International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Preserving Professional Skepticism in the Age of Algorithmic Auditing: A Conceptual Framework for AI Adoption in Accounting Practice

Pathan Rabiya Aiyubkhan[1], Vaidhya Uday Narendrabhai[2]

[1,2]Asst. Professor, Department of Commerce, Kadi Sarva Vishwavidyalaya, Gandhinagar, Gujarat,

Abstract

Generative artificial intelligence (AI) has moved beyond the transactional automation that characterized earlier accounting technologies such as robotic process automation and rule-based expert systems. Large language models and machine-learning-based analytics can now draft audit memoranda, flag anomalies in real time, and generate narrative explanations that resemble professional judgment rather than mere calculation. This shift raises a question that the accounting literature has only begun to address systematically: at what point does AI-generated output stop assisting professional judgment and start substituting for it, and what happens to auditors' professional skepticism along the way? Drawing on recent systematic reviews and empirical studies of AI in accounting and auditing, this paper develops a conceptual framework — the Skepticism-Preserving AI Adoption (SPAA) framework — that classifies accounting and audit tasks into three tiers according to the degree of contextual and ethical judgment they require, and links each tier to a specific set of human-control mechanisms: task decomposition, explainability thresholds, mandatory disconfirmation protocols, and accountability anchoring. The paper argues that automation bias, rather than technical unreliability, is the primary risk of AI adoption in high-judgment accounting tasks, and that professional skepticism must be deliberately engineered into workflows rather than assumed to survive automation by default. Implications for audit practice, standard-setting, and accounting education are discussed, along with testable propositions to guide future empirical research.

Keywords: artificial intelligence; auditing; professional skepticism; automation bias; generative AI; accounting profession; audit judgment

1. Introduction

The accounting profession has absorbed successive waves of technology without much disruption to the underlying logic of the work. Electronic spreadsheets replaced ledgers, enterprise resource planning systems replaced isolated departmental records, and robotic process automation replaced repetitive keystrokes with scripted rule execution. In each case, a human accountant or auditor remained the party who interpreted what the numbers meant, exercised judgment about materiality and risk, and stood accountable for the conclusion reached. Generative AI complicates this pattern in a way earlier tools did not, because it does not merely execute rules faster; it produces outputs — draft risk assessments, narrative explanations of anomalies, synthesized disclosures — that read as though a professional, rather than a machine, had produced them (Stratopoulos & Wang, 2025).

The empirical record on AI adoption in accounting is by now reasonably well developed. Fedyk et al. (2022) found that audit offices adopting AI-based tools examined a broader range of financial statement items and were associated with a lower incidence of subsequent misstatement, suggesting real gains in audit thoroughness. A recent systematic review by Abdo-Salloum and Chehade (2026), synthesizing studies published between 2021 and 2025, likewise concludes that AI improves operational efficiency, strengthens fraud-detection capability, and enhances the credibility of financial reporting, while noting that implementation costs, uneven technological readiness across firms and regions, and unresolved ethical questions continue to constrain adoption. What this literature has not yet resolved, however, is a more basic conceptual question: which accounting and audit tasks can be handed to AI with reasonable safety, and which cannot — not because the technology is unreliable, but because the task itself depends on a form of judgment that automation tends to erode rather than replicate.

This paper addresses that gap. Rather than adding another adoption study or efficiency estimate to an already substantial body of empirical work, it develops a conceptual framework intended to help practitioners, standard setters, and educators reason systematically about where AI belongs in the accounting workflow and what safeguards each placement requires. The central argument is that the chief risk of AI adoption in judgment-intensive accounting tasks is automation bias — the tendency of a human reviewer to defer to a system's output and to search less actively for disconfirming evidence — and that professional skepticism, far from being a fixed trait auditors either have or lack, is a workflow property that must be deliberately preserved through design choices once AI enters the process.

The paper proceeds as follows. Section 2 reviews the literature on AI's evolution in accounting practice, its documented effects on audit judgment, the behavioral concept of automation bias, and the regulatory and educational responses that have emerged. Section 3 introduces the Skepticism-Preserving AI Adoption (SPAA) framework, which classifies accounting tasks into three tiers and proposes four corresponding control mechanisms, along with four propositions intended to guide future empirical testing. Section 4 discusses implications for practice, regulation, and accounting education. Section 5 acknowledges the limitations inherent in a conceptual paper of this kind, and Section 6 concludes.

2. Literature Review

2.1 From Robotic Process Automation to Generative AI

Kokina and Davenport (2017) were among the first to distinguish, in the accounting literature, between automation that executes predefined rules and the pattern-recognition capabilities of machine learning, arguing that the latter would eventually allow audit tools to identify anomalies that a rules-based system could never have been programmed to catch. Their prediction has largely borne out: contemporary audit platforms use machine learning to flag unusual journal entries, cluster transactions by risk profile, and continuously monitor data outside the narrow window of a year-end sample. Stratopoulos and Wang (2025) describe the more recent shift toward generative AI — large language models capable of producing fluent text, code, and structured analysis — as a qualitatively distinct development, because these systems generate outputs in the same medium (natural language, narrative reasoning) that professionals have traditionally used to communicate judgment. A model that once flagged a transaction as statistically unusual can now also draft the explanatory memo describing why the transaction appears unusual and what follow-up procedures might be warranted, collapsing a step that used to require an auditor's independent synthesis.

This progression matters conceptually because it changes what is being delegated. Robotic process automation delegates execution; machine-learning anomaly detection delegates pattern recognition; generative AI delegates something closer to explanation and argument. Sundarasen et al. (2026), in a bibliometric mapping of the AI-enabled auditing literature, note that research attention has followed this same arc, moving from early studies of adoption and efficiency toward a more recent cluster of work concerned specifically with audit quality and the judgment implications of AI-generated content. Their analysis suggests the field itself has recognized that adoption and quality are not the same question, even though much practitioner discourse still treats them interchangeably.

2.2 AI-Enabled Audit Judgment and Risk Assessment

Risk assessment sits close to the center of the audit process, and it is precisely the area where AI's involvement is most consequential. Fedyk et al. (2022) provide some of the clearest large-sample evidence that AI adoption changes the substance of audit work: audit offices with greater AI investment examined more financial statement line items, suggesting that AI expanded the effective scope of testing rather than simply speeding up existing procedures. This is a genuinely positive finding, and it is easy to see why practitioner and vendor discourse has emphasized it. Yet expanded coverage is not the same as improved judgment about the items that are examined, and the literature on audit judgment more broadly cautions against assuming that broader data coverage automatically produces better conclusions, particularly when the analyst reviewing AI output has strong incentives — time pressure, budget constraints, client relationship considerations — to accept a plausible-looking answer without probing it further.

Abdo-Salloum and Chehade's (2026) systematic review reinforces this more cautious reading. Across the 31 studies they synthesize, AI's benefits for fraud detection and reporting credibility appear fairly consistently, but so does a recurring concern about the uneven readiness of firms and jurisdictions to use these tools responsibly, and about the ethical and governance gaps that adoption has outpaced. In other words, the literature does not describe a technology that is unreliable; it describes a technology that is being deployed faster than the professional and organizational safeguards needed to use it well.

2.3 Automation Bias and the Erosion of Professional Skepticism

Professional skepticism — an attitude that includes a questioning mind and a critical assessment of evidence — is the auditing profession's primary defense against being persuaded by a plausible but incorrect explanation. Brown-Liburd, Issa, and Lombardi (2015) anticipated, in the context of big data rather than generative AI specifically, that the sheer volume and apparent precision of algorithmically produced information could paradoxically reduce auditors' critical engagement with it, because a large, seemingly comprehensive dataset can create an impression of thoroughness that discourages further inquiry. Munoko, Brown-Liburd, and Vasarhelyi (2020) extend this concern directly to AI, arguing that opacity in how an AI system reaches a conclusion (the well-known "black box" problem), combined with auditors' limited technical ability to interrogate that reasoning, creates conditions under which skepticism is more likely to be suspended than exercised. Their analysis frames this as an ethical issue as much as a technical one: an auditor who defers to an AI-generated risk assessment without independent scrutiny has not necessarily done anything careless by the standards of current practice, yet the profession's traditional accountability structure assumes that a human, not a system, made the judgment call.

This dynamic is what the behavioural decision-making literature calls automation bias — the tendency to treat a system's output as a correct default and to expend less cognitive effort verifying it than one would for a colleague's judgment. The concern is not that AI systems are typically wrong. In many narrow tasks they may be more consistent than a fatigued or time-pressured human reviewer. The concern is that when

they are wrong, or right for the wrong reasons, a reviewer operating under automation bias is less likely to notice, precisely because the professional skepticism that would normally catch such an error has been quietly displaced by trust in the tool.

2.4 Regulatory and Standard-Setting Responses

Standard setters have begun to respond, though cautiously. The International Auditing and Assurance Standards Board (IAASB) has emphasized that its recently revised quality management standards were designed in part to reinforce professional skepticism as firms adopt new technologies, and the board has indicated it intends to pursue non-authoritative guidance on AI rather than prescriptive rules, reflecting both the pace of technological change and uncertainty about which specific practices will prove durable (International Auditing and Assurance Standards Board, 2025, 2026). This approach is defensible given how quickly the underlying tools are evolving, but it also means that, for the time being, firms and individual practitioners carry most of the responsibility for deciding how much independent verification an AI-assisted conclusion requires. That responsibility is difficult to discharge without a conceptual basis for distinguishing tasks where AI assistance is low-risk from tasks where it is not — which is precisely the gap this paper's framework is intended to address.

2.5 Educational and Workforce Implications

A parallel literature examines whether accounting education is preparing practitioners for this environment. Tiron-Tudor, Cordos, and Deliu (2025) compared employer skill demands, drawn from job-posting data, against accounting students' self-assessed competencies, and found a persistent mismatch: students consistently undervalued the digital, analytical, and AI-oriented skills that employers increasingly list as requirements, particularly in the categories the authors associate with Industry 5.0-style human-machine collaboration. This mismatch has a direct bearing on the argument developed here. A framework that assigns humans responsibility for maintaining skepticism over AI-assisted judgment is only useful if practitioners are actually trained to exercise that responsibility — to know what a disconfirmation check looks like, to recognize when an AI-generated explanation is being accepted too readily, and to understand enough about how the underlying model works to ask it meaningful follow-up questions. Where that training is absent, the safeguards proposed in Section 3 exist on paper but not in practice.

3. The Skepticism-Preserving AI Adoption (SPAA) Framework

3.1 Conceptual Foundations

The framework proposed here rests on a single premise drawn from the literature reviewed above: the risk that AI adoption poses to accounting and audit quality is not uniform across tasks, because tasks differ in how much of their difficulty lies in mechanical computation versus contextual, ethical, or relational judgment. A reconciliation task that matches thousands of transactions against a bank statement is difficult mainly because of volume; a going-concern assessment is difficult mainly because it requires weighing ambiguous, sometimes contradictory evidence about a client's future and forming a defensible opinion under uncertainty. AI is well suited to the first kind of difficulty and poorly suited, at least as currently understood, to bearing final responsibility for the second — not because it cannot generate a plausible-sounding going-concern narrative, but because the profession has no mechanism for holding a model accountable for that narrative being wrong.

The SPAA framework therefore organizes accounting and audit tasks into three tiers based on the type of judgment they require, and it pairs each tier with a control mechanism specifically designed to counteract the form of automation bias most likely to arise at that tier. The tiers are not meant to map cleanly onto

job titles or existing task taxonomies; they are meant to be applied at the level of the individual task or sub-task, since a single audit engagement typically contains work spanning all three tiers.

3.2 Three Tiers of Accounting and Audit Tasks

Table 1 summarizes the three tiers. Tier 1 tasks are mechanical and rule-bound: transaction matching, arithmetic recalculation, standard journal entry testing, and similar procedures where correctness can be verified against an objective, predetermined standard. Tier 2 tasks involve pattern recognition and risk flagging: anomaly detection, ratio analysis, fraud-risk scoring, and other procedures where AI identifies candidates for human attention but the significance of what it finds is not self-evident. Tier 3 tasks require contextual and ethical judgment: materiality determinations, going-concern assessments, related-party evaluations, and other conclusions that depend on interpreting ambiguous evidence within a specific organizational and regulatory context, and for which the auditor or accountant bears personal professional and legal responsibility.

Tier	Nature of task	Example procedures	Appropriate AI role
Tier 1 — Mechanical / rule-bound	Correctness is verifiable against an objective, predetermined standard	Transaction matching, recalculation, standard journal-entry testing, three-way match	Full automation with periodic human audit sampling
Tier 2 — Pattern recognition / risk flagging	AI identifies candidates for attention; significance requires interpretation	Anomaly detection, ratio and trend analysis, fraud-risk scoring, continuous monitoring alerts	AI-augmented; human reviews every flagged item before action
Tier 3 — Contextual / ethical judgment	Depends on interpreting ambiguous evidence within a specific context; carries personal professional accountability	Materiality determination, going-concern opinion, related-party evaluation, professional skepticism calls	AI-assisted drafting only; human forms and owns the conclusion

3.3 Four Control Mechanisms

Placing a task in a tier is only useful if it is paired with a control appropriate to that tier's risk. The framework specifies four such mechanisms.

Task decomposition requires firms to break composite workflows into their constituent tiers before deciding how AI will be used, rather than adopting a tool for an entire process (such as "the risk assessment") as a single unit. A risk assessment, decomposed, typically contains Tier 1 data-gathering steps, Tier 2 anomaly flagging, and a Tier 3 conclusion; automating all three identically invites the Tier 3 conclusion to inherit an unearned sense of reliability from the Tier 1 and Tier 2 work that preceded it.

Explainability thresholds set a minimum standard of interpretability that rises with tier: a Tier 1 tool needs only to be verifiably accurate, but a Tier 2 or Tier 3 tool must be able to show the reviewer which factors drove a given output, in terms a professional without a data-science background can evaluate. Munoko et al.'s (2020) concern about black-box opacity is, in this framework, treated as tolerable at Tier 1 and increasingly unacceptable moving up the scale.

Mandatory disconfirmation protocols require the human reviewer of Tier 2 and Tier 3 outputs to document at least one piece of evidence or one line of reasoning that would have supported a different conclusion than the one the AI produced, before that conclusion is accepted. This directly targets automation bias by making disconfirmatory search a required step rather than a matter of individual initiative, which behavioral research suggests is unlikely to occur reliably on its own once a plausible answer is already on the screen.

Accountability anchoring preserves the principle that a named human professional, not the system, is responsible for every Tier 2 and Tier 3 conclusion, and that this responsibility is documented at the point of decision rather than inferred after the fact. This is less a technical control than an organizational and cultural one, but the literature on IAASB's standard-setting direction suggests it is precisely the kind of firm-level practice that non-authoritative guidance is likely to recommend rather than mandate (International Auditing and Assurance Standards Board, 2025, 2026).

3.4 Propositions

The following propositions translate the framework into a form suitable for future empirical testing.

P1: Audit teams that decompose composite tasks into discrete tiers before selecting AI tools will exhibit lower rates of unexamined acceptance of AI-generated conclusions than teams that adopt AI at the level of a whole workflow.

P2: The relationship between AI use and audit quality at Tier 2 and Tier 3 tasks is moderated by the explainability of the AI tool, such that greater explainability strengthens the positive relationship between AI use and audit quality.

P3: Requiring a documented disconfirmation step before accepting an AI-generated Tier 2 or Tier 3 conclusion will be associated with a higher rate of conclusion revision than review processes without such a requirement.

P4: Practitioners who receive explicit training in interrogating AI-generated outputs will display lower levels of automation bias, measured as unexamined acceptance of AI recommendations, than practitioners without such training — consistent with the skills gap documented by Tiron-Tudor et al. (2025).

4. Discussion and Implications

4.1 Implications for Audit and Accounting Practice

For practicing firms, the SPAA framework offers a practical starting point that does not require waiting for authoritative regulatory rules that may be years away. A firm can, in principle, audit its own current AI deployments against the three tiers immediately: which tools are being used for which tasks, whether any Tier 3 conclusions are being generated with a degree of automation more appropriate to Tier 1, and whether disconfirmation and accountability documentation exist anywhere in the current workflow or only in engagement methodology manuals. Firms that find Tier 3 conclusions moving through review with Tier 1-level scrutiny have, in effect, allowed automation bias to expand the practical scope of automation beyond what the underlying technology or governance can support.

4.2 Implications for Standard Setters and Regulators

The IAASB's preference for non-authoritative guidance (International Auditing and Assurance Standards Board, 2025, 2026) is a reasonable response to a fast-moving technology, but it leaves a gap that a tiered framework can help fill without requiring prescriptive, tool-specific rules that would likely be outdated before they took effect. Rather than regulating specific AI products, standard setters could regulate the process discipline this framework describes — for instance, requiring firms to document which tier a given AI-assisted conclusion falls into and what disconfirmation evidence was considered, regardless of which vendor's tool produced the underlying output. Such a requirement would be technology-neutral and would age considerably better than guidance tied to a particular generation of tools.

4.3 Implications for Accounting Education

The skills-gap evidence from Tiron-Tudor et al. (2025) suggests that accounting curricula still emphasize the acquisition of technical accounting knowledge more than the specific skill of interrogating an AI-generated conclusion. Embedding the SPAA framework, or something like it, into audit and assurance coursework would give students a vocabulary and a concrete practice — tiering a task, then asking what disconfirmation evidence would change the answer — that is more transferable across specific AI tools than training on any single current platform is likely to be, given how quickly those platforms change.

4.4 Directions for Future Research

The propositions in Section 3.4 point toward an obvious next step: empirical testing, whether through experimental designs that manipulate explainability or disconfirmation requirements and measure resulting judgment quality, through field studies of firms at different stages of AI adoption, or through content analysis of actual audit workpapers to assess how often Tier 3 conclusions currently receive Tier 3-appropriate scrutiny. Sundarasen et al.'s (2026) observation that the field's research attention is already shifting from adoption toward quality suggests the timing for such studies is favorable.

5. Limitations

This paper is conceptual rather than empirical, and the framework it proposes has not yet been tested against field or experimental data; the propositions in Section 3.4 are offered as a starting point for that testing, not as established findings. The three-tier classification is also a simplification: some tasks will not fit neatly into a single tier, and the boundary between Tier 2 and Tier 3 in particular is likely to be contested in specific cases, such as fraud-risk scoring that shades into a judgment about a client's integrity. Finally, because generative AI capabilities are advancing quickly, some of the specific examples used to illustrate each tier may need revision even in the near term, though the underlying logic of the framework — that judgment-intensive tasks require different safeguards than mechanical ones — is intended to be more durable than any particular example.

6. Conclusion

Generative AI has given the accounting profession genuinely useful new capabilities, and the empirical literature reviewed here is consistent in finding real gains in efficiency, coverage, and fraud detection. The open question is not whether AI belongs in accounting and audit workflows, but how firms, regulators, and educators ensure that professional skepticism survives its arrival intact. This paper has argued that skepticism is not a personal trait that automatically persists once AI enters the process; it is a property of how work is organized, and it can be preserved only through deliberate design choices — decomposing tasks by the type of judgment they require, demanding explainability proportional to that judgment,

building in mandatory disconfirmation steps, and anchoring accountability to a named human professional. The Skepticism-Preserving AI Adoption framework offered here is one attempt to make those design choices explicit and testable. Whether or not this particular framework proves durable, the underlying claim — that automation bias, not model inaccuracy, is the central risk AI poses to judgment-intensive accounting work — is one the profession will need to take seriously well before regulators finish writing authoritative rules for it.

References

1. Abdo-Salloum, A. M., & Chehade, S. (2026). The role of artificial intelligence in transforming accounting and auditing practices: A systematic review. SAGE Open, 16(1). https://doi.org/10.1177/21582440251403296
2. Brown-Liburd, H., Issa, H., & Lombardi, D. (2015). Behavioral implications of Big Data's impact on audit judgment and decision making and future research directions. Accounting Horizons, 29(2), 451–468.
3. Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? Review of Accounting Studies, 27, 938–985. https://doi.org/10.1007/s11142-022-09697-x
4. International Auditing and Assurance Standards Board. (2025, July). IAASB highlights how revised standards reinforce professional skepticism [Press release]. IFAC. https://www.iaasb.org/news-events/2025-07/iaasb-highlights-how-revised-standards-reinforce-professional-skepticism
5. International Auditing and Assurance Standards Board. (2026, February). IAASB publishes global roundtable feedback on technology and quality management [Press release]. IFAC. https://www.iaasb.org/news-events/2026-02/iaasb-publishes-global-roundtable-feedback-technology-and-quality-management
6. Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. Journal of Emerging Technologies in Accounting, 14(1), 115–122.
7. Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. Journal of Business Ethics, 167, 209–234. https://doi.org/10.1007/s10551-019-04407-1
8. Stratopoulos, T. C., & Wang, V. X. (2025). Artificial intelligence and accounting research: A framework and agenda. International Journal of Accounting Information Systems, 56. https://www.sciencedirect.com/science/article/pii/ Sundarasen, S., Kamaludin, K., & Nakiran, D. (2026). From adoption to audit quality: Mapping the intellectual structure of artificial intelligence-enabled auditing. Journal of Risk and Financial Management, 19(3), Article 209. https://doi.org/10.3390/jrfm19030209
9. Tiron-Tudor, A., Cordos, A. L., & Deliu, D. (2025). Future-ready digital skills in the AI era: Bridging market demands and student expectations in the accounting profession. Technological Forecasting and Social Change, 215. https://www.sciencedirect.com/science/article/abs/pii/S0040162525001362