The Assurance Gap

The Case for Independent Conformance Assessment of Deployed AI Systems

John Francis · July 2026

The United States has AI standards but no AI assurance for deployed systems. This edition states the gap more precisely than earlier ones, and narrows the claim accordingly. Auditor accreditation is not missing: ANAB accredits ISO/IEC 42001 certification bodies, and Schellman, SGS, DEKRA, DQS, and BSI hold that accreditation. What is missing is conformance testing of system behavior, a defined certification unit that survives model updates, and a public registry. The right structure is therefore not a rival accreditation hierarchy but a conformance-assessment scheme operating inside the existing architecture. As of July 2026 the surrounding conditions have worsened: federal policy has shifted toward voluntary standards and state-law preemption, the most ambitious state AI law was repealed before it took effect, and Europe deferred high-risk enforcement by sixteen months. Drawing on Underwriters Laboratories, FedRAMP, and the NTSB, this paper specifies what gets certified and for how long, who pays and why publishing test methods defeats the issuer-pays capture that corrupted credit ratings, why parametric insurance makes AI harm priceable, why state procurement is the one demand lever federal preemption explicitly does not reach, and the seven ways the institution fails.

CONTENTS

Executive summary

The United States has AI standards. It does not have AI assurance.

The distinction sounds bureaucratic and it is not. Standards are documents describing what a trustworthy AI system should look like. Assurance is an operation, in which an independent examiner tests a specific system against the standard and signs the result, so that a buyer, an insurer, or a regulator can rely on that result without repeating the work.

The American economy runs on assurance in every other high-stakes domain. Electrical equipment carries an independent safety mark. Cloud services sold to the government are authorized once and reused everywhere. Accredited professionals audit financial statements against common standards. Artificial intelligence has no equivalent, while every enterprise, hospital, bank, and agency is being asked to adopt it at once.

The gap is not a lack of standards, and this edition is more precise than earlier ones about what it is. NIST's AI Risk Management Framework describes responsible AI management. ISO/IEC 42001 makes management systems certifiable. ISO/IEC 42006 specifies what a competent AI auditor looks like, and the accreditation machinery that qualifies those auditors is not hypothetical: the ANSI National Accreditation Board has accredited a growing roster of certification bodies against ISO/IEC 42001, including Schellman, SGS, DEKRA, DQS, and BSI. Any paper claiming the accreditation layer does not exist is wrong.

What does not exist is conformance testing of deployed system behavior, a definition of what unit gets certified and for how long, and a public registry in which a certificate's meaning is fixed. An ANAB-accredited body certifies that an organization operates a conforming AI management system. Nobody certifies that a particular model, in a particular configuration, serving a particular population, performs within stated bounds. That is the gap, and it is narrower and more tractable than "America needs a new institution."

The claim of this edition is correspondingly narrowed. The missing institution should not be a rival accreditation hierarchy. It should be a conformance-assessment scheme operating inside the existing accreditation architecture, seeking accreditation rather than competing with it, and contributing the three things that architecture does not supply: published test methods for deployed-system behavior, a defined certification unit with a lifecycle that survives model updates, and a public registry that records suspensions and withdrawals as well as grants.

The events of the past eighteen months make this more necessary rather than less. As of July 2026 the federal government's AI focus has shifted to voluntary standards and away from oversight of deployed systems, and Washington's most energetic AI initiative is a campaign to preempt state AI laws. Colorado repealed its landmark AI act and replaced it with a narrower disclosure law before the act took effect. California scoped its law to a handful of frontier-model developers. Europe deferred its high-risk enforcement deadline by sixteen months. Regulation is in retreat on both sides of the Atlantic, and the trust problem is not.

History offers a working blueprint. Underwriters Laboratories was not founded by government; fire insurers founded it because they could not price electrification risk, and that private independent institution became the trust infrastructure for a transformative technology. This paper describes the gap, shows precisely what the incumbent path can and cannot do, specifies what gets certified and for how long, states who pays and why that does not corrupt the mark, and sets out how the thing fails.

The assurance gap

Three words get used interchangeably in AI policy discussions, and the confusion between them is where the gap hides.

Standards are written descriptions of good practice. NIST's AI Risk Management Framework, released in January 2023, is the American reference point, a voluntary framework describing how organizations should govern, map, measure, and manage AI risk [4]. ISO/IEC 42001, published in December 2023, translates the same territory into a certifiable management-system standard, the AI analogue of ISO 9001 for quality or ISO 27001 for information security [18]. These documents are necessary and good, and they are also inert. A standard, by itself, verifies nothing.

Evaluation is what the government bodies with "AI safety" in their names practice. It means technical testing of frontier models, the small number of most capable systems, for national-security risks, which is what the U.S. Center for AI Standards and Innovation does, and what its British counterpart does. Evaluation is upstream and pre-deployment. It aims at a dozen laboratories rather than at the millions of AI systems being wired into commerce.

Assurance is the missing third thing: independent verification that a specific deployed system, in a specific context, meets a recognized bar, in a form third parties can rely on. It has three working parts, and the crucial point for this edition is that they are absent to different degrees.

1. **Certification.** An independent body examines an AI system, or its operator's management practices, against a recognized standard and issues a credential the market recognizes. For management practices this exists and is accredited. For system behavior it does not exist.
2. **Conformance testing.** Shared technical infrastructure evaluates systems against defined criteria, including bias and robustness testing, performance verification, and documentation review, so that certification rests on evidence rather than attestation. This does not exist at national scale in any form.
3. **Auditor accreditation.** A mechanism answers the question of who audits the auditors. This exists, operates, and works, which earlier editions of this paper understated.

The simplest way to see what remains missing is a test we call the deployment test. A hospital system is offered a diagnostic-support model. A mid-size insurer wants an underwriting assistant. A county government is procuring a benefits-eligibility screening tool. In each case, ask one question: what independent, nationally recognized verification can the seller present, and the buyer rely on, that this system meets an established bar for safety, fairness, and performance in this use?

As of July 2026 the honest answer is none. The buyer can request the vendor's own documentation. It can commission a custom audit from a consultancy at custom expense, producing a report no other buyer can reuse. It can look for an ISO/IEC 42001 certificate, which speaks to the vendor's management processes rather than to the behavior of the specific system being purchased, and which is a real and accredited credential answering a different question. The buyer cannot do what it does routinely for electrical equipment, cloud security, or financial controls, which is to point to a mark attesting that this artifact was tested and behaved thus.

Every sophisticated buyer in America therefore runs its own private version of AI due diligence. The work is duplicated and the results are not portable. More commonly the buyer skips the work and absorbs the risk unknowingly. The first outcome slows adoption to the speed of each buyer's in-house expertise, and the second accumulates a liability the system has no way to see, price, or reduce.

The 2026 landscape

The natural objection is that this gap is temporary, and that some existing body must be growing into the role. A survey as of July 2026 shows movement away from it in the public sector and, importantly, movement toward part of it in the private sector. The claims below are anchored to their sources deliberately, because in a landscape this fluid undated assertions age badly.

Federal: standards without conformance testing, and a preemption campaign

The federal government's designated home for AI trust work is CAISI, the Center for AI Standards and Innovation, formerly the U.S. AI Safety Institute. The rename, announced in June 2025, was substantive: the center's focus narrowed to "demonstrable risks, such as cybersecurity, biosecurity, and chemical weapons," and it gained an explicit mandate to protect U.S. AI exporters from "burdensome and unnecessary regulation" abroad [1].

CAISI's own description of its remit is instructive for what it omits. It includes voluntary guidelines, unclassified national-security evaluations of frontier models, interagency coordination, and international standards diplomacy. It does not include certification of deployed commercial AI systems, accreditation of auditors, or conformance-testing infrastructure open to the market [2]. Its most significant 2026 initiative, the AI Agent Standards Initiative launched in February, is standards development in the classic mold: industry-led, routed through ISO/IEC committees, and explicitly voluntary [3].

The White House's most energetic AI-governance initiative points the other direction. Executive Order 14365, "Ensuring a National Policy Framework for Artificial Intelligence," signed in December 2025, established a Department of Justice task force to challenge state AI laws in federal court and directed the Commerce Department to identify "onerous" state statutes and condition federal funding on their rollback [5]. Earlier editions of this paper cited the order under the title "Eliminating State Law Obstruction of National Artificial Intelligence Policy," which is the URL slug rather than the title, and the correction matters only because a cite-checker will find it.

Whatever one's view of preemption as policy, the order's premise matters here: it treats the problem in American AI governance as an excess of state rules, and proposes no federal assurance mechanism to replace the trust functions those rules attempted to serve. An executive order also cannot itself preempt state law, as contemporaneous legal commentary noted [5]. What it reliably signals is that federal AI-assurance legislation is not arriving soon.

The order contains one provision earlier editions omitted, and the omission cost this paper its most concrete recommendation. Alongside carve-outs for child safety and for AI compute and data-center infrastructure, the order **explicitly excludes state government procurement and use of AI from preemption** [5]. States therefore retain undisputed authority to demand assurance in what they themselves buy. That is a surviving demand lever that federal preemption does not reach, and it is developed below.

The federal government also acts as a buyer, and in that role could simply demand assurance. Instead its 2025 procurement memoranda require each agency to run its own AI risk monitoring internally, with no independent cross-agency certification body to rely on [6][22]. The federal government has the same problem as every private buyer and has so far chosen to solve it separately in each agency.

International: a regulatory regime, deferred, and not exportable

The European Union built the world's most comprehensive AI regulatory regime, and in 2026 its hardest parts slipped. The AI Act's prohibited-practice provisions took effect in February 2025 and its general-purpose AI obligations followed in August 2025 [7]. The centerpiece was mandatory conformity assessment for high-risk systems, originally due in August 2026; the Digital Omnibus package, politically agreed in May 2026, deferred it to December 2027, with product-embedded systems deferred to August 2028 [8]. The European AI Office is an enforcement and supervision body within the Commission rather than an independent assurance provider [9].

Two things follow for the United States. First, even the jurisdiction most committed to mandatory AI oversight concluded its assurance infrastructure was not ready on schedule, because notified bodies, harmonized standards, and testing capacity all lagged. Regulation outran the institutions needed to operationalize it, which is a warning about sequencing rather than about ambition. Second, EU conformity assessment is scoped to the EU market and generates no credential an American hospital or county government can use.

The United Kingdom completes the pattern. Its AI Safety Institute became the AI Security Institute in February 2025, with a mission to equip governments with scientific understanding of frontier-model risks, which is capability evaluation rather than certification of deployed commercial products [10].

States: the patchwork that unmade itself

The state story is best told through Colorado, because Colorado went first and furthest. The Colorado AI Act of 2024 was the first comprehensive U.S. law to govern high-risk AI systems, imposing a duty of reasonable care against algorithmic discrimination, requiring risk-management programs and impact assessments from deployers, and giving consumers notice and appeal rights [11]. It was scheduled to take effect in February 2026 and never did. An August 2025 special session postponed it to June 2026 [12], and before that date arrived the legislature repealed and replaced it outright. Senate Bill 26-189, signed in May 2026, abandons the duty of care, the risk-management mandate, and the impact-assessment requirement, substituting a disclosure regime with obligations from January 2027 [13].

The trajectory matters more than the destination. The most ambitious state AI law in America was delayed once and then substantially narrowed before its original effective date arrived, and the pressures that produced this outcome face any state acting alone: compliance-cost objections, preemption threats, and first-mover competitive anxiety. Texas took a different approach, with intent-based prohibitions enforced solely by the attorney general from January 2026 [14]. California took a third, imposing transparency and incident-reporting obligations on frontier-model developers above 500 million dollars in revenue [15]. None of the three produces a portable credential and none is interoperable with the others. Lawmakers in 45 states introduced more than 1,500 AI-related bills in the first months of 2026 [16][17].

For a company deploying AI nationally, this is a fifty-jurisdiction compliance problem that changes annually. For a buyer seeking assurance it is worse, because it is proof that trust infrastructure built on any single state's law is built on sand. Colorado enterprises stood up risk-management programs against the 2024 act and watched its requirements dissolve before enforcement began. The institution that fills the assurance gap must derive its authority from something more durable than a statute one special session can unwind.

The market: more built than earlier editions admitted

The private sector has assembled more of this than previous editions of this paper acknowledged, and the correction is substantial enough to change the argument.

ISO/IEC 42001 gives organizations a certifiable AI-management standard, with certifications issuing since late 2024 [18]. ISO/IEC 42006, published in July 2025, specifies requirements for the bodies that audit and certify against it [19]. Earlier editions called that "the accreditation layer, on paper." It is not on paper. The ANSI National Accreditation Board operates an accreditation program for ISO/IEC 42001 certification bodies, and a roster of bodies has been accredited under it: Schellman first, followed by SGS in April 2025, DEKRA in November 2025, DQS in February 2026, and BSI in March 2026, the last holding accreditation from UKAS and RvA alongside ANAB [28][29][30]. Accreditation bodies including ANAB in the United States and UKAS in the United Kingdom operate within the International Accreditation Forum's mutual-recognition machinery, which is what makes an accredited certificate legible across borders.

So the auditor-accreditation function of this paper's three-part scheme exists, is staffed, and is internationally recognized. What those accredited bodies certify is an organization's AI management system. What none of them certifies is the behavior of a deployed system.

The insurance market has begun to price AI risk directly. Munich Re's aiSure, expanded through a February 2026 partnership with Mosaic Insurance, offers AI vendors up to 15 million dollars in coverage against defined performance failures, underwritten on measured system behavior rather than as a variant of cyber or technology errors-and-omissions cover [20]. Independent estimates describe an AI audit-and-assurance services market well under 1 billion dollars in 2026 with projected growth of more than thirtyfold within a decade, and those figures are vendor-sourced and directional [21].

Each of these is a component. Management-system certification attests to an organization's processes rather than to a model's behavior in a use. Certification bodies compete on brand with no shared public registry, so a certificate's scope varies with its issuer even when the accreditation behind it is identical. Insurers underwriting AI performance must each build proprietary evaluation capability, privately recreating a verification function a shared institution would provide once. The market has produced standards, accreditation, incentives, and early demand. It has not produced published test methods for deployed behavior, a defined certification unit, or a registry.

Precedents: how America has closed assurance gaps before

The assurance gap is the recurring problem of transformative technology, in which adoption outruns the ability of buyers and risk-carriers to verify safety. Three American institutions show the available design space.

Underwriters Laboratories: the market builds its own assurance. In 1893 electricity was the artificial intelligence of its day, transformative, poorly understood, and prone to burning down buildings. William Henry Merrill Jr., an electrical engineer who had inspected wiring at the World's Columbian Exposition, saw the structural problem: fire insurers could neither price nor reduce electrification risk because no one could independently verify which equipment was safe. In 1894 he founded what became Underwriters Laboratories with two employees, 350 dollars of equipment, and the backing of two fire-insurance underwriters' associations [23]. UL was not a government agency and its mark was never legally mandatory, and it became universal anyway because the parties carrying the risk demanded it. That is the essential mechanism: assurance becomes infrastructure when risk-carriers recognize the mark, not when a statute compels it. UL also shows that independence and commercial sustainability can coexist, because manufacturers pay for testing while the standard-setting and the mark answer to the institution rather than to any client. The section below on who pays takes that structure apart, since it is the crux of the strongest objection to this proposal.

FedRAMP: authorize once, use many. By 2011 every agency was independently security-assessing the same cloud services, and FedRAMP's answer was architectural: a service authorized through one rigorous standardized assessment can be reused by every other agency without repeating the work [24]. The principle transfers directly. FedRAMP's limitation transfers too, because it works partly because the federal government is a captive buyer that mandated the program internally, and an institution serving the open market must earn reuse through recognition instead.

FedRAMP contributes something more specific that earlier editions cited without mining. It solved the problem of a certified artifact that changes after certification, through two mechanisms: continuous monitoring, in which an authorized provider reports on a fixed cadence against controls fixed at authorization, and significant-change procedures, under which material changes must be notified and may trigger reassessment. That is the template for the hardest unanswered question about AI certification, and it is developed next.

NTSB: independence is the source of authority. Congress moved the National Transportation Safety Board out of the Department of Transportation in 1974 on the rationale that no agency can credibly investigate failures connected to its own decisions, and that the investigator had to be "totally separate and independent" from the regulator [25]. Fifty years of findings are trusted by industry, courts, and the public precisely because the board has no stake in the outcome, since it does not regulate, license, or enforce [26]. The lesson cuts two ways. Independence from vendors, from any single buyer, and from the bodies whose standards are being tested against is what makes a mark meaningful. NTSB's investigate-only design also shows that independence without a market mechanism changes behavior slowly, because its recommendations can be declined and some are.

The synthesis is that the institution should be insurer-anchored like UL, reuse-architected like FedRAMP, and structurally independent like NTSB. None of the three can be copied whole, and together they define the corners of the design space.

What exactly gets certified, and for how long

UL certifies a toaster, and the toaster is the same toaster next year. An AI system is not. Models are updated, fine-tuned, retrieval-augmented, and in some architectures learn continuously. Any proposal for AI certification that does not define its unit and its lifecycle has not begun, and this is the first question a skeptical reader asks. Earlier editions of this paper contained no answer.

The certification unit is a deployment configuration bound to a use profile. Not a model checkpoint, which is insufficient because the same weights behave differently behind different prompts, retrieval corpora, and thresholds. Not an organization, which is what ISO/IEC 42001 already covers. The certified object names, at minimum: the model version or versions in use; the system prompt and policy layer; the class and provenance of any retrieval corpus; the guardrail and filtering configuration; and the declared use profile, meaning the task, the population served, the consequence tier of the decision, and the human-review posture. Change any element and the certified object has changed. The certificate is unreadable without all of them, which is why the registry described below must carry the full scope rather than a company name and a date.

Validity is bounded and short. A certificate expires twelve months from issue, and expiry is not a formality: re-testing at renewal is full rather than abbreviated, because the population and the environment move even when the system does not.

Significant change triggers re-testing before expiry. Following FedRAMP's structure but with AI-specific triggers, the following require notification and re-testing: any weight update, including fine-tuning; any change to the system prompt or policy layer; any change to the class or source of a retrieval corpus; any change to a decision threshold; any expansion of the declared use profile or the population served; and any reduction in the human-review posture. The following do not: infrastructure and hosting changes, latency optimization, logging changes, and changes that do not touch the inference path. That boundary will be argued about, and publishing it is what makes the argument possible.

Continuous monitoring covers the interval. Between certifications the operator reports agreed indicators on a fixed cadence against thresholds fixed at certification, including performance on a held-out reference set, subgroup performance where the use profile makes it relevant, and rates of human override and escalation. A threshold breach triggers review and can suspend the certificate. This is the mechanism that makes a twelve-month certificate meaningful for an artifact that could change next week, and it is also the mechanism that makes the whole scheme expensive, which the economics section takes up.

One class of system cannot be certified this way, and the proposal should say so. A system that updates continuously in production, with no stable configuration to bind a certificate to, has no certifiable artifact. For those systems the honest answer is that the certifiable object is the process rather than the artifact, which is what ISO/IEC 42001 already does through an accredited route, and this institution adds nothing. Naming that boundary costs the proposal some scope and buys it the credibility of having a boundary at all.

Who pays, and why that does not corrupt the mark

The institution proposes to be independent of AI vendors while being paid, as UL is, by the parties whose products it tests. That is the same structure that corrupted credit ratings, and a paper that does not confront the comparison has not made its case. Earlier editions noted approvingly that UL's manufacturers pay for testing, and separately ruled that no vendor may control the institution, without reconciling the two.

The fee structure. Fees are charged for testing, paid by the applicant, which may be the vendor or the deploying operator. As a design estimate rather than an observed cost, an initial certification of one deployment configuration against one use profile falls between 75,000 and 250,000 dollars depending on consequence tier, with annual renewal at 25 to 40 percent of initial and continuous-monitoring review included. The benchmark that matters is not the institution's cost but the buyer's alternative: a custom AI audit commissioned from a consultancy today costs a substantial fraction of that and produces a report no other buyer can use, and a FedRAMP authorization runs well into seven figures. If a certificate replaces private diligence at ten buyers, its price needs to sit nearer one buyer's diligence cost than ten, or nobody buys it.

Cross-subsidy is required, not optional. This paper argues elsewhere that a recognized mark lets a twelve-person company carry the trust credential of a trillion-dollar one. That argument fails if only the trillion-dollar company can pay, so a small-vendor tier priced below cost, funded from high-consequence-tier fees, is structural to the proposal rather than a courtesy.

Four structural answers to capture. Three are UL's and one is specific to this domain.

First, the institution owns the mark, and no client may license, sublicense, or use it outside the certified scope. Second, the institution earns no consulting or remediation revenue, ever, which is a bright line

that not every certification body in the current ISO 42001 market observes. Third, no single client exceeds a stated share of annual revenue, published annually, because a testing body dependent on one applicant will find reasons to pass it. Fourth, and decisively, **the test methods are public.** Anyone holding the artifact and the published method can rerun the test.

That fourth point is the answer to the credit-rating comparison, and it is a structural difference rather than a promise of better behavior. Issuer-pays corrupted ratings because a rating was an opinion about an unobservable future that could not be checked at the time it was issued, so a corrupted rating and an honest one were indistinguishable until the underlying assets failed. A conformance test result is a measurement against a published procedure, and it is reproducible immediately by any third party with access to the system. Publishing methods converts an opinion into a falsifiable claim, which changes the incentive rather than merely regulating it. It follows that the institution's methods must be public even where publishing them makes the tests easier to game, and that the residual gaming problem is handled by holding out reference sets rather than by keeping methods secret.

The incumbent path, and why this is a scheme rather than a rival

---

The cheapest available answer to everything in this paper is that the existing accreditation stack should simply be extended. That objection is strong, earlier editions never stated it, and engaging it properly narrows this paper's claim.

What the incumbent stack already does. ISO/IEC 42006 sets the competence and procedural requirements that accreditation bodies use to qualify certification bodies and their auditors [19]. ANAB operates the United States accreditation program for ISO/IEC 42001 certification bodies, and has accredited Schellman, SGS, DEKRA, DQS, and BSI among others [28][29][30]. UKAS performs the corresponding function in the United Kingdom, and the International Accreditation Forum's mutual-recognition machinery is what allows an accredited certificate issued in one jurisdiction to be recognized in another. This is a functioning, internationally connected conformity-assessment system with decades of institutional practice behind it, and it did not need to be invented for AI.

What it does not do, precisely. Three things.

The unit of certification is the management system. An accredited ISO/IEC 42001 certificate attests that an organization has established, implemented, and maintains a conforming AI management system. It does not attest that a named model, in a named configuration, serving a named population, performs within stated bounds, and no amount of accreditation rigor applied to management-system auditing produces that attestation, because it is a different claim about a different object.

There is no shared conformance-testing infrastructure. Accreditation verifies that an auditor is competent and that a certification body's procedures are sound. It does not supply validated test methods for measuring the behavior of a deployed model, and in their absence each body that attempts behavioral evaluation builds its own, which reproduces at the method level the fragmentation that accreditation solved at the auditor level.

There is no public registry with fixed scope semantics. Certificates exist, and a buyer cannot query one place to learn what a given certificate covers, whether it is current, or whether it has ever been suspended.

The consequence, which revises this paper's earlier claim. The right structure is not a new accreditation hierarchy competing with ANAB and the IAF, which would fragment recognition in exactly the

way this paper complains about and would take a decade to earn international standing. It is a new **conformance-assessment scheme** for deployed AI system behavior, operated by an independent institution that seeks accreditation within the existing architecture rather than around it, and whose distinctive contributions are the three things listed above: published and validated test methods, a defined certification unit with a lifecycle, and a registry.

That is a materially smaller claim than "the United States needs a new national AI assurance institution," and it is the claim the evidence supports. It also makes the proposal cheaper, faster, and far more likely, because it inherits recognition instead of building it. Readers of earlier editions should treat this as a correction rather than a refinement.

Blueprint principles

Earlier editions opened this section by declining to specify institutional design and then offered eight principles, of which six restated analysis already presented. The principles below are limited to those that constrain a design decision, and each is stated as a choice with an alternative rejected.

1. **Operate inside the existing accreditation architecture.** Seek accreditation from an established accreditation body rather than asserting authority independently. The rejected alternative, a parallel hierarchy, buys autonomy at the cost of the international recognition that makes a certificate portable.
2. **Certify configurations bound to use profiles, on twelve-month terms with significant-change triggers and continuous monitoring.** The rejected alternative, certifying models or organizations, is either meaningless or already available.
3. **Publish the test methods.** The rejected alternative, proprietary methods, protects revenue and forfeits the only structural defense against issuer-pays capture.
4. **Take no consulting or remediation revenue and cap single-client concentration, both published.** The rejected alternative, a mixed-revenue body, is the current market and is the reason a certificate's meaning varies by issuer.
5. **Anchor demand in risk-carriers and in state procurement, not in a statute.** Insurers price against the mark and states specify it in what they buy, which is authority no special session can repeal [13] and which federal preemption explicitly does not reach [5]. The rejected alternative, seeking a mandate, imports the political fragility Colorado demonstrated.
6. **Run a registry that records suspensions and withdrawals, not only grants.** The rejected alternative, publishing successes, produces a marketing asset rather than trust infrastructure, and a registry that quietly removes failures is worse than none because it misleads.
7. **Federate testing capacity, keep the credential singular.** The reason is specific rather than architectural fashion: conformance testing of deployed behavior requires domain-specific test environments, including clinical data under the governance that clinical data requires, financial data under its own, and sector-specific expertise in what constitutes an acceptable error. No single facility can hold those environments or that expertise, so capacity must be distributed to where the data and the domain competence already sit, while the criteria and the mark stay in one place. An earlier edition justified this by analogy to lab-embedded entrepreneurship programs, which share a topology with conformance laboratories and share nothing else, and that analogy is withdrawn.

A state-scale design study, and why its foundation failed

An earlier edition titled this section “Proof at state scale” and closed by saying that Colorado proved the model’s feasibility. That claim was too strong and the section is retitled accordingly. A concept paper that was never funded, never staffed, and never operated cannot establish that a conformance laboratory is buildable, and asserting otherwise is the move this paper condemns elsewhere when it observes that a standard by itself verifies nothing.

What the Colorado episode does establish is narrower and still useful.

The demand side is real and it appeared immediately. The Colorado AI Act of 2024 demonstrated a sequence: the moment a jurisdiction articulates duties for high-risk AI, deployers need exactly the three functions this paper describes. The law even anticipated the mechanism, since compliance with recognized risk frameworks could support a rebuttable presumption of reasonable care, which was a statutory bridge inviting certification against NIST-aligned criteria [11]. Businesses facing the act asked less what the law required than who could verify that they met it, and they found no one to answer.

A design study was produced, and its status should be stated plainly. In 2025, work in which the author participated produced a concept for the Center for AI Sustainable Advancement, a state-level assurance and access institution combining open-access compute and evaluation infrastructure, a conformance laboratory, and AI safety certification aligned to NIST’s framework, operated as a public-private consortium [27]. The reference is a concept white paper available from the author, which means a reader cannot check it, and a claim resting on an unverifiable source should carry no weight in an argument. What the work produced was an institutional design and a judgment that no scientific breakthrough was required. It produced no throughput figures, no staffing model, no cost per certification, and no validated test methods, which are precisely the things that would make feasibility checkable. The unit economics sketched earlier in this paper are the beginning of that work and are labeled as estimates for the same reason.

Then the foundation moved. The 2025 special session delayed the act [12] and the 2026 session repealed and replaced it with a disclosure-only regime [13], so the statutory demand driver dissolved eighteen months after enactment and before enforcement. A state-scoped institution would have leaned on the duty of care, the impact assessments, and the presumption mechanism, and all three are gone.

The episode therefore resolves a design question rather than proving a capability. Assurance infrastructure should not be built on a single state’s statute, because a statute is subject to annual revision and, since December 2025, to active federal preemption pressure [5]. Demand for assurance is national and market-driven and survived Colorado’s repeal intact, since every hospital and insurer that needed verification in 2025 still needs it in 2026. Colorado’s contribution was to demonstrate cheaply which foundation not to build on.

The economics: assurance as adoption infrastructure

“Safety” framing systematically understates the case. Assurance is not a brake on AI adoption but a missing component of it, and its absence currently rations deployment to organizations wealthy enough to self-verify.

The historical analogy is exact. Electricity did not scale because it became safer in the abstract; it scaled because UL's mark let insurers price electrification, priced risk let capital flow, and buyers could then adopt without each becoming electrical engineers [23]. The mark converted diffuse technological risk into a legible, priceable quantity, and every transformative infrastructure technology has required that conversion.

Why AI harms are underwritable at all

The standard objection to the insurance argument is that AI harm is diffuse, contested, and frequently unattributable, which is what kept software errors-and-omissions coverage thin for decades. If AI failure cannot be attributed, it cannot be priced, and the entire risk-carrier flywheel this paper depends on does not turn.

The structure of the one product in market suggests why the objection is weaker than it looks. aiSure underwrites on what the system does and how its outputs are used, drawing on measurable performance data in what is described as a parametric-like claims model, and it is explicitly distinct from cyber and technology errors-and-omissions cover [20]. A parametric structure pays on a measured deviation from agreed performance rather than on proof that a failure caused a particular harm, which sidesteps attribution entirely. That is the mechanism that makes AI performance priceable where software defect liability was not: the insured event is defined as a measurement rather than as a causal chain.

The dependency runs in both directions and the paper should say so. Parametric coverage requires an agreed measurement, and an agreed measurement is exactly what a conformance-testing institution would supply. So the insurance market cannot scale without the assurance institution, and the assurance institution's demand case rests on the insurance market. That circularity is real. It resolves in one direction only, because the measurement standard has to exist before coverage can be written against it at scale, which is an argument for building the institution and simultaneously an admission that the insurance demand cited as evidence here is currently one product, from two carriers, with a 15 million dollar limit. That is an existence proof and not a market, and readers should weigh it as such.

Three further effects

Assurance is a startup subsidy in competitive clothing. Enterprise AI procurement currently defaults to incumbent vendors because brand substitutes for verification, and no buyer gets fired for choosing the large laboratory. A recognized certification lets a twelve-person company carry the same credential as a trillion-dollar one, and analogous marks historically lowered barriers to entry in electrical equipment and cloud services alike [23][24]. The assurance gap is a moat for incumbents and an assurance institution is a moat-leveler, which is also why the small-vendor fee tier is structural rather than charitable.

Certification collapses transaction costs on both sides. Enterprise sales cycles currently spend months on custom security and AI review, which the vendor pays for in sales cost and the buyer pays for again in diligence cost, repeated for every vendor-buyer pairing and leaving no residual value to anyone else. Authorize-once-use-many converts that repeated bilateral cost into a single amortized one [24].

The patchwork makes assurance more valuable rather than less. State regimes diverge [13][14][15] and federal preemption is contested [5]. A credential anchored in recognized technical standards rather than in any single statute is the only compliance asset that survives every legislative outcome, so regulatory volatility, usually quoted as a cost of this moment, is the value proposition.

How this fails

An institution described only through its benefits has not been designed. Seven failure modes, in rough order of likelihood.

Insufficient demand, which is the load-bearing risk. If insurers do not price against the mark and buyers do not specify it, nothing else in this paper matters. The evidence for that demand today is one insurance product and a nascent audit market, and the honest statement is that the central assumption of this paper is supported by the thinnest evidence in it.

The mark becomes a process checkbox. This is the failure the paper attributes to management-system certification, and the new institution can reproduce it by certifying that a test was run rather than what the test found. The mitigation is that certificates state measured outcomes, so that a certificate a reader cannot parse as "this configuration achieved X on method Y for population Z" has already failed.

A race to the bottom among accredited testing laboratories. Federated capacity means laboratories compete, and in a market for certificates the way to win business is speed and leniency. The specific mitigation is proficiency testing, borrowed from clinical laboratories and metrology: the institution submits blind reference systems with known properties to each laboratory and publishes laboratory-level agreement rates. A laboratory whose results diverge is visible before its certificates are relied upon.

Certifier capture. The four structural defenses above reduce this and do not eliminate it. The residual exposure is concentration in a thin market, because in the early years a small number of large applicants will represent most revenue, and the published concentration cap will bind hardest exactly when the institution can least afford to lose a client.

Liability when a certified system causes harm. The institution will be sued, and the first suit will define its posture if the institution has not. The mark is not a warranty of outcomes and states what was measured under what method, and that position needs to be documented, insured, and communicated before the first certificate issues rather than after the first incident.

A certified system fails publicly, or its vendor is breached. Suspension and withdrawal procedures must exist, be published, and be fast, and the registry must show withdrawn and suspended certificates rather than removing them. An institution whose registry only shows current grants is concealing its own error rate.

Success converts into regulation. If the mark becomes a de facto requirement, pressure will follow to make it a statutory one, at which point the institution acquires precisely the political fragility that Colorado demonstrated and that this paper's whole design is meant to avoid. The defense is a governance commitment against seeking enforcement authority, made early, when it is cheap.

What happens next

Institutions like this are not legislated into existence and are not incubated inside any single company. They are convened, and the practical question is who sits at the table first.

The candidates disqualify themselves in an instructive order. The federal government has organized itself around voluntary standards and frontier evaluation [2][3], and its present energy is directed at preemption rather than construction [5]; federal convening would be welcome and waiting for it is

a decision to wait indefinitely. Standards bodies cannot convene it, because an institution certifying against a standard must be independent of the body writing the standard, or the conflict NTSB was built to escape is reproduced at the foundation [25]. Individual AI vendors cannot convene it for the mirror reason, since a mark controlled by a producer is marketing. Consultancies sell remediation, and the accreditation function cannot originate inside firms whose revenue depends on the findings.

That leaves the parties whose money sits where the risk is.

- **Insurers and reinsurers.** They already underwrite AI performance and need shared verification to scale it [20], and they were the founding constituency of UL [23].
- **Institutional buyers.** Hospital systems, financial institutions, and universities pay the duplicated-diligence tax today and would specify a recognized mark the moment one existed.
- **Research institutions.** They hold the evaluation science and the public-interest credibility, and their distributed laboratories are the natural federated testing capacity.
- **State and local procurement offices.** This is the addition earlier editions forfeited by omitting the carve-out in EO 14365. States retain explicit authority over their own AI procurement and use notwithstanding the preemption campaign [5], which makes them the one demand channel that federal action does not reach. A state that cannot impose duties on private deployers can still require a recognized certification in everything it buys, and state and local government is a large enough buyer to move a market on its own. For a jurisdiction like Colorado, which has just lost its regulatory instrument, procurement is the instrument that remains, and it needs no new legislation to use.
- **Investors.** They carry AI risk across portfolios, and for them assurance is both protection and the unlock on the next generation of deployable AI companies. This paper's publisher is one of them, which is addressed directly below.

The sequence has run before. Convene the risk-carriers, adopt the standards that exist, seek accreditation within the architecture that exists, stand up testing against a first sector's use cases where the deployment test bites hardest, publish the methods, issue the first certificates, and let the registry, the insurers, and the state procurement offices do the rest. Underwriters Laboratories began with two employees and a mark the market learned to demand [23].

Disclosure

Stout Street Capital is a deep-technology venture investor and is not a disinterested party to this argument.

The interest is specific rather than atmospheric. If the institution described here is built, the market for AI evaluation and audit tooling becomes materially larger, and companies supplying that tooling become more valuable. This firm invests in early-stage deep technology and would be a plausible investor in such companies. The third paper in this series applies an evaluation framework to a venture in exactly that category, and the scenario in which that venture scores best is the scenario in which this paper's institution exists. A reader is entitled to note that the three papers form a mutually reinforcing set published by a single interested firm, and to discount accordingly.

Two things are worth saying alongside that. The argument here does not depend on the firm's interest being served, and its load-bearing claims are checkable against sources that have nothing to do with us: the deployment test can be run by any reader against any vendor, the accreditation roster is public, the statutory history is public, and the insurance product either exists at the stated limit or does not. Where this paper's evidence is thin, as with the demand case resting on one insurance product, it now

says so, and where an earlier edition overclaimed, as with state-scale feasibility and the accreditation layer, this edition retracts the claim rather than softening it.

The alternative to disclosure is not neutrality. It is an undisclosed interest, which a reader will eventually find and will reasonably treat as worse than the interest itself.

About this series

Stout Street Capital publishes this paper as №01 of a three-part series on the institutions the deep tech era requires.

- **№01. The Assurance Gap** (this paper). The case for an independent conformance-assessment institution for deployed AI systems: test methods, a defined certification unit, and a public registry, operating inside the existing accreditation architecture.
- **№02. LEVCAP.** A lab-embedded venture capital access program: the commercialization engine that moves deep technology from laboratory to market.
- **№03. A Holistic Evaluation Framework for Innovation and Technology Commercialization.** The evaluation instrument: a refined framework that joins technical readiness with commercial metrics to judge what is real.

References

1. Broadband Breakfast, “AI Safety Institute Renamed Center for AI Standards and Innovation,” June 6, 2025. https://broadbandbreakfast.com/ai-safety-institute-renamed-center-for-ai-standards-and-innovation/
2. National Institute of Standards and Technology, “Center for AI Standards and Innovation (CAISI),” undated agency page, accessed July 28, 2026. https://www.nist.gov/caisi
3. Cloud Security Alliance, “CSA Research Note: NIST CAISI AI Agent Standards,” February 17, 2026. https://labs.cloudsecurityalliance.org/research/csa-research-note-nist-caisi-ai-agent-standards-compliance-2/
4. National Institute of Standards and Technology, “AI Risk Management Framework,” January 26, 2023. https://www.nist.gov/itl/ai-risk-management-framework
5. The White House, Executive Order 14365, “Ensuring a National Policy Framework for Artificial Intelligence,” December 11, 2025. Carve-outs from preemption include child safety, AI compute and data-center infrastructure, and state government procurement and use of AI. https://www.whitehouse.gov/presidential-actions/2025/12/eliminating-state-law-obstruction-of-national-artificial-intelligence-policy/
6. Hunton Andrews Kurth, “OMB Issues Revised Policies on AI Use and Procurement by Federal Agencies” (OMB M-25-21 and M-25-22, April 3, 2025). https://www.hunton.com/privacy-and-cybersecurity-law-blog/omb-issues-revised-policies-on-ai-use-and-procurement-by-federal-agencies
7. European Commission, “Regulatory framework on AI,” page last updated July 27, 2026. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
8. Ibid. Digital Omnibus deferral of Annex III high-risk obligations to December 2, 2027 (politically agreed May 7, 2026).
9. Ibid. European AI Office role.

10. Infosecurity Magazine, “UK AI Safety Institute Rebrands as AI Security Institute,” February 14, 2025, https://www.infosecurity-magazine.com/news/uk-ai-safety-institute-rebrands/; AI Security Institute, https://www.aisi.gov.uk/, accessed July 28, 2026.
11. Colorado General Assembly, SB 24-205, “Consumer Protections for Artificial Intelligence,” signed May 17, 2024. https://leg.colorado.gov/bills/sb24-205
12. Colorado General Assembly, SB 25B-004 (special session), signed August 28, 2025. https://leg.colorado.gov/bills/sb25b-004
13. Colorado General Assembly, SB 26-189, “Automated Decision-Making Technology,” signed May 14, 2026, https://leg.colorado.gov/bills/sb26-189; Troutman Pepper, “Colorado Legislature Passes Bill to Repeal and Replace Colorado AI Act,” May 2026. https://www.troutmanprivacy.com/2026/05/colorado-legislature-passes-bill-to-repeal-and-replace-colorado-ai-act/
14. Norton Rose Fulbright, “The Texas Responsible AI Governance Act” (HB 149, effective January 1, 2026), December 2025. https://www.nortonrosefulbright.com/en/knowledge/publications/c6c60e0c/the-texas-responsible-ai-governance-act
15. Future of Privacy Forum, “California’s SB 53: The First Frontier AI Law, Explained,” October 3, 2025. https://fpf.org/blog/californias-sb-53-the-first-frontier-ai-law-explained/
16. National Conference of State Legislatures, “New Trends Emerge as States Refine AI Legislation,” January 22, 2026. https://www.ncsl.org/resources/details/new-trends-emerge-as-states-refine-ai-legislation
17. MultiState, “Artificial Intelligence (AI) Legislation” tracker, accessed July 28, 2026 (1,561 bills across 45 states as of March 2026). https://www.multistate.ai/artificial-intelligence-ai-legislation
18. KPMG Australia, “KPMG first company to achieve AI Management System standard certification by BSI” (ISO/IEC 42001:2023), October 17, 2024. https://kpmg.com/au/en/media/media-releases/2024/10/kpmg-first-company-to-achieve-ai-management-system-standard-certification-by-bsi.html
19. GDPR Local, “ISO/IEC 42006:2025, Requirements for bodies providing audit and certification of AI management systems” (published July 11, 2025). https://gdprlocal.com/iso-iec-42006-2005/
20. Reinsurance News, “Mosaic and Munich Re introduce AI-specific insurance for developers” (aiSure, up to $15M coverage, underwritten on measured system behavior in a parametric-like claims model, distinct from cyber and technology errors-and-omissions cover), February 27, 2026. https://www.reinsurancene.ws/mosaic-and-munich-re-introduce-ai-specific-insurance-for-developers/
21. Fact.MR, “AI Agent Audit and Assurance Services Market,” June 18, 2026. Directional vendor estimate; treat market-size figures as indicative. https://www.factmr.com/report/ai-agent-audit-and-assurance-services-market
22. Hunton Andrews Kurth, op. cit. OMB M-25-22 procurement provisions.
23. UL Research Institutes, “Our History,” accessed July 28, 2026. https://ul.org/about/our-history/
24. Kiteworks, “What is FedRAMP?” (Risk & Compliance Glossary), accessed July 28, 2026. https://www.kiteworks.com/risk-compliance-glossary/fedramp/
25. National Transportation Safety Board, “History of the NTSB,” accessed July 28, 2026. https://www.ntsb.gov/about/history/Pages/default.aspx
26. National Transportation Safety Board, “About the NTSB,” accessed July 28, 2026. https://www.ntsb.gov/about/Pages/default.aspx
27. Stout Street Capital, “Center for AI Sustainable Advancement (CASA),” concept white paper, 2025. Available from the author. Not independently verifiable by the reader; treated in this edition as a design study rather than as evidence of feasibility.

28. ANSI National Accreditation Board, “Artificial Intelligence Management Systems, ISO/IEC 42001 Certification Bodies,” accreditation program page, accessed August 3, 2026. https://anab.ansi.org/accreditation/iso-iec-42001-artificial-intelligence-management-systems/
29. Schellman, “Schellman Becomes First ISO 42001 ANAB Accredited Certification Body,” accessed August 3, 2026. https://www.schellman.com/blog/news/schellman-becomes-1st-iso-42001-anab-accredited-certification-body
30. BSI, “BSI secures ANAB accreditation to certify ISO/IEC 42001,” March 2026, accessed August 3, 2026 (BSI holding accreditation from UKAS, RvA, and ANAB). https://www.bsigroup.com/en-US/insights-and-media/media-center/press-releases/2026/march/bsi-secures-anab-accreditation-to-certify-isoiec-42001/

Recorded at Denver, Colorado · MMXXVI