The Verification Gap

Why Independent Assurance Will Decide Whether Artificial Intelligence Earns Institutional Trust

Mubien Ahsan

July 2026

1. Introduction

Every era of business gets the assurance industry it needs. Railway booms produced the chartered accountant; the corporate scandals of the early 2000s produced Sarbanes–Oxley and an industry of internal control attestation; the cloud era produced SOC 2 reports as the price of admission to enterprise procurement. Each time, the pattern is the same: a technology or practice becomes economically indispensable before it becomes verifiably trustworthy, and a market emerges for independent parties who can close that gap.

Artificial intelligence has now reached exactly this point. Organizations are deploying AI in credit decisions, hiring, medical triage, fraud detection and customer service—domains where errors carry legal, financial and human consequences—while boards, regulators, insurers and customers have no standardized way to know whether a given system is safe, fair, secure or even effective. The response is the emergence of *AI assurance*: independent evaluation and, increasingly, formal certification of AI systems and the management structures around them. The United Kingdom's government has sized this market at just over £1 billion across 524 firms in 2024, with a projection of up to £18.8 billion by 2035 if quality and skills barriers are addressed—and has published a national roadmap to build it deliberately.[1] The Big Four professional services firms are racing to define and capture the category, seeing in it the most significant new assurance franchise since internal control attestation.[2]

The views expressed in this essay are the author's own and do not represent the position of any employer or organization. Researched and drafted in collaboration with AI; all claims were independently verified by the author against the primary sources cited.

[1] UK Department for Science, Innovation and Technology, *Trusted Third-Party AI Assurance Roadmap* (September 2025); market figures per DSIT's November 2024 study *Assuring a Responsible Future for AI*, summarized in Burges Salmon, "AI assurance: the UK market and government actions," `https://www.burges-salmon.com/articles/102jph1/ai-assurance-the-uk-market-and-government-actions-dsit-report/`, and the Written Ministerial Statement of 3 September 2025, `https://questions-statements.parliament.uk/written-statements/detail/2025-09-03/hcws903`.

[2] See reporting originating with the *Financial Times*, summarized in City A.M., "Big Four giants eye AI assurance tools" (June 2025), `https://www.cityam.com/big-four-giants-eye-ai-assurance-tools-to-tap-in-on-client-concerns/`, and Consultancy.uk, "Big Four weighing up new AI audit offerings" (June 2025), `https://www.consultancy.uk/news/40424/big-four-weighing-up-new-ai-audit-offerings`.

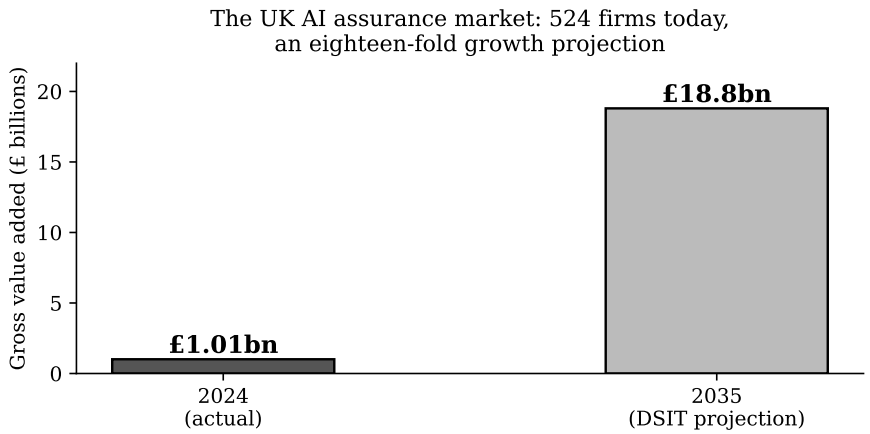

Figure 1. The UK government's sizing of its third-party AI assurance market: £1.01 billion across 524 firms in 2024, projected at up to £18.8 billion by 2035 (DSIT, 2024–2025).

This essay examines AI assurance as an emerging discipline rather than a marketing category. It asks four questions. What, precisely, is being assured when a firm "audits" an AI system? What is driving demand—and how durable are those drivers, given that the flagship regulatory deadline in Europe has just been pushed back? Who is competing to supply assurance, and with what credibility? And, most importantly, what are the unsolved methodological problems—criteria, evidence, competence, independence—that will determine whether AI assurance matures into a discipline with the standing of financial audit, or degenerates into a rubber-stamping exercise that manufactures false confidence at scale? The analysis is anchored in Canada's regulatory position, particularly OSFI's Guideline E-23 on model risk management, with comparison to the European Union, the United Kingdom and the United States.

2. What AI Assurance Actually Is

The term "AI assurance" is used loosely, and the looseness is itself a market problem. It helps to distinguish three layers of activity that are frequently conflated.

The first layer is *AI governance consulting*: helping organizations design policies, inventories, risk assessments and oversight structures for their AI use. This is advisory work—valuable, but it produces recommendations, not independent conclusions, and the provider works for management.

The second layer is *management system certification*: independent audit of an organization's AI *governance apparatus* against a published standard, most prominently ISO/IEC 42001:2023, the first international standard for AI management systems, which specifies re-

quirements for establishing, implementing, maintaining and continually improving how an organization governs its AI.[3] Accredited certification bodies now operate globally, and—notably for this essay's context—a Big Four member firm in Switzerland became one of the first accredited certification bodies in Europe for ISO/IEC 42001, an early signal of how the largest professional services networks intend to institutionalize the offering.[4] Crucially, certifying a management system says the organization has sound *processes*; it does not, by itself, say any particular AI system is safe or accurate.

The third layer—the frontier—is *system-level assurance*: independent examination of a specific AI system's properties, such as accuracy on defined tasks, bias across protected groups, robustness to adversarial inputs, security of the model and its data pipeline, and the operating effectiveness of human oversight. PwC's US firm moved first among the Big Four in mid-2025, announcing independent assurance over whether AI systems have been designed, deployed and operated responsibly, transparently and within regulatory expectations, delivered by teams pairing machine learning specialists with attest professionals.[5] Deloitte has publicly described AI assurance as critical to adoption, arguing that companies will demand assurance over AI managing their critical functions, while EY has acknowledged its own offering remains further from market.[6]

The reason this layered distinction matters is that public trust will attach to the strongest claim in circulation. If "AI-audited" comes to mean anything from a two-week policy review to a rigorous system examination, the market will price all of it at the credibility of the weakest version. The UK government's roadmap identifies precisely this problem—inconsistent quality and fragmented terminology—as the central obstacle to market growth, and one study it cites found that 38% of assessed AI governance tools incorporated metrics problematic enough to potentially cause harm.[7]

---

[3] ISO/IEC 42001:2023, *Information technology — Artificial intelligence — Management system*; on the accreditation landscape see BSI, `https://www.bsigroup.com/en-US/products-and-services/standards/iso-42001-ai-management-system/`, and Schellman (first ANAB-accredited certification body), `https://www.schellman.com/services/ai-governance/iso-42001`.

[4] See, e.g., "ISO/IEC 42001: AI Management System for Governance," `https://kpmg.com/ch/en/insights/artificial-intelligence/iso-iec-42001.html`.

[5] Accounting Today, "PwC offering independent assurance on AI systems" (June 2025), `https://www.accountingtoday.com/news/pwc-offering-independent-assurance-on-ai-systems`.

[6] City A.M. (June 2025), cited above.

[7] DSIT roadmap analysis and the World Privacy Forum finding, as summarized in Burges Salmon (2025), cited above; on the roadmap's quality-assurance models, see ICAEW, "Trusted AI assurance roadmap outlines three quality assurance models," `https://www.icaew.com/insights/viewpoints-on-the-news/2025/sep-2025/trusted-ai-assurance-roadmap-outlines-three-quality-assurance-models`.

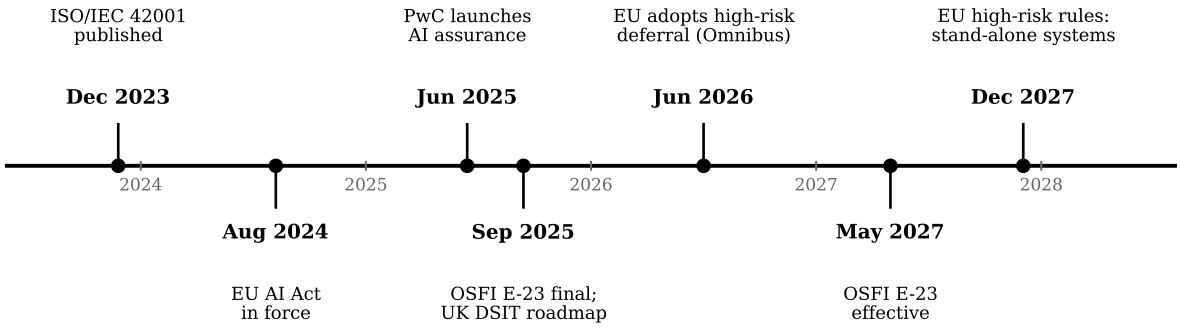

Figure 3. The compressed arc of the AI assurance market: from the first international AI management standard to binding prudential deadlines in under five years. The December 2027 EU date for stand-alone high-risk systems reflects the Digital Omnibus amendments given final approval by the Council on 29 June 2026.

3. Demand: Why Organizations Will Pay to Be Checked

3.1 Regulation—the compliance engine, with a complication

The EU Artificial Intelligence Act is the demand driver most often cited, and with reason: it is the world's first comprehensive horizontal AI law, imposing on providers of high-risk systems a battery of obligations—risk management systems, data governance, technical documentation, logging, human oversight—capped by *conformity assessment* before market placement and penalties reaching EUR 35 million or 7% of global turnover.[8] Conformity assessment is assurance by another name: an evidence-based judgment that a system meets defined requirements, in some cases performed by accredited third-party "notified bodies."

Honesty requires acknowledging the complication. The high-risk obligations were scheduled to bite on 2 August 2026; instead, through the Digital Omnibus on AI—politically agreed in May 2026, endorsed by the European Parliament on 16 June and given final approval by the Council on 29 June 2026—the EU has deferred obligations for stand-alone high-risk systems to 2 December 2027, and for AI embedded in regulated products to 2 August 2028. Notably, the August 2026 dates for transparency obligations and general-purpose AI enforcement remain in force; only the hardest tier moved.[9] Superficially the deferral looks

[8]Regulation (EU) 2024/1689; obligations and penalty structure summarized in Legiscope, "EU AI Act Deadlines 2026–2027," `https://www.legiscope.com/blog/eu-ai-act-timeline-deadlines.html`, and Cloud Security Alliance, "EU AI Act High-Risk Deadline: Enterprise Readiness Gap" (March 2026), `https://labs.cloudsecurityalliance.org/research/csa-research-note-eu-ai-act-high-risk-compliance-deadline-20/`.

[9]Council of the European Union, "Artificial intelligence: Council gives final green light to simplify and streamline

like demand deferred. In practice it changes timing more than trajectory: enterprises with multi-year AI investments cannot build compliance architecture in the final quarter before a deadline (realistic conformity preparation runs eight to fourteen months even with mature governance),[10] and notified-body assessment capacity was already a documented bottleneck well before the revision, with assessment slots booking out quarters in advance—evidence that the assurance supply side, not client appetite, has been the binding constraint.

Canada's demand engine is quieter but, for financial services, arguably more certain. OSFI's final Guideline E-23 on Model Risk Management, published in September 2025 and effective 1 May 2027, extends enterprise-wide model risk expectations to *all* models at *all* federally regulated financial institutions—explicitly including AI and machine learning, explicitly covering third-party and vendor models, and explicitly demanding that the maturity of governance scale with the sophistication of the techniques used.[11] Unlike the EU's product-regulation model, E-23 does not mandate third-party certification; it mandates internal capability—validation independent of development, lifecycle governance, documented accountability. But regulatory expectations of this kind reliably generate external demand through two channels: institutions buying expertise they cannot staff, and boards seeking independent comfort that the framework their management attests to actually functions. Canada's parallel policy work—the Financial Industry Forum on AI convened by OSFI, the Department of Finance and the Global Risk Institute—has catalogued the risk surface (data poisoning, model extraction, adversarial manipulation, AI-enabled cyber threats) that such assurance must address.[12]

---

rules" (29 June 2026), https://www.consilium.europa.eu/en/press/press-releases/2026/06/29/artificial-intelligence-council-gives-final-green-light-to-simplify-and-streamline-rules/; on the two-tier deferral and surviving August 2026 obligations, see Covington, "EU AI Act Update: Timeline Relief, Targeted Simplification, and New Prohibitions" (May 2026), https://www.insideprivacy.com/artificial-intelligence/eu-ai-act-update-timeline-relief-targeted-simplification-and-new-prohibitions/, and Travers Smith (8 May 2026), https://www.traverssmith.com/knowledge/knowledge-container/eu-agrees-to-delay-key-ai-act-compliance-deadlines/.

[10] Modulos, "352 Days to Compliance" (2025), https://www.modulos.ai/blog/eu-ai-act-high-risk-compliance-deadline-2026/.

[11] OSFI, *Guideline E-23 – Model Risk Management (2027)*, https://www.osfi-bsif.gc.ca/en/guidance/guidance-library/guideline-e-23-model-risk-management-2027; OSFI Backgrounder (September 2025), https://www.osfi-bsif.gc.ca/en/news/backgrounder-guideline-e-23-model-risk-management; legal analysis in Blakes, https://www.blakes.com/insights/osfi-releases-final-guideline-e-23-for-model-risk-management-and-ai-use-by-frfis/, and Torys, https://www.torys.com/en/our-latest-thinking/publications/2025/10/osfi-updates-and-expands-scope-of-guideline-e-23.

[12] OSFI, Department of Finance Canada and Global Risk Institute, *Financial Industry Forum on Artificial Intelligence* (2025), as summarized in Protiviti, https://www.protiviti.com/gl-en/insights-paper/strengthening-decision-making-with-osfi-e-23-model.

3.2 The patchwork effect

The United States, by contrast, has no horizontal federal AI statute, and its posture has oscillated with administrations—yet this absence generates rather than suppresses assurance demand, through a mechanism worth naming: the *patchwork effect*. State-level regimes, sectoral regulators applying existing law to AI (fair lending, employment discrimination, unfair and deceptive practices), and enforcement actions collectively confront national firms with overlapping, non-identical obligations. When requirements diverge, the rational enterprise response is to build to a defensible common denominator and evidence it once—and independent assurance against a recognized framework such as the NIST AI Risk Management Framework or ISO/IEC 42001 is precisely that evidence. The same dynamic operates internationally: a Canadian bank subject to OSFI E-23, serving EU clients within reach of the AI Act, and procuring from US vendors, cannot run three unrelated compliance programs. It runs one governance framework and cross-walks it—and practitioner analysis already treats the E-23/ISO 42001/EU AI Act cross-walk as a standard supervisory artifact, with mature model-risk programs covering the substantial majority of E-23 once AI-specific treatment is added.[13] Fragmentation, in short, is not an obstacle to the assurance market; it is one of its most reliable engines, because independent verification is the cheapest portable proof across jurisdictions.

3.3 Commercial and reputational drivers

Regulation is the floor, not the ceiling, of demand. Three market mechanisms operate independently of any statute. First, *procurement*: exactly as SOC 2 became a de facto requirement for selling software to enterprises, AI-intensive vendors are beginning to treat independent certification as a sales asset—the AI video company Synthesia's pursuit of ISO 42001 certification, motivated explicitly by differentiation and regulatory anticipation, is an early template.[14] Second, *insurance and liability*: insurers have begun offering coverage for AI malfunction losses, and underwriting discipline will eventually demand the same evidence assurance produces.[15] Third, *internal confidence*: boards approving AI capital expenditure increasingly want an answer to the blunt question—does this actually work?—that neither the vendor nor the sponsoring executive can answer disinterestedly. The FT-reported framing of the Big Four opportunity was precisely this: clients want proof their AI systems are

[13] On regime cross-walks and overlap between SR 11-7-style model risk programs, ISO/IEC 42001 and the EU AI Act relative to E-23, see VerifyWise, "OSFI Guideline E-23: how AI and ML models fit the new regime" (June 2026), `https://verifywise.ai/blog/osfi-e-23-ai-model-risk-management-canada`.

[14] A-LIGN, "Understanding ISO 42001" (2025), `https://www.a-lign.com/articles/understanding-iso-42001`.

[15] Consultancy.uk (June 2025), cited above.

effective, not merely compliant.[16]

4. Supply: The Race to Own the Category

The competitive field has four camps. The Big Four bring board access, attest credibility, global delivery and—not to be underestimated—the muscle memory of having industrialized a novel assurance product once before, under Sarbanes–Oxley. Accredited certification bodies (BSI, TÜV SÜD, DNV, SGS, Schellman and others) bring the ISO machinery: accreditation, surveillance audit discipline, and international recognition of certificates.[17] Specialist AI governance and testing firms (Credo AI, Holistic AI, red-team boutiques) bring technical depth—continuous monitoring, bias measurement, adversarial testing—that generalist auditors currently lack; market analysis characterizes competition as centring on evidence quality and the repeatability of methods.[18] And governments are acting as market-makers: the UK roadmap commits to professionalizing the field through a code of ethics, a skills framework, defined information-access expectations for assurers (system boundaries, inputs and outputs, model parameters, change-management mechanisms), and an £11 million innovation fund.[19]

For the Big Four specifically, the strategic logic is double-edged. AI assurance is a natural extension of the trust franchise—and it is also a defensive necessity, because the firms are simultaneously embedding AI agents into their own audit delivery (multi-agent audit platforms with built-in decision logging, launched across the Big Four in 2025, are representative),[20] which means their credibility in assuring clients' AI will be judged partly by the demonstrable governance of their own. A firm that cannot evidence control over its internal agents will struggle to sell control evaluation as a product. This reflexivity is new: no prior assurance product required the auditor to be a sophisticated operator of the very technology under audit.

---

[16] City A.M. (June 2025), cited above.

[17] See the certification-body landscape at BSI, TÜV SÜD (`https://www.tuvsud.com/en-us/services/auditing-and-system-certification/iso-iec-42001-artificial-intelligence-management-system`), DNV and SGS, all cited above.

[18] Fact.MR, *AI Agent Audit & Assurance Services Market* (2026), `https://www.factmr.com/report/ai-agent-audit-and-assurance-services-market`.

[19] DSIT roadmap commitments as analyzed in Burges Salmon, "Trusted Third-Party AI Assurance – DSIT Roadmap," `https://www.burges-salmon.com/articles/102l5fo/trusted-third-party-ai-assurance-dsit-roadmap/`, and ICAEW (September 2025), cited above.

[20] On the Big Four's agent platforms, see Unity Connect, "The Big 4 AI Agents of 2025," `https://unity-connect.com/our-resources/blog/big-4-ai-agents/`.

5. The Hard Problems

The commercial narrative is straightforward; the professional one is not. Four unsolved problems will determine whether AI assurance earns the standing its promoters claim.

The criteria problem. Assurance is only as meaningful as the benchmarks the subject matter is measured against. Professional standards require criteria to be objective, complete, relevant and available to users—attributes codified in the attestation literature for decades.[21] For financial statements, criteria took a century to stabilize. For AI, they are contested at the root: there is no single accepted operationalization of "fairness" (indeed, several statistical fairness definitions are mutually incompatible), no consensus threshold for "acceptable" hallucination rates, and sector standards are still in development. An opinion that a system is "responsible" without disclosed, testable criteria is marketing, not assurance. The realistic near-term path is criteria pluralism with radical transparency: engagements anchored to named frameworks—ISO/IEC 42001 controls, the NIST AI Risk Management Framework's functions, EU AI Act article-level requirements, OSFI E-23 lifecycle expectations—with the practitioner's report stating exactly which benchmarks were applied and which properties were *not* examined.

The evidence problem. Financial audit examines a closed population of past transactions; AI assurance examines a system whose behaviour is probabilistic, input-dependent and—where models are retrained or fine-tuned—non-stationary. A test passed in March may not describe the system in June. Practitioner commentary has captured this crisply: auditors of AI are not checking one version of anything; they are chasing a moving target.[22] The methodological consequences are significant. Point-in-time opinions must be scoped honestly (a conclusion about a model version and configuration, not about "the AI"); change management and retraining controls become as central to the opinion as the model's measured performance, because they are what make the opinion durable; and the long-run product is likely continuous assurance—automated monitoring of drift, performance and incident indicators between formal examinations—rather than the annual-cycle model inherited from financial reporting.

The competence and independence problem. Credible system-level assurance requires a genuinely hybrid bench: attest professionals who understand evidence, materiality and

[21] See AICPA/PCAOB attestation standards on suitable criteria, AT §101, `https://pcaobus.org/oversight/standards/attestation-standards/details/AT101`; internationally, ISAE 3000 (Revised) defines criteria as the benchmarks used to evaluate the subject matter, `https://www.ifac.org/_flysystem/azure-private/publications/files/ISAE%203000%20Revised%20-%20for%20IAASB.pdf`.

[22] The Finance Story, "Big 4 firms introduce a new kind of audit" (2025), `https://thefinancestory.com/big-4-to-launch-ai-assurance-services`.

reporting discipline, alongside machine learning engineers who can interrogate training data lineage, evaluation design and adversarial robustness. Neither community alone suffices, and both are scarce—the UK counts roughly 12,500 people employed across its entire AI assurance market, against demand projected to grow eighteen-fold.[23] Independence pressures are equally structural. The same firms selling AI assurance also sell AI implementation, and the profession has lived through the consequences of that combination before; the emerging conflict rules of the EU AI Act's notified-body regime and the UK roadmap's professionalization agenda are early attempts to draw the boundary. There is also a subtler threat: firms using AI tools to *perform* assurance over AI systems must guard against correlated blind spots, where the evaluating model shares failure modes with the evaluated one.

The expectations-gap problem. Financial audit's oldest pathology—users believing an opinion guarantees more than it does—arrives in AI assurance amplified. A certificate that an organization's AI *management system* conforms to ISO/IEC 42001 will be read by procurement teams, journalists and juries as "this company's AI is safe." The profession's own skeptics have made the point bluntly: absent agreement on what fair AI even means, AI audits risk becoming another compliance checkbox.[24] Managing this gap is not a communications afterthought; it is a design requirement for reports—plain-language scope, named exclusions, explicit statements of what a reasonable-assurance conclusion on a probabilistic system can and cannot mean.

6. What Credible AI Assurance Will Look Like

Synthesizing the regulatory architecture and the methodological constraints, a defensible engagement architecture is already visible, in three tiers that echo—but do not merely copy—existing practice (Figure 2).

Tier one is *governance assurance*: examination of the organization's AI management system, whether via ISO/IEC 42001 certification or an ISAE 3000-style examination of the AI governance framework against E-23, NIST AI RMF or equivalent criteria. This is the broadest and most immediately deliverable tier, and it maps naturally onto board reporting.

Tier two is *system-level examination*: deep evaluation of individual high-stakes systems—model documentation and data lineage review, performance and bias testing against disclosed metrics, security and red-team assessment, and testing of human-oversight controls in oper-

[23] DSIT market study employment figures, per Burges Salmon (2025), cited above.

[24] For a representative practitioner-skeptic view of PwC's launch, see the discussion at `https://www.teamblind.com/post/big-4-now-selling-ai-audits-here-we-go-357w2nen`.

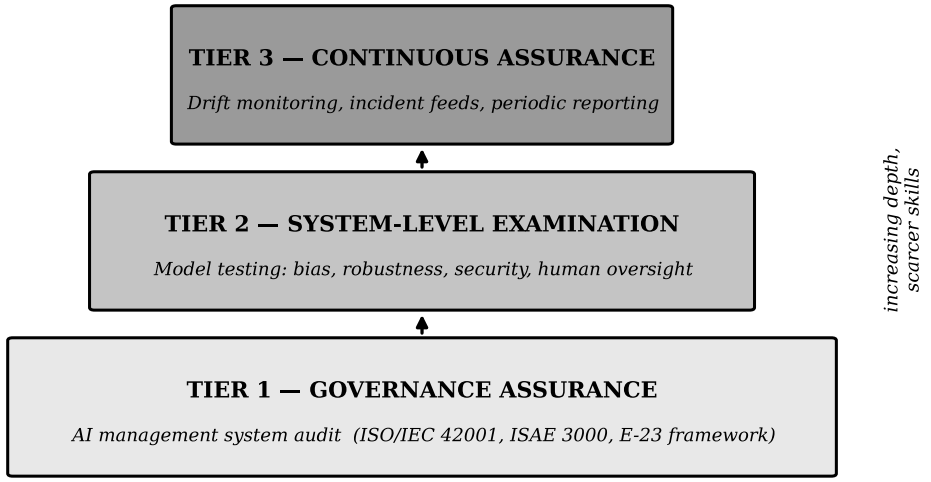

Figure 2. Three tiers of AI assurance. Breadth narrows and technical depth increases moving upward; each tier depends on the integrity of the one beneath it.

ation. This is where hybrid teams and the moving-target discipline matter most, and where OSFI E-23's validation expectations give Canadian financial institutions a ready-made internal analogue: external system assurance is, in effect, independent model validation elevated to an attest discipline.

Tier three is *continuous assurance*: standing monitoring arrangements—drift metrics, incident feeds, control telemetry—that keep tier-two conclusions alive between examinations, plausibly reported through SOC-style periodic reports to relying parties. The DSIT roadmap's insistence on standardized assurer access to system information (functionality boundaries, inputs and outputs, parameters, change mechanisms) is best read as the information architecture this tier requires.[25]

Two cross-cutting requirements bind the tiers. The first is proportionality: not every chatbot warrants a red team, and a credible market needs graduated products the way financial reporting has audit, review and compilation. The second is professional infrastructure: the standard-setters are moving—the AICPA's assurance apparatus is actively monitoring AI-specific projects, and the international framework of ISAE 3000 provides a serviceable vessel for AI subject matter today—but engagement-level methodology, quality management expectations and practitioner accreditation for AI assurance remain under construction, and firms entering now are, in effect, writing the profession's future standards through their

[25]ICAEW summary of DSIT information-access expectations (September 2025), cited above.

engagement files.[26]

7. Conclusion

The comparison that best illuminates AI assurance is not, in the end, financial audit—it is the two decades of assurance history since Sarbanes–Oxley. SOX demonstrated that a new attest product can scale from nothing to universal within a few years when regulation, board anxiety and procurement align; it also demonstrated the costs of scaling before methodology matures: mechanical checklists, expectation gaps, and years spent rebuilding proportionality. AI assurance now stands where internal control attestation stood in 2003, with one difference that cuts both ways: the subject matter is harder—probabilistic, non-stationary, criteria-contested—but the profession is entering with open eyes, published roadmaps, an international management-system standard, and regulators (OSFI among the most methodical) specifying lifecycle expectations before, rather than after, the crisis.

For the Big Four, the strategic claim is easy to state and hard to earn: the franchise will belong to whoever can pair genuine technical evaluation capability with the evidential discipline of attest practice, at global scale, while visibly governing its own AI to the standard it certifies in others. For the profession, the stakes are larger than a revenue line. If AI assurance is done rigorously—disclosed criteria, honest scoping, hybrid competence, real independence—it becomes the mechanism by which societies extend verifiable trust to the most consequential technology of the era. If it is done cheaply, it will manufacture confidence without foundation, and the correction, when it comes, will be paid in the profession's oldest and least renewable asset. The market opportunity is real; the discipline is the product.

[26] On the AICPA Auditing Standards Board's monitoring of AI assurance projects, see Thomson Reuters, "AICPA's Auditing Board Plans to Issue 4 Standard-Setting Proposals in 2025," `https://tax.thomsonreuters.com/news/aicpas-auditing-board-plans-to-issue-4-standard-setting-proposals-in-2025/`; on ISAE 3000 as the umbrella framework, IFAC, cited above.