© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-17122848

The Importance of Auditing Artificial Intelligence

SYED RIZWAN SHAHID
ISACA - USA (Information Systems Audit and Control Association)

Abstract- ***Artificial intelligence (AI) has moved from experimental deployment to embedded infrastructure across finance, operations, healthcare, retail, and public administration. As organizations delegate consequential decisions — credit scoring, fraud detection, hiring, medical triage, dynamic pricing, and regulatory reporting — to machine learning models and generative AI systems, the assurance function faces a widening gap between what these systems do and what boards, regulators, and stakeholders can verify. This paper argues that auditing AI is no longer a specialized niche but a core extension of internal audit's mandate to provide independent, objective assurance over governance, risk management, and control. It examines the drivers behind AI audit — financial, regulatory, ethical, and reputational — surveys the emerging standards landscape (IIA's Global Technology Audit Guide series, ISO/IEC 42001, NIST AI RMF, and the EU AI Act), and proposes a practical audit approach spanning governance, data, model, and monitoring layers. The paper concludes that organizations which fail to build AI-specific audit capability expose themselves to undetected model drift, algorithmic bias, regulatory non-compliance, and erosion of stakeholder trust — risks that are financially and reputationally material.***

I. INTRODUCTION

Artificial intelligence differs from traditional IT systems in a way that matters directly to auditors: it learns, adapts, and produces probabilistic rather than deterministic outputs. A conventional application executes a fixed set of rules that can be traced line by line; an AI model, particularly one built on machine learning or large language model (LLM) architectures, infers patterns from data and can behave in ways its own developers did not explicitly program. This shift from deterministic to probabilistic systems is the central reason AI requires its own audit discipline rather than a simple extension of general IT audit checklists.

Internal audit's traditional value proposition — independent assurance that risks are identified, controls are designed appropriately, and management's representations are reliable — applies with equal or greater force to AI. Where AI differs is in the mechanics of assurance: auditors must now evaluate training data lineage, model validation methodologies, explainability, bias testing, and continuous monitoring regimes that did not exist in traditional control frameworks a decade ago.

II. WHY AUDITING AI MATTERS

2.1 Financial and Operational Risk

AI systems increasingly drive decisions with direct financial consequence: automated underwriting, algorithmic trading, dynamic pricing, demand forecasting, and fraud detection. A miscalibrated or drifted model can generate losses at a speed and scale that manual processes never could — a flawed pricing algorithm can misprice thousands of transactions before anyone notices, and a fraud model that degrades silently can let material losses accumulate for months. Internal audit's role is to test whether management has controls that would catch this kind of failure before it becomes material, not after.

2.2 Regulatory and Legal Exposure

The regulatory perimeter around AI is expanding rapidly and unevenly across jurisdictions. The EU AI Act introduces risk-tiered obligations with substantial penalties for non-compliant “high-risk” AI systems. In the Kingdom of Saudi Arabia, the Saudi Data and Artificial Intelligence Authority (SDAIA) has issued AI ethics principles and generative AI guidelines, and data protection obligations under the Personal Data Protection Law (PDPL) directly constrain how AI systems may process personal data. Sector regulators — financial, healthcare, and telecom — are layering additional model risk management expectations on top. An organization that cannot demonstrate a defensible audit trail over its AI systems is exposed

not only to regulatory penalty but to the much harder-to-quantify cost of losing a regulator's confidence.

2.3 Algorithmic Bias and Ethical Risk

Models trained on historical data can encode and amplify historical bias — in lending, hiring, insurance pricing, and healthcare triage. Because bias in AI is often statistical rather than visible in any single decision, it evades traditional sample-based testing unless auditors specifically design for it. This is a distinct assurance problem: the absence of a single "bad" transaction does not mean the population is fair. Auditing AI for bias requires disaggregated outcome testing across protected classes and proxies for those classes, which is a fundamentally different technique from traditional substantive testing.

2.4 Model Opacity and the "Black Box" Problem

Complex models — particularly deep learning and generative AI — are frequently not interpretable even to their own developers. This opacity directly threatens two things auditors care about most: management's ability to explain a decision when challenged, and the auditor's ability to independently verify that a control is operating as designed. Explainability tooling (SHAP, LIME, counterfactual analysis) has matured enough that "the model is a black box" is no longer an acceptable answer to an audit inquiry — it is now itself an audit finding.

2.5 Third-Party and Vendor AI Risk

Most organizations do not build their core AI capability from scratch; they embed vendor AI — Microsoft Copilot, SAP Joule, Google Gemini, and similar platforms — directly into business processes. This introduces an assurance gap: the organization is accountable for outcomes it did not fully design and cannot fully inspect. Internal audit must extend its third-party risk methodology to cover AI-specific due diligence: what training data was used, what guardrails exist, how the vendor handles model updates, and what contractual right the organization retains to audit or exit.

2.6 Reputational and Trust Risk

AI failures are disproportionately newsworthy. A biased hiring algorithm, a hallucinating customer-service chatbot, or a discriminatory pricing model can generate reputational damage far exceeding the direct financial loss. Boards increasingly ask internal audit not "did we lose money" but "could this embarrass us publicly" — and AI is now one of the highest-probability sources of that kind of exposure.

III. THE EMERGING STANDARDS AND GOVERNANCE LANDSCAPE

A credible AI audit program should be anchored to recognized frameworks rather than built ad hoc:

- IIA Global Technology Audit Guides (GTAGs) on AI — provide internal-audit-specific guidance on auditing AI governance, risk, and controls, mapped to the IIA's International Professional Practices Framework (IPPF).
- ISO/IEC 42001 — the first international management-system standard for AI, giving auditors a certifiable framework analogous to ISO 27001 for information security, covering AI governance, risk assessment, and lifecycle controls.
- NIST AI Risk Management Framework (AI RMF) — organizes AI risk management around four functions (Govern, Map, Measure, Manage) and is widely used as a benchmarking framework even outside the United States.
- EU AI Act — establishes a risk-tiered regulatory regime (unacceptable, high, limited, minimal risk) with binding obligations for high-risk AI systems, including conformity assessments and post-market monitoring.
- National frameworks — including SDAIA's AI ethics principles and generative AI guidelines in Saudi Arabia, which increasingly shape what "reasonable AI governance" looks like for organizations operating in the region.

The convergence of these frameworks gives internal audit a defensible basis for scoping AI audits even in the absence of a single universal standard — auditors

© SEP 2026 | IRE Journals | Volume 10 Issue 3 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I3-17122848

can triangulate expectations across GTAG, ISO/IEC 42001, and NIST AI RMF, and calibrate for local regulatory requirements such as PDPL and SDAIA guidance.

IV. A PRACTICAL FRAMEWORK FOR AUDITING AI

An effective AI audit program can be organized into four layers, each with distinct test objectives:

4.1 Governance Layer

- Is there a board-approved AI policy and a defined accountable owner (AI governance committee, Chief AI/Data Officer, or equivalent)?
- Are AI use cases inventoried and risk-classified (e.g., low/medium/high impact) before deployment?
- Is there a documented approval gate before a model moves from development to production?

4.2 Data Layer

- Is training and fine-tuning data lineage documented and traceable to source systems?
- Are data quality, completeness, and representativeness assessed and evidenced before model training?
- Are personal data processing activities within the model lifecycle mapped against PDPL/GDPR-equivalent obligations?

4.3 Model Layer

- Was the model independently validated (performance, robustness, and bias testing) before go-live, and is that validation refreshed on a defined cycle?
- Is there documented explainability sufficient to support a decision challenge (regulatory, customer, or legal)?
- Are model risk tiers assigned, with proportionally deeper testing for high-impact models (credit, hiring, medical, safety)?

4.4 Monitoring and Change Layer

- Is model performance monitored in production for drift, degradation, and anomalous output patterns?
- Is there a defined incident response process for AI-specific failures (hallucination, bias detection, adversarial manipulation)?
- Are model retraining and version changes subject to change management controls equivalent to those applied to financial systems?

This four-layer structure allows internal audit to scale test depth to risk: a low-impact internal productivity tool (e.g., an AI meeting summarizer) warrants a lighter governance-layer review, while a high-impact model driving credit or medical decisions warrants full-depth testing across all four layers.

V. INTERNAL AUDIT'S EVOLVING ROLE

Internal audit functions face a capability decision: build in-house AI audit competency, partner with specialist third parties, or some hybrid of both. Certifications such as the IIA's Artificial Intelligence Auditing Certificate (AAIA) are emerging specifically to close this skills gap, giving auditors a structured grounding in AI risk taxonomy, testing techniques, and governance expectations without requiring a data-science background. Regardless of the staffing model chosen, the strategic point is the same: AI assurance cannot remain solely the responsibility of the teams that build and deploy the models. Independence is the entire value internal audit adds, and that value applies with particular force to AI, where the builders' incentives (speed to deployment, model performance metrics) can diverge from the organization's risk appetite.

VI. CONCLUSION

Auditing AI is not an optional specialization for forward-looking internal audit functions — it is an extension of the same assurance mandate that has always applied to financial, operational, and IT controls, adapted to a class of systems that is probabilistic, adaptive, and often opaque. The financial, regulatory, ethical, and reputational risks AI introduces are material and growing, and the standards landscape — from IIA's GTAG series to ISO/IEC 42001 and the EU AI Act — now gives auditors a workable basis to scope and execute this work. Organizations that treat AI governance as solely an IT or data-science concern, without

independent audit oversight, are accumulating risk that will surface as regulatory findings, biased outcomes, or public failures precisely when the organization can least afford it. The case for auditing AI is, ultimately, the same case that has always justified internal audit's existence: independent assurance is most valuable exactly where risk is least visible.

REFERENCES

[1] Institute of Internal Auditors (IIA) — Global Technology Audit Guide series on Artificial Intelligence

[2] International Organization for Standardization — ISO/IEC 42001:2023, Information technology — Artificial intelligence — Management system

[3] National Institute of Standards and Technology (NIST) — AI Risk Management Framework (AI RMF 1.0)

[4] European Union — Regulation (EU) 2024/1689 (EU AI Act)

[5] Saudi Data and Artificial Intelligence Authority (SDAIA) — AI Ethics Principles and Generative AI Guidelines