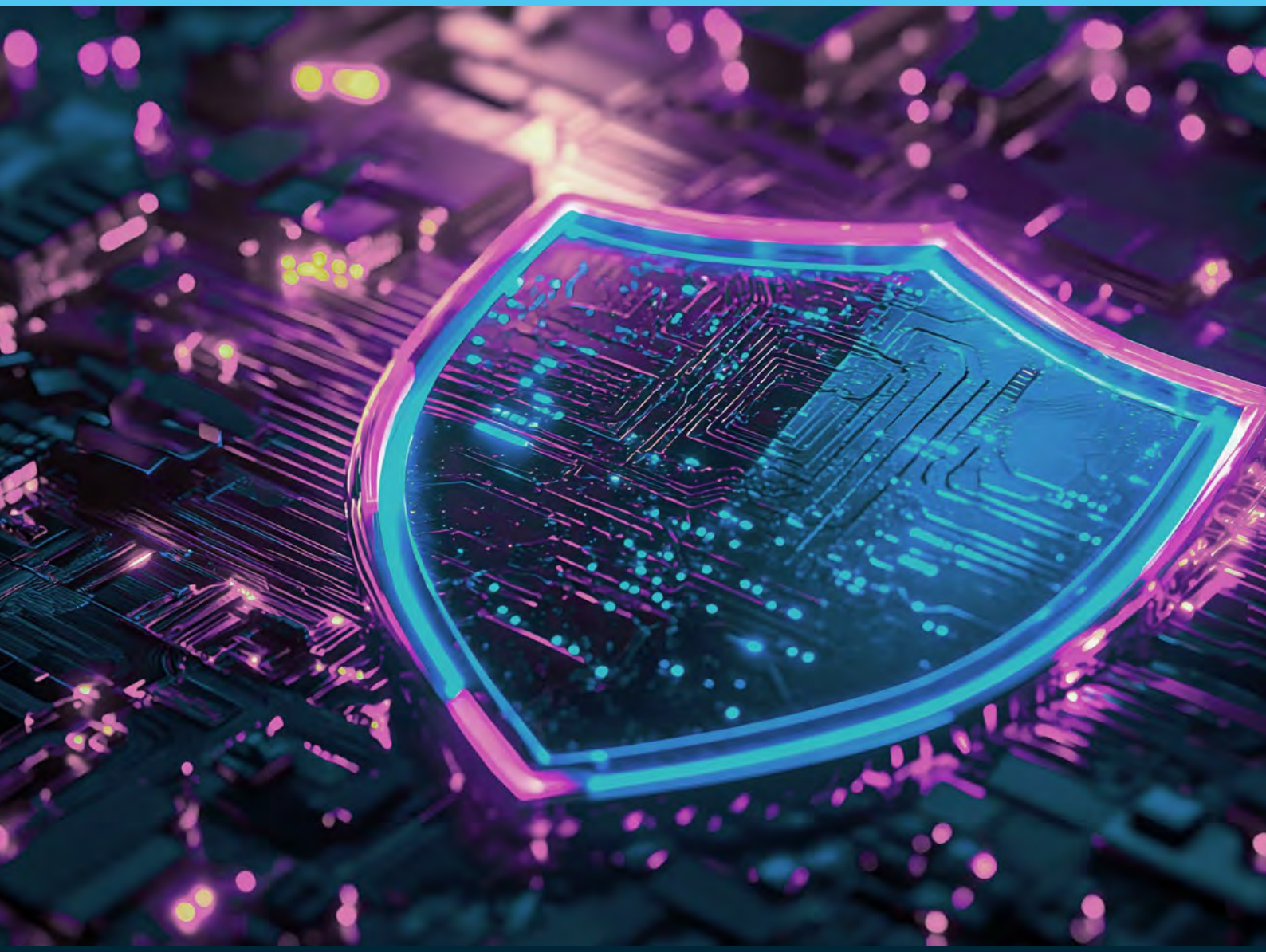
ISO 42001 in Practice:

A Unified Approach to AI Governance

From Documentation to Ongoing Compliance



A Joint White Paper by

DRATA | RAIDS AI™ | PRESCIENT SECURITY

MARCH 2026

Contents

Executive Summary

Global AI spending is projected to reach $2.5 trillion in 2026.[1] Organizations are deploying AI across every function, from customer service and supply chain management to financial forecasting and clinical decision-making. The technology is no longer experimental. It is operational, embedded, and expanding.

Governance has not kept pace. Only 25% of organizations have fully implemented an AI governance program, despite the vast majority now running AI in production. [2] Among U.S. public companies, formal AI risk disclosure is rising rapidly but remains uneven; fewer than half or 48% of all 10-K filers specifically cited AI risk as part of their board's oversight of risk factor disclosures as recently as 2024, and only 27% of boards have formally added AI governance to their committee charters.[3] The gap between deployment velocity and governance maturity has created a structural vulnerability that regulators, auditors, and enterprise buyers are now forcing organizations to close.

The regulatory catalyst is the EU AI Act, which entered into force in August 2024 and follows a phased enforcement timeline.[4] Under the

[1] Gartner, "Gartner Says Worldwide AI Spending Will Total $2.5 Trillion in 2026," January 2026.

[2] AuditBoard / Panterra Research, survey of 400+ GRC/audit professionals (US, Canada, Germany, UK), July 2025.

[3] NACD 2025 Board Governance and AI Survey; EY analysis of Fortune 100 AI risk disclosures, 2025.

[4] European Commission, AI Act implementation timeline. EU Regulation 2024/1689, entered into force August 1, 2024. The Digital Omnibus package (proposed November 19, 2025) proposes extending high-risk AI system obligations to December 2, 2027 at the latest, contingent on availability of harmonized standards. The original August 2, 2026 deadline remains in force until the Omnibus is formally adopted.

GO TO CONTENTS

including high-risk AI system obligations, become enforceable in late 2027 or early 2028. In November 2025, the European Commission proposed extending high-risk deadlines by up to 16 months through the Digital Omnibus package; that proposal remains under legislative review and has not been adopted. Penalties reach €35 million or 7% of global turnover for prohibited practices. Regardless of the final enforcement date, organizations that have not operationalized their AI governance face regulatory, commercial, and reputational exposure.

ISO/IEC 42001:2023, the world's first certifiable AI Management System standard, provides the operational blueprint for closing this gap. But certification is not achieved through documentation alone, nor sustained through monitoring alone, nor meaningful without independent validation. It requires all three.

This white paper presents a unified, three-layer approach to AI governance built around ISO 42001. Pillar 1 covers the documentation and governance framework that establishes the structural foundation. Pillar 2 addresses continuous monitoring and evidence generation, the operational discipline that keeps governance alive in production. Pillar 3 examines certification and audit readiness, the independent validation that converts internal rigor into external trust. Together, the three pillars form an integrated lifecycle: build it, monitor it, certify it.

GO TO CONTENTS

Introduction

The history of enterprise compliance follows a recognizable pattern. A standard emerges. Early adopters treat it as a competitive advantage. Procurement teams begin requiring it. Within a few years, it becomes table stakes.

SOC 2, introduced in 2011, took roughly eight years to become a non-negotiable requirement for SaaS vendors. ISO 27001, published in 2005, followed a similar trajectory, reaching mainstream adoption by 2013–2015 and surpassing 96,000 valid certificates worldwide by 2024.[5] Both standards matured gradually because the regulatory environment allowed it. Organizations had time to build, test, and refine their management systems before external pressure demanded proof.

ISO 42001 does not have that luxury. Published in December 2023, the standard began appearing in enterprise procurement and supplier programs within its first year. Microsoft's SSPA v10 Data Protection Requirements, implemented in September 2024, introduced AI-specific requirements and expects ISO 42001 certification for suppliers with sensitive AI use cases. Accordingly, over 100 organizations achieved ISO 42001 within the first 18 months. More than a dozen certification bodies have applied for or received ANAB accreditation. The adoption curve that took SOC 2 and ISO 27001 a decade to traverse is being compressed into two to three years, driven by the urgency of the EU AI Act enforcement timeline and the speed at which AI systems are being deployed.

[5]Secureframe, "History of SOC 2," 2025; ISO Survey 2024 data.

GO TO CONTENTS

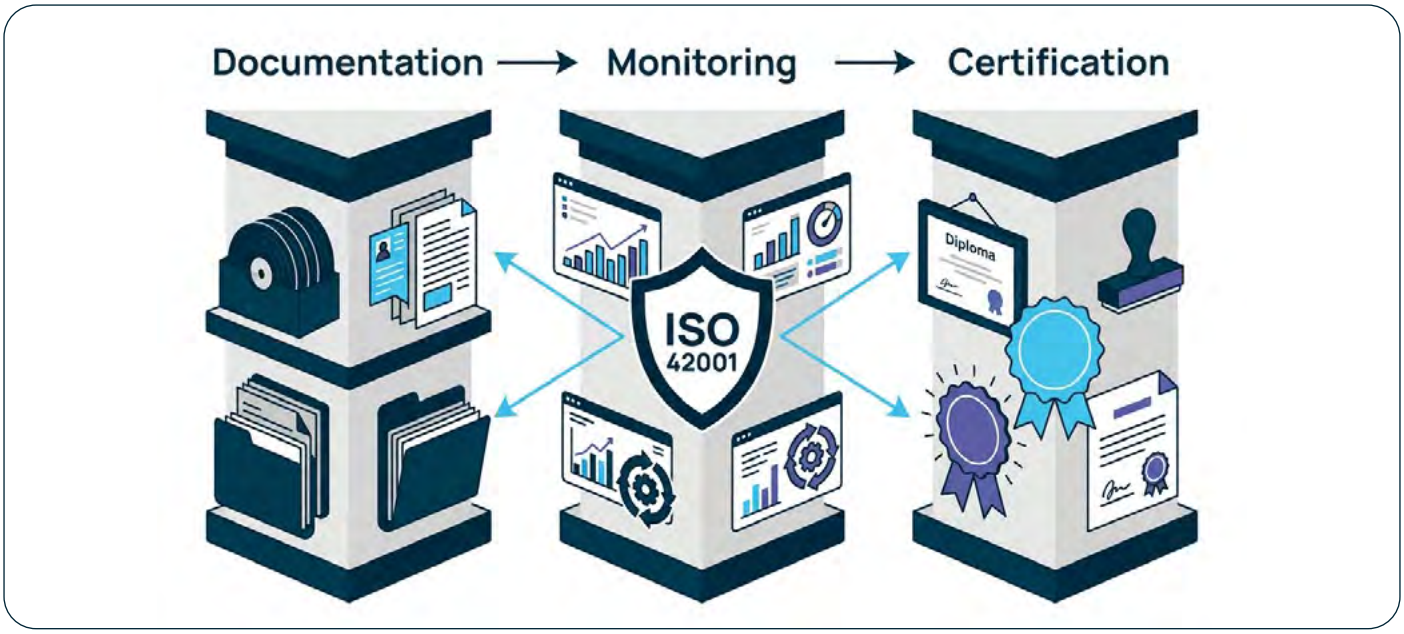


Figure 1: Visual overview of the three-pillar approach showing how Documentation (Pillar 1), Continuous Monitoring (Pillar 2), and Certification (Pillar 3) form an integrated lifecycle.

The speed of this adoption curve exposes a fundamental limitation of traditional compliance thinking. Each layer of governance, taken in isolation, is structurally incomplete.

Static policies don't govern dynamic, non-deterministic systems that learn, adapt, and drift over time. The Deloitte Australia incident illustrates this precisely: despite organizational AI policies, AI-generated fabrications (including fake citations and non-existent court references) were included in a government deliverable, ultimately requiring a partial contract refund. The governance framework existed on paper. It failed in production.

Monitoring alone is insufficient. Detection without a governance structure to interpret, escalate, and remediate findings is operationally incomplete. Alerts that fire into a vacuum do not constitute compliance. Monitoring generates the evidence, but that evidence must map to documented policies, defined risk appetites, and assigned responsibilities to have audit value.

Certification alone is insufficient. An audit is a point-in-time evaluation within a three-year cycle. It confirms that the management system was operational at the moment of assessment. It cannot guarantee that the system remains operational between audits. Without continuous monitoring, the gap between certification and the next surveillance audit becomes a period of unobserved risk.

The thesis of this white paper is that AI governance requires all three layers working in concert. Documentation establishes the structural intent. Monitoring operationalizes it. Certification validates it. The market is already moving in this direction; AI governance platform spending is projected to reach approximately $500 million in 2026 and surpass $1 billion by 2030.[6] Organizations that build this integrated infrastructure now will be positioned not only for compliance, but for competitive advantage as AI governance transitions from differentiator to prerequisite.

[6] Forrester, "Global Commercial AI Software Governance Market Forecast, 2024 to 2030," November 2024. 30% CAGR.

GO TO CONTENTS

Pillar 1: Documentation and Governance Framework

Building the AI Management System foundation with Drata Agentic Trust Management Platform.

1.1 Why Documentation Matters: The Foundation of AI Governance

ISO 42001 mandates 15+ documented artifacts across its core clauses, making documentation the structural backbone of any AI Management System (AIMS).[7] This is not administrative overhead; it is the mechanism through which an organization demonstrates that its AI systems are governed with accountability, repeatability, and transparency. Without formal documentation, governance exists only as intention. In the eyes of auditors, intention without evidence is indistinguishable from absence.

The documentation gap is stark. As of 2025, only 31% of organizations had established a formal AI policy.[8] Only 28% had formally defined oversight roles for AI governance.[9] These figures describe a landscape where the vast majority of organizations deploying AI have not yet documented the most basic governance structures: who is responsible, what policies apply, and how risk is assessed.

Figure 2: The documentation hierarchy for an AI Management System, showing the relationship between AI policy, risk assessment, statement of applicability, procedures, and evidence artifacts.

[6] ISO/IEC 42001:2023, Clauses 4–7 (Context, Leadership, Planning, Support).

[8] ISACA AI Pulse Poll 2025, survey of 3,200+ digital trust professionals. ISACA, May 2025.

[9] IAPP AI Governance in Practice Report, 2024 (670+ respondents, 45 countries).

ISO 42001 closes this gap by requiring a complete, traceable documentation chain. Clause 4 requires a documented AIMS scope statement defining which AI systems, datasets, and teams fall under governance. Clause 5 requires a published AI Policy approved by senior management, with clear organizational roles and responsibilities. Clause 6 introduces the standard's most distinctive requirements: a formal AI risk assessment process, a risk register, a Statement of Applicability mapping every Annex A control, and a risk treatment plan approved by leadership. Clause 7 mandates documentation of competencies, communication protocols, and document control procedures.

1.2 Building the AI Management System (AIMS)

Organizations already certified under ISO 27001 will find the AIMS architecture familiar. Both standards follow the Annex SL harmonized structure, sharing identical clause numbering (Clauses 4–10) and overlapping management processes: risk methodology, Statement of Applicability, document control, internal audit, and management review.[10] This structural alignment means that ISO 27001-certified organizations can typically achieve ISO 42001 readiness significantly faster than those starting from scratch, with practitioners reporting 30–40% reductions in timeline by reusing existing governance infrastructure as a foundation.

The critical distinction is scope. ISO 27001 governs the confidentiality, integrity, and availability of information. ISO 42001 governs the behavior, ethics, and societal impact of AI systems. The AIMS must extend beyond traditional information security to address AI-specific dimensions: algorithmic bias, explainability, model lifecycle governance, data quality management, and the documentation of design intent for each AI system.

ISO 42001's Annex A contains 38 controls organized across nine AI-specific topics, from policies and internal organization to data governance, AI lifecycle management, and third-party supplier relationships. Where ISO 27001 asks whether access to data is controlled, ISO 42001 asks whether the AI system's decisions are fair, whether its behavior is explainable, and whether its impact on individuals and society has been assessed and documented.[11]

1.3 AI-Specific Risk Assessment

Clause 6.1 introduces a paradigm shift in how organizations must evaluate risk. Traditional information security risk assessments are deterministic: they evaluate the likelihood of a known vulnerability being exploited. AI risk assessment must address the probabilistic, non-deterministic nature of machine learning systems, where outputs can change over time without any modification to the underlying code.[12]

The AI-specific risk categories are fundamentally different from traditional cyber threats. They include algorithmic bias and discrimination, transparency and explainability deficits, model drift, adversarial attacks (data poisoning, prompt injection), and accountability gaps across diffused AI lifecycles. Most distinctively, Clause 6.1.4 requires organizations to conduct AI system impact assessments evaluating consequences on individuals, groups, and society.[13] This provision has no equivalent in ISO 27001 and reflects the unique societal implications of AI deployment.

[10] A-LIGN, "The Intersection of ISO 42001 and ISO 27001," 2025. ANAB-accredited certification body.

[11] InfosecTrain, "15 Must-Have Documents & Evidence for an ISO/IEC 42001 Audit," 2025.

[12] BARR Advisory, "ISO 42001 Requirements Explained," 2025.

[13] ISO/IEC 42001:2023, Clause 6.1.4 (AI system impact assessment).

The governance challenge is not theoretical. Only 21% of organizations report having a mature governance model for agentic AI, despite nearly three-quarters planning deployment.[14] Boards hold regular AI discussions (62% do), but only 27% have formally added AI governance to committee charters. The risk assessment mandate in ISO 42001 forces organizations to close the gap between awareness and action, converting boardroom discussions into documented, repeatable processes with assigned ownership and defined risk appetites.

1.4 Policy Architecture for ISO 42001

A robust policy architecture begins with the overarching AI Policy (Clause 5.2), which articulates the organization's commitment to responsible AI development and deployment. Subordinate to this are specialized policies governing data quality, algorithmic fairness, human oversight, and third-party vendor management. Each policy must connect to the Plan-Do-Check-Act (PDCA) cycle that ISO 42001 inherits from the harmonized structure.

The PDCA cycle maps directly to the standard's clause architecture. Planning (Clauses 4–6) establishes context, scope, policy, risk assessment, and objectives. Doing (Clauses 7–8) implements controls, training, and operational procedures. Checking (Clause 9) monitors performance, conducts internal audits, and executes management reviews. Acting (Clause 10) addresses nonconformities, drives corrective action, and ensures continual improvement.

This cyclical structure is particularly well-suited to AI governance. Unlike static IT infrastructure, AI systems evolve continuously. Models retrain, data distributions shift, user behavior changes, and new regulations emerge. A policy architecture built on PDCA treats governance as an iterative process rather than a one-time compliance exercise.

1.5 Evidence Collection as Infrastructure

Establishing a policy architecture is sound in theory but impractical at scale without technological support. This is where compliance automation platforms transition from optional tools to critical infrastructure. In the context of ISO 42001, manual evidence collection through spreadsheets is unscalable, error-prone, and incapable of capturing the dynamic, high-frequency changes in AI system behavior that auditors will expect to see.

Compliance automation platforms serve as the centralized system of record for AI governance. They integrate directly with an organization's technical environment to gather governance artifacts automatically: risk assessments, model lifecycle documentation, access logs, vendor security questionnaires, and control evidence. They provide a unified Statement of Applicability framework, cross-mapping ISO 42001 controls with existing ISO 27001 or SOC 2 evidence to eliminate duplicate effort. And they enable continuous compliance visibility rather than periodic scrambles before audit season.

[14] Deloitte State of AI in the Enterprise 2026, survey of 3,235 leaders across 24 countries, Aug–Sep 2025.

GO TO CONTENTS

The market context reinforces the shift. Organizations deploying AI governance platforms are 3.4 times more likely to achieve high effectiveness in AI governance.[15] For small and mid-size organizations (the primary audience for this approach), automation can compress the path from initial gap analysis to certification readiness to as little as four to six months.[16]

1.6 From Static Documentation to Living Governance

The ultimate goal of the documentation framework is to ensure that policies do not become artifacts reviewed only in the weeks before an audit. ISO 42001 requires a living, adaptive management system capable of responding to shifting operational realities.

The risk of static documentation is well documented. Despite organizational AI policies, Deloitte Australia's consulting arm delivered a government report containing AI-generated fabrications, including citations to non-existent court cases, in a deliverable worth AU$440,000.[17] Separately, 48% of organizations do not monitor AI systems after deployment, undermining any governance documentation that was created during the planning phase.[18]

Living governance requires integration between documentation platforms and operational monitoring tools. When a monitoring system detects that a model has drifted beyond acceptable thresholds, the integration should trigger an incident response workflow documented within the AIMS, forcing the execution of the policy in real time. This transforms the risk register from a static spreadsheet into an active defense mechanism, generating the evidence that proves governance is operational, not aspirational.

This is the bridge to Pillar 2. Documentation establishes what the organization intends to do. The question that follows is whether the organization is actually doing it. Answering that question requires continuous monitoring.

[15] Gartner, "Global AI Regulations Fuel Billion-Dollar Market for AI Governance Platforms," February 17, 2026.

[16] Pacific AI / Gradient Flow, 2025 AI Governance Survey (n=351 participants).

[17] Computerworld, "Deloitte's AI governance failure exposes critical gap in enterprise quality controls," 2025.

[18] Pacific AI / Gradient Flow, 2025 AI Governance Survey (n=351 participants).

Pillar 2: Continuous Monitoring and Evidence Generation

Operationalizing governance through continuous monitoring with RAIDS AI

AI systems move too quickly, too unpredictably, and at too large a scale to be monitored by human oversight alone. Anomalies and inaccuracies are easy to overlook, especially as organizations scale from a handful of AI tools to dozens of interconnected systems operating across business functions. Pillar 2 represents the operational discipline needed to safeguard AI systems by catching and ameliorating their errors, overreach, and degradation in time to make corrections.

In Pillar 1, we have already established the governance foundation: policies, risk registers, and the documented controls that form the backbone of an AI Management System (AIMS). That foundation is necessary, but it is not sufficient. ISO 42001, the EU AI Act, and the NIST AI Risk Management Framework all emphasize the same operational requirement: AI systems must be monitored continuously, throughout their lifecycle, with evidence generated automatically to support audits, management reviews, and regulatory reporting.

This section examines why continuous monitoring is structurally unavoidable, what global frameworks require, and how organizations can implement it at scale without requiring access to proprietary model internals.

2.1 The Monitoring Imperative: Adoption Has Outpaced Oversight

Enterprise AI adoption has crossed the "default" threshold. McKinsey's 2025 Global AI Survey reports that 88% of organizations now use AI in at least one business function, up from 78% a year prior.[19] Generative AI specifically has reached 79% enterprise adoption, with companies spending an estimated $37 billion on GenAI in 2025 alone.[20]

[19] McKinsey & Company, "The State of AI: Global Survey," November 2025 (n=1,993 respondents, 105 countries).

[20] Menlo Ventures, "2025: The State of Generative AI in the Enterprise," January 2026 (approx. 500 U.S. enterprise decision-makers).

This deployment pace has dramatically outstripped monitoring infrastructure. Only 24% of current generative AI projects incorporate security measures.[21] And 95% of senior data leaders admit they lack full visibility into how their AI systems make decisions.[22]

The gap between AI adoption and active risk mitigation represents exactly the exposure that Pillar 2 is designed to close. At 88% adoption, AI is no longer optional; but monitoring it remains the exception, not the rule.

2.2 The Inevitability of Drift

AI systems degrade in production. This is not a theoretical risk; it is a structural property of machine learning.

A peer-reviewed 2022 study in Scientific Reports analyzed 128 model-dataset pairs across four industries and four model types. The researchers observed temporal quality degradation in 91% of cases.[23] In practice, drift can emerge rapidly once production data diverges from training distributions, making continuous monitoring a lifecycle necessity, not a one-time checkpoint.

Drift manifests in three well-documented categories: data drift (changes in input statistical properties), concept drift (evolving relationships between inputs and outputs), and prediction drift (shifting outputs despite stable inputs, often from feedback loops). Each category requires distinct detection strategies, but all share the same prerequisite: continuous observation of the system's behavior in production.

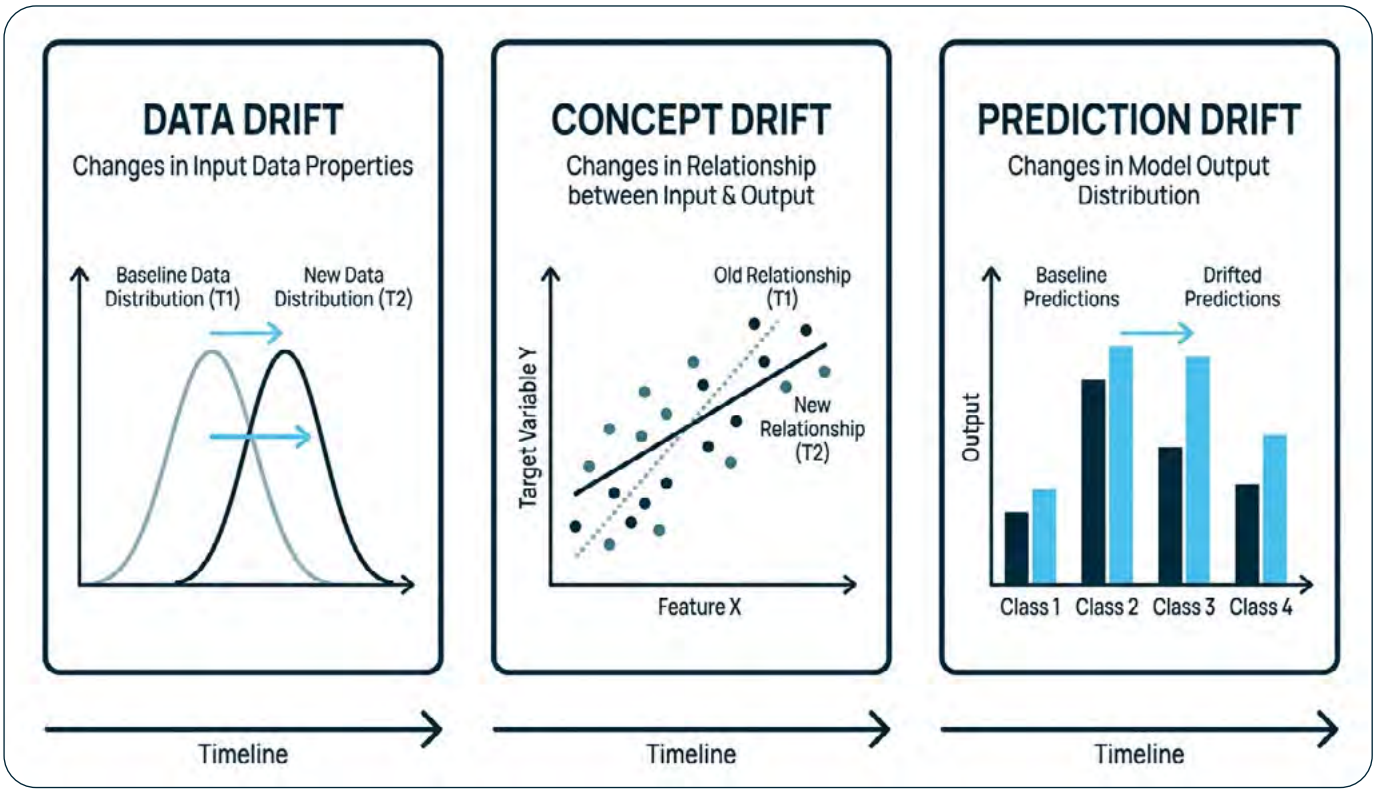


Figure 3: Visual representation of the three drift categories (data drift, concept drift, prediction drift) with simple before/after distribution curves showing how each type manifests over time.

[21] IBM Institute for Business Value, "CEO decision-making in the age of AI," 2024 (survey of enterprise AI deployments).

[22] Dataiku / Harris Poll, "Global AI Confessions Report: Data Leaders Edition," October 2025 (n=812 senior data executives, 8 countries).

[23] Vela et al., "Temporal quality degradation in AI models," Scientific Reports, 2022 (128 model-dataset pairs, 4 model types, 32 datasets, 4 industries).

The 2020 pandemic provided a dramatic global demonstration. Demand forecasting models, credit risk models, and consumer behavior models trained on pre-pandemic data all failed across industries as underlying data distributions shifted overnight. JPMorgan's credit risk models, for example, required emergency recalibration as default patterns diverged from every historical precedent. Retailers saw demand prediction accuracy collapse within days as consumer spending patterns shifted to categories and channels their models had never prioritized. Organizations without monitoring infrastructure had no way to detect or quantify the failure. Those with continuous monitoring activity were able to retrain or adjust within weeks. This represented one of the largest simultaneous model failure events in the history of applied machine learning.

2.3 What Happens When No One Is Watching

The Stanford HAI AI Index recorded 233 AI-related incidents in 2024; a 56.4% increase over 2023 and the highest count on record.[24] The average cost of a data breach reached $4.88 million in 2024.[25]

The pattern across high-profile AI failures is consistent, spanning market cap destruction, legal liability, brand damage, compliance violations, and physical safety.

Market Cap Wipeouts: Google's Bard demo hallucinated a factual answer, erasing approximately $100 billion in market capitalization within hours.

Legal Liability: Air Canada was ordered to pay damages after its chatbot provided incorrect bereavement fare information, with the tribunal ruling that the company was responsible for information provided by its AI.

Brand Damage: McDonald's terminated its AI drive-thru partnership with IBM after the system consistently misinterpreted customer orders, generating viral social media exposure.

Compliance Violations: New York City's MyCity chatbot advised small business owners to take actions that would violate local law, creating direct regulatory liability for the city.

Physical Safety: Waymo recalled over 1,200 robotaxis after its software failed to detect gates, chains, and similar barriers, resulting in low-speed collisions.

In every case, the failure was detectable through behavioral monitoring before it reached customers or regulators. The common thread was not technical complexity; it was the absence of continuous observation.

2.4 Regulatory Convergence on Continuous Monitoring

Three foundational frameworks now mandate monitoring as an ongoing lifecycle obligation, not a pre-deployment checkpoint.

The EU AI Act treats risk management as a

[24] Stanford HAI AI Index Report 2025, drawing from the AI Incidents Database. April 2025.

[25] IBM, Cost of a Data Breach Report, 2024.

> "continuous iterative process planned and run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic review and updating." [26]

Article 72 requires providers of high-risk AI systems to establish a post-market monitoring system that

> "actively and systematically collect[s], document[s] and analyse[s] relevant data"

throughout the system's lifetime. [27]
Article 12 mandates that high-risk systems be designed to automatically record events (logs) over their operational life. [28]

ISO 42001 Clause 9.1 requires organizations to determine

> "appropriate methods of monitoring, measurement, analysis and evaluation"

for their AI systems, including what to monitor, when, and how results are analyzed and retained.[29] Clauses 9.2 and 9.3 require internal audits and management reviews that draw on monitoring data as input. Clause 10.1 requires that monitoring results feed back into the continual improvement cycle.[30]

The NIST AI Risk Management Framework connects production monitoring to ISO 42001 through its GOVERN and MANAGE functions,

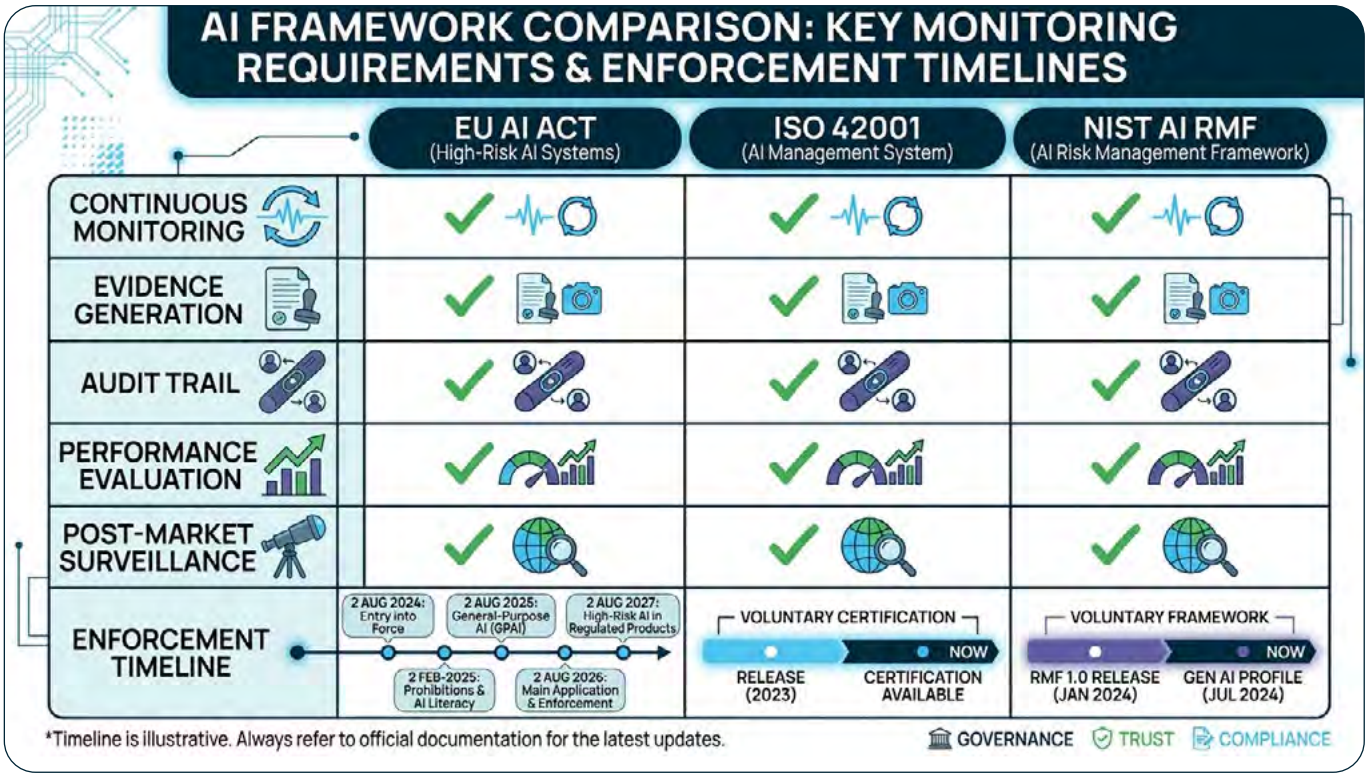


Figure 4: Side-by-side comparison of EU AI Act, ISO 42001, and NIST AI RMF monitoring requirements, showing how all three converge on the same operational obligations.

[26] EU Regulation 2024/1689 (EU AI Act), Article 9(2).

[27] EU Regulation 2024/1689, Article 72.

[28] EU Regulation 2024/1689, Article 12.

[29] ISO/IEC 42001:2023, Clause 9.1 (Monitoring, measurement, analysis and evaluation).

[30] ISO/IEC 42001:2023, Clauses 9.2, 9.3, and 10.1.

[31] NIST AI Risk Management Framework 1.0, GOVERN 1.5 and MANAGE functions.

explicitly requiring that deployed AI systems be monitored for performance, fairness, and alignment with intended use.[31]

Many industries have primarily converged on the same operational conclusion: post-deployment monitoring is not optional. It is a prerequisite for compliance with every major AI governance framework currently in force or approaching enforcement.

2.5 The Case for Black-Box Monitoring

Most organizations using AI in production do not build their own models. They consume AI through APIs, embedded features, and third-party platforms where access to model internals is contractually and technically infeasible. This is the operational reality that any monitoring strategy must accommodate.

Black-box monitoring observes what an AI system does without requiring access to how it works internally. It analyzes inputs, outputs, and behavioral patterns at the system boundary, much like a smoke detector monitors for danger without needing to understand the chemistry of combustion. The detector does not know what caused the fire.

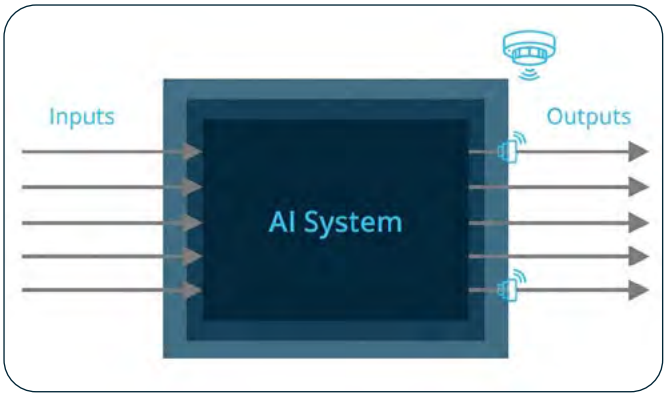


Figure 5: Shows an AI system as a 'black box' with inputs entering and outputs exiting, with a monitoring layer observing the behavioral boundary.

It knows something is wrong, and it alerts you to act on time.

This approach mirrors established practice in regulated industries. Financial institutions monitor transactions for anomalies without needing to reverse-engineer client behavior. Healthcare systems flag adverse drug interactions through outcome observation, not by analyzing molecular mechanisms. Aviation uses flight data recorders to detect deviations from expected parameters. In each case, the monitoring layer sits outside the system it observes, generating evidence through behavioral analysis rather than internal inspection.

For ISO 42001 compliance, the practical implication is significant. Clause 9.1 requires monitoring of AI system performance using appropriate methods. It does not prescribe that organizations access model weights, training data, or internal architectures.[32] The NIST AI RMF similarly emphasizes outcome-based measurement.[33] Black-box monitoring satisfies both frameworks while remaining deployable across proprietary, open-source, and third-party AI systems.

2.6 Bridging the Certification Maintenance Gap

Organizations familiar with ISO 27001 or SOC 2 understand the distinction between achieving certification and maintaining it. Surveillance audits, evidence collection, and continuous compliance are familiar concepts. ISO 42001 follows the same pattern, but the dynamic nature of AI makes the maintenance challenge significantly harder.

[31] NIST AI Risk Management Framework 1.0, GOVERN 1.5 and MANAGE functions.

[32] ISO/IEC 42001:2023, Clause 9.1 (Monitoring, measurement, analysis and evaluation).

[33] NIST AI RMF to ISO/IEC 42001 Crosswalk, linking production monitoring to ISO 42001 Clause 9.1.

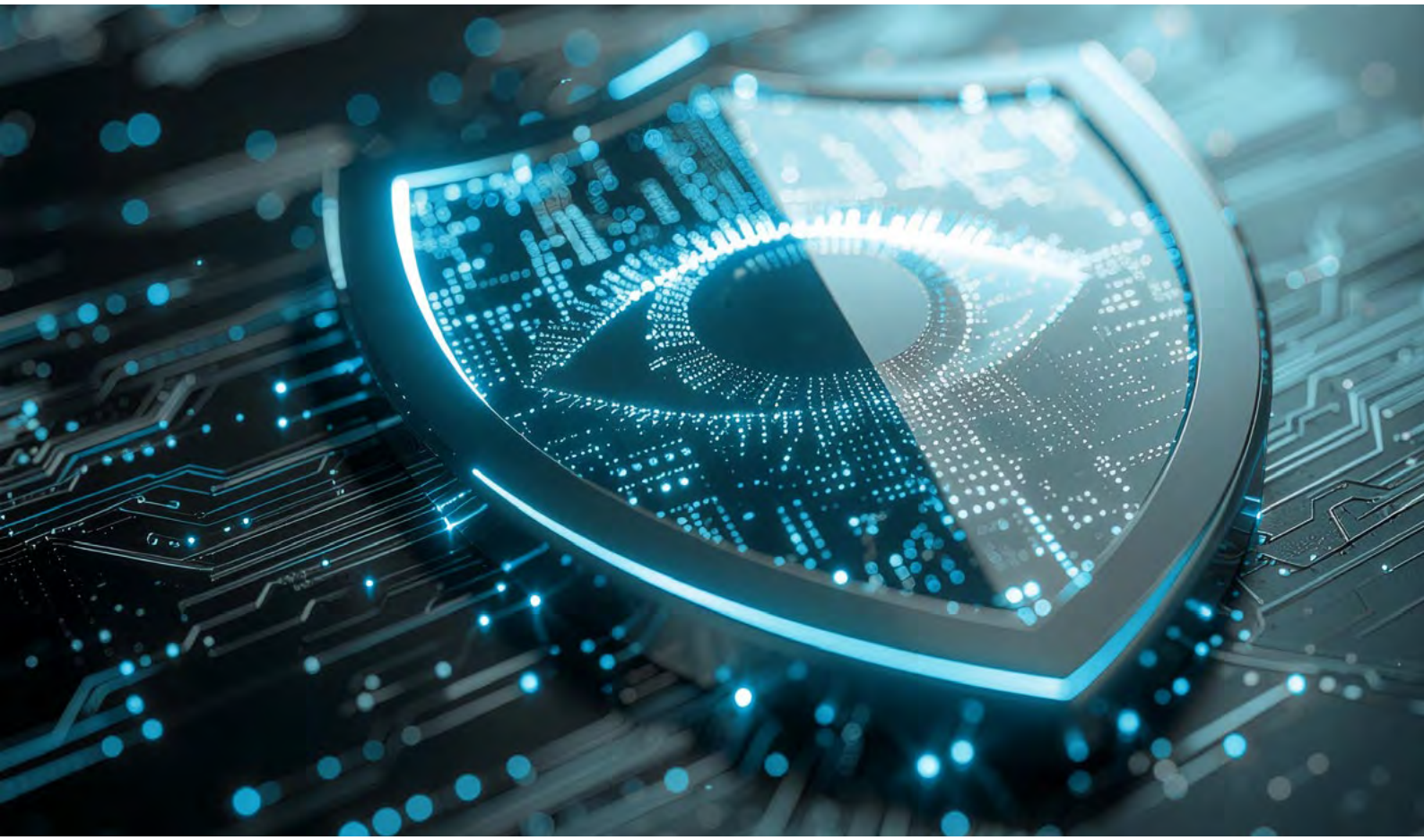
At initial certification, an organization may succeed with well-crafted documentation of policies and controls. By the first surveillance audit, however, auditors will expect to see that the organization has been doing what it documented: results from quarterly AI performance evaluations, logs of incidents and how they were addressed, evidence of management review, and a clear baseline for improvement. These questions can be answered with data from continuous monitoring.

The operational reality makes this even more urgent. 66% of workers admit to using AI outputs without verifying their accuracy.[34] If humans are not consistently verifying AI outputs, the monitoring system must act as a critical safeguard.

This is where Pillar 2 connects directly to Pillar 1 and Pillar 3. Documentation and policy management (Pillar 1) create the governance framework. Continuous monitoring (Pillar 2) generates the evidence that proves the framework is operational. Third-party validation and certification (Pillar 3) confirms that the evidence meets the standard. Without the middle layer, organizations face a gap between their documented objectives and their production reality; a gap that auditors will find, regulators will scrutinize, and incidents will expose.

Continuous monitoring is the connective tissue of operational AI governance, it shows when deployed AI systems start to drift away from what you designed and documented.

[34] Aristek Systems, citing global workforce surveys, 2025.



Pillar 3: Certification and Audit Readiness

Achieving and maintaining ISO 42001 certification with Prescient Security

3.1 The Certification Landscape: Why Organizations Are Pursuing ISO 42001

ISO/IEC 42001:2023 was published in December 2023 as the world's first certifiable AI Management System standard. Within 18 months, over 100 organizations achieved certification, including AWS, Microsoft, Google Cloud, KPMG, Anthropic, and IBM.[35] More than a dozen certification bodies have applied for or received ANAB accreditation, and BSI became the first CB accredited by both UKAS and RvA.[36]

The growth trajectory is being driven by three forces.

First, regulatory alignment. The EU AI Act and ISO 42001 share substantial overlap in high-level

Figure 6: Timeline visualization of the certification process from readiness assessment through Stage 1, Stage 2, and ongoing surveillance audits, showing where documentation and monitoring evidence is evaluated at each stage.

[34] Aristek Systems, citing global workforce surveys, 2025.

[35] Schellman, "What to Expect in the ISO 42001 Certification Process," 2025. First ANAB-accredited CB for ISO 42001.

[36] ANAB Blog, "ISO/IEC 42001 AI Management Systems," 2025. Multiple certification bodies accredited as of early 2026, with the count growing.

requirements. While ISO 42001 is not a legally mandated compliance pathway, it provides the most structured operational framework available for demonstrating the governance practices that regulators will scrutinize once high-risk provisions take effect.

Second, supply chain requirements. Microsoft's SSPA v10 Data Protection Requirements, implemented in September 2024, include AI-specific requirements. For suppliers with sensitive AI use cases, Microsoft expects ISO 42001 certification or equivalent independent assessment.[37] This is not aspirational guidance; it is a procurement condition. As enterprise buyers embed AI governance into vendor evaluation, certification moves from a differentiator to a qualifier.

Third, board and investor pressure. Boards are discussing AI governance (62% hold regular discussions), but few have formalized it. Certification provides a verifiable, externally validated artifact that boards, investors, and customers can point to as evidence of responsible AI deployment.

3.2 Inside the Audit: What Evaluators Actually Assess

The ISO 42001 certification process follows a two-stage methodology conducted by independent, accredited certification bodies.[38]

The Stage 1 audit is a gap-identification checkpoint and readiness assessment, typically lasting one to two days. Auditors evaluate the foundational design of the AIMS: scope statement, AI policy, risk methodology, Statement of Applicability, roles and governance structures. The objective is to identify potential weaknesses in the management system design and confirm that the organization is ready for operational evaluation. By highlighting areas of concern before Stage 2, it ensures a clear path to certification. This checkpoint is only required in Year 1 of the three-year certification cycle.[39]

The Stage 2 audit evaluates implementation effectiveness. Lasting three to nine days depending on scope and complexity, it requires auditors to verify that documented procedures match actual practice.[40] Evaluators conduct interviews with AIMS owners, control owners, and engineering leads. They sample evidence across Annex A controls. They expect to see empirical artifacts: monitoring dashboards, bias test logs, incident response records, management review minutes with documented decisions and action items.

Clause 9 (Performance Evaluation) is the focal point of Stage 2. Auditors expect defined metrics and measurement intervals for AI system performance, evidence of internal audits with independence from audited activities,

[37] Microsoft SSPA v10 Data Protection Requirements, implemented September 23, 2024. ISO 42001 certification accepted for sensitive AI use cases per Schellman analysis, 2025.

[38] Schellman & risk3sixty, "ISO 42001 Lessons Learned from Auditing and Implementing the Framework," CSA blog, May 2025.

[39] Schellman, "What to Expect in the ISO 42001 Certification Process," 2025. First ANAB-accredited CB for ISO 42001.

[40] Schellman & risk3sixty, "ISO 42001 Lessons Learned from Auditing and Implementing the Framework," CSA blog, May 2025.

and management review records showing that monitoring results are feeding back into governance decisions.[41] The distinction between compliance and certification is this: auditors do not just check that documents exist. They validate that the system works. Findings in this area often arise when monitoring is actively in place, but the outputs cannot be clearly tied to documented actions and follow-up. This ensures that monitoring results are not just collected, but are actively feeding back into governance decisions.

3.3 Common Gaps and How to Close Them

Analyzing audit outcomes across early ISO 42001 certifications reveals a consistent pattern of readiness gaps.[42]

The most common gap is an incomplete AI system inventory. Organizations frequently include flagship AI models in their scope but miss embedded tools, third-party API integrations, and internal productivity tools that also qualify as AI systems under the standard. A second persistent gap is the absence of AI-specific risk assessments; organizations rely on generic IT risk templates that do not address bias, drift, explainability, or societal impact.[43]

The third gap, and often the most consequential, is the disconnect between policy and practice. Organizations produce well-crafted AI policies during the documentation phase but fail to operationalize them. Nonconformities typically arise when a lack of recorded evidence, such as logs, outputs, or tickets, prevents auditors from verifying that policies are actually in use. For instance, fairness testing policies may exist on paper, but if no logs demonstrate that tests were conducted in production, or if incident response procedures lack records showing they were executed, the system fails the certification standard. [44]

Additional gaps include missing AI impact assessments (Clause 6.1.4), insufficient post-deployment monitoring evidence, and lack of competence records for staff responsible for AI governance. Each of these is addressable, but all require deliberate planning and, critically, the integration of documentation, monitoring, and evidence collection infrastructure before the audit begins.

3.4 The Stage 1 and Stage 2 Journey

The timeline from decision to certification depends heavily on existing governance maturity. Organizations already certified under ISO 27001 can typically achieve ISO 42001 readiness in four to six months by reusing existing management system infrastructure.[45] Organizations starting from scratch should generally plan for six to twelve months.[46] IBM's Granite models achieved certification through Schellman in under three months with zero nonconformities, demonstrating that mature AI governance practices can accelerate the process significantly.[47]

---

[41] ISO/IEC 42001:2023, Clause 9.1 (Monitoring, measurement, analysis and evaluation).

[42] Sprinto, "ISO 42001 Certification: Steps, Cost and Timelines," 2026.

[43] Glocert International, "ISO 42001 FAQ," 2025.

[44] StackAware, interview with Sammy Chowdhury (Prescient Security), 2025.

[45] Polimity, "ISO 42001 Certification Steps, Cost and Timelines for AI Compliance," 2025.

[46] Cycore Secure, "ISO 42001 Certification Cost, Timeline & Requirements FAQ," 2026.

[47] IBM, ISO 42001 certification announcement for IBM Granite models (Schellman), completed in under 3 months with zero nonconformities.

A critical requirement that organizations frequently underestimate is the operational evidence window. Auditors generally expect to see at least three months of live operation, including completed internal audits and a management review, before the Stage 2 assessment.[48] This means the clock starts when the AIMS is operational, not when the documentation is complete.

Nonconformities identified during audits are classified as major or minor (certification granted with a corrective action plan). Major findings affect the capability of the management system to achieve intended results and require a follow-up review to confirm the gap is closed. Opportunities for improvement (OFIs) are noted but do not block certification. Organizations rarely fail ISO 42001 audits outright; more commonly, audits result in nonconformities that require corrective action, ensuring each small gap a reviewer would notice is effectively addressed.

3.5 Beyond Certification: Surveillance and Continual Improvement

Certification is not a destination. It is the beginning of a three-year commitment. Certificates are valid for three years, contingent on successful annual surveillance audits.[49]

Surveillance audits use a sampling approach, reviewing approximately half of the controls each year and focusing on Clauses 8–10 (operation, performance evaluation, and improvement). They evaluate whether the organization is maintaining the management system, responding to incidents and nonconformities, updating risk assessments, and demonstrating continual improvement. Certificates can be withdrawn if major nonconformities are found during surveillance and not resolved.

Clause 10 (Improvement) is the engine of the surveillance cycle. Auditors expect to see a complete loop: nonconformity identification, root cause analysis, corrective action, and verification that the action was effective. They also expect evidence that management review outputs are leading to concrete improvements, not just documented acknowledgments. Reduced incident rates, updated controls, improved training completion, and documented corrective action closure rates all serve as evidence of a management system that is maturing rather than stagnating.

Recertification occurs in Year 4, with a full assessment of the complete AIMS and all Annex A controls. Organizations can strategically align ISO 42001 recertification with ISO 27001 audit cycles to reduce overhead through integrated audits.

[48] Sprinto, "ISO 42001 Certification: Steps, Cost and Timelines," 2026.

[49] Advisera, "ISO 42001 Certification," 2025.

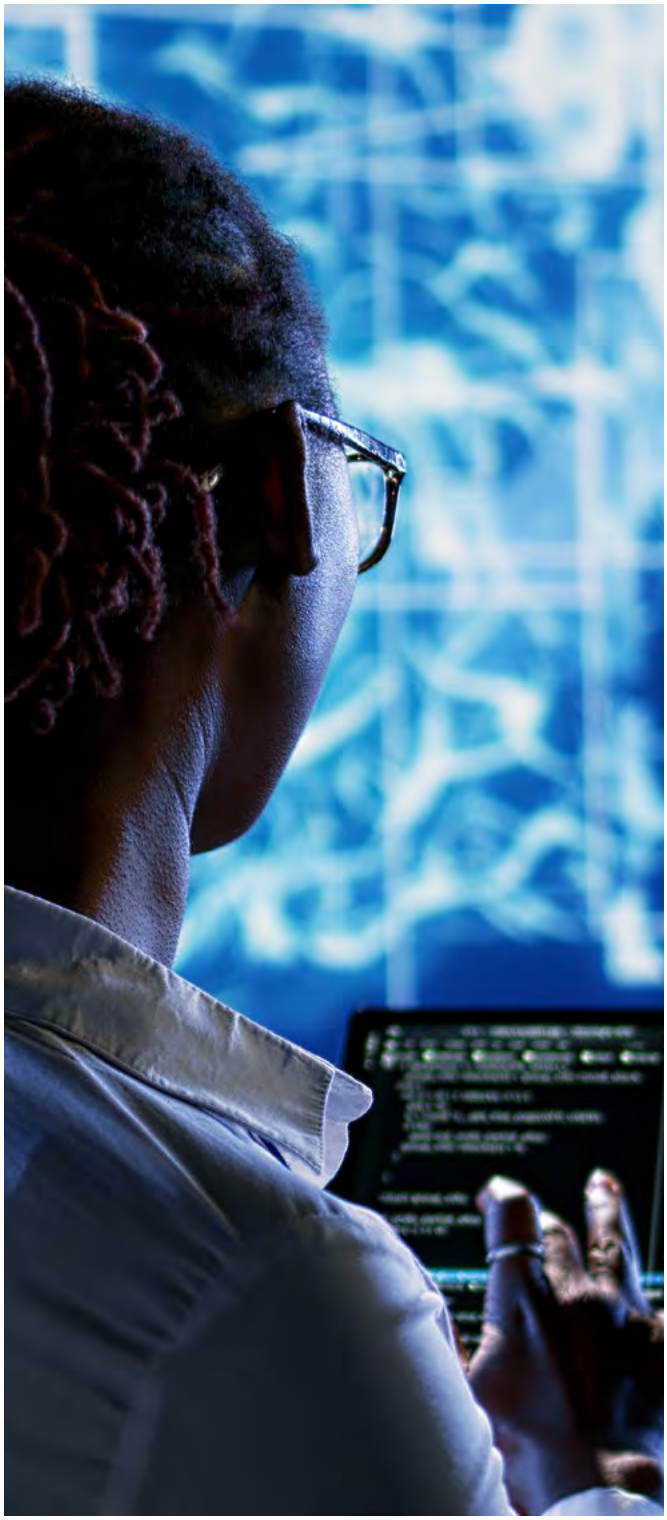
3.6 Building an Audit-Ready Organization

A truly audit-ready organization does not scramble to compile evidence in the weeks before an auditor arrives. It operates with governance embedded in its daily workflow, generating audit artifacts as a byproduct of normal operations.

This requires three capabilities. First, instant traceability: the ability to surface any control change, risk assessment update, or incident response record on demand. Second, immutable evidence: logs and records that are timestamped and tamper-evident, ensuring auditors can trust the integrity of the evidence chain. Third, unified access: a single source of truth that connects documentation, monitoring data, and compliance status across security, engineering, and governance teams.

The integration of trust management infrastructure (Pillar 1) with continuous monitoring (Pillar 2) is what makes this possible. When monitoring tools detect a behavioral anomaly, the alert flows into the documented incident response workflow. When a corrective action is taken, the evidence is captured in the compliance platform. When the auditor arrives, the evidence already exists, mapped to the relevant clauses, without manual compilation.

This is the operational definition of the three-pillar framework in action: governance documented, compliance monitored, and certification maintained through integrated infrastructure rather than periodic heroics.



Conclusion

The AI governance landscape in 2026 is defined by a single structural tension: the speed of AI deployment has outpaced the maturity of the systems designed to govern it. Organizations are running AI in production at scale while governance infrastructure remains, for the majority, aspirational rather than operational.

The regulatory environment has responded. The EU AI Act's high-risk requirements are scheduled to take effect from mid-2026, with the final timeline subject to the outcome of the Digital Omnibus legislative process.[50] Penalties reach €35 million or 7% of global turnover for prohibited practices. 86% of business leaders with cybersecurity responsibilities have already experienced AI-related security incidents.[51] The era of voluntary, self-directed AI governance is ending. What follows is auditable, enforceable, and consequential.

ISO 42001 provides the operational framework for this transition, but the standard's value is realized only when all three layers of governance work in concert.

Documentation (Pillar 1) establishes what the organization intends to do: the policies, risk assessments, and governance structures that define how AI is managed. It creates the architectural blueprint. Without it, governance has no structure.

Continuous monitoring (Pillar 2) generates the evidence that the organization is doing what it documented: detecting drift, flagging anomalies, and producing audit-ready records in real time. It operationalizes the blueprint. Without it, governance has no proof.

Certification (Pillar 3) validates the entire system through independent, third-party assessment. It translates internal rigor into external trust. Without it, governance has no credibility.

Build it. Monitor it. Certify it.

The organizations that assemble this infrastructure now will have a measurable advantage: faster sales cycles driven by verifiable compliance credentials, reduced incident exposure through continuous detection, and a governance posture that satisfies auditors, regulators, investors, and enterprise procurement teams simultaneously. AI governance platform spending is projected to surpass $1 billion by 2030.[52] The market is moving. The standard exists. The enforcement timeline is fixed.

The question isn't whether to pursue AI governance. It's whether you'll be the organization that already has continuous trust in place when auditors, customers, and regulators come looking for proof.

---

[50] EU Regulation 2024/1689, Articles 6, 9, 26, 50. High-risk AI system requirements originally scheduled from August 2, 2026; subject to proposed Digital Omnibus extension (see footnote 4).

[51] Cisco 2025 Cybersecurity Readiness Index, April 2025. 86% of companies experienced AI-related security incidents.

[52] Gartner, February 2026. AI governance platform spending projected to reach approximately $500 million in 2026, surpassing $1 billion by 2030.

Contacts

DRATA

Drata provides the trust network that enables businesses to operate, scale, and partner with confidence. Powered by AI and designed to operationalize trust, the Drata Agentic Trust Management platform continuously interprets controls, risk, and assurance signals—reducing repetitive manual work while improving visibility into internal and third-party risk, enabling always-on audit readiness across frameworks, and accelerating security reviews. Purpose-built for enterprise complexity, Drata unifies governance, risk, compliance, and assurance to deliver faster time-to-value, reduce operational overhead, and enable continuous trust for 8,000+ organizations worldwide.
For more information, visit drata.com

Hala Alsagheer
halaalsagheer@drata.com

PRESCIENT SECURITY

Prescient Security provides audit, certification, and risk assessment services, ensuring governance systems operate effectively by aligning documented controls with real-world practices, helping organizations achieve compliance and build trusted, resilient security frameworks.

Sammy Chowdhury
sammy.chowdhury@prescientsecurity.com

RAIDS AI™

RAIDS AI is an AI safety and monitoring platform that provides continuous, real-time detection of anomalous or "rogue" AI behavior. Using non-invasive, black-box monitoring, it helps organizations maintain compliance, detect risks early, and ensure AI systems operate safely and reliably at scale.

Nikolas Kairinos
nik@raidsai.ai