Working Paper · AI Governance & Financial Regulation · June 2026 · Preprint, not peer reviewed

# The Independence Gap: Why Hyperscaler AI Governance Fails the OSFI E-23 Test

*Native provider governance cannot satisfy a model-review regime that requires independence from model development.*

Mirza M. Baig, Kinetic AI Labs · Correspondence: mirza@kineticlabs.ai

---

**Abstract.** On 11 September 2025, Canada's Office of the Superintendent of Financial Institutions (OSFI) released the final version of Guideline E-23, *Model Risk Management*, effective 1 May 2027. The revised guideline expands its scope to artificial intelligence and machine learning models and, critically, applies in full to vendor and third-party models. As federally regulated financial institutions (FRFIs) adopt hyperscaler AI services such as Microsoft Copilot, Google Vertex, and Amazon Bedrock, a structural problem emerges: the same provider that develops and operates the model also supplies the tooling that attests to its governance. E-23 states plainly that the model review process should be independent from model development. This paper argues that native, in-process governance offered by the model provider cannot, as a matter of structure rather than product quality, satisfy that independence requirement. We define the *independence gap*, map it against the specific principles of E-23 and the parallel Bermuda Monetary Authority (BMA) framework, and argue that closing it requires an evidence layer that is verifiable by an examiner without the cooperation of the governed system. A bank cannot sign its own audit opinion. A provider's agent cannot generate independent proof that its own execution was properly governed.

**Keywords:** AI governance, model risk management, OSFI E-23, independence, agentic AI, third-party risk, guardian agents, auditability, financial regulation.

## 1. Introduction

Financial institutions are deploying AI agents into regulated workflows faster than their model risk frameworks were designed to absorb. The agents now draft credit memos, summarize suspicious-activity reviews, triage claims, and increasingly take actions rather than merely produce text. Most of these capabilities arrive through a small number of hyperscale providers, bundled with native governance dashboards that promise observability, policy enforcement, and audit logging inside the same platform that runs the model.

This paper examines a question that the procurement conversation tends to skip: can governance supplied by the model's own developer satisfy a supervisory regime that explicitly demands independence from development? The Canadian answer, encoded in the final 2025 update to OSFI Guideline E-23, is the cleanest test case available, because E-23 is unusually explicit on two points at once. First, it brings AI and ML models, including vendor and third-party models, squarely into its scope.[1,2] Second, it retains the long-standing model-risk principle that the review of a model must be

independent of the people and processes that built it.[3] Put those two provisions side by side and a gap appears that no amount of dashboard sophistication can close.

We call this the **independence gap**: the distance between governance that a provider performs *on its own model, inside its own trust boundary*, and the independent, examiner-verifiable evidence that a model-risk regime requires.[*] Our claim is structural, not a criticism of any specific vendor's engineering. The native tooling from major providers is often excellent at what it does. It simply sits on the wrong side of the independence line that E-23 draws.

[*] A note on the title. "Fails the E-23 test" refers to native provider governance used *on its own* as the independent record of governed execution. E-23 does not prohibit vendor tooling, and an FRFI may layer independent review on top of it (see Section 3, Objection 3, and Section 7). The claim throughout is that provider-native governance, standing alone, does not satisfy the independence-from-development requirement; it is necessary but not sufficient.

## 2. What OSFI E-23 Actually Requires

OSFI published the final E-23 on 11 September 2025 with an 18-month transition to an effective date of 1 May 2027.[1,4] Three features of the final guideline matter for the present argument.

### 2.1 Scope now reaches AI, ML, and third-party models

The revised guideline applies to all models used by an FRFI regardless of whether they are developed internally or sourced externally, and regardless of whether the underlying data is internal or external.[2] Principle 2.1 requires that model identification include vendor and third-party models, and the guideline establishes heightened expectations for models and data sourced from third-party vendors, read together with OSFI's Guideline B-10 on third-party risk management.[2,3] Legal commentators have noted the practical consequence directly: many AI and ML vendors do not yet have governance, validation, and reporting capabilities consistent with E-23, and FRFIs must define precisely where vendor accountability ends and their own begins.[2,5] When a bank runs a hyperscaler model in a regulated process, that model is a third-party model under E-23, and the obligation to govern it rests with the bank, not the provider.

### 2.2 Model review must be independent from model development

E-23 frames the model lifecycle as a set of components: design, review, deployment, monitoring, and decommission.[3] The review component carries the load for our argument. The guideline states that the model review process should be independent from model development, should validate that the model is properly specified and working as intended, and that policies, procedures, and controls should ensure appropriate independence and objectivity are maintained.[3] Institutions may rely on internal reviewers or objective third parties, but the separation between the act of building and the act of reviewing is not optional.[3]

### 2.3 Governance is a lifecycle obligation, not a one-time gate

E-23 requires ongoing monitoring to ensure models remain fit-for-purpose, with explicit attention to AI-specific risks such as autonomous decision-making, autonomous re-parametrization, and elevated model drift.[3] A point-in-time certificate does not discharge this duty. Monitoring must be continuous, and the evidence of monitoring must persist across the life of the model. This is why a static attestation, however

well produced, is insufficient on its own terms: the obligation refreshes continuously and so must the evidence.

## 3. The Architecture of the Independence Gap

Native provider governance follows a recognizable pattern. The model executes inside the provider's platform. The same platform observes that execution, applies policy, and writes the audit log. The provider then offers the customer a governance attestation generated by the provider's own framework. Every link in that chain sits inside one trust boundary, controlled by one party, who is also the developer of the model under review.

Hold that pattern against Section 2.2. The reviewer (the provider's governance layer) is not independent of the developer (the provider). It is the developer. The evidence of governance is produced by the system being governed, about itself, using tooling the governed party controls. This is the precise configuration that the independence principle exists to prohibit. The analogy that supervisors reach for is the external audit: a public company cannot issue its own audit opinion, because an opinion about one's own conduct, produced with one's own tooling, carries no independent assurance value no matter how rigorous the internal process. The same logic transfers without modification to AI execution.

> *A bank cannot sign its own audit opinion. A provider's AI agent cannot generate cryptographic proof that its own execution was properly governed.*

Three objections are worth meeting directly.

*Objection 1: the provider's logs are immutable and tamper-evident, so they are trustworthy.* Tamper-evidence and independence are different properties. A log can be perfectly tamper-evident and still fail the independence test, because the question an examiner asks is not only "was this record altered?" but "who attests that the record is complete and faithful, and are they independent of the party being examined?" When the attesting party is the model's own developer, the answer to the second question defeats the assurance regardless of the answer to the first.

*Objection 2: the FRFI, not the provider, owns the configuration, so the governance is the bank's.* Configuration ownership does not relocate the trust boundary. If the evidence is generated by the provider's runtime and can only be verified with the provider's cooperation and tooling, the bank cannot demonstrate to an examiner that the evidence is independent of the system that produced it. The bank owns the policy intent; it does not own independent proof of execution.

*Objection 3: E-23 permits reliance on the vendor's own validation.* E-23 permits the use of objective third parties, and it permits leveraging vendor work, but it places the residual risk and the accountability on the FRFI and requires that independence be maintained.[2,3] Reliance on the developer's self-assessment is the one form of "third-party" validation that is not, in fact, independent, because the developer is a party to the matter under review.

## 4. Why Agentic AI Sharpens the Problem

The independence gap exists for any model, but agentic systems make it acute. A predictive model produces an output that a human reviews before acting. An agent takes the action. The window between decision and consequence collapses, and the governance question shifts from "was the output accurate?" to "did the agent do only what it was permitted to do, and can we prove it after the fact to someone who does not trust us?"

Gartner's research community has named the emerging control category for this problem. Guardian agents are described as AI agents that manage and contain other AI agents, sitting between users, agents, and the systems agents touch, performing oversight, blocking or redirecting actions, and producing audit evidence.[6,7] Gartner has forecast that guardian-agent technology will represent 10 to 15 percent of the agentic AI market by 2030, and its analysts, including Avivah Litan, have argued that AI governance needs more than policies: it needs operational controls and validation mechanisms embedded across the lifecycle.[7,8] The market is converging on the recognition that something must watch the agents. The unresolved design question is whether that something can live inside the same platform as the agents it watches. The independence principle in E-23 answers no.

## 5. The Gap Is Not Canada-Specific

OSFI E-23 is the sharpest articulation, but the independence principle is not parochial. The Bermuda Monetary Authority's July 2025 discussion paper on the responsible use of AI in financial services organizes supervisory expectations around board accountability, risk assessment, model validation, and transparency and disclosure, and it calls expressly for supervisory access to auditable model trails.[10] That last phrase is the independence requirement stated in the regulator's own words: an examiner expects to inspect a faithful trail of what the model did, and a trail the supervised entity can curate about itself is not the assurance the supervisor is asking for. Independent validation and verifiable transparency appear in the BMA framework for the same reason they appear in E-23: a supervisor cannot accept the supervised entity's self-description of its own controls as a substitute for independent evidence. The United States, through the Federal Reserve's model-risk supervisory guidance lineage, has long held effective challenge and independent validation as core tenets of model risk management. The vocabulary differs across jurisdictions; the structural requirement does not. Any regime that asks "who, independent of the builder, verifies that this model behaved?" exposes the same gap when the answer is "the builder."

| Requirement | E-23 principle | Native provider governance | Independent evidence layer |
|---|---|---|---|
| Review independent of development | Mandatory (§2.2) | Fails: reviewer is the developer | Satisfies: separate trust boundary |
| Third-party model in scope | Yes (Principle 2.1; B-10) | Provider is the third party | Verifies the third-party model |
| Ongoing lifecycle monitoring | Required, continuous | Possible, but self-attested | Continuous, externally verifiable |
| | | | Verifiable independently |

| | | | |
|---|---|---|---|
| Examiner can verify without the vendor | Implied by independence | Requires vendor cooperation | |

## 6. What Independent Governance Has to Look Like

If native governance fails on structure, the remedy must be defined on structure. Four properties follow directly from E-23 rather than from any product roadmap.

**First, a separate trust boundary.** The evidence of governed execution must be generated and held outside the runtime that executes the model. If the governed system can rewrite, suppress, or selectively present its own governance record, the record is not independent. Out-of-process generation is the architectural expression of Section 2.2.

**Second, examiner-verifiable proof.** An examiner, an internal auditor, or a third-party reviewer should be able to confirm the integrity and completeness of the execution record without the cooperation of the party being examined. Cryptographic methods, such as a hash-chained ledger of execution events that any reviewer can independently recompute and verify, convert "trust our logs" into "check the math yourself." The assurance no longer depends on trusting the provider.

**Third, continuity.** Because E-23 frames monitoring as a lifecycle obligation, the evidence must be continuous and durable, not a point-in-time snapshot. The record must span design changes, re-parametrization, drift events, and eventual decommission.

**Fourth, no privileged access to the governed system's secrets.** An independent verifier should not require the provider's internal credentials to do its work, because dependence on those credentials reintroduces the provider into the trust boundary it was meant to sit outside of. Independence verified through dependence is not independence.

None of these four properties is exotic. Each is the direct engineering consequence of taking the word "independent" in E-23 literally. The point is that they are properties of *where* governance sits, not of *how good* the provider's governance product is. A better dashboard inside the same boundary does not move the boundary.

## 7. Implications for Practitioners

For a Chief Risk Officer or Chief Compliance Officer at an FRFI, three practical conclusions follow before the May 2027 effective date. The model inventory required by E-23 should explicitly enumerate hyperscaler AI services as third-party models, with named accountability for their review.[2,3] The validation plan for each such model should identify the independent source of execution evidence, and should treat provider-native logs as supporting material rather than as the independent record. And procurement should ask vendors a specific question that native governance cannot answer in the affirmative: can an examiner verify your governance evidence without your cooperation? Where the honest answer is no, an independent evidence layer is not an optional enhancement. It is the part of the control that makes the rest of it admissible.

For supervisors and standard-setters, the convergence of E-23, the BMA framework, and the guardian-agent category suggests an opportunity to state the independence requirement for AI execution as

explicitly as it has long been stated for credit and market-risk models. The clearer that statement, the smaller the space for self-attestation to masquerade as assurance.

## 8. Conclusion

The independence gap is not a temporary product shortcoming that the next release closes. It is a fixed consequence of asking the developer of a model to also serve as the independent reviewer of that model's execution. OSFI E-23 makes the requirement explicit and gives it a date: 1 May 2027. Institutions that treat native provider governance as sufficient are, in effect, planning to present a self-signed audit opinion to their examiner. The institutions that close the gap early will be the ones that build their evidence on the right side of the independence line, where an examiner can check the proof without being asked to trust the party who produced it.

---

### Disclosure

The author leads Kinetic AI Labs, which develops independent AI governance technology and therefore has a commercial interest in the conclusions advanced here. This is a position paper, not legal or compliance advice. Statements about OSFI Guideline E-23, the BMA framework, and Gartner research summarize publicly available sources cited below; readers should consult the primary texts and qualified counsel for application to their institution. Direct paraphrases of E-23 reflect OSFI's published guideline and reputable legal analyses of it; no quotation should be attributed to any named individual.

### References


1. Office of the Superintendent of Financial Institutions. *Guideline E-23 – Model Risk Management (2027).* osfi-bsif.gc.ca/en/guidance/guidance-library/guideline-e-23-model-risk-management-2027.
2. Torys LLP. "OSFI updates and expands scope of Guideline E-23 for AI governance." 2025. torys.com/en/our-latest-thinking/publications/2025/10/osfi-updates-and-expands-scope-of-guideline-e-23.
3. OSFI. *Guideline E-23 – Model Risk Management (2027)*, full text and letter. osfi-bsif.gc.ca/en/guidance/guidance-library/guideline-e-23-model-risk-management-2027-letter.
4. Blakes. "OSFI Releases Final Guideline E-23 for Model Risk Management and AI Use by FRFIs." September 2025. blakes.com/insights/osfi-releases-final-guideline-e-23-for-model-risk-management-and-ai-use-by-frfis.
5. BLG. "OSF