Corrosion Management **ISSN: 1355-5243**
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

Artificial Intelligence and Internal Control Systems: A Theoretical Examination of Risk Detection, Monitoring, Compliance, and Organizational Accountability in Businesses

Arowolo, Isiaka O (Ph.D)
Department of management and Accounting,
faculty of Management and Social Science,
Lead City University.
Email: arowologbenga@yahoo.com **Phone No**: +234 803 308 8681

Abstract

Internal control systems exist to give organizations reasonable assurance that operations are effective, financial reporting is reliable, and applicable laws and regulations are followed — an assurance that has historically rested on periodic testing performed by people who can be asked, directly, why a particular control judgement was made. Artificial intelligence (AI) is now performing a growing share of the risk-detection and monitoring work internal control systems depend on, and this paper offers a theoretical examination of what that shift means for the control environment as a whole, not merely for the efficiency of individual control activities. Integrating the COSO Internal Control framework, agency theory, the Institute of Internal Auditors' Three Lines Model, and the NIST AI Risk Management Framework with recent empirical evidence on AI-enabled monitoring, automation, and algorithmic decision authority, the paper argues that AI strengthens two of internal control's traditional components — risk assessment and monitoring activities — more reliably than it strengthens a third, less often theorized requirement: organizational accountability for control failures. Because AI systems increasingly perform the detection and monitoring work that control frameworks assume a human occupies, the traditional assumption that a specific, identifiable person can explain why a control operated as it did becomes harder to satisfy exactly as detection and monitoring become more comprehensive and more automated. Drawing on the distinction between post-hoc explanation and interpretability by design, the paper proposes that this accountability gap is not simply a communication problem to be solved with better explanation tools, but a structural design choice organizations make when they select complex, opaque models over simpler, inherently interpretable ones for control-relevant decisions. The paper develops a four-dimensional theoretical model, compares the governance vocabulary of COSO, the Three Lines Model, and the NIST framework directly, and considers what the resulting account implies for how internal control systems, and the professionals who operate them, need to adapt.

Keywords: artificial intelligence; internal control; risk detection; continuous monitoring; regulatory compliance; accountability; COSO framework; three lines model; explainability.

Introduction

Internal control systems are built on a specific and, until recently, largely unexamined assumption: that the people responsible for designing, operating, and testing controls can be asked to explain their judgements, and that this explainability is itself part of what makes a control system trustworthy. The COSO Internal Control — Integrated Framework, the most widely adopted reference model for internal control globally, organizes this assumption around five interacting components — control environment, risk assessment, control activities, information and communication, and monitoring activities — each of which has historically been operated and evaluated by identifiable people occupying identifiable organizational roles. Artificial intelligence disrupts this assumption not by replacing internal control wholesale, but by increasingly performing the risk-detection and monitoring work that two of these five components specifically require, while remaining, in important respects, less explainable than the human judgement it supplements or replaces.

This is not the first time accounting scholarship has confronted claims about AI's transformative potential for control and assurance work. Sutton, Holt, and Arnold (2016), tracing three decades of AI research in accounting, find that predictions of AI's imminent transformation of the profession have circulated since at least the 1980s expert-systems era, and that the profession's actual encounter with AI has been considerably more gradual and cumulative than earlier waves of commentary anticipated. That historical pattern is a useful caution against treating the present moment as unprecedented, but it is not a reason to treat it as unremarkable: Murphy, Feeney, Rosati, and Lynn (2024), mapping the contemporary accounting-and-AI research literature through topic modelling across 930 peer-reviewed abstracts, find informational reliability, automation, and fraud detection among the field's most active current research clusters — precisely the concerns internal control systems exist to address, and evidence that the questions this paper takes up sit at the center of current scholarship rather than at its margins.

The evidence that AI is already performing internal-control-relevant work at scale is substantial. Ashraf (2025), examining the relationship between automation and financial reporting quality specifically through the lens of internal controls, finds automation associated with measurable improvements in control effectiveness — direct evidence that AI-enabled automation is not merely adjacent to internal control but is actively reshaping how control objectives are achieved. Eulerich, Waddoups, Wagener, and Wood (2024a), examining robotic process automation specifically, find that automated systems are frequently deployed without

commensurate attention to the control and governance principles that should accompany them, identifying control and security risks, underestimated costs, and difficult governance as recurring problems. Read together, this evidence suggests AI's relationship to internal control is neither straightforwardly positive nor straightforwardly negative; it is improving specific control objectives while simultaneously straining the governance assumptions those objectives were designed around.

This paper develops a theoretical account of that strain. Its central argument is that **AI strengthens internal control's risk-assessment and monitoring components more reliably than it strengthens organizational accountability for control failures, because accountability within existing control frameworks presumes an identifiable human judgement that AI-enabled detection and monitoring increasingly displaces or obscures**. This is not an argument against AI-enabled internal control — the evidence reviewed in this paper indicates real and substantial gains in detection and monitoring capability — but an argument that these gains create a specific governance problem that existing frameworks, including COSO and the Three Lines Model, were not designed to address and have not yet been systematically adapted to address. The paper reviews the theoretical foundations relevant to this argument, including a distinction between post-hoc explanation and interpretability by design that much of the applied internal-control literature has not yet absorbed; examines AI's contribution to risk detection, monitoring, and regulatory compliance in turn; places COSO, the Three Lines Model, and the NIST AI Risk Management Framework in direct comparison; and develops an account of the resulting accountability gap and what it implies for internal control governance going forward.

Conceptualizing AI's Role in Internal Control Systems

Internal control, in the COSO framework's formulation, is a process affected by an entity's governing body, management, and other personnel, designed to provide reasonable assurance regarding the achievement of objectives relating to operations, reporting, and compliance. The framework's five components interact rather than operate independently: risk assessment identifies what could prevent objective achievement; control activities are the policies and procedures that address identified risks; monitoring activities evaluate whether controls are operating effectively over time; and the control environment and information-and-communication components provide the organizational context within which the other three function. AI's growing role touches at least two of these five components directly — risk

assessment, where AI-enabled analytics increasingly identify risks a purely manual review would miss, and monitoring activities, where AI-enabled continuous monitoring increasingly supplements or replaces periodic manual testing.

Davenport and Ronanki (2018), writing for a general business audience, propose a simple typology for what AI applications in organizations are actually for: automating a process, generating insight from data, or engaging with people through language. This typology transfers usefully to internal control specifically, since AI's contribution to risk detection functions primarily as insight generation, its contribution to monitoring functions primarily as automation, and its emerging contribution to control-related communication — flagging exceptions, summarizing control test results, drafting compliance narratives — increasingly functions as the engagement category Davenport and Ronanki describe. This paper treats AI's contribution to internal control as operating along three practically distinct but conceptually related dimensions built on this typology: risk detection, the identification of conditions that threaten control objectives; monitoring, the ongoing evaluation of whether controls continue to operate as designed; and compliance, the demonstration that operations satisfy applicable legal and regulatory requirements. A fourth dimension, accountability — the capacity to identify who is responsible when a control fails and why — cuts across all three and is the dimension this paper argues is most poorly served by AI's current trajectory, precisely because accountability was never itself listed as a discrete COSO component; it was assumed to follow automatically from the human occupancy of the roles the five components describe.

Theoretical Foundations

The COSO Internal Control Framework and all its AI-Era Relevance

The Committee of Sponsoring Organizations of the Treadway Commission's Internal Control — Integrated Framework (COSO, 2013) remains the dominant reference model against which internal control systems are designed and evaluated, including for regulatory purposes such as compliance with the Sarbanes-Oxley Act's internal control reporting requirements. The framework's monitoring component specifically anticipates that monitoring may be performed through ongoing evaluations, separate evaluations, or a combination of both, and does not specify that monitoring must be performed by humans — a flexibility that has allowed AI-enabled continuous monitoring to be absorbed into existing control frameworks without requiring wholesale revision of the framework itself. This flexibility is also, this paper argues, precisely what has allowed the accountability question to go under-examined: because

the framework does not specify who or what performs monitoring, only that monitoring occurs, an AI system can satisfy the letter of the monitoring component while leaving open exactly the accountability question the framework's broader purpose — assurance to a governing body that can be held responsible — depends on resolving.

Agency Theory and Control as a Monitoring Mechanism

Jensen and Meckling's (1976) agency theory provides the governance rationale underlying internal control as an organizational function, treating monitoring mechanisms as a direct response to the agency costs created when ownership and control are separated. Internal controls, on this account, exist specifically to constrain the divergence between what agents (management and employees) might otherwise do and what principals (owners, and by extension the governing body acting on their behalf) require. This framing implies that a monitoring mechanism's value depends not only on its technical effectiveness at detecting agent behavior, but on its capacity to support principal accountability — the ability of the governing body to determine, when a control fails, whether the failure reflects an agent's misconduct, an agent's error, or a flaw in the control system itself. AI-enabled monitoring complicates this determination in a way traditional, human-operated monitoring did not: when an AI system fails to flag a risk it should have detected, the resulting agency-cost analysis has to account for a new category of failure — model error — that does not map cleanly onto the misconduct-versus-error distinction agency theory was originally built to resolve.

The Three Lines Model and the Question of Who Controls the Controller

The Institute of Internal Auditors' Three Lines Model (IIA, 2020), an update of the earlier Three Lines of Defense model, assigns risk-management responsibility across three organizational roles: first-line roles that own and manage risk directly within business operations, second-line roles that provide oversight and specialist support such as compliance and risk-management functions, and third-line roles, occupied by internal audit, that provide independent assurance to the governing body. Eulerich (2021), critically examining the transition from the Three Lines of Defense to the Three Lines Model, notes that the updated model deliberately broadens the earlier framework's rigid role assignments to allow more flexible reporting relationships between management and the governing body — a flexibility Eulerich argues was necessary given the increasing complexity of organizational risk environments, but one that does not, on its own, resolve where an AI-based detection or monitoring system actually sits within the model's three

lines. An AI system deployed within first-line operations to flag risk exposures is performing a function historically associated with first-line risk ownership, but it lacks the accountability characteristics — the capacity to be questioned, to exercise judgement recognized as such, to bear consequences — that the model's allocation of responsibility across lines presumes each line's occupants possess.

Explainability by Design Versus Post-Hoc Explanation

A fourth theoretical foundation, largely absent from the internal-control literature to date, comes from the machine-learning field's own internal debate about how model opacity should be addressed. Rudin (2019) draws a sharp and, for this paper's purposes, essential distinction between two different responses to the problem of black-box AI models used in high-stakes decisions: building supplementary tools that attempt to explain a complex model's behavior after the fact, and choosing, at the design stage, a model that is inherently interpretable in the first place, even at some cost to raw predictive accuracy. Rudin argues that the first approach — post-hoc explanation — is systematically unreliable for high-stakes decisions, because explanations generated after the fact can be unfaithful to what the underlying model actually computed, creating an illusion of accountability without the substance of it. Applied to internal control specifically, this distinction reframes the accountability gap this paper identifies as a design choice rather than an unavoidable cost of AI adoption: organizations that select complex, high-accuracy but opaque models for risk detection and monitoring are choosing, whether deliberately or by default, to rely on post-hoc explanation for control-relevant decisions, precisely the approach Rudin's analysis suggests is least trustworthy for the kind of high-stakes determination internal control exists to support.

AI and Risk Detection

AI’s contribution to risk detection, the first of the three practical dimensions this paper examines, is well evidenced across multiple risk categories relevant to internal control. Bao, Ke, Li, Yu, and Zhang (2020) demonstrate that machine-learning models trained on raw financial statement data substantially outperform established benchmark models in detecting fraud among publicly traded firms — a specific, high-consequence risk category internal control systems are directly responsible for addressing through the control-activities and monitoring components. Bertomeu, Cheynel, Floyd, and Pan (2021) extend this evidence to the related problem of ongoing misstatement detection, finding that machine learning combining accounting, audit, and market variables meaningfully improves detection relative to accounting

information alone — with the notable finding that accounting variables in isolation predict misstatements poorly, a result with direct relevance to internal control design, since it suggests risk-detection systems confined to a single functional data source, however sophisticated, are likely to underperform systems that integrate data across the first- and second-line functions the Three Lines Model distinguishes.

Ashraf (2025) extends this evidence beyond fraud and misstatement detection specifically to the broader relationship between automation and internal control quality, finding automation associated with measurable improvements in financial reporting quality attributable in part to strengthened internal controls, rather than automation and control quality operating as independent developments. This risk-detection capability is not confined to financial statement risk. Cooper, Holderness, Sorensen, and Wood (2019), examining robotic process automation across Big Four accounting firms, find RPA deployed across a range of processes with clear control implications, including reconciliation and exception-flagging tasks that function as de facto risk-detection mechanisms even where they were originally implemented primarily for efficiency reasons. The theoretical implication worth drawing out is that AI's risk-detection contribution to internal control has often arrived as a byproduct of automation initiatives pursued for other reasons — efficiency, cost reduction — rather than as a deliberate internal-control design choice, which has consequences for how systematically that contribution has been evaluated against control-framework requirements specifically, as opposed to evaluated on efficiency grounds alone.

AI and continuous Monitoring

From Periodic Testing to Continuous Control Monitoring

Trtaditional internal control monitoring has historically operated on a periodic cycle- quarterly control testing, annual management assessments, external audit procedures performed once a year — a cadence poorly matched to the pace at which control failures can actually occur and compound. Chan and Vasarhelyi (2011), tracing the development of continuous auditing as a successor paradigm to traditional periodic audit, argue that continuous auditing methodology fundamentally changes the timing relationship between when a control failure occurs and when it is detected, shifting from after-the-fact discovery toward near-real-time identification. Alles, Brennan, Kogan, and Vasarhelyi (2006), documenting a pilot implementation of continuous control monitoring at Siemens, provide an early and concrete demonstration that this shift is organizationally feasible at scale, finding that automated, continuous monitoring of business process controls could be implemented within an existing large organizational control

environment rather than requiring a wholesale redesign of that environment.

Evidence on AI-Enabled Monitoring Effectiveness

More recent, AI-specific evidence extends this continuous-monitoring paradigm considerably further than the largely rule-based systems Alles et al. (2006) examined. Choi and Xie (2026), studying generative AI use inside real accounting workflows, find AI adoption associated with increased general-ledger granularity and accountants selectively intervening when AI-generated confidence scores are low — a pattern consistent with AI performing genuine, ongoing monitoring rather than simply automating periodic testing at greater frequency. Kokina, Blanchette, Davenport, and Pachamanova (2025), interviewing professionals working with AI in accounting and assurance settings, find that this expanded monitoring capability comes bundled with a specific and unresolved cost: transparency and explainability problems that make it difficult for the professionals responsible for the control environment to fully understand why an AI-based monitoring system flagged, or failed to flag, a particular condition.

Human –AI Interaction as an Emerging Control Competency

Krakowski, Luger and Raisch (2023), extending resource-based theory to AI adoption through a controlled study of competitive chess, find that AI adoption triggers substitution and complementation dynamics operating simultaneously: certain traditional human capabilities become obsolete even as new, persistent sources of advantage emerge from the specific way particular individuals interact with AI systems. Applied to internal control monitoring specifically, this framework suggests that the professionals who operate AI-enabled monitoring systems are developing a genuinely new competency — not simply the technical skill of configuring a monitoring tool, but the judgement-intensive skill of knowing when an AI-generated flag warrants escalation and when it does not, a skill that does not reduce to either traditional manual control-testing expertise or generic data-science competency. This finding connects directly to the theoretical concern raised earlier in this paper: monitoring effectiveness, in the narrow sense of detection capability, is improving, while monitoring's contribution to the accountability the broader control framework depends on is not improving at the same pace, and the emerging human–AI interaction competency Krakowski et al. describe is precisely the skill organizations need to cultivate deliberately if that gap is to close rather than widen.

Comparing Governance Frameworks: COSO, thee Three Lines Model and the NIST AI Risk Management Framework

The accountability gap this paper identifies can be seen more clearly by placing COSO and the Three Lines Model directly alongside a framework developed specifically for AI risk

rather than adapted to it after the fact. The National Institute of Standards and Technology's AI Risk Management Framework (NIST, 2023) organizes AI governance around four functions — govern, map, measure, and manage — and, unlike COSO's monitoring component, explicitly and repeatedly addresses transparency and accountability as first-order design requirements rather than as implicit assumptions. The govern function specifically requires organizations to establish policies, processes, and procedures for accountability structures before an AI system is deployed, a sequencing that inverts the pattern this paper has identified in internal-control practice, where AI-enabled monitoring has generally been adopted first, with accountability questions addressed, if at all, only after control weaknesses or failures surface.

This comparison is instructive rather than merely descriptive. COSO's five components and the Three Lines Model's three roles were both developed to organize human responsibility within a control environment assumed to be operated by people; the NIST framework was developed after AI systems capable of opaque, high-consequence decision-making already existed, and its structure reflects that difference directly — accountability is a named, primary function, not a property assumed to follow from the framework's other components. The practical implication for internal control specifically is that organizations relying solely on COSO and the Three Lines Model to govern AI-enabled control systems are applying frameworks whose accountability assumptions were built for a different technological context, while frameworks built for the correct context, such as the NIST AI RMF, have not yet been systematically integrated into internal-control practice or into the auditing standards that evaluate it. Eulerich et al. (2024b), proposing governance principles specifically for robotic process automation, can be read as an early, domain-specific attempt at exactly this integration — but one that, on this paper's analysis, would benefit from more direct engagement with the NIST framework's explicit accountability-by-design sequencing rather than developing RPA-specific principles in relative isolation from the broader AI-governance literature.

AI and Regulatory Compliance

Regulatory compliance, the third practical dimension this paper examines, benefits from many of the same AI capabilities driving risk-detection and monitoring improvements, but introduces a distinct governance consideration: compliance failures typically carry direct legal and regulatory consequences that attach to specific, identifiable parties, in a way that internal risk-detection failures alone may not. Eulerich et al. (2024b) argue that control frameworks designed for manual and rule-based automated processes do not transfer cleanly to

AI-enabled ones, precisely because AI-enabled systems' probabilistic and adaptive character does not fit neatly within compliance frameworks built around fixed, auditable rules. This gap has direct consequences for the compliance function specifically: a compliance program built to demonstrate adherence to a fixed rule set has to be substantially reconceived when the underlying detection and monitoring systems generating the evidence of compliance are themselves adaptive systems whose behavior can change as they encounter new data.

Munoko, Brown-Liburd, and Vasarhelyi (2020), examining the ethical implications of AI in auditing more broadly, identify accountability for AI-generated conclusions as a distinct risk at the level of the firm — directly relevant to compliance specifically, since a firm that cannot clearly establish accountability for an AI-influenced compliance determination is poorly positioned to defend that determination to a regulator, even where the determination itself was substantively correct. Zhang, Zhu, Dai, and Chen (2023), examining the ethics of AI in managerial accounting more broadly, reach a structurally similar conclusion regarding transparency and accountability. Dwivedi et al. (2023), addressing generative AI's risks and opportunities across disciplines more broadly, reach a consistent conclusion from outside accounting specifically: productivity and detection gains from AI adoption are genuine, but so are risks around accuracy, bias, and accountability that organizations adopting the technology cannot assume away — a pattern this paper's analysis suggests applies with particular force to compliance functions, where the cost of an unaccountable determination is not merely reputational but directly regulatory.

Organizational Accountability in AI-Enabled Control Systems

The Accountability Gap

Bringing the preceding sections together, this paper identifies a specific accountability gap opening between AI's demonstrated contribution to risk detection and monitoring on one hand, and the accountability structures internal control frameworks depend on for their broader governance purpose on the other. Kim, Glaeser, Hillis, Kominers, and Luca (2024) supply directly relevant evidence on the human side of this gap: examining a municipal inspections context in which algorithms delivered genuine, measurable prediction improvements, they find that officials retaining override authority frequently exercised it, often without documenting reasons clearly connected to information the algorithm lacked — a pattern that, translated into an internal-control context, suggests human control operators may exercise override authority over AI-generated risk flags in ways that are themselves difficult to hold accountable, precisely

because the override decision is treated as an exercise of professional judgement not requiring the same documentation the underlying AI system's output does.

Kellog et al.(2020), examining algorithmic control more broadly across occupational contexts, add a further dimension to this gap: AI-based systems function simultaneously as tools that detect and monitor risk on an organization's behalf and as mechanisms through which the organization exercises control over the humans whose work those systems monitor. Applied to internal control specifically, this duality means an AI-based monitoring system is not a neutral instrument sitting outside the control environment it monitors; it is itself part of that environment, shaping the behavior of the people it monitors in ways the traditional, more clearly external conception of monitoring embedded in the COSO framework did not anticipate.

Risk Perception, Skill Change, and the Human Side of Accountability

The accountability gap this paper identifies is not solely a technical or structural problem; it has a documented behavioral dimension as well. Ross and Zhang (2025), surveying accounting practitioners on generative-AI use, find that perceived risk, rather than availability of the technology, is what most limits practitioners from extending their use of AI beyond low-stakes administrative applications toward the more consequential, judgement-intensive applications internal control monitoring increasingly requires — a finding this paper interprets as evidence that practitioners themselves may be implicitly aware of the accountability gap this paper theorizes, and responding to it by limiting their reliance on AI in exactly the high-stakes settings where that gap matters most. Bou Reslan and Jabbour Al Maalouf (2024), surveying 454 accountants on AI's effect on their work, find efficiency and skill effects occurring together, with AI adoption changing the skills required of accounting professionals rather than simply reducing the work available to them — evidence consistent with this paper's argument that the professionals responsible for internal control need a new, specifically accountability-oriented competency, not merely general AI literacy, if the gap this paper identifies is to close.

Algorithmic Control and the Three Lines Model Revisited

This paper's accountability-gap argument has a direct implication for how the Three Lines Model should be understood in an AI-enabled control environment. Eulerich's (2021) critical analysis of the model's transition from a rigid defense-line structure toward a more flexible role-based structure was motivated by organizational complexity generally; this paper argues that AI-enabled detection and monitoring introduces a specific form of complexity the model's flexibility does not, on its own, resolve — namely, that an AI system can simultaneously perform functions historically associated with more than one line. An AI-based continuous monitoring system

embedded in first-line operations, generating risk flags that second-line risk-management functions rely on, and producing an audit trail that third-line internal audit examines for assurance purposes, is not clearly assignable to any single line, and the accountability the model is designed to ensure — someone, somewhere, answerable to the governing body for each line's function — becomes correspondingly harder to trace to a specific accountable party when the same underlying system performs functions traditionally distributed across all three.

Toward an Integrated Model

This paper proposes that AI's effect on internal control be evaluated along the four dimensions developed throughout — risk detection, monitoring, compliance, and accountability — rather than through the single, aggregate judgement ("does AI improve internal control?") that much of the applied literature implicitly poses. Figure 1 summarizes this model. On the evidence reviewed here, risk detection and monitoring are improving in ways that are increasingly well documented empirically (Bao et al., 2020; Bertomeu et al., 2021; Ashraf, 2025; Choi & Xie, 2026). Compliance is improving unevenly, constrained by the mismatch Eulerich et al. (2024b) identify between fixed-rule compliance frameworks and adaptive AI systems, and by the gap this paper has identified between COSO's implicit accountability assumptions and the NIST framework's explicit ones. Accountability, this paper's central theoretical contribution, is not improving at the same pace as the other three, and in specific respects — the Kim et al. (2024) override-without-documentation pattern, the Kellogg et al. (2020) dual role of AI as monitor and controller, the cross-line ambiguity this paper identifies in the Three Lines Model — may be actively deteriorating even as the underlying detection technology improves.

Corrosion Management ISSN: 1355-5243
(https://corrosion-management.com/)
Volume 35, Issue 02 – 2025
SCOPUS

Figure 1. The AI-Enabled Internal Control Accountability Framework

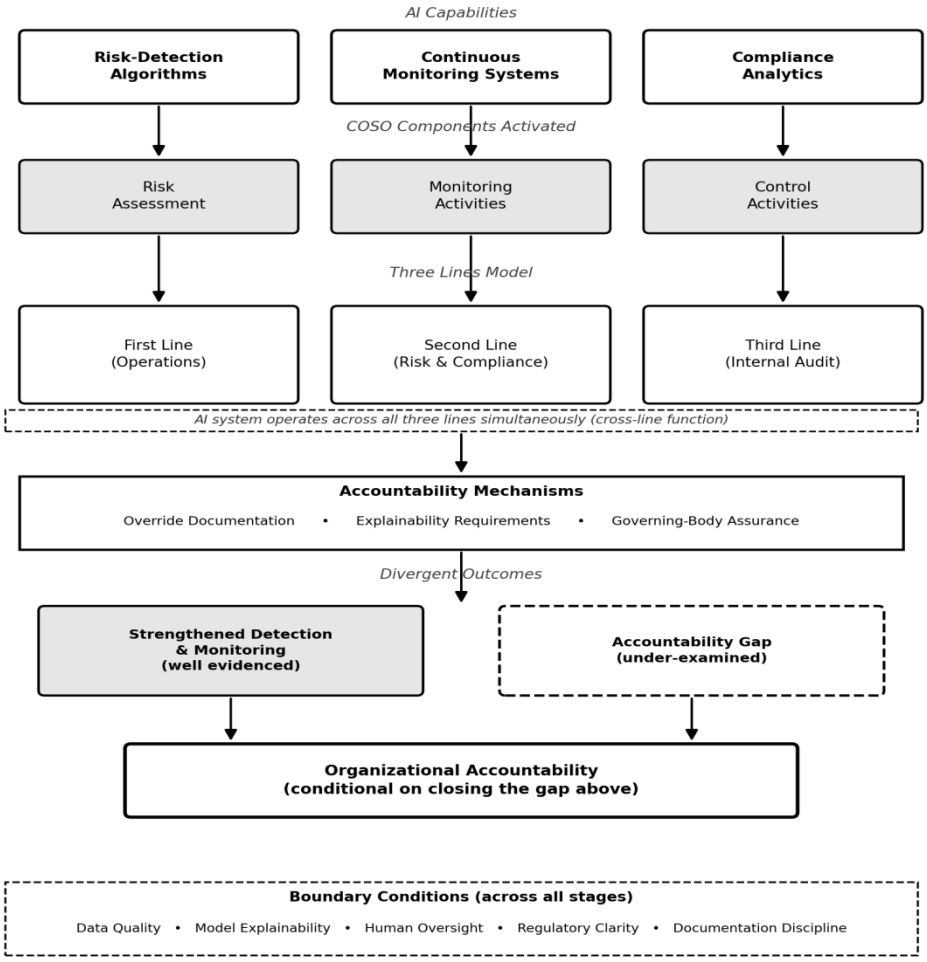

Figure 1. The AI-Enabled Internal Control Accountability Framework, showing AI capabilities flowing through COSO's components and the Three Lines Model toward divergent outcomes, conditioned on accountability mechanisms and boundary conditions.

The figure's central design feature is the divergence it depicts between two outcomes that follow from the same underlying AI capabilities: strengthened detection and monitoring, which the evidence reviewed in this paper supports as a well-evidenced and largely realized outcome, and an accountability gap, which this paper argues remains under-examined and, absent deliberate intervention, unresolved. Both outcomes pass through the same accountability mechanisms — override documentation, explainability requirements, governing-body assurance — but the evidence reviewed throughout this paper suggests these mechanisms are, at present, better developed for supporting the first outcome than for closing the second. Rudin's (2019) interpretability-by-design distinction offers one way to strengthen the mechanisms themselves: an organization that selects inherently interpretable models for its highest-stakes control

decisions, even at some cost in raw detection accuracy, is choosing an accountability mechanism built into the technology itself, rather than relying solely on override documentation and post-hoc explanation applied to an opaque system after the fact.

The practical significance of treating this four dimensions separately, rather than aggregating them into a single assessment, is that an organization can score well on the first three dimensions while its governing body's actual capacity to hold specific parties accountable for control failures deteriorates — a combination existing internal-control reporting, built around COSO's five components without a discrete accountability component, is not well positioned to detect. This paper's model suggests internal control frameworks need an explicit accountability component, distinct from but related to the existing five, specifically addressing whether the organization can trace AI-influenced control outcomes to a party capable of explaining and being held responsible for them — a requirement the framework's current monitoring component does not make explicit, the NIST framework's govern function already anticipates in the broader AI-governance context, and a gap the evidence reviewed in this paper suggests is widening rather than narrowing as AI-enabled detection and monitoring capability improves.

Implications for Practice

For Organizations, the model developed here suggests internal audit and compliance functions should evaluate AI-enabled control components against accountability criteria explicitly, not only against detection-accuracy criteria. Eulerich et al.'s (2024b) governance principles for RPA offer one starting template, but this paper's analysis suggests those principles need extension specifically to address cross-line ambiguity in the Three Lines Model — assigning clear ownership for AI-based systems that perform functions spanning more than one line, rather than assuming existing role assignments transfer automatically. For governing bodies and audit committees, the override-documentation gap Kim et al. (2024) identify suggests a specific, actionable control activity: requiring documented justification whenever a human overrides an AI-generated risk flag, creating exactly the accountability trail this paper argues current practice frequently lacks.

For model selection specifically, Rudin's (2019) distinction suggests a concrete design principle internal control functions can apply directly: reserving complex, high-accuracy but opaque models for lower-stakes, exploratory analytics, while requiring inherently interpretable models for control decisions that carry direct accountability consequences, such as decisions that could trigger a reportable control deficiency. For standard-setters, the gap between COSO's flexible, medium-agnostic monitoring component and the accountability problem this paper identifies

suggests the framework itself may benefit from more explicit guidance on AI-specific monitoring, drawing directly on the NIST framework's govern function and on the domain-specific principles Eulerich et al. (2024b) propose for RPA, generalized across the broader range of AI applications now embedded in control environments.

Directions for Future Research

Several gaps follow from the theoretical account developed above. First, the accountability gap this paper identifies has been developed conceptually rather than measured directly; research operationalizing accountability as a distinct, measurable construct — perhaps through documented-override rates, of the kind Kim et al.'s (2024) evidence suggests, or through audit-trail completeness for AI-influenced control decisions — would give this paper's central claim an empirical foundation it currently lacks.

Second, the cross-line ambiguity this paper identifies in the Three Lines Model has not been examined empirically; case-study research tracing how specific organizations have assigned ownership for AI-based detection and monitoring systems across first-, second-, and third-line roles would test whether the ambiguity this paper identifies theoretically is being resolved in practice, and if so, how.

Third, this paper's comparison of COSO, the Three Lines Model, and the NIST AI Risk Management Framework has been conducted at the level of framework design rather than organizational implementation; research examining how organizations that have formally adopted the NIST framework alongside COSO structure the resulting internal-control environment, relative to organizations relying on COSO alone, would test whether the accountability-by-design sequencing this paper attributes to the NIST framework actually produces the accountability improvements this paper's theory predicts.

Fourth, Rudin's (2019) interpretability-by-design argument was developed primarily with criminal-justice and healthcare applications in mind; research testing whether the accuracy cost of choosing inherently interpretable models over black-box alternatives is as significant in internal-control-relevant prediction tasks specifically — fraud detection, misstatement prediction — as it is in the domains Rudin examined would clarify whether this paper's proposed design principle is a genuinely low-cost recommendation for internal control or one that requires organizations to accept a meaningful accuracy trade-off.

Finally, comparative research examining whether the accountability gap this paper identifies varies systematically with regulatory environment — comparing jurisdictions with more prescriptive AI-governance requirements to those without — would help establish whether the

gap is primarily a technological problem awaiting better tools, or primarily a regulatory and governance problem awaiting clearer external requirements, a distinction with significant implications for where organizational and policy attention should be directed.

Conclusion

Artificial intelligence is demonstrably strengthening internal control systems' capacity to detect risk and monitor control effectiveness, and the evidence reviewed in this paper — spanning fraud and misstatement detection, automation-linked control quality improvements, and continuous monitoring capability — leaves little doubt that these are genuine and increasingly well-documented gains. This paper has argued, however, that evaluating AI's contribution to internal control on these terms alone significantly understates what is actually changing, because internal control's governance purpose depends on organizational accountability that AI-enabled detection and monitoring, for all their technical improvement, do not automatically preserve. Integrating the COSO framework, agency theory, the Three Lines Model, and the NIST AI Risk Management Framework with recent evidence on algorithmic override behavior, algorithmic control, and the distinction between post-hoc explanation and interpretability by design, this paper has proposed that accountability be treated as a fourth, explicit dimension of AI-enabled internal control, alongside risk detection, monitoring, and compliance, precisely because the evidence suggests it is the dimension least likely to improve on its own as the other three continue to advance.

The practical and theoretical implication is the same: organizations, standard-setters, and researchers should stop evaluating AI's contribution to internal control primarily by asking whether detection and monitoring have improved, and start asking the more demanding question this paper's model makes explicit — **whether the organization can still identify, and hold accountable, a specific party responsible for each control outcome AI now materially influences**. Internal control's value has always depended on this accountability as much as on detection accuracy; the evidence reviewed in this paper suggests AI is improving the latter considerably faster than it is preserving the former, and closing that gap deliberately — through explicit accountability components in control frameworks, through interpretable-by-design model choices for the highest-stakes decisions, and through governance structures that borrow the accountability-first sequencing frameworks such as the NIST AI RMF already provide — is the central governance challenge this paper's synthesis leaves for practice and future research alike.

References

Alles, M. G., Brennan, G., Kogan, A., & Vasarhelyi, M. A. (2006). Continuous monitoring of business process controls: A pilot implementation of a continuous auditing system at Siemens. *International Journal of Accounting Information Systems, 7*(2), 137–161. https://doi.org/10.1016/j.accinf.2006.05.001

Ashraf, M. (2025). Does automation improve financial reporting? Evidence from internal controls. *Review of Accounting Studies, 30*(1), 436–479.

Bao, Y., Ke, B., Li, B., Yu, Y. J., & Zhang, J. (2020). Detecting accounting fraud in publicly traded U.S. firms using a machine learning approach. *Journal of Accounting Research, 58*(1), 199–235. https://doi.org/10.1111/1475-679X.12292

Bertomeu, J., Cheynel, E., Floyd, E., & Pan, W. (2021). Using machine learning to detect misstatements. *Review of Accounting Studies, 26*(2), 468–519. https://doi.org/10.1007/s11142-020-09563-8

Bou Reslan, F., & Jabbour Al Maalouf, N. (2024). Assessing the transformative impact of AI adoption on efficiency, fraud detection, and skill dynamics in accounting practices. *Journal of Risk and Financial Management, 17*(12), Article 577. https://doi.org/10.3390/jrfm17120577

Chan, D. Y., & Vasarhelyi, M. A. (2011). Innovation and practice of continuous auditing. *International Journal of Accounting Information Systems, 12*(2), 152–160. https://doi.org/10.1016/j.accinf.2011.01.001

Choi, J. H., & Xie, C. L. (2026). Human + AI in accounting: Early evidence from the field. *Journal of Accounting Research, 64*(3), 1333–1373. https://doi.org/10.1111/1475-679X.70052

Committee of Sponsoring Organizations of the Treadway Commission. (2013). *Internal control—Integrated framework.* COSO.

Cooper, L. A., Holderness, D. K., Sorensen, T. L., & Wood, D. A. (2019). Robotic process automation in public accounting. *Accounting Horizons, 33*(4), 15–35. https://doi.org/10.2308/acch-52466

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review, 96*(1), 108–116.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy.

International Journal of Information Management, 71, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Eulerich, M. (2021). The new three lines model for structuring corporate governance—A critical discussion of similarities and differences. *Corporate Ownership & Control, 18*(2), 180–187.

Eulerich, M., Waddoups, N., Wagener, M., & Wood, D. A. (2024a). The dark side of robotic process automation (RPA): Understanding risks and challenges with RPA. *Accounting Horizons, 38*(2), 143–152. https://doi.org/10.2308/HORIZONS-2022-019

Eulerich, M., Waddoups, N., Wagener, M., & Wood, D. A. (2024b). Development of a framework of key internal control and governance principles for robotic process automation (RPA). *Journal of Information Systems, 38*(2), 29–49. https://doi.org/10.2308/ISYS-2023-067

Institute of Internal Auditors. (2020). *The IIA's three lines model: An update of the three lines of defense.* IIA.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals, 14*(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Kim, H., Glaeser, E. L., Hillis, A., Kominers, S. D., & Luca, M. (2024). Decision authority and the returns to algorithms. *Strategic Management Journal, 45*(4), 619–648. https://doi.org/10.1002/smj.3569

Kokina, J., Blanchette, S., Davenport, T. H., & Pachamanova, D. (2025). Challenges and opportunities for artificial intelligence in auditing: Evidence from the field. *International Journal of Accounting Information Systems, 56*, Article 100734. https://doi.org/10.1016/j.accinf.2025.100734

Krakowski, S., Luger, J., & Raisch, S. (2023). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal, 44*(6), 1425–1452. https://doi.org/10.1002/smj.3387

Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics, 167*(2), 209–234. https://doi.org/10.1007/s10551-019-04407-1

Murphy, B., Feeney, O., Rosati, P., & Lynn, T. (2024). Exploring accounting and AI using topic modelling. *International Journal of Accounting Information Systems, 55*, Article 100709.

https://doi.org/10.1016/j.accinf.2024.100709

National Institute of Standards and Technology. (2023). *Artificial intelligence risk management framework (AI RMF 1.0).* U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1

Ross, M., & Zhang, Y. (2025). ChatGPT is ready for the profession—But is the profession ready for ChatGPT? *Accounting Horizons, 39*(2), 185–204.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Sutton, S. G., Holt, M., & Arnold, V. (2016). "The reports of my death are greatly exaggerated"—Artificial intelligence research in accounting. *International Journal of Accounting Information Systems, 22*, 60–73. https://doi.org/10.1016/j.accinf.2016.07.005

Zhang, C., Zhu, W., Dai, J., & Chen, X. (2023). Ethical impact of artificial intelligence in managerial accounting. *International Journal of Accounting Information Systems, 49*, Article 100619. https://doi.org/10.1016/j.accinf.2023.100619