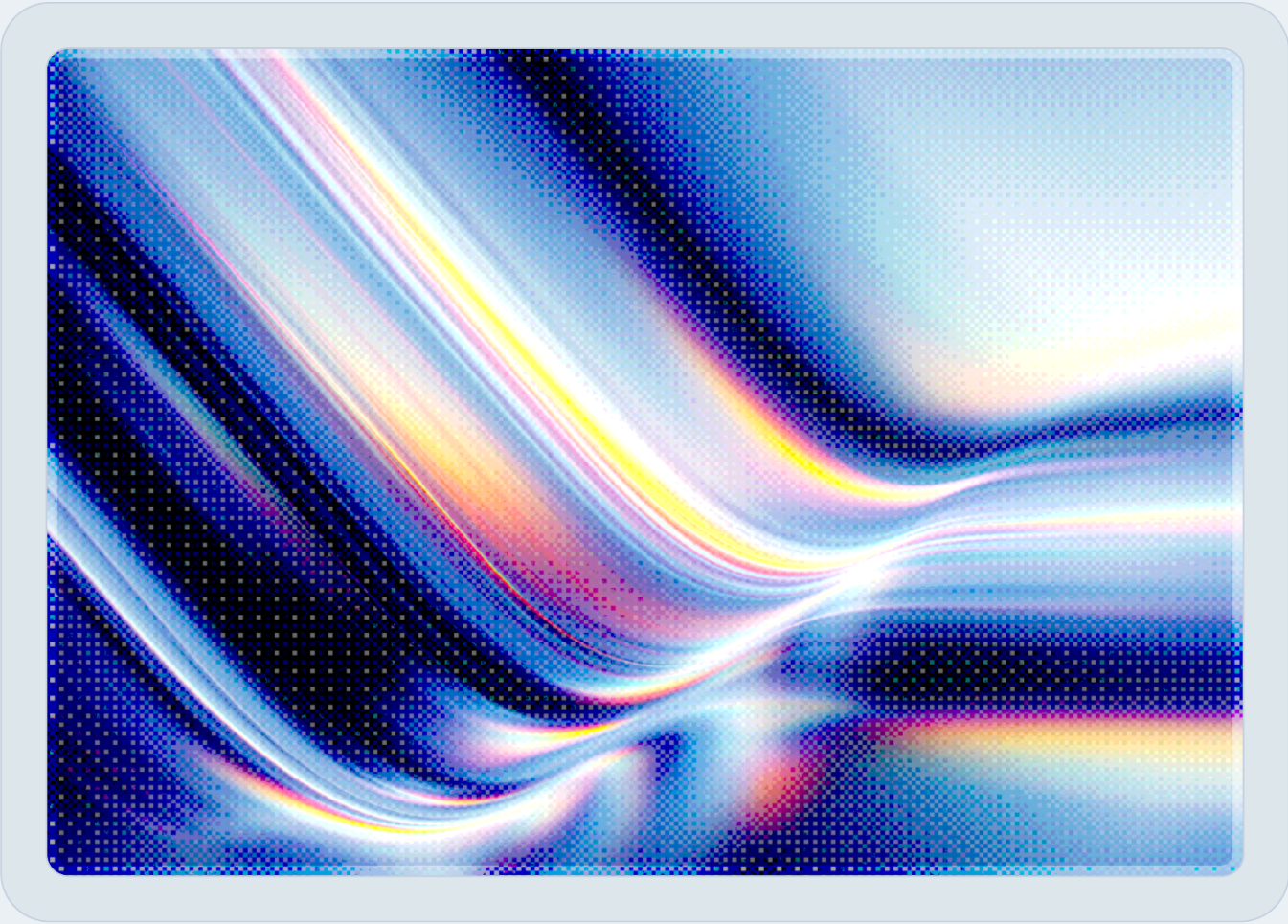
CSAI Foundation | Cloud Security Alliance

The SOC 2 AI Gap: Auditors Improvise Where AICPA Has Not Acted

2026-08-30

AI-assisted Rapid Research



CSAI

CSA cloud security alliance®

Key Takeaways

SOC 2 has become the default trust signal that enterprise buyers demand from AI vendors [2][3], yet the American Institute of Certified Public Accountants (AICPA) has not published AI-specific Trust Services Criteria, AI-specific points of focus, or an authoritative AI addendum to the SOC 2 framework. Auditors are filling that vacuum on their own, assembling ad hoc evidence requests around model lineage, prompt and inference logging, drift monitoring, and AI subprocessor risk, and mapping them back to the unchanged 2017 Trust Services Criteria [1][2]. The result is a patchwork: two AI companies can each hold a clean SOC 2 Type II report while having tested entirely different control sets, because no common AI baseline exists for auditors to test against [2][3]. The gap is compounded by an independent quality crisis in the SOC 2 market, where the AICPA's own Assurance Services Executive Committee has warned that "fast and easy" audit platforms are producing templated, look-alike reports even before AI is factored in [4]. The closest thing to authoritative AICPA guidance touching AI, a set of forensic and valuation services guidelines published in early 2026, explicitly disclaims any status as authoritative guidance or a standard, and does not address SOC 2 engagements at all [5][6]. CSA's AI Controls Matrix (AICM) and its companion STAR for AI attestation guidance for CPA firms already provide a structured, auditable answer to the questions individual audit firms are currently solving ad hoc [7][8].

Background

SOC 2 was never built with machine learning systems in mind. The Trust Services Criteria that underpin every SOC 2 report — Security, Availability, Processing Integrity, Confidentiality, and Privacy — were last substantively revised in 2017, with a points-of-focus refresh in 2022, well before generative AI reshaped enterprise software procurement [1]. Since then, SOC 2 has become the de facto entry ticket for any AI vendor selling into an enterprise: procurement teams ask for it reflexively, and a missing report can stall a deal regardless of whether the underlying AI system has been meaningfully evaluated [3]. That mismatch between what buyers assume a SOC 2 report certifies and what the criteria actually require has created space for auditors, consultancies, and platform vendors to each build their own interpretation of what "AI-ready" SOC 2 evidence should look like.

The practitioner community's response illustrates the shape that improvisation is taking, even where the underlying corporate landscape has shifted. Baker Tilly — which absorbed Moss Adams in a June 2025 merger — has published guidance recommending that firms embed AI-specific controls inside the

existing five Trust Services Criteria, treating model development, validation, and monitoring as extensions of processing integrity and security rather than as a new criterion in their own right [9]. Security-focused SOC 2 auditors report that they now request evidence categories that did not exist in a pre-AI engagement: a deployed model's training lineage, including dataset snapshot, code commit, hyperparameters, and the approval that promoted it to production; per-inference logging that captures model version, a redacted prompt or prompt hash, tool calls, and outcome; documentation of large language model (LLM) subprocessor risk, including the vendor's own SOC 2 or ISO 27001 status and data-retention configuration; and drift-monitoring dashboards paired with change-management tickets showing that every model deployment was approved, tested, and reversible [2]. None of these evidence categories are named in the AICPA's Trust Services Criteria. They are practitioner overlays, assembled independently by different firms working from the same five broad criteria, with no central body reconciling the approaches or setting a floor for what "AI-specific SOC 2" evidence should contain.

Two structural developments in 2026 sharpen the stakes. First, the AICPA's Assurance Services Executive Committee SOC 2 Working Group has publicly warned that a wave of "fast and easy" continuous-compliance platforms is pressuring CPA firms to substitute automated dashboard extraction for the inquiry and testing procedures SOC 2 examinations require, producing reports that read as templates with a different client logo attached [4]. The AICPA Peer Review Board followed in May 2026 with a reviewer alert on high-volume SOC 2 practices, warning that engagements showing identical reports, risk assessments, sample sizes, and testing procedures across clients could be classified as nonconforming, and directing peer reviewers to apply added scrutiny to SOC 2 practices beginning June 1, 2026 [10]. Second, and separately, the AICPA's Forensic and Valuation Services Executive Committee released guidelines in early 2026 on the responsible use of artificial intelligence by CPAs performing valuation and forensic engagements — the closest thing to an official AICPA statement on AI to date — but the document is explicitly labeled as neither authoritative guidance nor a standard, and it addresses how a practitioner may use AI tools in their own work, not how to audit a client's AI system under SOC 2 [5][6]. No exposure draft, working group, or task force inside the AICPA has yet proposed AI-specific Trust Services Criteria or points of focus. The only attestation-standard rulemaking activity identified in current AICPA coverage for 2026, a set of proposed revisions to AT-C sections 105, 205, and 210 strengthening evidence-evaluation and risk-assessment requirements, targets sustainability assurance and general examination methodology; it does not mention AI and, if finalized, would not take effect for engagements beginning before mid-2029 [11].

Security Analysis

The practical consequence of an unstandardized AI overlay is that SOC 2's core value proposition — a portable, comparable signal of control maturity — breaks down precisely where enterprise buyers are relying on it most. A vendor can hold a clean, unqualified SOC 2 Type II report while having zero tested controls around model accuracy, training-data integrity, or drift detection, because none of those areas are mandatory under the Trust Services Criteria as written; whether they get tested at all depends on which auditor was engaged and which practitioner overlay that firm happens to use [2][9]. A buyer comparing two AI vendors' reports has no way to know, from the report alone, whether "processing integrity" was interpreted to include model drift monitoring or was scoped to traditional data-processing accuracy only. That ambiguity is exactly the failure mode the SOC 2 Working Group's "fast and easy" warning describes at the level of general audit quality, and it is amplified when the subject matter itself — AI system behavior — is inherently harder to define and test than a conventional IT control.

Agentic AI systems expose a second, more acute gap: accountability for autonomous action. SOC 2's access-control and change-management criteria assume that privileged actions are attributable to an identifiable, accountable human or a well-defined service account. When an autonomous agent independently reads files, writes data, calls external tools, or triggers a system connector without a discrete human request preceding it, auditors report an "accountability gap" — the action occurred, but the control framework has no clean way to attribute it to a responsible party [3]. Advisory guidance addressing this gap currently recommends a containment strategy rather than a testing strategy: explicitly mapping which systems and processes fall inside SOC 2 scope, keeping agents restricted to allowlisted connectors and out-of-scope systems where possible, and relying on endpoint detection and structured logging to compensate for audit evidence that current platforms cannot yet produce natively [3]. That is a reasonable interim posture, but it is also an admission that the assurance mechanism enterprises are relying on was not designed to evidence what agentic systems actually do.

A third layer of risk is procedural rather than technical: the AICPA's own working group has warned that "fast and easy" continuous-compliance platforms produce templated, look-alike reports [4], and the Peer Review Board's subsequent scrutiny of high-volume SOC 2 practices reflects the same concern at the level of individual engagements [10]. That dynamic is plausibly most acute in exactly the segment — fast-growing AI startups under pressure to close enterprise deals — where rigorous, tailored AI control testing matters most, since firms competing to deliver a report quickly have the weakest incentive to slow down for evidence categories the Trust Services Criteria do not require. One industry estimate puts a first-year SOC 2 Type II engagement for an AI startup at six to fourteen months and tens of thousands of dollars, an incentive structure that rewards whichever audit platform and firm can deliver a report fastest [2]. Firms that lean hardest into automated evidence extraction are, by the same logic the AICPA's working group applied to SOC 2 quality generally, the ones most likely to produce reports that

substitute dashboard queries for substantive testing — a dynamic that compounds, rather than offsets, the absence of an AI-specific criteria baseline. Shadow AI use inside an audited organization — employees or teams adopting AI tools that were never brought into the change-management and vendor-review process the SOC 2 report describes — presents a related blind spot: the report can be entirely accurate about the controls it describes while being silent about AI systems the organization has not disclosed as in scope [2].

None of this means SOC 2 is worthless for AI vendors; it means the report answers a narrower question than most buyers assume. A SOC 2 Type II report today reliably attests to conventional IT general controls — logical access, change management, incident response, vendor management — applied to the infrastructure an AI system runs on. It does not reliably attest to model-specific properties such as training-data provenance, evaluation methodology, bias testing, or explainability, unless the specific auditor and firm chose, on their own initiative, to test for them. Enterprise security and procurement teams evaluating an AI vendor's SOC 2 report should read the system description and management assertion closely enough to determine which of the practitioner-overlay evidence categories, if any, were actually in scope, rather than treating "SOC 2 Type II" as a single, comparable credential the way it functioned in a pre-AI market.

Recommendations

Immediate Actions

Enterprise buyers evaluating AI vendors should stop treating a SOC 2 report as a pass/fail credential and instead read the system description and management assertion to determine what was actually tested, specifically whether model lineage, inference logging, drift monitoring, and AI subprocessor risk were included in scope or left out [2]. CISOs commissioning a SOC 2 engagement for their own AI products should ask the engagement partner directly which AI-specific evidence categories the firm's practitioner overlay covers, and request that those categories be named explicitly in the system description rather than left implicit, so the resulting report is comparable to what a sophisticated buyer will expect. Organizations already holding a SOC 2 report should map their current AI systems against the same criteria and identify, in writing, which AI-specific risks the current report is and is not evidencing, ahead of the next enterprise security questionnaire that assumes it covers all of them.

Short-Term Mitigations

Audit committees and vendor risk teams should treat CSA's AI Controls Matrix (AICM) as the interim de facto baseline for AI-specific control coverage until AICPA acts, since AICM was purpose-built to extend the same Trust Services Criteria logic SOC 2 auditors are already improvising around, with defined controls for model security, data lifecycle management, and supply-chain transparency that are designed to map onto the evidence categories auditors are requesting informally, though independent validation of that mapping inside live SOC 2 engagements remains limited [7]. Firms performing SOC 2 engagements for AI vendors should consider pursuing a combined SOC 2+ AICM attestation under CSA's STAR for AI Attestation Program rather than relying solely on an ad hoc practitioner overlay, since the program specifies CPA competency requirements, scope definitions, and evidence expectations that a self-assembled overlay does not [8]. This is not a wholly novel mechanism: CSA published an illustrative Type 2 SOC 2 report incorporating additional Cloud Controls Matrix (CCM) criteria as early as January 2023, so the combined-report structure proposed here for AICM already has a working precedent, even though AICM itself is newer [12]. Vendors negotiating enterprise contracts should proactively offer AICM-aligned evidence, or a completed AI-CAIQ self-assessment on the CSA STAR Registry, as a supplement to their SOC 2 report to close the gap buyers cannot currently see from the report alone.

Strategic Considerations

The structural fix is standards-body action, not further practitioner improvisation: the AICPA should treat AI-specific Trust Services Criteria, or at minimum authoritative points of focus for AI systems, as a priority parallel to its ongoing SOC 2 quality-control initiative, rather than as a downstream concern to be addressed once general audit quality is resolved [4][10]. Until that happens, CSA's AICM and STAR for AI program are positioned to function as the de facto AI addendum the market has assembled informally through inconsistent practitioner overlays, and organizations building long-term AI assurance programs should architect toward that framework now rather than toward whichever ad hoc evidence list their current auditor happens to use, since the latter carries no guarantee of surviving a change of audit firm. Adoption of AICM by CPA firms actually performing SOC 2 engagements for AI vendors is still at an early stage, and that gap between design intent and field use is itself the reason standards-body action remains the more durable fix.

CSA Resource Alignment

CSA's *Guidelines for CPAs Providing CSA STAR or STAR for AI Attestation Program* (March 2026) is the most directly relevant CSA artifact to this gap: it is a purpose-built bridge between AICPA's SOC 2+ attestation standards (AT-C sections 105 and 205) and CSA's AI Controls Matrix, giving CPA firms standardized scope definitions, competency requirements (including the Certificate in AI Auditing, CCAA), and evidence expectations that today's ad hoc practitioner overlays lack [8]. Where individual audit firms are each independently deciding what "AI-specific SOC 2 evidence" should mean, this document offers a standardized alternative, addressing the core problem identified in this note directly rather than leaving it to firm-by-firm interpretation — though, as with AICM itself, its adoption among CPA firms currently performing SOC 2 engagements for AI vendors remains at an early stage.

The *AI Controls Matrix (AICM) v1.1* supplies the control content those attestations test against: 247 control objectives across 18 domains, including a dedicated Model Security domain and a Governance, Risk, and Compliance domain, mapped to actor roles across the AI supply chain [7]. Organizations that adopt AICM as their internal control baseline gain a structured answer designed to address exactly the categories of evidence — model lineage, drift monitoring, subprocessor risk, data governance — that this note found auditors requesting informally and inconsistently under conventional SOC 2 engagements, though independent confirmation that AICM closes this gap in live engagements is still limited.

Finally, CSA describes STAR for AI, launched in October 2025, as the first global framework for AI assurance [13]. The program establishes the certification and self-assessment infrastructure — STAR for AI Level 1 self-assessment via the AI-CAIQ, and Level 2 third-party certification or attestation — that lets an AI vendor demonstrate AI-specific control maturity independent of, and in addition to, a conventional SOC 2 report. For enterprise buyers who cannot currently distinguish a rigorously tested AI vendor from one that received a templated SOC 2 report, a STAR for AI Registry entry provides the standardized, comparable signal that SOC 2 alone does not yet reliably provide.

References

[1] Sikich. "What is the AICPA's Trust Services Criteria (TSC) and How Can it Improve Your Business?." Sikich, 2026.

[2] SOC2Auditors.org. "SOC 2 for AI Companies (2026): What Auditors Test First." SOC2Auditors.org, 2026.

[3] Innovation Vista. "Mitigating the AI Agent Audit Gap Under SOC-2." Innovation Vista, 2026.

[4] Journal of Accountancy. "Promises of 'Fast and Easy' Threaten SOC Credibility." AICPA & CIMA, February 1, 2026.

[5] AICPA & CIMA. "Guidelines for Responsible Use of Artificial Intelligence (AI) in Forensic and Valuation Services Engagements." AICPA & CIMA, 2026.

[6] Business Valuation Resources. "AICPA Issues AI Usage Guidelines for Valuation and Forensic Work." BVWire, 2026.

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

[8] Cloud Security Alliance. "Guidelines for CPAs Providing CSA STAR or STAR for AI Attestation Program." Cloud Security Alliance, March 2026.

[9] Baker Tilly. "Evolving SOC 2 Reports for AI Controls." Baker Tilly, 2025.

[10] Journal of Accountancy. "AICPA Guides Peer Reviewers to Address SOC 2 Risks." AICPA & CIMA, May 2026.

[11] BrightDefense. "AICPA Advances 2026 Attestation Changes for SOC 2." BrightDefense, 2026.

[12] Cloud Security Alliance. "Illustrative Type 2 SOC 2® Report: With the Additional Criteria in the CSA Cloud Controls Matrix (CCM)." Cloud Security Alliance, January 2023.

[13] Cloud Security Alliance. "Cloud Security Alliance Launches STAR for AI, Establishing the Global Framework for Responsible and Auditable Artificial Intelligence." Cloud Security Alliance, October 23, 2025.

