Independent AI Assurance for Boards

Idea In Short

Boards should require independent verification for material decisions, incidents and investigations involving Artificial Intelligence [AI] or automated systems. A management explanation may be accurate, even when it was assembled from the same systems that produced the outcome. That overlap creates a governance risk:

> the organization may investigate itself through the digital environment under review

Directors should therefore establish assurance paths that can examine source data, challenge assumptions, test controls and assign a named human owner. The objective is not for directors to rebuild models or replicate every automated decision. It is to ensure that material conclusions can be tested through evidence, methods or reviewers sufficiently separate from the systems that generated the original result.

Artificial Intelligence [AI] and automated systems now move information, coordinate activity, identify risk and, in some cases, make operational decisions across the enterprise. Their outputs often become inputs for other systems, whether through model scores, automated recommendations, cloud services or vendor-provided data. That architecture can improve speed and coordination, but it changes the board's ability to establish what happened after a material event. Directors should treat independent verification as a governance requirement wherever a system-generated account could influence accountability, remediation or a decision to preserve automated authority.

The company spent years building an increasingly connected enterprise. AI and automated systems move information, coordinate activity, identify risk and, in some instances, make decisions across the business. Increasingly, those systems also act on outputs produced

elsewhere in the digital environment:

> a calculation from one model, a recommendation from another system or data supplied by a vendor platform. Some operate inside the company; others depend on cloud platforms, outside models, vendors and data sources.

The result can be a faster, more responsive and, in many ways, more capable organization. The difficulty becomes visible when a serious issue reaches the board.

The directors begin with familiar questions:

> What happened? What caused it? Why was it not detected sooner? Who or what was responsible? And who now owns the result?

Management answers many of them. But as decisions increasingly pass between systems, no single person may be able to fully reconstruct the event. Management may therefore have to rely heavily on records generated by those systems, tracing how they interacted and identifying the sequence that produced the outcome.

Establish an independent route to facts

A board should require management to demonstrate how it can test a material system-generated explanation without depending entirely on the digital environment being examined. This does not mean directors must understand every technical detail, reproduce each automated decision or operate as forensic engineers. It means the organization needs a defensible route from a board-level conclusion back to underlying facts, evidence and assumptions.

The requirement becomes urgent after a serious failure. A customer may receive an incorrect decision, a fraud-control system may miss a pattern, a pricing engine may act on

corrupted data or an automated workflow may create a regulatory breach. Management will need to explain the event quickly, often by reconstructing a chain of activity across several systems. The reconstruction may be correct, but the board should distinguish between a persuasive narrative and independently verified evidence.

The National Institute of Standards and Technology [NIST] structures AI risk work around four functions:

1. govern
2. map
3. measure, and
4. manage

Its framework places governance across the lifecycle and calls for processes that identify, assess and manage AI risk rather than treating controls as a one-time technical exercise.AI RMF Core This provides a practical frame for boards:

> define accountability, map dependencies, measure the reliability of explanations and manage residual risk.

The board's core question should remain straightforward:

> How do we know this explanation is right?

That question shifts the discussion from whether management has an answer to whether the organization can test the answer through a sufficiently independent route.

Recognize the closed loop

A closed assurance loop emerges when connected systems generate an outcome, create

the records used to explain it, support the reconstruction of the event and supply the primary evidence used to validate that reconstruction. Each stage may work as designed. The governance concern lies in the lack of separation between the object under examination and the mechanism that verifies it.

Consider an automated credit process. One model ranks applications, a rules engine applies lending policy, a vendor service provides external data and a workflow platform issues the decision. When a complaint reaches the board, management may combine each system's logs into a coherent explanation. That account could accurately identify the sequence of events, yet still fail to expose a shared data problem, a faulty integration or an assumption embedded across several components.

A system can process millions of records and still deliver a wrong conclusion if its data are incomplete, mislabeled or inherited from a flawed source. The apparent precision of the reconstruction can make the account more difficult to challenge. Directors may assume that an explanation supported by extensive digital records deserves more confidence than an executive's account, even when the records depend on the same compromised environment.

The Organisation for Economic Co-operation and Development [OECD] calls for transparency, responsible disclosure and traceability across datasets, processes and decisions so that AI outputs can be analyzed and challenged in context.Recommendation Of The Council On Artificial Intelligence Traceability is necessary, but it is not sufficient. The board must also determine whether another person, method or evidence base can test what the trace reveals.

Separate evidence from explanation

Event logs, model outputs and workflow histories are evidence, but they also reflect the systems that generated them. A sound assurance process retains these artifacts while identifying evidence that does not depend on the same failure path. That evidence might include source-system records, independently preserved data extracts, external reconciliations, sample-based transaction testing or customer-level documentation.

A separate route does not require complete duplication of the production environment. It requires enough independence to identify correlated error. If two reviews rely on the same source data, transformation logic and evaluation criteria, agreement between them may

simply repeat the original mistake. The board should ask management to identify which inputs are common across the production system and the assurance process.

For major decisions, management should preserve an investigation package before making material changes to systems or data. The package should include event identifiers, relevant model versions, configuration details, source-data lineage, decision thresholds, vendor dependencies and records of human intervention. Without that preservation, later assurance may be forced to rely on reconstructed evidence rather than the state of the environment at the time of the event.

European Union rules for high-risk AI systems reinforce the operational value of traceability. They require deployers to assign human oversight to people with appropriate competence, training and authority, and to retain automatically generated logs under their control for an appropriate period, generally at least six months.Article 26: Obligations Of Deployers Of High-Risk AI Systems Boards should use those obligations as a minimum operational reference when designing evidence retention for material automated activity.

Treat vendor opacity as a board issue

Vendor services can extend the closed loop beyond the enterprise. A company may depend on an external model, cloud platform, data source or orchestration service that management cannot fully inspect. The absence of visibility does not remove accountability. It increases the importance of contractual controls and independent evidence.

Management should identify third-party dependencies in every material automated decision path. The inventory should state what the provider supplies, what data enters and leaves the service, which records management can access, what changes the provider can make and how the organization can investigate an incident. A board cannot oversee a dependency that remains invisible until it fails.

Contracts should support assurance, not merely procurement. Relevant provisions may include access to audit artifacts, incident-notification obligations, retention of technical records, change-management commitments, service-level controls and the ability to obtain independent assessments. The suitable design will vary by use case, but the principle remains consistent:

> a provider's internal explanation should not become the customer's only practical

proof.

The board should also define escalation triggers for vendor opacity. Where management cannot access material evidence, cannot establish data lineage or cannot obtain timely incident information, it should classify the dependency as a governance limitation. The decision to continue using the service then becomes a deliberate risk acceptance by an accountable executive rather than an unexamined technical constraint.

Design assurance beyond a second opinion

The medical second-opinion analogy is useful because the second doctor does not simply ask the first doctor whether the diagnosis was correct. The value comes from examining scans, test results and clinical reasoning independently. In automated decision-making, a second AI system may provide a challenge, but it does not automatically create independent assurance.

A second model can share the first model's data sources, assumptions, feature definitions or vendor infrastructure. If it does, agreement between the models may be less informative than it appears. A board should therefore focus on the independence of the assurance method, not the number of systems involved.

Internal audit, risk, compliance, external specialists and technical review teams can each contribute to assurance. Their roles should be defined according to the materiality of the use case, the speed of the decision process and the nature of the potential harm. The assurance design should state who may challenge management, what access they receive and how disagreements reach the board.

The Institute of Internal Auditors [IIA] describes governance, risk management and control as linked responsibilities that depend on clear organizational roles. Its standards and guidance provide a basis for positioning internal audit as an independent evaluator rather than as an operator of management controls.IPPF And Global Internal Audit Standards That distinction matters when internal audit reviews AI-enabled processes:

it should assess whether controls work, while management remains responsible for operating and improving them.

Test assumptions, not only outputs

An assurance review that checks only whether a system produced the expected output can miss the cause of a failure. Reviewers should examine the assumptions embedded in the decision process. These can include data definitions, thresholds, model objectives, fallback procedures, override rules, performance tolerances and the criteria used to label an outcome as successful.

For example, a fraud model might perform according to its selected accuracy metric while still producing unacceptable customer outcomes because its threshold is too aggressive for a particular segment. A review limited to model performance could approve the result. A review that examines the business decision, data distribution and customer-impact evidence may reach a different conclusion.

Management should identify assumptions that, if wrong, would materially change the board's view of the event. Those assumptions deserve targeted testing. The test may use a different data sample, an alternative calculation, manual case review, an external benchmark or a red-team challenge. The method must be proportionate, but it must be capable of producing a meaningful disagreement.

This approach aligns with the Committee of Sponsoring Organizations of the Treadway Commission [COSO] emphasis on monitoring as a component of effective internal control. Monitoring evaluates whether controls continue to operate as intended and provides a basis for identifying and correcting deficiencies.Internal Control - Integrated Framework In AI-enabled environments, monitoring should include the quality of the evidence used to explain important outcomes.

Preserve human accountability

Automation can distribute activity across multiple systems, teams and providers. It cannot distribute responsibility until no one owns the result. A board should require a named

executive to own each material automated process, including its intended purpose, delegated authority, risk thresholds, incident response and decision to continue or suspend its operation.

The accountable owner does not need to have written the model or operated every component. The role requires authority to obtain evidence, commission reviews, correct defects and accept or escalate residual risk. If no one has this authority, the board may receive an explanation without a party capable of acting on its findings.

Clear ownership also improves incident reporting. Management can explain which person accepted the system's operating authority, who approved changes, who validated the controls and who decided how to remediate the event. This avoids the unhelpful conclusion that "the systems did it," which describes a mechanism but not an accountable decision.

The owner should report material limitations candidly. These may include unavailable vendor evidence, uncertain data lineage, incomplete logs, inability to recreate a decision or assurance methods that share dependencies with the system under review. Naming the limitation allows directors to decide whether to accept it, demand compensating controls or reduce the system's authority.

Make challenge routine

Independent verification should not appear only after a crisis. Boards should embed challenge requirements into the lifecycle of material automated systems, starting with approval and continuing through monitoring, change management and incident response. Routine challenge reduces the pressure to improvise assurance when an event has already become consequential.

Management can begin by classifying automated processes according to financial exposure, customer impact, regulatory relevance, operational dependency and reversibility. The most consequential processes should have documented assurance plans before deployment. Lower-risk applications can use lighter controls, provided the board has confidence that the classification itself is reviewed as the use case evolves.

A usable board pack should not drown directors in technical detail. It should state the decision or incident under review, the systems and providers involved, the evidence

retained, the assumptions tested, the independent challenge performed, unresolved limitations and the named owner. Directors can then focus their attention on whether the assurance route is credible rather than attempting to interpret every technical artifact.

The board should expect management to answer five questions for every material system-generated explanation:

1. What outcome or event requires explanation
2. Which systems, models, data sources and providers contributed to it
3. What evidence supports the reconstruction
4. What independent method tested the reconstruction and its assumptions
5. Who owns remediation, residual risk and the system's continued authority

This discipline turns a broad governance concern into a repeatable operating practice. It also makes it harder for a persuasive digital narrative to substitute for verification. When management cannot answer one of these questions, the gap itself becomes relevant information for the board.

The World Economic Forum has emphasized governance practices that account for AI risks across design, deployment and use, including organizational roles and oversight arrangements.Empowering AI Leadership: AI C-Suite Toolkit Boards can adapt that principle by ensuring their reporting design links technical evidence to management accountability and board challenge.

Keep oversight credible

Corporate governance has long protected the independence of people who make consequential judgments. In an enterprise shaped by connected systems, boards must also protect independence in the routes used to verify facts. The issue is not whether automation should participate in investigation or assurance. It often must, because the systems hold relevant records and can analyze evidence at scale.

The issue is whether those systems become the only proof available to support their own account. If they do, directors may receive a detailed explanation without a practical way to test it. The explanation could be accurate, but the board would have no reliable basis for knowing.

A credible assurance design preserves evidence, tests assumptions, identifies shared dependencies, enables genuinely separate challenge and assigns responsibility to people with authority to act. It gives the board a way to ask whether the enterprise is governing automated activity or merely governing the enterprise's account of it.

> A system can provide an answer. It should not become the only proof that the answer is right.

For directors, the enduring questions remain simple:

> How do we know, what was independently tested and who owns the outcome?

Summary

Connected systems can accelerate decisions, expose patterns and coordinate work across complex enterprises. They can also create a closed loop when the same environment produces an outcome, records it, reconstructs it and validates the reconstruction. Boards can break that loop by defining materiality thresholds, preserving source evidence, separating assurance methods and keeping accountability with named people. Independent assurance does not require perfect visibility into every model or vendor platform. It requires a credible way to challenge the explanation presented to the board. When directors can ask what was tested, who tested it and whether the testing followed a genuinely different route to the facts, automated accountability remains governable.