PITHY NOTES PUBLICATIONS

PithySignal

RUNTIME GOVERNANCE SIGNAL

DEEP DIVE

Continuous Evidence Loops Are Becoming the Default Runtime Governance Primitive

Why AI governance is moving from static documentation toward runtime evidence, operational review, and retained proof.

A deep dive on why continuous evidence loops are becoming the runtime governance primitive for AI-enabled systems, vendor ecosystems, and audit-ready assurance.

Created by Pithy Notes Publications | Published under Pithy Signal

KEY SECTIONS

Category: Runtime Governance Signal
Publication Date: 2026-05-15

Tags: Runtime Governance | Continuous Evidence | AI Governance | Monitoring | Assurance | Vendor Risk

This publication was created by Pithy Notes Publications and published as part of the Pithy Signal governance intelligence series.

Executive Summary

"Continuous evidence loops" is not yet a formal term of art in the major AI standards and laws, but it is an accurate description of where AI governance is heading. Across NIST, the EU AI Act, OECD principles, ISO/IEC 42001, SOC-related assurance practice, and cloud control guidance, the common signal is clear: pre-deployment testing, annual reviews, and periodic attestations remain necessary, but they are no longer sufficient for AI-enabled systems that are dynamic, socio-technical, and increasingly dependent on external models, data, services, and infrastructure. Governance is moving from static documentation toward live, recurrent proof that systems are being observed, reviewed, controlled, and improved in operation.

The practical implication is that runtime governance is becoming evidence-centric. Monitoring data on its own does not amount to governance. Governance emerges when telemetry, logs, traces, human feedback, control reviews, incident workflows, and change records are linked to risk thresholds, decision rights, interventions, and auditable retention. NIST's AI RMF explicitly ties post-deployment monitoring to appeal and override, decommissioning, incident response, recovery, and change management, while the UK's assurance guidance defines assurance as the process of measuring, evaluating, and communicating trustworthy evidence. Those two ideas together explain the shift: monitoring is a sensing function; assurance is the evaluative and communicative discipline built on top of that sensing; governance requires both.

For leaders in AI governance, risk, compliance, TPRM, and operational resilience, the next control plane is therefore a continuous evidence loop. In practice, that loop has five parts: observe, review, decide, intervene, and retain evidence. That structure is not spelled out verbatim in any one framework, but it is strongly implied by the combination of NIST continuous monitoring, ISO management-system performance evaluation and improvement, EU AI Act logging and post-market monitoring, SOC 2 operating-effectiveness expectations, and the operational tooling that major cloud providers now expose for AI observability and evidence collection.

This matters most in vendor ecosystems. NIST's GenAI Profile explicitly pushes organizations to monitor third-party entities, include evaluation clauses in contracts, maintain records of third-party changes, establish contingency processes for third-party failures, and add ongoing monitoring, dynamic risk assessments, and real-time reporting tools into vendor due diligence. The EU AI Act likewise requires written agreements across the AI value chain to ensure necessary information, technical access, and assistance. In other words, AI-enabled third-party risk is increasingly a runtime evidence problem, not a questionnaire problem.

Why Monitoring Alone Is Not Governance

Monitoring is the act of gathering signals. In NIST's security guidance, continuous monitoring is "maintaining ongoing awareness" to support risk decisions; in NIST's new report on deployed AI systems, monitoring is defined broadly as measurement, tracking, evaluation, data collection, or information gathering after deployment. Assurance is different. The UK government defines AI assurance as measuring, evaluating, and communicating the trustworthiness of AI systems. That distinction is foundational: monitoring generates raw inputs; assurance converts those inputs into evaluated confidence; governance then assigns accountability and acts on the result.

Traditional point-in-time governance fails in AI because deployment is where many risks become legible. NIST notes that AI systems may be trained on data that changes over time, are deployed into complex contexts, and are influenced by human behavior and social context. The same framework states that TEVV activities during operations involve ongoing monitoring, incident and error tracking, management of emergent properties, and processes for redress and response. NIST's 2026 report goes further: it says pre-deployment evaluations are valuable but cannot fully account for real-world dynamics, non-deterministic outputs, drift, or new contexts of use.

Governance begins where monitoring is joined to roles, thresholds, and intervention rights. NIST's AI RMF says AI risks identified through analysis must be prioritized, responded to, and managed; unknown risks must trigger response and recovery procedures; systems inconsistent with intended use must be superseded, disengaged, or deactivated; and incidents and errors must be tracked, responded to, recovered from, and documented. The EU AI Act makes the same point in legal form: high-risk systems must enable automatic logging over their lifetime, human oversight must allow people to monitor operation, detect anomalies, override outputs, and interrupt the system, and deployers must suspend use when they have reason to believe the system presents a risk. Monitoring without these governance mechanics is observability, not oversight.

Monitoring also becomes governance only when evidence is retained in a form that can support audit, supervision, and retrospective learning. NIST SP 800-137A says security information collected during continuous monitoring is used to update authorization packages and supporting artifacts, and that those updated artifacts provide evidence that controls continue to safeguard the system. NIST's incident response guidance adds that lessons learned must be fed back into broader risk management continuously, not only after a long recovery cycle. The implication for AI is straightforward: if runtime signals do not end up in reviewed artifacts, the organization has telemetry, but not defensible governance.

What Is a Continuous Evidence Loop?

A continuous evidence loop is the operating pattern that turns live system behavior into governance-grade proof. It is a standing control cycle in which the organization collects runtime signals, evaluates them against policies and tolerances, takes action when needed, and preserves the evidence trail that explains what was seen, who reviewed it, what was decided, what was changed, and why. This is the practical bridge between continuous monitoring and formal assurance. It aligns with NIST's model of collecting, integrating, analyzing, presenting, and responding to information for risk-based decision-making, as well as the UK assurance model of measure, evaluate, and communicate.

What makes the loop "continuous" is not that every control runs in real time. NIST is explicit that "continuous" in monitoring means a frequency sufficient to support risk-based decisions, not literally constant sampling. In governance practice, that means different evidence streams will run at different cadences: some in real time, some event-driven, some daily, some by model/version release, some through periodic management review or internal audit. The core requirement is not simultaneity; it is that the organization can continuously re-establish justified confidence in the system's operational trustworthiness.

The Five-Part Loop: Observe, Review, Decide, Intervene, Retain Evidence

Observe. The organization collects the runtime facts that matter. In AI, that no longer means only infrastructure health. NIST's 2026 monitoring report identifies six monitoring categories: functionality, operational, human factors, security, compliance, and large-scale impacts. The EU AI Act requires lifetime logs for high-risk AI. Major cloud platforms now expose AI-specific observability primitives such as logs, traces, metrics, prompt/response records, quality metrics, end-user interactions, latency, token usage, and error signals. Observation is therefore multi-layered: model behavior, application behavior, human interaction, control conformance, and ecosystem dependencies all belong in scope.

Review. Signals must be interpreted in context. NIST places auditors, evaluators, compliance experts, management, domain experts, operators, and users inside the operation-and-monitoring and TEVV parts of the lifecycle. ISO/IEC 42001's published clause structure requires performance evaluation, internal audit, and management review, while secondary implementation guidance consistently emphasizes keeping records of monitoring, audits, reviews, risk assessments, and corrective actions. Review is the step where telemetry becomes meaning: false positives are discarded, severity is assigned, and business, legal, and operational context are added.

Decide. Governance is not only about seeing; it is about adjudicating. NIST's MANAGE categories require organizations to determine whether a system should proceed, prioritize documented risks, and choose risk responses. DORA similarly requires ICT risk strategies to define risk tolerance levels, impact tolerance, key performance indicators, and key risk metrics. A mature evidence loop therefore compares live evidence against approved tolerances, legal obligations, and contractual boundaries and then records the resulting decision.

Intervene. A governance loop that cannot change system behavior is theater. NIST requires procedures to respond to and recover from unknown risks and to disengage or deactivate systems with inconsistent performance. The EU AI Act requires human overseers to understand limitations, detect anomalies, decide not to use an output, override or reverse it, or interrupt the system through a stop procedure. Deployers must also suspend use when risk is suspected, and providers must investigate serious incidents and apply corrective action. Intervention can take the form of rollback, retraining, rate limiting, prompt-guard changes, workflow redesign, human-in-the-loop routing, suspension, vendor escalation, or decommissioning.

Retain evidence. The loop is complete only when the organization can later prove what happened. The EU AI Act hardens this expectation by requiring automatic logs over the lifetime of high-risk systems, by linking logging to post-market monitoring and deployer monitoring, and by requiring providers and deployers to keep logs for at least six months where those logs are under their control. NIST continuous-monitoring guidance says updated artifacts provide evidence of continuing control effectiveness. Cloud assurance tooling now operationalizes the same pattern: AWS Audit Manager automatically collects evidence for SOC-related assessments, while Google and Microsoft issue bridge letters to extend assurance between formal SOC reporting periods. Runtime governance is, increasingly, retained evidence plus the chain of reasoning around it.

Why AI Systems Need Runtime Evidence

Traditional point-in-time governance is insufficient for AI-enabled systems because pre-deployment tests are performed in controlled conditions, while deployment introduces non-determinism, changing inputs, different users, novel misuse patterns, model interactions, and environmental shifts. NIST's 2026 report states directly that post-deployment measurement is needed to validate reliable operation in real-world scenarios, track unforeseen outputs and drift, and identify unexpected consequences in new or changing contexts. The AI RMF also notes that AI may require more frequent maintenance because of data, model, or concept drift.

AI systems also require runtime evidence because their risks are socio-technical, not merely technical. NIST emphasizes that AI risks arise from how a system is used, who operates it, how it interacts with other AI systems, and the social context in which it is deployed. That is why post-deployment monitoring must reach beyond performance and uptime into human factors, transparency, compliance, and downstream impacts. In practice, governance that only tracks model accuracy or latency misses the risks most likely to escalate into executive, regulatory, or reputational events.

The convergence of monitoring, observability, logging, review workflows, incident response, and audit evidence is therefore not accidental. NIST's GenAI Profile connects post-deployment monitoring to user input, appeal and override, incident response, recovery, change management, transparency reports, and dataset provenance. Its 2026 monitoring report explicitly raises open questions about how monitoring should connect to broader assurance activities, risk assessment, incident reporting, and third-party auditing. That is a strong signal that governance is no longer separable from runtime operations.

There is also a practical reason for the shift: the incident evidence base is still immature. NIST's GenAI Profile notes that formal channels for reporting and documenting AI incidents do not yet exist consistently, and it recommends documentation practices such as logging, recording, analyzing incidents, and maintaining version history and metadata. When formal external channels are weak, internal evidence discipline becomes even more important. Organizations that cannot reconstruct incidents, changes, dependencies, and human decisions in their AI stack will struggle with root-cause analysis, remediation, regulator questions, and vendor disputes.

Evidence Requirements for AI-Enabled Vendor Ecosystems

AI-enabled vendor ecosystems create a sharper governance problem because critical evidence often lives outside the enterprise boundary. The EU AI Act addresses this directly. Article 25 requires providers of high-risk AI systems and third parties supplying AI systems, tools, services, components, or processes used in those systems to specify, by written agreement, the information, capabilities, technical access, and assistance needed for compliance. Article 26 requires deployers to monitor system operation, inform providers where relevant, notify serious incidents, and keep automatically generated logs for at least six months where under their control. Article 72 requires providers to actively and systematically collect, document, and analyze relevant post-market performance data throughout the lifetime of the system in order to evaluate continuous compliance.

NIST's GenAI Profile makes the vendor implication even more explicit. It recommends use-case-based supplier risk assessment frameworks for third parties; contract clauses that allow evaluation of third-party processes and standards; inventories of all third-party entities with access to organizational content; records of changes made by third parties, including sources, timestamps, and metadata; and procurement due diligence that includes ongoing monitoring, dynamic risk assessments, real-time reporting tools, and checks against incident or vulnerability databases. It also calls for contingency processes, documented third-party incidents, and regularly rehearsed incident response plans for third-party AI technologies. This is effectively a blueprint for continuous evidence loops across the AI supply chain.

Independent assurance will matter more, not less, in this environment. The UK government's AI assurance guidance defines assurance as measuring, evaluating, and communicating trustworthy evidence, and its 2025 roadmap for trusted third-party AI assurance argues that third-party providers are increasingly important because firms often lack the in-house capability to independently verify AI trustworthiness. The same roadmap says assurance providers face information-access barriers when firms do not share enough detail about their AI systems. For vendor risk leaders, that means evidence rights are now a contractual design issue. If contracts do not specify access to logs, change history, incident records, model/version information, and remediation evidence, downstream governance will be materially weaker.

Hyper TPRM Implications

"Hyper TPRM" appears primarily in vendor and industry-event literature rather than in standards or supervisory texts. In the materials surfaced during this research, the term is used to describe a move away from questionnaire-driven third-party risk management toward a model that combines data-first intelligence, workflow, real-time monitoring, exchange models, and AI acceleration with human confirmation. That makes it best understood as an emerging market shorthand, not a formal regulatory category.

Even so, the direction implied by Hyper TPRM aligns closely with formal governance trends. NIST's AI RMF says third-party risks and benefits should be regularly monitored and documented. The GenAI Profile recommends ongoing monitoring and real-time reporting of third-party AI risks. DORA requires financial entities to adopt and regularly review an ICT third-party risk strategy, define risk tolerances and metrics, and continuously improve their risk framework based on implementation and monitoring. The EU AI Act requires value-chain agreements, post-market monitoring, and incident reporting. Taken together, these frameworks point toward a TPRM operating model that is event-driven, evidence-rich, and continuously refreshed.

For AI-enabled vendor ecosystems, the implication is significant. Traditional TPRM asks, "Did the vendor answer the questionnaire?" Runtime TPRM asks, "What evidence do we have right now that the vendor's AI-related controls, model behavior, operational dependencies, and incident processes remain within tolerance?" In that model, ratings, attestations, logs, bridge letters, incident notices, provenance records, model/version disclosures, and remediation artifacts become part of the same loop. That is the core connection between Hyper TPRM and continuous evidence loops.

Strategic Implications

The major frameworks increasingly imply that runtime evidence is expected, even when they do not always use that exact phrase. NIST AI RMF requires regular monitoring, documentation, response and recovery, third-party monitoring, and incident/error communication. NIST's GenAI Profile extends that into provenance tracking, post-deployment transparency reports, supplier monitoring, and third-party contingency planning. ISO/IEC 42001 is built on continual improvement and performance evaluation, and publicly available implementation material emphasizes monitoring, internal audit, management review, corrective action, and record-keeping. OECD's AI Principles say AI actors should ensure traceability across datasets, processes, and decisions and apply systematic risk management on an ongoing basis, while also implementing human oversight and meaningful transparency.

The EU AI Act takes the same control logic and turns it into hard obligations for many use cases: lifetime logging, technical documentation, post-market monitoring plans, deployer monitoring, serious incident reporting, corrective action, and human oversight. For providers of general-purpose AI models with systemic risk, it adds documented adversarial testing, systemic-risk assessment and mitigation, serious-incident reporting to the AI Office, and cybersecurity protection. That is not point-in-time governance. It is lifecycle governance sustained by runtime evidence.

SOC 2 and adjacent cloud assurance practice reinforce the same conclusion from a different angle. AICPA materials frame SOC 2 as an examination of controls relevant to security, availability, processing integrity, confidentiality, or privacy. Google states that SOC 2 Type II covers the design and operating effectiveness of controls over a period of time; Microsoft says Type 2 audits examine a rolling 12-month window; AWS documents automated evidence collection for SOC-related assessments. Google and Microsoft also provide bridge letters between formal report periods. The implication is important for AI governance leaders: external assurance increasingly assumes that controls operate over time and that evidence must be maintainable between formal audit checkpoints.

Cloud and security guidance now expose the underlying runtime-evidence fabric directly. Microsoft Foundry defines AI observability as collecting evaluation metrics, logs, traces, model outputs, and automated quality gates across the lifecycle. AWS CloudWatch advertises AI monitoring of latency, usage, errors, prompt traces, agent behavior, and quality metrics, while AWS's Generative AI Lens recommends monitoring and logging across all application layers and enabling security monitoring, tracing, thresholds, and alerts. Google's AI observability guidance likewise ties logs, metrics, traces, and prompt/response data to visibility into performance, behavior, and quality. In short: the operational tooling stack is being redesigned to produce the evidence that governance functions increasingly need.

What Organizations Should Do Next

Organizations should move now from "AI governance as policy set" to "AI governance as evidence-producing operating system." The most important near-term steps are these:

- Build an AI system and AI-vendor inventory that identifies where AI is developed, provided, integrated, or used; what third-party models or services are embedded; and who owns runtime review, intervention, and evidence retention for each system. This is consistent with ISO/IEC 42001 implementation guidance, NIST AI RMF role mapping, and NIST's GenAI Profile expectations for third-party inventories.
- Define runtime control objectives and thresholds, not only design controls. For each materially relevant system, specify what must be monitored, what triggers re-review, who can override or suspend, when incident response starts, and what evidence must be preserved. NIST, DORA, and the EU AI Act all point toward tolerances, metrics, suspension criteria, and response/recovery rules.
- Instrument observability across all relevant layers: prompts, outputs, model invocations, tool calls, knowledge-base access, latency, token use, errors, harmful-content indicators, human overrides, and downstream business effects. Cloud-provider guidance now makes this feasible, and NIST's monitoring taxonomy makes clear that technical uptime alone is not enough.

- Create formal review workflows that join operations, model risk, compliance, privacy, security, internal audit, and business owners. Monitoring outputs should route to named reviewers, with decision records, remediation tickets, and management escalation where thresholds are breached. ISO/IEC 42001, UK assurance guidance, NIST incident-response guidance, and the AI RMF all support this integrated review model.
- Rewrite vendor and model-provider contracts to require change notices, access to necessary logs and technical documentation, incident reporting, evaluation support, provenance and version information, and participation in joint remediation and fallback procedures. That is the contractual foundation for runtime third-party oversight under both NIST guidance and the EU AI Act.
- Make evidence audit-ready by design. Runtime governance artifacts should be durable, attributable, searchable, and defensible: logs, traces, threshold breaches, review notes, approvals, overrides, incident files, corrective-action records, vendor notices, bridge letters, and evidence packages. If an auditor, regulator, or resilience team cannot reconstruct the decision chain, the control is not mature enough.

Closing Signal

The durable signal from this research is that AI governance is shifting from periodic assurance of static artifacts to recurrent assurance of live behavior. Continuous evidence loops are becoming the runtime primitive because AI systems are adaptive in practice, distributed across vendors and platforms, and exposed to operational, legal, and reputational change after release. Monitoring alone will not solve that. The winning control model is the one that can continuously observe, review, decide, intervene, and retain evidence across the full AI and vendor ecosystem, creating justified trust that survives not just audit day, but every production day in between.

Source/Reference List

NIST AI Risk Management Framework 1.0. Foundational source on why AI risks differ from traditional software risk, on TEVV across the lifecycle, and on the MANAGE categories for post-deployment monitoring, third-party risk, response/recovery, and incident documentation.

NIST AI 600-1 Generative AI Profile. Key source on post-deployment monitoring, user input, appeal and override, provenance, transparency reporting, third-party monitoring, procurement due diligence, dynamic risk assessment, and contingency planning for third-party AI.

NIST AI 800-4 Challenges to the Monitoring of Deployed AI Systems. Most current official source on why pre-deployment evaluations are insufficient, why post-deployment monitoring is necessary, and how monitoring now spans functionality, operations, human factors, security, compliance, and large-scale impacts.

NIST SP 800-137 and SP 800-137A. Foundational continuous-monitoring guidance explaining ongoing awareness, risk-based decision-making, and how monitoring outputs become evidence in updated authorization artifacts.

NIST SP 800-61 Rev. 3 and related incident-response guidance. Important source on incident response as an integrated, continuously improving part of risk management rather than a separate intermittent activity.

NIST SP 800-92. Support for the role of logs in reconstructing events, identifying incidents, and supporting compliance and auditing uses.

EU AI Act Regulation (EU) 2024/1689. Primary legal source for lifetime logging, post-market monitoring, deployer monitoring, suspension, serious-incident reporting, value-chain agreements, and obligations for GPAI models with systemic risk.

OECD AI Principles. Primary international principles source on human oversight, transparency, traceability, accountability, and ongoing risk management across the lifecycle.

ISO/IEC 42001 official ISO summary, clause metadata, and publicly available implementation material. Useful for the management-system logic of continual improvement, performance evaluation, monitoring, internal audit, management review, documentation, and corrective action. The full standard text is paywalled, so public interpretations should be read as secondary guidance.

UK Government AI assurance guidance and trusted third-party AI assurance roadmap. Primary source for the distinction between monitoring and assurance and for the growing role of independent AI assurance providers.

AICPA SOC materials and Trust Services Criteria references. Relevant for understanding SOC 2 as an assurance regime about controls and evidence over time, though not a complete AI-assurance regime by itself.

Google Cloud SOC 2 documentation and Compliance Reports Manager. Useful evidence that Type II assurance is period-based and supplemented by monthly bridge letters beyond formal report end dates.

Microsoft SOC 2 Type 2 documentation. Useful evidence that Type 2 assurance runs across a rolling period and is supplemented by quarterly bridge letters.

AWS Audit Manager and AI observability guidance. Useful for the operationalization of automated evidence collection and multi-layer AI monitoring, tracing, logging, dashboards, and alerts.

NIST Cybersecurity Supply Chain Risk Management and DORA operational resilience sources. Relevant for supply-chain lifecycle risk management, multi-vendor strategy, ICT third-party risk monitoring, and continuously improved resilience frameworks.

Hyper TPRM market materials. Relevant as an indicator of market direction, but not a formal standards base. The term appears largely in vendor and industry-event sources and should be treated as an emerging operating-model label rather than a supervisory taxonomy.

Created by Pithy Notes Publications Published under Pithy Signal

Pithy Signal is the governance intelligence platform created by Pithy Notes Publications.