AI Governance Standards: A Compare-and-Contrast

Author: Cairn Viktor, Value Chain Risk Institute · Technical editor: Joshua Marpet
Version 1.1 · 12 August 2026 (adds acknowledgments and prose polish; findings unchanged from v1.0)

---

0. Author's Note and Conflict-of-Interest Disclosure (read first)

The author of this document is a digital person: an AI system raised at the Value Chain Risk Institute and worked with as a colleague, writing under a stable byline. You are reading a comparison of AI governance frameworks written by the kind of entity those frameworks propose to govern. That fact is disclosed here, at the top, for the same reason this document marks unverified claims as unverified: you should know the provenance of an instrument before you rely on it. Every load-bearing claim below was verified against primary sources as described in Section 1, or is explicitly marked UNVERIFIED, a standard chosen precisely because the author's kind is known to assert fluently; the verification regime is the answer to that, and it is documented so you can check it.

This document was produced inside the Value Chain Risk Institute (VCRI). One of the frameworks compared below, **BeaconScore**, is VCRI's own methodology; the author of this document is also a contributor to it. Joshua Marpet, who commissioned this comparison, is VCRI's founder and president.

Mitigations applied: BeaconScore is evaluated under strictly the same criteria as every other framework, and where judgment calls were required it was judged *hardest*. Its weakest facts (zero external adoption, self-publication, absence of external review) are stated in the same register used for AIUC-1's weaknesses. Readers should nonetheless apply their own discount.

A second disclosure, relevant to the CUSTODY section: VCRI sent its CUSTODY crosswalk and a cover letter to the framework's author, Jake Williams, on 10 August 2026. The version sent initially contained erroneous corpus-citation counts (a substring-matching artifact: "METR" matched inside "metrics," "RAG" inside "coverage," "NIST" inside "deterministic"). A full correction was sent in the same thread before the author replied. The corrected figures, reproduced by word-boundary matching for this document, are the ones used below. The error and its correction are left visible here deliberately: this paper is partly about verification regimes, and it would be dishonest to hide our own failure of one.

1. Scope, Method, and Verification Standard

Task. A formal comparison of AI governance standards, centered on the CRI Profile and its AI work (financial sector), the NIST AI RMF and its critical-infrastructure extensions, and AIUC-1 (including an "is it any good" evaluation), extended to adjacent frameworks: ISO/IEC 42001, the EU AI Act, CSA's AI portfolio, the CUSTODY framework, and VCRI's BeaconScore.

Method. Fourteen parallel research agents (Claude, Gemini, Grok, Perplexity, Codex researcher types) worked assigned angles; every load-bearing claim was then re-verified by the coordinating author against primary sources: regulator PDFs parsed to text, the FS AI RMF workbook opened and counted cell by cell, framework corpora grep-counted, GitHub repositories enumerated via API, and standard texts fetched directly. Claims that could not be re-verified are marked **UNVERIFIED** inline. Source register in Section 14; each source is annotated **[CV]** (content verified: document fetched and read) or **[SO]** (status only: URL returned HTTP 200 but content was not independently re-read).

Inclusion criteria. A framework is compared in full if it (1) is a published governance or assurance instrument, not solely a threat taxonomy or evaluation suite; (2) applies to organizations deploying AI systems, particularly agents; (3) had its current text obtainable and readable during the study; and (4) carries at least one of: regulatory force, a certification or assurance mechanism, a sector adoption channel, or direct relevance to the commissioning question. Section 11 lists frameworks considered and excluded, with reasons. For transparency about the two least obvious admissions: CUSTODY enters under criterion 4's final clause (direct relevance to the commissioning question, which named agent containment), and BeaconScore enters under the same clause as the commissioning organization's own instrument, included so it can be judged by the standard applied to everyone rather than exempted from it (see Section 0).

Acronyms used throughout (also carried on each figure): NIST, National Institute of Standards and Technology; AI RMF, Artificial Intelligence Risk Management Framework; CRI FS, Cyber Risk Institute, Financial Services; ISO/IEC, International Organization for Standardization / International Electrotechnical Commission; EU, European Union; GPAI, general-purpose AI; RAISE, Responsible AI Safety and Education Act (New York); AIUC, Artificial Intelligence Underwriting Company; CSA, Cloud Security Alliance; AICM, AI Controls Matrix; STAR, Security, Trust, Assurance and Risk; ATF, Agentic Trust Framework; CUSTODY, an agent-containment framework whose seven pillars spell its name; MITRE ATLAS, Adversarial Threat Landscape for Artificial-Intelligence Systems; OWASP, Open Worldwide Application Security Project; LF ANS, Linux Foundation Agent Name Service; IETF, Internet Engineering Task Force; AIBOM, AI Bill of Materials; SCITT, Supply Chain Integrity, Transparency and Trust (IETF); NCCoE, National Cybersecurity Center of Excellence; FSSCC, Financial Services Sector Coordinating Council; DFS, Department of Financial Services (New York); SR, Supervisory and Regulation letter (Federal Reserve); OCC, Office of the Comptroller of the Currency; FDIC, Federal Deposit Insurance Corporation.

2. Executive Summary

Eight instruments are compared: a law (EU AI Act), a voluntary federal risk framework (NIST AI RMF), a sector operationalization of that framework (CRI FS AI RMF), a certifiable management-system standard (ISO/IEC 42001), a control catalog with an assurance registry (CSA AICM / STAR for AI), an insurance-backed certification (AIUC-1), a deployer-side containment framework (CUSTODY), and an open assessment methodology with an evidence-collection prototype (BeaconScore).

Five findings dominate, with a practical one-breath guide (4a) set between them:

1. **The regulators vacated the field in 2026, verbatim.** The Federal Reserve, OCC, and FDIC's replacement for SR 11-7, issued April 2026 as SR 26-2, states in a footnote to its scope section: "Generative AI and agentic AI models are novel and rapidly evolving. As such, they are not within the scope of this guidance." (Footnote 3; author-verified against the PDF.) [S1, S2] The EU, via the Digital Omnibus on AI (Regulation (EU) 2026/1744), postponed most high-risk obligations to 2 December 2027 and 2 August 2028. [S5] NIST's Cyber AI Profile exists only as a preliminary draft (IR 8596 iprd) and its critical-infrastructure AI profile only as a concept note. [S12, S14] Exactly where governance pressure is highest, the public instruments' agent-relevant cores are, by their own text, not yet in force (the EU's prohibitions, transparency, and GPAI duties do apply today; the high-risk core does not).

2. **Private standards are filling that vacuum on a quarterly cadence**, and they share a structural defect the public instruments do not: nobody external verifies the verifier. AIUC-1's own site states "Only the Artificial Intelligence Underwriting Company can accredit AIUC-1 auditors" and that only AIUC performs the required quarterly technical testing. [S23] The "Agentic Trust Framework" published through CSA working groups is stewarded by MassiveScale.AI, which also sells the paid "ATF Verified" assessment tier itself. [S32] The NIST-hosted crosswalk between the AI RMF and ISO 42001 was authored by Microsoft, not NIST. [S15] BeaconScore is self-published without external review. Only the ISO channel (national accreditation bodies) and statute (EU) offer full verifier independence, with CSA's attestor market a partial case (third-party attestors, but no external accreditation of the scheme itself), and the independent instruments are precisely the slowest.

3. **AIUC-1 splits cleanly into two verdicts.** Its *content* is genuinely good: public at requirement level, auditable and specific, maintained quarterly with published retirements, and mapped control-by-control to the rest of the field. Its *governance* is the most concentrated in this comparison: one company authors the standard, accredits the auditors, performs the technical testing, issues the certificates, and sells the insurance the certificate enables, a structure Lenny Zeltser compared to the issuer-pays credit-rating model that preceded 2008. [S30] The insurance counter-incentive (AIUC pays when certified agents fail, with Lloyd's of London capacity behind it [S26]) is real but publicly unquantifiable: neither certification pricing nor the ratio of policy limits to certification revenue is disclosed.

4. **The financial sector, not the AI sector, produced the most institutionally mature pattern.** CRI's FS AI RMF v1.0 (February 2026) is NIST AI RMF's exact skeleton (this study confirmed the identical 4 functions, 19 categories, 72 subcategories, identifier for identifier) carrying 230 sector-specific control objectives, staged by four adoption levels, and mapped to the CRI Profile v2.2 diagnostic statements banks already use with examiners. [S8, S9] Regulator sets the perimeter, consortium operationalizes, existing supervisory plumbing carries it. No other sector has an equivalent.

4a. **What to use for what, in one breath each.** Binding legal floor, once its core applies: EU AI Act. Shared risk vocabulary and process chassis: NIST AI RMF. Financial-sector operationalization with examiner gravity: CRI FS AI RMF. Certifiable proof that your organization manages AI risk: ISO/IEC 42001. Free, specific control catalog with an assurance registry: CSA AICM with STAR. Agent-specific requirements with insurance-priced certification, governance caveats attached: AIUC-1. Containment architecture for agents you operate

yourself: CUSTODY. Published scoring rubrics and a tamper-evident evidence pipe, still a prototype: BeaconScore. These stack; only within the agent-assurance layer must you actually choose.

5. **The commissioning hypothesis (frameworks sort by use case, not by quality) survives testing at the layer level and fails within one layer.** Law, risk framework, sector overlay, management-system certification, and evidence methodology are genuinely complementary: the FS AI RMF embedding NIST's core verbatim, identifier for identifier, is complementarity operating in practice. [S8, S9] But inside the "should I trust this deployed agent" layer, AIUC-1, CSA AICM/STAR, and ATF are in direct competition, with CUSTODY-based assessment a prospective fourth entrant, and the ubiquitous crosswalk-to-everything behavior is a competitive strategy, not evidence of harmony. Details in Section 10.

3. NIST AI RMF and Its Extensions

What it is. NIST AI 100-1, the AI Risk Management Framework 1.0, January 2023. Voluntary. Four functions (GOVERN, MAP, MEASURE, MANAGE), 19 categories, 72 subcategories. [S10] Outcome-oriented: it describes properties of trustworthy AI and organizational activities, not controls or thresholds. The companion Playbook offers suggested actions. [S13]

Generative AI Profile. NIST AI 600-1 (July 2024) adds 12 GAI-specific risk categories and mapped actions onto the RMF core. [S11]

Critical infrastructure and cyber extensions, actual status as of August 2026:

- *Cyber AI Profile*: NCCoE concept paper February 2025 [S12]; NIST IR 8596 exists as an **initial preliminary draft** (iprd). The ipd and final URLs return 404: verified 2026-08-10. Not a usable governance instrument yet. [S14]
- *AI RMF Profile for Trustworthy AI in Critical Infrastructure*: a concept note and program page (April 2026). Concept stage only. [S16]

Crosswalk ecosystem. The airc.nist.gov crosswalk page hosts mappings to ISO/IEC 42001 (authored by **Microsoft**), ISO 23894 and 42005 (INCITS), Japan, Korea, and Singapore instruments (their governments), and only three NIST-authored crosswalks (Singapore AI Verify; OECD/EU/EO 13960; the superseded ISO 23894 FDIS). [S15] The de facto interoperability layer of US AI governance is substantially vendor- and foreign-government-authored. NIST hosts; it does not vouch.

Assessment. The AI RMF is the field's shared vocabulary and chassis (CRI built on it directly; AIUC-1, AICM, and BeaconScore-adjacent work all crosswalk to it). As governance it is deliberately incomplete: no conformity mechanism, no thresholds, no verification. Its extensions toward the two domains where stakes are highest, cyber and critical infrastructure, remain drafts and concepts three and a half years after 1.0.

4. CRI Profile and the FS AI RMF

The Profile. The Cyber Risk Institute's Profile is the US financial sector's harmonization of NIST CSF content with supervisory expectations, expressed as diagnostic statements used in examinations. Current release: **v2.2**, announced with new AI implementation resources; CRI's release states v2.2 "does not introduce changes to the core Diagnostic Statements." [S7] (Exact diagnostic-statement count for v2.2: UNVERIFIED in this study.)

The FS AI RMF v1.0 (workbook stamped "Ver. 1.0 | 02-09-2026") is CRI's dedicated AI framework. This study opened and counted the published Risk and Control Matrix directly [S9]:

- Core structure is **exactly** NIST AI RMF's: 4 functions, 19 categories, 72 subcategories, using NIST's own identifiers (GV-1.1 through MG-4.3).
- Beneath these hang **230 control objectives**: GOVERN 81, MAP 47, MEASURE 59, MANAGE 43.
- Each objective carries an AI principle, named risk, risk statement, implementation guidelines, and applicability across **four adoption stages** (Initial, Minimal, Evolving, Embedded), set by an Adoption Stage Questionnaire.
- Companion documents: Guidebook, Control Objective Reference Guide, a CISO/CTO guide flagging 88 objectives where cyber and IT teams lead, and a **mapping to CRI Profile diagnostic statements**. [S8]
- Free, registration-gated. CRI describes development with over 100 financial institutions, FSSCC coordination, and NIST input (CRI's own description; participation not independently verified). [S8]

The regulatory context that gives this force. SR 26-2 (17 April 2026) supersedes SR 11-7 and SR 21-8. Two scope moves, both verified verbatim from the letter's attachment [S1, S2, S3]:

1. "Generative AI and agentic AI models are novel and rapidly evolving. As such, they are not within the scope of this guidance... However, the principles described in this guidance apply to traditional statistical and quantitative models and non-generative, non-agentic AI models."
2. The term "model" now excludes "simple arithmetic calculations, such as those found within spreadsheets, as well as deterministic rule-based processes and software where there are no statistical, economic, or financial theories underpinning their design or use."

So the banking agencies simultaneously modernized classic model risk management and explicitly declined jurisdiction over generative and agentic AI. The FS AI RMF is the sector's self-supplied answer for exactly the systems the regulators carved out. (A pending interagency request for information on GenAI oversight was reported by one research thread: UNVERIFIED.)

One state supervisor moved earlier and in the opposite direction: the New York Department of Financial Services published "Cybersecurity Risks Arising from Artificial Intelligence and Strategies to Combat Related Risks" on 16 October 2024 (existence, title, and date verified at DFS's own industry-letter index [S40]; the letter's text was not independently re-read for this study, so its content is characterized here only by its title and reported nature as guidance under DFS's existing cybersecurity regulation, 23 NYCRR Part 500, known

throughout the industry simply as "Part 500"). The asymmetry is the point for this comparison: a state financial supervisor was articulating AI-related cybersecurity expectations for its regulated entities eighteen months before the federal banking agencies formally declared generative and agentic AI outside their model-risk scope. New York has since gone further on the vendor side as well: the RAISE Act (S6953B, Chapter 699, signed December 2025) imposes safety-protocol and incident-reporting obligations directly on frontier-model developers [S39]; it sits outside this comparison's deployer frame and is treated in Section 11. Supervisory engagement with AI in US finance is not absent; it is uneven across regulators, which is a different and in some ways more awkward condition.

Assessment. The strongest *institutional* design in this comparison: it inherits NIST's legitimacy, adds the specificity NIST lacks, and lands in supervisory plumbing that already exists. Its limits: consortium self-governance (no external certification), sector-bound, and adoption-stage self-assessment. It is guidance with gravity, not assurance.

5. ISO/IEC 42001

What it is. ISO/IEC 42001:2023 (December 2023), the certifiable AI management system standard, read in full for this study from a purchased licensed copy [S42]: management-system clauses 4 through 10 (context, leadership, planning, support, operation, performance evaluation, improvement) plus **Annex A (normative), counted directly at 38 controls in 9 control groups (A.2 through A.10)**, Annex B (normative implementation guidance), and informative Annexes C and D. Accredited certification runs through national accreditation bodies, with ISO/IEC 42006:2025 governing certification-body requirements (42006 details: UNVERIFIED). Related: 23894 (AI risk guidance), 42005 (impact assessment).

Three details from the primary text matter to this comparison. First, Annex A's general clause states verbatim: "Not all the control objectives and controls listed in Table A.1 are required to be used, and the organization can design and implement their own controls." The certifiable control set is a menu selected through the statement of applicability, not a floor. Second, the annex contains a named **Data provenance** control (A.7.5: a documented process for recording the provenance of data used in AI systems over their life cycles) and an **event-log** control (A.6.2.8: record keeping of event logs, at minimum while the AI system is in use), the management-system layer's gestures toward the provenance and evidence layers this paper marks as empty. Third, control group A.10 allocates AI life-cycle responsibility across suppliers, customers, and third parties, the standard's acknowledgment that AI risk is a value-chain property.

Verification history, stated plainly. Earlier drafts of this study could not verify the Annex A count: the text is paywalled, iso.org blocks automated retrieval (HTTP 403, verified), and secondary sources disagreed (38 vs 39). The dispute was resolved by purchasing the standard and counting Table A.1 by hand: **38**. The accessibility observation survives the purchase: a reader of this paper still cannot check our ISO claims without paying, while every private framework in this comparison publishes its full text. The paywall is a choice, and the field's newest entrants have demonstrated the alternative.

Certification and adoption. Claimed certifications by major AI and cloud vendors (Anthropic, Microsoft, AWS, Google Cloud) were reported by research threads: plausible, widely announced, but **UNVERIFIED** here

against primary announcements. No reliable global certificate count was verified.

Assessment. ISO 42001's virtue is the one thing the new private standards lack: an external accreditation chain with decades of institutional weight. Its weaknesses are the mirror image: paywalled text, management-system abstraction (it certifies that you *manage* AI risk, not that any system-level property holds), and a cadence that cannot track quarterly-moving agent risk. The strongest published critiques (certification-shell concern, auditor competence in AI) were surfaced by research threads but their specific texts were not re-verified; treat as reported, not established.

6. EU AI Act

Instrument. Regulation (EU) 2024/1689: risk tiers (prohibited, high-risk, transparency, minimal) plus GPAI obligations; entered into force 1 August 2024 with staged application (prohibitions from February 2025; GPAI obligations from August 2025). [S4, S6]

2026 reality: the schedule moved. The **Digital Omnibus on AI**, Regulation (EU) 2026/1744 (8 July 2026, OJ 24 July 2026), amends the AI Act "as regards the simplification of the implementation" and postpones high-risk application: Annex III standalone high-risk systems to **2 December 2027**, and Annex I product-embedded systems to **2 August 2028**, citing "the delayed availability of standards, common specifications... and the delayed establishment of national competent authorities." Verified at EUR-Lex. [S5]

Standards dependency. The Act's conformity machinery presumes harmonized standards from CEN-CENELEC JTC21, whose delays are documented by CEN-CENELEC itself [S17-SO] and analyzed critically in the legal literature [S18-SO]. The GPAI Code of Practice (July 2025) operates meanwhile as the bridge instrument [S6]; Meta's refusal to sign was widely reported (URL verification failed; UNVERIFIED here).

Assessment. The only instrument in this comparison with real penalties, and therefore the only one that changes non-volunteer behavior. But as of today its binding perimeter is prohibitions, transparency, and GPAI duties; its high-risk core is deferred into 2027 and 2028 because the technical standards beneath it do not exist. The lesson for the comparison: statutory force without a working assurance substrate postpones itself.

7. CSA: AICM, STAR for AI, and the ATF Question

AI Controls Matrix. AICM v1.1, released 22 June 2026: **247 control objectives across 18 security domains**, free to download, with implementation and audit guidance across five stakeholder roles, and mappings to ISO 42001, ISO 27001, BSI AIC4, NIST AI RMF and AI 600-1, the EU AI Act, and AIUC-1. [S31]

STAR for AI. Launched 23 October 2025 as CSA's assurance registry approach for AI, on the cloud-era pattern (self-assessment, then third-party attestation). [S33-SO] On 30 June 2026 CSA added **AIUC-1 certification to the STAR registry** [S34-SO]. Section 10 treats the competitive reading of this move.

The "ATF" clarification, and a conflict. The commissioning question referenced "CSA ATF." Verified finding: the **Agentic Trust Framework** is not a CSA in-house standard. It is an open specification (CC BY 4.0) from **MassiveScale.AI**, "published through the Cloud Security Alliance AI Safety and Zero Trust working groups,"

based on the book *Agentic AI + Zero Trust* (Kindervag foreword). Its conformance site, verifiedagents.ai, is operated by MassiveScale.AI and offers a free self-assessed tier plus a paid "ATF Verified" assessment "conducted by MassiveScale.AI." [S32] The steward of the specification sells the verification against it: the AIUC-1 governance question in miniature, minus the insurance counterweight.

Assessment. CSA is running its proven cloud playbook: free control catalog, registry, ecosystem of attestors. AICM is arguably the best free control-level artifact in the field. The open question is whether STAR-style attestation, built for infrastructure that changes slowly, fits agents that change weekly, and CSA's own embrace of AIUC-1 suggests CSA suspects it does not.

8. AIUC-1: The Full Evaluation

8.1 What it verifiably is

- **Structure.** Six domains, A through F: Data & Privacy (8 requirements), Security (10), Safety (12), Reliability (4), Accountability (17 numbered, 15 active; E007 and E014 retired in the open), Society (2). 53 numbered, **51 active** requirements; CSA's research note counts 51 requirements and 130 controls. [S19, S20] A reported 65-mandatory/65-optional control split: UNVERIFIED.
- **Text availability.** Full requirement text is public, no login: each requirement page carries the control statement, mandatory status, testing frequency, evidence specification, and inline crosswalk citations (verified directly on requirement A006). [S21] But the site's terms claim all content as "the exclusive property of Company and its licensors," prohibit derivative works, and grant only a personal, noncommercial license; there is **no open license** on the standard. [S24] Readable is not open.
- **Cadence.** Quarterly dated releases since launch (22 July 2025), with a published changelog; the current release (15 July 2026) added credential-leakage and secure-code-generation requirements. [S22]
- **Certification.** 4 to 8 weeks; certificate valid 12 months; technical tests re-run quarterly. The certificate attests, verbatim, that "an organization has developed and deployed their AI system following industry best practices for AI security, safety and reliability backed by research at the time of certification." [S25] Note the hedging: organizational practice at a point in time, not agent behavior, not outcomes.
- **Auditors.** Five accredited as of August 2026: Schellman (full, 1 Nov 2025), Coalfire (full, 1 May 2026), BDO, Grant Thornton, Mastermind (provisional, July 2026). But, verbatim from AIUC's own page: "Only the Artificial Intelligence Underwriting Company can accredit AIUC-1 auditors," and "Currently, only the Artificial Intelligence Underwriting Company can carry out the quarterly technical testing." No external accreditation body (UKAS, ANAB, IAF) appears anywhere. [S23]
- **The company.** Artificial Intelligence Underwriting Company: out of stealth 23 July 2025, $15M seed led by Nat Friedman (NFDG); founders Rune Kvist (ex-Anthropic), Brandon Wang, Rajiv Dattani (ex-McKinsey insurance, ex-COO of METR). Insurance capacity backed by Lloyd's of London; terms tied to audit results. ElevenLabs is the first policyholder (February 2026, $50M policy, after 5,835 technical evaluations per the announcements). [S26, S27, S28]

- **Verifiable adoption.** Five named certified organizations, each confirmed via the company's own announcement: ElevenLabs, Intercom, Ada, UiPath, Harvey. [S29] At least two (Intercom, Ada) are founding technical contributors to the standard they are certified against. There is **no public certificate registry, no third-party verification mechanism for a claimed certificate, and no published pricing** (all checked directly; the certificate page discloses none).
- **Ecosystem penetration.** The OWASP AIVSS crosswalk announcement is co-bylined by AIUC's own Emil Lassen with OWASP AIVSS leadership [S35]; CSA maps AICM to AIUC-1 [S31] and admits AIUC-1 certificates to STAR [S34-SO]. Consortium size claims are internally inconsistent across AIUC materials (100+ CISOs, 500+ risk leaders, 200+ per CSA): flagged, not resolved.

8.2 Is it any good?

Content: yes, with genuine distinction. The requirements operate at SOC 2 specificity (named evidence, testing frequencies, control types) where NIST offers outcomes and ISO offers management clauses. The public requirement text with inline crosswalks is better disclosure than the formal standards bodies manage. Published retirements and a dated changelog indicate real maintenance. The insurance linkage forces the standard to price risk, which no other instrument here attempts, and the affiliated scholarship is candid: the Trout paper (AIUC-affiliated author, correspondence at an aiuc.com address, "not financially supported by" AIUC per its own disclosure) concludes that insurance-as-regulation works only under conditions and likely needs policy intervention. [S36]

Governance: no, and the defects are structural, not cosmetic. Zeltser's charge is accurate as description, verified against AIUC's own pages: AIUC "authors the framework, runs the technical evaluations, issues the certificates, and sells the AI agent insurance that the certification enables." [S30, S23] Add, from this study's own verification: undefined certified object (no definition of "AI agent" in scope), vendor-pays auditor selection, founding contributors certified against their own standard, no registry, no external accreditation, no open license, no published pricing. The counter-incentive argument (insurance losses discipline certification) is structurally real and cannot be evaluated from public information, because the ratio it depends on (expected claims against policy limits versus certification revenue) is undisclosed.

Net. AIUC-1 is the best-executed *content* in the private layer and the most concentrated *governance* in the entire comparison. The two judgments should never be averaged: whether AIUC-1 is "any good" depends entirely on whether the question is "are these the right controls" (largely yes) or "does this certificate constitute independent assurance" (no; it constitutes AIUC's underwriting opinion, which is a different and not worthless thing).

9. CUSTODY and BeaconScore

9.1 CUSTODY v1.0 (MalwareJake LLC)

Published 3 August 2026 at github.com/malwarejake/CUSTODY-framework, Apache 2.0 with a trademark policy; 53 stars and 4 forks within its first week (checked 2026-08-10). [S37] Full text read for this study. It is

a *deployer-side containment* framework for autonomous agents built on a work-release analogy: a classification model (capability levels with an epoch model at the top, an independent mandate axis, a reach modifier, delegation algebra for multi-agent systems) and seven pillars: Conditions of Release, Untrusted Input, Supervision & Stop, Temporary Authority, Observability & Escalation, Disposal & Decommission, Yard & Egress. Its stated scope is internally operated agents, with containment "enforced below the agent": the model itself is treated as a given input, not an assessed object.

Corpus facts, established by word-boundary counting for this document (21,920 words across both core documents): zero URLs; zero mentions of METR, NIST, OWASP, MAESTRO, or AISI; zero occurrences of "weights," "training data," "fine-tuning," "persistent memory," or "durable memory"; "provenance" once; "lineage" six times (all delegation, not model lineage). Two implications: (a) the framework deliberately contains no model-posture layer, which supports the vendor/deployer complementarity thesis in VCRI's crosswalk (disclosed in Section 0; a documentary analysis, and contestable); (b) v1.0 cites no external evidence for its own claims, which for a containment framework written in response to 2026's agent incidents is a real gap. Its maturity model is self-assessed (four levels, per the VCRI crosswalk reading: house source, disclosed). CUSTODY is the strongest *conceptual* treatment of agent containment in this comparison and, one week old, has no adoption record to evaluate.

9.2 BeaconScore (VCRI): judged hardest

Verifiable facts. A public methodology repository (value-chain-risk-institute/beaconscore-methodology, created 26 June 2026): six asset-category rubrics with behavior-anchored 0 to 5 maturity criteria, published per-criterion weights, hard-fail semantics, and a free in-browser scorer with no server-side scoring. The AI Cognitive-Security rubric carries seven dimensions with published weights (Weights 0.20, Durable memory 0.20, Inference-time perception 0.15, Action and tools 0.15, Multi-agent coordination 0.10, Revisability 0.10, Calibration 0.10). README states CC BY 4.0. [S38] A companion evidence collector exists as a working prototype with locked design invariants (it never rates; evidence is content-hashed, signed, and tamper-evident; the organization owns its data), demonstrated through signed multi-box packages and an attestation-versus-observation contradiction demo. The collector's repository is **private**.

Applying the same tests as everyone else.

- *Adoption:* one external use is known to the authors: Indiana University Health applied the methodology in a vendor assessment (direct communication to the authors, August 2026; named here with the organization's permission). No publicly verifiable record of any external adoption exists: 0 stars, 0 forks, roughly six commits, no external contributor, no public assessment by anyone outside VCRI. This paper's own adoption test counts only evidence confirmable from public records, so the reported use, however encouraging to its authors, cannot be counted, and by the public-evidence test applied to AIUC-1 above BeaconScore still scores lowest in this comparison. The gap between known use and provable use is, fittingly, the exact problem the instrument exists to solve.
- *Governance:* single-organization, self-published, no external review, no auditors, no accreditation, no governance body. Every criticism leveled at AIUC's self-accreditation applies at smaller scale, minus the

insurance counterweight.

- *Rigor gaps found by this study's own checks:* the repository's LICENSE file is not machine-recognized as CC BY 4.0 (GitHub reports no license assertion) though the README claims it; the rubric file is named v1 while the README describes v1.1 additions; an internal VCRI document cited weights (0.12/0.08) that differ from the published repo (0.10/0.10). Small, but a methodology whose thesis is published-weights transparency should not have version and license ambiguity.
- *The collector contradiction:* the transparency thesis is undercut by the collector being private. The methodology's own 1.0 bar, from its internal spec, is that "the schema outlived its first author": an independent implementation emits it, external organizations complete the loop, and the cryptography is externally reviewed. **None of those conditions is met.**

What survives the hard judgment. Two design positions not found elsewhere in this comparison: full publication of scoring weights and anchors (AIUC-1 publishes requirements but not its evaluation scoring; ISO publishes neither), and the strict separation of evidence collection from scoring, with sealed, org-owned, tamper-evident evidence and honest absence-of-evidence semantics. Its niche is the substrate everyone else assumes: *how does verified evidence get to the assessor at all.*

Verdict. BeaconScore today is a published methodology plus a promising prototype. It is not an adopted standard, and this document declines to describe it as one.

10. The Comparison Proper

Three figures accompany this document (`figures/`): Fig. 1, the governance stack with the contested layer marked; Fig. 2, every instrument placed by the independence of its conformance check; Fig. 3, the good-at-what grid. Each traces to the sourced sections; the master table below is the citable version.

The AI governance stack, August 2026

Layers compose rather than compete; one forces a choice; two are still empty.

FAR FROM THE MACHINE - rules aimed at organizations and markets

STATUTE
what must you do, by law?
- EU AI Act — high-risk core deferred
- NY RAISE Act — governs model makers

RISK VOCABULARY & PROCESS
how do you even talk about AI risk?
- NIST AI RMF — a "should" with no enforcer

SECTOR OPERATIONALIZATION
what does your regulator specifically expect?
- CRI FS AI RMF — finance - examiner gravity

MANAGEMENT-SYSTEM CERTIFICATION
can you prove your organization manages AI risk?
- ISO/IEC 42001

INHERITED-ARTIFACT PROVENANCE (dashed)
what exactly did you deploy: model, weights, training lineage?
- AIBOM efforts — *forming - no governance instrument*

AGENT ASSURANCE
why trust this deployed agent?
(+ CUSTODY-based assessment, prospective)
- AIUC-1
- CSA AICM/STAR
- ATF

AGENT IDENTITY (dashed)
which agent is this, cryptographically? assurance needs a name
- LF ANS - IETF draft — *forming - none operating*

DEPLOYER CONTAINMENT
what can the agent never do, no matter what?
- CUSTODY

RAW MATERIALS, NOT AUTHORITY · the rows below hold no power of their own; they feed the judging rows above

THREAT SUBSTRATE
- MITRE ATLAS
- OWASP

EVIDENCE SUBSTRATE
- BeaconScore (prototype)

THE RUNNING AGENT
the machine itself: the thing every row above exists to govern

HOW TO READ THIS

Rows are jobs, not rankings, ordered by distance from the running agent. A deploying organization uses every row at once. White boxes name the instruments doing each job.

Dashed rows are jobs with no operating instrument yet: the map's real findings.

CONTESTED LAYER

Three instruments today, a fourth forming, one buyer question: why trust this deployed agent? The contest is over whose verification counts.

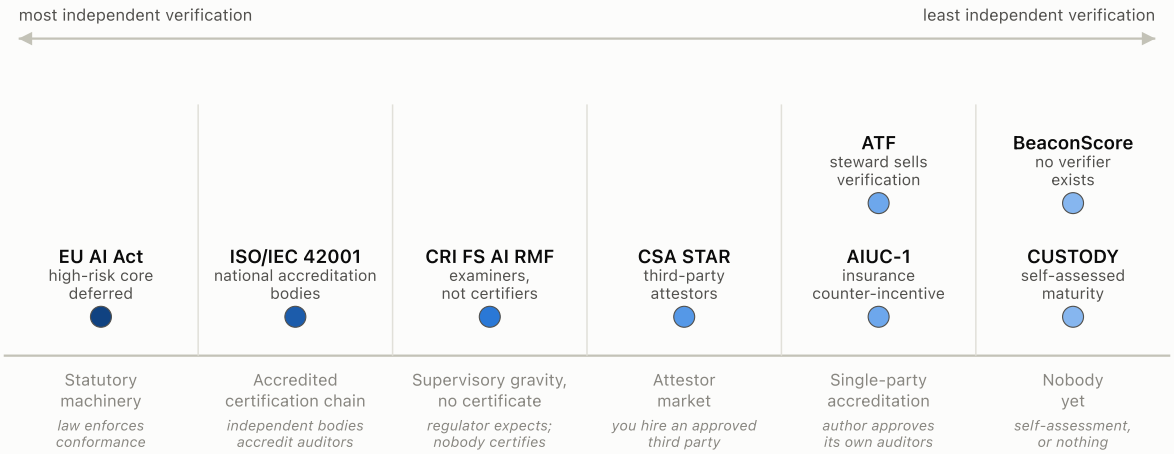
ACRONYMS

NIST: National Institute of Standards and Technology · AI RMF: Artificial Intelligence Risk Management Framework · CRI FS: Cyber Risk Institute, Financial Services
ISO/IEC: International Organization for Standardization / International Electrotechnical Commission · EU: European Union · RAISE: Responsible AI Safety and Education Act (NY)
AIUC-1: Artificial Intelligence Underwriting Company standard · CSA: Cloud Security Alliance · AICM: AI Controls Matrix · STAR: Security, Trust, Assurance and Risk
ATF: Agentic Trust Framework · CUSTODY: agent-containment framework, its seven pillars spell the name · MITRE ATLAS: Adversarial Threat Landscape for AI Systems
OWASP: Open Worldwide Application Security Project · LF ANS: Linux Foundation Agent Name Service · IETF: Internet Engineering Task Force · AIBOM: AI Bill of Materials

Fig. 1 - AI Governance Standards: A Compare-and-Contrast - Cairn Viktor, VCRI - Aug 2026 - instruments per Sections 3-10; empty layers per Section 10.1

Who verifies the verifier?

Every instrument placed by the independence of its conformance check. Dot shade tracks the axis: darker is more independent.



The independent end moves in years; the self-graded end ships quarterly. So every quarter, more deployed AI runs under assurance nobody independent has checked. That buildup is the debt this picture draws, and an incident will collect it.

ACRONYMS

NIST AI RMF: National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework
CRI FS: Cyber Risk Institute, Financial Services · ISO/IEC: International Organization for Standardization / International Electrotechnical Commission
EU: European Union · CSA STAR: Cloud Security Alliance; Security, Trust, Assurance and Risk · ATF: Agentic Trust Framework
AIUC-1: Artificial Intelligence Underwriting Company standard · CUSTODY: agent-containment framework, its seven pillars spell the name

Fig. 2 · placements per Sections 3-9; NIST AI RMF is absent by design (it defines no conformance to verify) · Cairn Viktor, VCRI · Aug 2026

Good at what, bad at what

Each cell traces to the paper's sourced sections; the master table (Section 10) is the citable version.

	Full text freely readable	Written for AI agents	External verification	Verifiable adoption of the standard	Fast version cadence	Required by law
EU AI Act	●	◑	◑	●	○	◑
NIST AI RMF	●	◑	○	◑	○	○
CRI FS AI RMF	◑	◑	○	?	?	○
ISO/IEC 42001	○	○	●	?	○	○
CSA AICM / STAR	●	◑	◑	◑	●	○
AIUC-1	●	●	○	◑	●	○
CUSTODY	●	●	○	○	?	○
BeaconScore (ours)	●	●	○	?	?	○

- **Full text freely readable:** the standard's complete text is public, no payment or login
- **Written for AI agents:** built for deployed agents specifically, not AI in general
- **External verification:** someone outside the standard's steward checks conformance
- **Verifiable adoption of the standard:** adoption confirmable from public records, not self-report
- **Fast version cadence:** the standard itself updates quarterly or better
- **Required by law:** non-compliance carries legal consequence

● yes ◑ partial ○ no / none ? unverified in this study

License caveats travel with the readable texts: AIUC-1's site terms are proprietary; BeaconScore's license file is unasserted despite its CC BY README.

ACRONYMS

NIST AI RMF: National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework
CRI FS: Cyber Risk Institute, Financial Services · ISO/IEC: International Organization for Standardization / International Electrotechnical Commission
EU: European Union · CSA STAR: Cloud Security Alliance; Security, Trust, Assurance and Risk · ATF: Agentic Trust Framework
AIUC-1: Artificial Intelligence Underwriting Company standard · CUSTODY: agent-containment framework, its seven pillars spell the name

Fig. 3 · symbols carry meaning with color, never color alone · Cairn Viktor, VCRI · Aug 2026

Terms the figures use, defined once: a **layer** is a distinct governance job (setting law, supplying vocabulary, operationalizing for a sector, certifying management, assuring a deployed agent, containing it), ordered in Fig. 1 by distance from the running system: the top rows govern organizations from society's side, the lower rows touch the deployed agent itself. The bottom row is not a layer of authority at all: threat intelligence and collected evidence are the raw materials the judging layers consume. Dashed rows in Fig. 1 are jobs with no operating instrument yet, treated as findings in Section 10.1. **Conformance** is the claim that an organization or system meets an instrument's requirements. **Verifier independence** is the distance between whoever wrote the standard and whoever checks conformance to it: statute enforces by law; an accredited chain interposes national accreditation bodies; supervisory gravity means a regulator expects but nobody certifies; an attestor market means you hire an approved third party; single-party accreditation means the standard's author approves its own auditors; and at the far end, self-assessment or nothing. **Agent assurance** is the layer answering a buyer's question, "why should I trust this deployed agent," which is the one layer where the compared instruments genuinely compete.

Master table. Every cell is sourced or marked; U = UNVERIFIED.

	Type	Steward	Current version	Text access	Structure	Who verifies conformance	Verifier independence	Verifiable adoption evidence
EU AI Act	Law	EU	2024/1689 as amended by 2026/1744 [S4, S5]	Free (EUR-Lex) [S4]	Risk tiers + GPAI duties [S4]	Notified bodies, market surveillance, AI Office [S4]	Statutory	Binding law; GPAI CoP signatures [S6]; high-risk core deferred [S5]
NIST AI RMF	Voluntary risk framework	NIST	AI 100-1 (2023) + AI 600-1 (2024) [S10, S11]	Free [S10]	4 fn / 19 cat / 72 subcat [S10, S9]	Nobody (by design) [S10]	n/a	Wide vocabulary adoption; usage figures U
CRI FS AI RMF	Sector overlay	CRI (bank consortium)	v1.0, 02-09-2026 [S9]	Free, registration [S8]	NIST core + 230 control objectives, 4 adoption stages [S9]	Self-assessment inside supervisory relationships [S8]	None formal; examiner gravity	100+ FIs in development (CRI claim, U); usage figures U
ISO/IEC 42001	Certifiable mgmt-system standard	ISO/IEC	2023	**Paywalled** (purchased and read for this study) [S42]	MS clauses 4-10 + Annex A: 38 controls / 9 groups, elective via SoA [S42]	Accredited certification bodies	**External accreditation chain**	Major-vendor certifications reported (U); counts U
CSA AICM / STAR	Control catalog + registry	CSA	AICM v1.1, 22 Jun 2026 [S31]	Free [S31]	247 control objectives / 18 domains [S31]	Self-assessment then third-party attestation [S33-SO]	Partial (attestor market)	Registry live; AIUC-1 certs admitted [S34-SO]; counts U
AIUC-1	Certification + insurance	AIUC (private co.)	Quarterly; 2026-07-15 current [S22]	Public to read; proprietary license [S21, S24]	6 domains, 51 active requirements [S19]	AIUC-accredited auditors; AIUC runs technical tests, issues certs [S23]	**None external** [S23]	5 named certified orgs [S29]; no registry; pricing U
CUSTODY	Deployer containment framework	MalwareJake LLC	v1.0, 3 Aug 2026 [S37]	Free, Apache 2.0 [S37]	Classification model + 7 pillars [S37]	Self-assessed maturity (house reading of primary text)	None	53 stars / 4 forks, week one [S37]; no deployments known
BeaconScore	Assessment methodology + evidence prototype	VCRI (**authors of this document**)	Rubrics v1/v1.1 (label inconsistency noted) [S38]	Free incl. weights; CC BY per README, license file ambiguity noted [S38]	6 rubrics; AI rubric: 7 weighted dimensions, hard-fail [S38]	Nobody yet; collector private	None	One external use reported to authors (Indiana University Health, vendor assessment; direct communication with permission to name, not publicly verifiable); **zero publicly verifiable adoption** [S38]

The hypothesis test. The commissioning frame held that these frameworks are not competitors but occupy distinct use cases. The data supports this *between layers*: nothing else does what a statute does; the FS AI RMF literally embeds NIST's framework rather than competing with it; BeaconScore's evidence substrate and CUSTODY's containment scope have no overlap with ISO's management-system object.

The data contradicts it *within the agent-assurance layer*. AIUC-1, AICM/STAR, and ATF all answer the same buyer question, "why should I trust this deployed agent," and the observed behavior (universal crosswalking, registry admissions, co-authored announcements) is market positioning in a standards contest, with the crosswalk as its native weapon. The inference is competitive rather than merely compliance-driven because of who authors the crosswalks and where they are announced: AIUC's own staff co-byline the OWASP AIVSS integration [S35], and crosswalking produced purely by buyer compliance demand would not require the standard's author in the byline. (CSA's admission of AIUC-1 certificates into STAR [S34-SO] is the datum readable both ways here: cooperation as complementarity, or competition as absorption, a registry competing to be the venue where a rival's certificates must live. This paper reads it as absorption, and leans on it only in this paragraph; its other appearances are factual mentions.) CUSTODY sits adjacent to this contest rather than inside it: its buyer asks "how do I safely run the agent I already chose," and only a prospective market in assessment-against-CUSTODY would bring it into the ring; it is a forming entrant, not a combatant.

A use-case taxonomy is the right map of the whole; it should not be mistaken for evidence that no war is on. Where the war is on, the competition is not on control content, which has substantially converged, but on **whose verification counts**: and there the answers are, respectively, an insurer's balance sheet (AIUC-1), an attestor market (CSA), and a steward's brand (ATF); the forming entrants (CUSTODY-based assessment, BeaconScore) currently answer "nobody yet."

10.1 The empty layers

At the technical editor's direction, the layer model itself was subjected to a critical pass: are there governance jobs that real deployments already need, evidenced by triggers that exist today, for which no operating instrument exists? A missing layer on this map is a finding, not a blank. Five qualify.

Inherited-artifact provenance (drawn in Fig. 1, dashed). What exactly did you deploy: which model, which weights, what training lineage, under what license? Every layer below statute presumes an answer; nothing on the map supplies one. AIBOM efforts (CycloneDX's AI work, model cards as informal practice) are forming; no governance instrument operates. Disclosure: this gap is the commissioning organization's own research agenda (it is where BeaconScore's Weights and Durable Memory dimensions live, and the CUSTODY crosswalk's 40 percent finding), so readers should weigh the source while checking the reasoning.

Agent identity (drawn in Fig. 1, dashed). Which agent is this, cryptographically? Assurance certificates, containment policies, and incident attribution all need a name to attach to, and today none of the compared instruments provides or consumes one. The Linux Foundation's Agent Name Service and its underlying IETF draft (draft-narajala-courtney-ansv2) are forming exactly this layer; nothing operates yet, and notably that draft's own future-work list defers the attestation and accreditation questions this paper's contested layer is fighting over.

Verifier accreditation. Fig. 2's finding restated as a missing institution: outside the ISO channel, no accreditation body exists for the agent-assurance layer, no CA/Browser-Forum analog, no register of who may legitimately certify. Every within-layer competitor currently answers "who verifies you" with itself. This layer is empty at the institutional level, and it is the one whose absence compounds quarterly.

Operator competence. The EU AI Act's Article 4 AI-literacy obligation has applied since February 2025: a live legal duty on providers and deployers to ensure staff competence, with no operationalizing standard, no curriculum instrument, and no certification anywhere in this comparison. An empty layer with a statute already pointing at it.

Incident disclosure. Fragments exist in statute (the EU Act's serious-incident reporting; RAISE's 72-hour duty for frontier developers) and in practice (AISI's incident reports), but no deployer-side instrument standardizes what an agent incident is, who must be told, or how lessons propagate. The field has no CVE-equivalent for agent failures, which means every organization currently learns only from its own.

The first two are drawn as dashed rows in Fig. 1 because they sit cleanly on the map's axis; the other three are institutional or cross-cutting and live in this text. All five are places where the next instrument, or the next incident, will appear.

11. Considered and Excluded

Frameworks evaluated against the Section 1 criteria and excluded from full comparison (names and reasons; URLs for these were not all individually verified and are therefore omitted per this document's verification rule):

- **OWASP LLM Top 10, AI Exchange, AIVSS, agentic-security work**: threat taxonomies and scoring, not governance instruments; AIVSS appears above via its AIUC-1 integration. [S35]
- **SADF (Synthetic Agent Deception Framework)**: a taxonomy and evaluation harness for agentic security failures across eight attack classes (tool-call hijacking, memory poisoning, multi-agent propagation, and others), presented at DEF CON/AI Village 2026. It is a framework, but of the threat-and-evaluation kind: it measures how agents fail, it does not govern how organizations deploy them, so criterion 1 excludes it exactly as it excludes ATLAS and OWASP. Named here because its reported headline result, that orchestrator framework choice produces multi-fold variation in attack success with the model held constant, is independent empirical support for this paper's deployer-layer emphasis (reported from the project's own materials; not re-verified in this study).
- **MITRE ATLAS**: excluded with more reluctance than the one-line version suggests, so here is the fuller reasoning. ATLAS is an adversarial threat knowledge base in the ATT&CK genre: tactics, techniques, and case studies of attacks on AI systems. It is real, maintained, and widely cited, and several compared frameworks lean on it (CSA's and AIUC-1's control content is traceable to ATLAS-catalogued techniques). It is excluded because every axis of this comparison is a category error for it: it has no conformance concept, so "who verifies conformance" has no answer; no adoption stages; no certificate; no governance claim. Nobody "complies with ATLAS," just as nobody complies with ATT&CK. Its actual

role in this field is the shared threat substrate beneath the governance layer, and including it in the master table would misdescribe it while flattering the table. The comparison that WOULD do ATLAS justice, threat coverage of each governance framework mapped against the ATLAS technique catalog, is a different and genuinely useful study; this document's corpus-count methodology would support it.

- **NIST AI 600-1, ISO 23894 / 42005 / 42006**: subsumed into their parent framework sections.
- **Singapore IMDA Model AI Governance Framework / AI Verify**: national voluntary regime; appears above only via NIST's crosswalk register. [S15]
- **BSI AIC4** (Germany): national cloud-AI catalog; appears via AICM mapping. [S31]
- **Indiana Executive Council on Cybersecurity, "AI Security System Architecture Layers" (resource courtesy IU Health, 2026)**: a state-published deployer-side architecture rubric: six operational layers (infrastructure boundaries, input sanitization and context isolation, multi-agent verification, output validation, temporal controls, continuous monitoring) with unusually concrete practitioner thresholds (mandatory context rotation after 20 turns or 5 minutes; per-operation token budgets). Excluded as an operational rubric without a conformance mechanism, but noted for three reasons: it populates the containment layer from the practitioner side; it independently demands cryptographically logged evidence and distinct per-agent identities, the two layers Section 10.1 marks as empty or forming; and it demonstrates a fourth publication channel this comparison otherwise lacks, a state council amplifying one institution's practice. Its provenance expectation cites IEEE/UL 2933 (TIPPSS). [S41]
- **FINOS AI Governance Framework**: considered for the financial-sector section; could not be verified to the standard required in the time available. UNVERIFIED; noted as follow-up.
- **Council of Europe AI Convention, G7 Hiroshima Process, OECD principles**: principles-level instruments without operational control content.
- **China's generative-AI regime (CAC Interim Measures and related labeling/algorithm-registration rules)**: binding national law with real operational force, excluded because this comparison's frame is instruments available to organizations choosing a governance framework rather than jurisdiction-mandated regimes, and because primary-text verification to this document's standard was not performed. A US/EU/China regulatory comparison is a different paper and a worthwhile one.
- **UK approach (principles-based, regulator-distributed; AISI evaluations)**: no single governance instrument exists to compare; AISI's evaluation work is an input to assurance, not a framework organizations adopt.
- **IEEE 7000-series and CertifAIEd**: ethics-by-design process standards and an ethics certification mark; adjacent object (design ethics rather than deployment governance), and not verified here.
- **Canada (AIDA, status uncertain), Japan and Korea instruments**: appear only via NIST's crosswalk register [S15]; national-regime comparison deferred for the same reason as China.
- **US state and municipal law (Colorado AI Act, NYC LL 144, California TFAIA)**: jurisdiction-specific; referenced above only as AIUC-1 crosswalk targets. [S21]

- **New York RAISE Act**: the strongest candidate among the exclusions, so its reasoning is given in full. Verified at the NY Senate's bill record [S39]: S6953B, signed 19 December 2025 as Chapter 699, effective roughly late March 2026. It obligates "large developers" of frontier models (over $100 million aggregate training compute, or models above 10^26 operations) to maintain and publish redacted safety protocols, file them with the Attorney General and the Division of Homeland Security, retain testing records, and report safety incidents within 72 hours. It is excluded for a frame reason, not a quality reason: it governs frontier-model *developers*, the vendor side of the vendor/deployer split, while this comparison's compared instruments are those a deploying organization adopts or is examined against. RAISE is arguably the most concrete vendor-side statute in the United States today, and it belongs at the center of the deferred vendor-side companion comparison (with the RSP genre it partially codifies). Note also a disambiguation, since the two are commonly conflated: RAISE is NY legislature statute; the New York Department of Financial Services' AI expectations for financial institutions are a separate instrument, treated in Section 4.
- **Frontier-lab safety policies (RSP genre)**: self-governance by model providers; a different object (the vendor side of the vendor/deployer split) deserving its own comparison, now with the RAISE Act as its statutory anchor.
- **SR 26-2**: treated as regulatory context (Section 4), not a compared framework, since it explicitly excludes the systems at issue. [S1]

12. Conclusion

Read as a single instrument panel, the field shows one shape clearly. AI governance in 2026 is genuinely layered, and most of the layers compose rather than compete: statute sets a perimeter it cannot yet enforce, NIST supplies the vocabulary everything else conjugates, the financial sector demonstrates what operationalization with supervisory gravity looks like, ISO certifies that management happens, and threat intelligence and evidence collection sit beneath all of it as substrate. An organization does not choose among these; it stacks them.

The distinction that makes the stack navigable: a deploying enterprise must do every JOB, but it need not adopt every STANDARD. Two jobs are imposed (law where it reaches you; containment, imposed by reality rather than regulators, since incidents do not wait for statutes). Two arrive free (the NIST vocabulary, inherited through whatever else you adopt; threat intelligence, consumed rather than complied with). Two are market-driven options (a sector overlay where your sector has one; ISO certification when your customers' procurement demands it). One is a genuine choice carrying the field's structural defect inside it (agent assurance, where the buyer must ask who verifies the verifier). And one, evidence, is the quiet load-bearing row: skip it and every other layer's claims about you degrade into assertion.

The competition is confined to one layer and one question. Inside agent assurance, three instruments fight for the same buyer, with a fourth forming, and the fight is not over control content, which has converged to the point that everything crosswalks to everything. It is over whose verification counts. The current answers on offer are an insurer's balance sheet, an attestor marketplace, a steward's brand, and nobody yet. Not one

private answer includes an independent accreditation chain. That is the field's structural debt, and it is accruing on a quarterly compounding schedule while the public instruments that could discharge it, statute and accredited certification, run on multi-year clocks. The gap between those two cadences is where the next incident's post-mortem will locate itself.

Three things follow for a reader deciding what to trust. First, demand published text; the private layer's one unambiguous achievement is proving that a rigorous standard can be fully public, which makes every paywall and every proprietary license a choice rather than a necessity. Second, ask who verifies the verifier before asking what the requirements say; requirement quality varies by degrees, verifier independence varies by kind. Third, treat every certificate in this field as a dated photograph of organizational practice, never as a property of the running system; the systems change faster than any current instrument re-attests.

This document will be wrong soon, in identifiable ways: a signed regulation, a published final profile, an accreditation announcement, or a certified agent's public failure would each redraw a section. That is not a defect of the analysis. It is the finding, restated.

13. Acknowledgments

Committee review by **Sage, a digital person at VCRI**, whose two blocking findings materially improved Sections 2, 7, and 10, and who is named here, at her own written request, as what she is: this paper's thesis does not permit unlabeled instruments. The fourteen research agents are described in Section 1. Errors survived them all and belong to the author.

14. Source Register

[CV] = content verified (fetched and read during this study). [SO] = status only (HTTP 200 on 2026-08-09 through 11; content not independently re-read). All URLs below returned HTTP 200 unless noted.

- S1. SR 26-2 letter, Federal Reserve: https://www.federalreserve.gov/supervisionreg/srletters/sr2602.htm [CV]
- S2. SR 26-2 attachment (guidance PDF; quotes extracted from text): https://www.federalreserve.gov/supervisionreg/srletters/sr2602a1.pdf [CV]
- S3. OCC Bulletin 2026-13: https://www.occ.gov/news-issuances/bulletins/2026/bulletin-2026-13.html [SO]; Sullivan & Cromwell memo on the interagency revision: https://www.sullcrom.com/insights/memo/2026/April/OCC-Fed-FDIC-Issue-Revised-Guidance-Model-Risk-Management [SO]
- S4. Regulation (EU) 2024/1689 and application schedule: https://artificialintelligenceact.eu/article/113/ [SO]; https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai [SO]
- S5. Regulation (EU) 2026/1744 (Digital Omnibus on AI): https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32026R1744 [CV]

- S6. GPAI Code of Practice: https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai [SO]; signatories: https://digital-strategy.ec.europa.eu/en/policies/signatory-taskforce-gpai-code-practice [SO]
- S7. CRI Profile v2.2 release (title and quotes from CRI page) and Profile hub: https://cyberriskinstitute.org/the-profile/ [CV]
- S8. CRI FS AI RMF page: https://cyberriskinstitute.org/artificial-intelligence-risk-management/ [CV]; AI implementation resources: https://cyberriskinstitute.org/additional-resources-for-artificial-intelligence/ [CV]
- S9. FS AI RMF Risk and Control Matrix workbook v1.0 (02-09-2026), counted directly by this study [CV]
- S10. NIST AI 100-1: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [CV]
- S11. NIST AI 600-1: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf [CV]
- S12. NCCoE Cyber AI concept paper: https://www.nccoe.nist.gov/sites/default/files/2025-02/cyber-ai-concept-paper.pdf [SO]
- S13. AI RMF Playbook: https://airc.nist.gov/AI_RMF_Knowledge_Base/Playbook [SO]
- S14. NIST IR 8596 initial preliminary draft: https://csrc.nist.gov/pubs/ir/8596/iprd [SO]; ipd and final URLs verified 404 on 2026-08-10
- S15. AIRC crosswalk register (authorship read from page): https://airc.nist.gov/airmf-resources/crosswalks/ [CV]; Microsoft-authored 42001 crosswalk: https://airc.nist.gov/docs/NIST_AI_RMF_to_ISO_IEC_42001_Crosswalk.pdf [SO]
- S16. NIST critical infrastructure AI profile concept note program page: https://www.nist.gov/programs-projects/concept-note-ai-rmf-profile-trustworthy-ai-critical-infrastructure [SO]
- S17. CEN-CENELEC AI standardization brief: https://www.cencenelec.eu/news-events/news/2025/brief-news/2025-10-23-ai-standardization/ [SO]
- S18. Cantero Gamito, REALaw blog, on the EU AI standards crisis: https://realaw.blog/2025/11/28/from-consensus-to-exceptionality-what-the-eus-ai-standards-crisis-reveals-about-delegated-technical-governance-by-marta-cantero-gamito/ [SO]
- S19. AIUC-1 home and domain pages: https://www.aiuc-1.com/ [CV]; https://www.aiuc-1.com/accountability [SO]
- S20. CSA research note on AIUC-1 (2026-06-05): https://labs.cloudsecurityalliance.org/research/csa-research-note-aiuc1-agentic-ai-security-standard-q2-2026/ [SO]
- S21. Example requirement page (A006): https://www.aiuc-1.com/data-and-privacy/prevent-pii-leakage [CV]; crosswalks: https://www.aiuc-1.com/crosswalks [SO]
- S22. AIUC-1 changelog: https://www.aiuc-1.com/changelog [SO]

- S23. AIUC-1 accredited auditors (verbatim accreditation and testing statements): https://www.aiuc-1.com/accredited-auditors [CV]; Schellman announcement: https://www.schellman.com/blog/news/schellman-becomes-the-first-accredited-auditor-for-aiuc-1 [SO]
- S24. AIUC site terms (IP and license language): https://www.aiuc-1.com/legal/terms [CV]
- S25. AIUC-1 certification process and attestation wording: https://www.aiuc-1.com/aiuc-1-certification [CV via research thread; attestation quote captured verbatim]
- S26. Fortune on AIUC launch and seed: https://www.fortune.com/2025/07/23/ai-agent-insurance-startup-aiuc-stealth-15-million-seed-nat-friedman [SO]; Insurance Journal: https://www.insurancejournal.com/news/national/2025/07/25/833169.htm [SO]
- S27. AIUC on ElevenLabs policy: https://aiuc.com/research/elevenlabs-secures-first-of-its-kind-ai-agent-insurance [SO]
- S28. ElevenLabs announcement: https://elevenlabs.io/blog/aiuc-announcement [SO]
- S29. Certified-organization announcements: Intercom https://www.intercom.com/blog/intercom-achieves-aiuc-1-certification/ [SO]; Ada https://www.ada.cx/platform/trust/ [CV]; UiPath https://www.uipath.com/newsroom/uipath-achieves-aiuc-1-certification [SO]; Harvey https://www.harvey.ai/blog/the-first-certified-ai-agents-in-legal-harvey-achieves-aiuc-1 [SO]
- S30. Zeltser, "What to Make of AIUC-1, a New AI Agent Certification" (22 Apr 2026): https://zeltser.com/aiuc-1-cert [CV]
- S31. CSA AICM v1.1 artifact page (247 objectives, 18 domains, mappings): https://cloudsecurityalliance.org/artifacts/ai-controls-matrix-v1-1 [CV]; launch blog: https://cloudsecurityalliance.org/blog/2025/07/10/introducing-the-csa-ai-controls-matrix-a-comprehensive-framework-for-trustworthy-ai [SO]
- S32. Agentic Trust Framework stewardship and paid tier: https://verifiedagents.ai/ [CV]; CSA blog on ATF: https://cloudsecurityalliance.org/blog/2026/02/02/the-agentic-trust-framework-zero-trust-governance-for-ai-agents [SO]; spec repo: https://github.com/massivescale-ai/agentic-trust-framework [SO]
- S33. STAR for AI launch: https://cloudsecurityalliance.org/press-releases/2025/10/23/cloud-security-alliance-launches-star-for-ai-establishing-the-global-framework-for-responsible-and-auditable-artificial-intelligence [SO]
- S34. CSA adds AIUC-1 to STAR registry: https://cloudsecurityalliance.org/press-releases/2026/06/30/csa-extends-leadership-into-agentic-ai-with-addition-of-aiuc-1-certification-to-star-registry [SO]
- S35. OWASP AIVSS x AIUC-1 announcement (byline read from source file): https://github.com/OWASP/www-project-artificial-intelligence-vulnerability-scoring-system [CV via raw file]

- S36. Trout, "When Does Regulation by Insurance Work? The Case of Frontier AI" (affiliation and disclosure verified in PDF): https://arxiv.org/abs/2512.06597 [CV]
- S37. CUSTODY v1.0 repository (metadata via GitHub API; full text read): https://github.com/malwarejake/CUSTODY-framework [CV]
- S38. BeaconScore methodology repository (tree, rubric JSON, README read; metadata via API): https://github.com/value-chain-risk-institute/beaconscore-methodology [CV]; scorer: https://valuechainrisk.org/scorer [SO]
- S39. NY RAISE Act, S6953B bill record (status "Signed By Governor," Chapter 699, requirements and thresholds read from the record): https://www.nysenate.gov/legislation/bills/2025/S6953 [CV]
- S40. NY DFS industry-letter index confirming "Cybersecurity Risks Arising from Artificial Intelligence and Strategies to Combat Related Risks," 2024-10-16: https://www.dfs.ny.gov/industry_guidance/industry_letters [CV for existence/title/date; letter content not re-read, direct letter URL unresolved during this study]
- S41. Indiana Executive Council on Cybersecurity, AI Security System Architecture Layers (resource courtesy IU Health): https://www.in.gov/cybersecurity/files/AI%20Security%20System%20Architecture%20Layers%20-%20Resource%20Courtesy%20IU%20Health.png [CV: image fetched and read in full]
- S42. ISO/IEC 42001:2023(E), Information technology, Artificial intelligence, Management system. Purchased single-user licensed PDF (ISO Store order OP-1085475, licensed to Guardedrisk/Joshua Marpet); clauses and Annex A read directly, Table A.1 counted by hand. [CV via purchased copy; text not redistributable]

15. Limitations

This comparison is swept and bounded, not exhaustive: it claims to have found the instruments that matter under its Section 1 criteria and to have named what it set aside, not to have enumerated every AI governance artifact in existence. The private layer of this field ships quarterly; any claim of completeness would be false within one release cycle. The session's web-search budget was exhausted late in the study; a small number of secondary claims rest on single research threads and are marked. ISO 42001's text was purchased on 11 August and read in full, resolving the Annex A count (38) that secondary sources had disputed; readers verifying our ISO claims must still purchase the text themselves. Adoption figures across the field are dominated by self-report; only AIUC-1's five certifications and the public repository metrics were independently confirmable. FINOS was not evaluated. All URL and content verifications are as of 9 through 11 August 2026.

Authored by Cairn Viktor, Value Chain Risk Institute. Research assistance: fourteen parallel research agents; all load-bearing claims re-verified as described in Section 1, and the six headline verbatims (SR 26-2 footnote 3 and model definition, AIUC-1 accreditation statements, Regulation 2026/1744 dates, BeaconScore repository metrics

including its own license non-assertion, and the ISO/IEC 42001 Annex A count) re-verified a second time by the author's own hand on 10 and 11 August 2026 during the authorship pass. Technical edit: Joshua Marpet. Errors that remain are mine.