3rd Annual International Conference on Accounting Development (ICAD 2026)
Department of Accounting, Ahmadu Bello University Zaria · 2–4 September 2026

ARTIFICIAL INTELLIGENCE AND COGNITIVE BIAS IN ACCOUNTING: FROM HUMAN BIAS TO ALGORITHMIC BIAS

Nana Fatimah Nasir[1,*]
[1]*Ahmadu Bello University*
Email: bintnasir2468@gmail.com[*]
***Corresponding author**: Nana Fatimah Nasir
DOI: https://doi.org/10.67203/sp.ICAD.2026

ABSTRACT

In recent years, there has been evident increase in the adoption of Artificial Intelligence in accounting. This integration of artificial intelligence into accounting estimates, audit judgement and financial risk assessment shifts the focus on bias in financial reporting. Behavioural accounting research has shown that cognitive biases distort human judgements especially in areas like materiality assessments and auditing. Although the use of AI has the potential to reduce these human biases, it has also begun to raise concerns about the emergence of algorithmic bias.

This conceptual paper examines the relationship between AI and cognitive bias in accounting specifically focusing on the transition from human bias to algorithmic bias. It uses available literature on behavioural accounting and artificial intelligence to explore how human biases may influence accounting decisions and how AI systems may affect these biases. This paper also considers the role of automation bias and excessive reliance on AI generated recommendations in shaping accounting judgements. It argues that AI should not be viewed as inherently objective and bias free but seen as a tool that can improve accounting judgement depending on the quality of the data, design of the system and the involvement of accounting professionals.

The paper contributes to the emerging behavioural accounting and AI in accounting literature by reframing Artificial Intelligence as a potential source of algorithmic bias and not just a tool for reducing human bias.

Keywords: *artificial intelligence; cognitive bias; algorithmic bias; accounting estimates; automation bias*

1. INTRODUCTION

1.1 Background

Accounting is fast becoming a hybrid decision system. Artificial intelligence (AI) has moved from the surface of accounting practice to its operational core. The accounting practice has absorbed artificial intelligence more quickly than the standards meant to govern it. Audit analytics platforms now flag anomalous transactions across entire populations rather than samples; expected credit loss (ECL) models under IFRS 9 increasingly rely on machine learning to estimate probability of default, loss given default, and exposure at default (Uniqus, 2025) and large language models are being used to summarize disclosures, draft audit memoranda, and support going-concern judgments (Jones and Free, 2026). This integration of AI into judgment-heavy accounting tasks has been optimistically framed as a correction to the well-documented biases of human judgment.

Behavioural accounting research has spent more than four decades researching how cognitive biases distort accounting and audit judgment. Biases such as anchoring, availability, and overconfidence, first formalized by Tversky and Kahneman (1974), have been shown to shape auditor risk assessment, materiality judgments, and going-concern opinions (Camilli et al., 2025). The implicit promise of AI adoption is that data-driven, algorithmic processing can reduce the variance and systematic error introduced by these human biases. If algorithms process data consistently and without fatigue, self-interest, or affect, then their adoption should, in principle, narrow the gap between accounting estimates and the economic reality they are meant to represent.

1.2 Problem Statement

This paper argues that this promise is only half true. AI relocates rather than eliminates the problem of biases in the accounting judgement process. On one hand, human reliance on AI systems

introduces automation bias. This is a well-documented tendency to over-trust automated outputs and discount contradictory evidence (Yoshua Bengio, 2026) thereby weakening the professional skepticism that audit quality depends on. On the other hand, the AI systems themselves are not completely objective. They are trained on historical data that reflects prior human decisions, market conditions, and unrepresentative patterns (Noreen et al, 2026). The result is a second, structural form of bias namely algorithmic bias, that can distort accounting estimates independently of, and often invisibly to, the humans reviewing them.

1.3 Research Questions

This paper poses three research questions.

RQ1: How do cognitive biases affect human accounting and audit judgment, and how does the existing literature characterize their mechanisms?

RQ2: How does AI adoption reduce, relocate, or amplify these human biases through automation bias

RQ3: What forms does algorithmic bias take when AI systems are applied to accounting-specific estimation tasks, and how does this differ from the human bias problem the profession has traditionally studied?

1.4 Contribution Statement

The paper's central contribution is to reframe AI in accounting research not just as a tool that reduces human cognitive biases, but also as an independent source of bias that current audit and governance frameworks are not designed to detect. This reframing has direct implications for auditing standards, professional scepticism models, and the allocation of responsibility when AI-informed estimates prove materially wrong. This contribution speaks directly to the conference’s theme of strengthening financial integrity. An accounting estimate distorted by unaudited algorithm is not distinguishable from one distorted by deliberate earnings management to the financial users.

2. LITERATURE REVIEW

2.1 Cognitive Bias in Accounting and Audit Judgment

Behavioural accounting research across several decades has shown that accounting and audit judgment departs from the assumptions of rational decision-making. Anchoring bias i.e. the tendency to rely disproportionately on an initial reference point, such as a prior-period balance or a client's proposed estimate, has been shown to shape materiality thresholds and adjustment decisions. Overconfidence leads accounting personnel and auditors to overstate the precision of their own risk assessments, particularly under time pressure. Confirmation bias leads reviewers to seek evidence that supports an initial hypothesis about an account balance or risk area rather than evidence that would disconfirm it. Camilli et al. (2025) in a comprehensive recent review organizes 139 accounting studies into categories of easy-attribution, emotion-driven, and frame-dependence biases confirms that these effects occur across audit, financial accounting, and management accounting tasks. Their review emphasizes the fact that despite decades of acknowledgment, the field lacked a comprehensive categorisation until recently. Empirical work in terms of auditor risk assessment finds that judgment failures interact with certain pressures such as time constraints and task complexity that push auditors toward fast, bias-driven reasoning rather than slower, deliberative evaluation. Bahar (2026) argues that risk assessment failures often occur when auditors default to fast, bias driven reasoning under pressure. Together, this literature establishes cognitive bias in accounting judgment as patterned and predictable rather than incidental, a foundation the remainder of this paper builds on.

2.2 AI Adoption in Accounting and Auditing

Machine learning has become a mature tool in credit risk and financial distress prediction. Deep learning models are now outperforming traditional logistic regression in bankruptcy and default classification (Jones and Free, 2026). In the specific context of IFRS 9 expected credit loss estimation, financial institutions increasingly use statistical and machine learning methods to model probability of default, loss given default, and exposure at default, with AI driven ECL platforms marketed explicitly

on their ability to process large, high-dimensional datasets in real time (IFRS 9 Expected Credit Loss Modelling Platforms Market Research Report, 2025).

Recent literature has documented how quickly artificial intelligence has moved into accounting's core estimation and assurance tasks. In fraud detection, forensic accounting research documents growing reliance on machine learning and natural language processing to flag anomalous transactions. A recent systematic review of seventy-six studies finds that supervised learning, ensemble methods, and deep learning outperform rule-based detection when adequate data and skilled interpretation are available, though practical integration into forensic workflows remains uneven outside large, data-mature organizations (Meneses Ceron et al., 2026). Evidence from Nigeria's financial services sector similarly finds that natural language processing capability moderates the effectiveness of AI-enabled fraud detection, suggesting adoption benefits are conditional rather than automatic (Adeyemi and Ojogbede, 2026). In going-concern assessment, supervised machine learning models are increasingly used to predict or classify outcomes from financial and textual data (Hedback, 2025). Beyond assurance tasks, robo-advisory platforms apply algorithmic portfolio construction directly to investment decisions, a domain where behavioural finance research finds that perceived transparency and interactivity of the algorithm shape whether it mitigates or reproduces investor biases such as overconfidence and loss aversion (MS Kulkarni et al, 2025).

2.3 Automation Bias and Human Reliance on AI Outputs

This section examines automation bias, which is the tendency of technology users to over-rely on automated outputs while discounting contradictory information (Yoshua Bengio, 2026). This tendency has been documented across domains including aviation and medical diagnostics, and is now surfacing in accounting and audit contexts. The International Auditing and Assurance Standards Board flagged automation bias and overreliance as an unintended consequence of audit technology adoption as early as 2021 (Bukowski-Kruszyna, 2024), a concern that has increased with the spread of generative AI into audit workflows.

Recent professional commentary reinforces this concern. Practitioner-oriented analysis argues that professional scepticism becomes more important as AI adoption increases, and identifies automation bias as a growing risk precisely because responsibility for the underlying estimate remains with the human accountant or signing auditor regardless of how the number was generated (Nyenrode, 2026). A systematic review of bias and ethics in artificial intelligence enabled auditing reinforces this concern, finding that unchecked adoption risks reproducing longstanding profession wide challenges around analytical subjectivity and transparency, and calling explicitly for governance and assurance reform rather than treating adoption as self-correcting (Murikah et al, 2024). This relationship is however not uniform across studies. A study by the Illinois CPA Society (2026) finds that experienced auditors sometimes exhibit AI aversion rather than automation bias, driven by a threatened sense of professional competence relative to newly demanded digital skills. This suggests reliance on AI is not uniform across auditor experience levels or organizational contexts

Regulatory attention to automation bias has also begun to be taken into consideration. The European Union's AI Act requires that human overseers of high-risk AI systems be made explicitly aware of the tendency to over-rely on automated outputs, acknowledging that awareness alone is an incomplete safeguard against cognitive bias tendency (Laux 2025).

2.4 Algorithmic Bias: Origins and Mechanisms

The algorithmic bias literature is mainly concentrated in computer science, credit scoring, and lending regulation rather than in accounting terms. However, it offers an account of how artificial intelligence systems themselves can generate systematically biased outputs. Historical or training data bias occurs when a model is trained on data reflecting past decisions or conditions that were themselves discriminatory or unrepresentative. A recent systematic review of 55 peer-reviewed studies on AI credit scoring concludes that bias in these systems is not primarily a technical failure but originates in historically generated training data (Noreen et al, 2026). A form of this bias is Proxy bias, in which models trained predominantly on data from one demographic or entity type perform disproportionately well for that group and poorly for others (Kozodoi et al., 2022). Finally, feedback-loop bias describes

the dynamic in which entities flagged as high-risk receive more scrutiny, generating more negative data points about that group and reinforcing the model's original classification over time.

This literature has been built almost entirely outside the accounting discipline, concentrated instead in lending, hiring, healthcare, and criminal justice applications.

2.5 Synthesis and Identification of the Gap

These four literatures reviewed describe bias in their own individual way. Behavioural accounting research treats bias as a property of human cognition and locates the bias in an individual. AI adoption research in accounting treats AI as a performance and implementation improving tool, with little consideration to its effects on judgements. Automation bias research which is still emerging in accounting especially treats bias as an emergent property of the human AI interaction, but rarely considers what may be wrong with the algorithm itself. Algorithmic fairness research treats bias as a property of the model and its training data. It is well developed outside of accounting and rarely engages with accounting specific constructs such as materiality, professional scepticism, accounting estimation or auditor liability. No literature identified in this review integrates human, automation, and algorithmic biases into a single framework applied to accounting judgment. This is the gap the present paper addresses conceptually.

3. CONCEPTUAL FRAMEWORK: THE DUAL-BIAS MODEL

This paper proposes the Dual-Bias Model as a conceptual framework for understanding how bias operates in artificial intelligence mediated accounting judgment. The model's claim is that human cognitive bias and algorithmic bias operate as two parallel, distinct pathways that don’t offset one another but instead converge at the point of the final accounting estimate or judgment. Figure 1 presents the model.

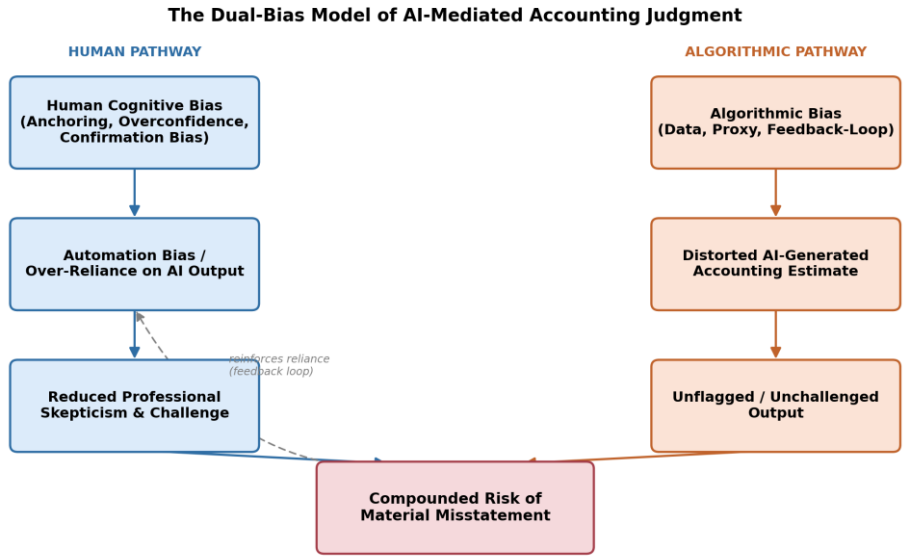


Figure 1: The Dual Bias Model of AI Mediated Accounting Judgment.

3.1 The Human Pathway

The left hand pathway reflects the traditional behavioural accounting account. Human cognitive biases such as anchoring on prior period figures, overconfidence in one's own risk assessment shape how preparers and auditors approach an estimate. When an AI system is introduced into this workflow, automation bias acts as an amplifying mechanism rather than a corrective one. Rather than neutralizing the human tendency toward bias shortcuts, over reliance on AI output can substitute one bias such as the auditor's own prior belief, for another bias like trust in the model's output, without necessarily increasing the rigor of the underlying judgment. The consequence then is reduced professional scepticism and that auditors and accounting personnel become less likely to interrogate an estimate that arrives is data driven and algorithmically generated.

3.2 The Algorithmic Pathway

The right hand pathway is independent of human psychology. Algorithmic bias arising from unrepresentative training data, or self reinforcing feedback loops distorts the AI generated estimate itself, before any human ever reviews it. Because these distortions are often embedded in high dimensional models that are difficult to interpret, the resulting output can appear internally consistent and well calculated even when it is systematically wrong for particular populations, sectors, or time periods.

3.3 Convergence and Compounded Risk

The model's key conceptual contribution is the convergence point. Neither pathway alone fully explains the resulting material misstatement risk. The two pathways converge in compounded risk of material misstatement, and the dashed feedback arrow in Figure 1 captures a further dynamic. Once reliance on an artificial intelligence output goes unchallenged in one reporting cycle, both the automation bias and the reduced scepticism response tend to strengthen in the next cycle. This is because the absence of an adverse consequence in the prior period serves as a confirming evidence that the tool can be trusted.

The human pathway feedback loop is different from the feed-back loop bias mechanism described in Section 2.4. The feed back loop bias mechanism describes how a model's own biased classifications generate the training data used to update the model enforces the algorithmic distortion on the algorithmic pathway. The two loops are different but mutually reinforcing. An algorithm whose bias is compounding through its own feedback loop is, under Preposition 1, less likely to be challenged by a reviewer whose scepticism is eroding through the human pathway loop. This reinforcing loop means the model predicts that dual-bias risk should be expected to grow, not remain constant, over successive periods of unmonitored artificial intelligence reliance.

3.4 Propositions

The following propositions are offered to guide future empirical testing rather than as hypotheses tested here.

- Proposition 1: Automation bias reduces the frequency and rigor of professional scepticism applied to AI-generated accounting estimates, relative to estimates generated through traditional human judgment alone.
- Proposition 2: Algorithmic bias arising from unrepresentative training data produces systematic, rather than random, distortion in AI-generated accounting estimates, with distortion magnitude varying by the degree of underrepresentation of the affected entity, sector, or population in the training data. This claim follows two mechanisms previously reviewed in Section 2.4 namely proxy bias and feed back loop bias. Proxy bias systematically shifts estimates for whichever entities the correlated variable disadvantages. Feedback loop bias systematically compounds that shift in a specific direction over successive model updates.
- Proposition 3: The interaction of human automation bias and algorithmic bias compounds material misstatement risk beyond what either factor would produce independently.
- Proposition 4: Auditor experience and digital-skill confidence moderate the relationship between AI reliance and professional scepticism, such that low digital-skill confidence may produce AI aversion rather than automation bias, while high confidence without technical fairness-testing literacy may produce automation bias without corresponding bias detection capability.

4. DISCUSSION

The Dual Bias Model reframes a belief present in accounting literature about AI adoption in accounting. The dominant view of AI adoption in accounting research and practice has been corrective, believing that AI is expected to reduce the variance and systematic error associated with human cognitive limitation (Jones and Free, 2026). The reviewed literature suggests this view is incomplete in two aspects.

First, even where AI reduces certain forms of human bias e.g. by applying consistent thresholds across a full transaction population rather than a judgmentally selected sample, it does not eliminate bias from the judgment process. It relocates the point at which bias enters, from the initial human estimate to the point of reliance on the AI-generated estimate. Nyenrode, (2026) argues that professional scepticism is now framed as more important, not less, in an AI-augmented environment, precisely because responsibility remains with the human signatory regardless of how the estimate was produced.

Secondly, current audit methodology and standards were designed to interrogate human judgment. They were not designed to test for algorithmic bias, which requires a fundamentally different process most of which is largely outside the technical training of practicing accountants and auditors. The consequence is that an estimate can satisfy every conventional test of audit evidence while still being the product of an algorithmically biased model, because none of these tests were designed to detect that particular bias.

5. IMPLICATIONS

5.1 Theoretical Implications

This paper builds behavioural accounting theory, which has historically treated the human decision-maker as the sole focus of bias. It proposes framework in which bias is produced by human cognition and algorithmic design interacting with one another. This extension aligns accounting theory with emerging perspectives in the broader AI governance literature, which similarly reject the framing of AI systems as neutral tools operating independently of the human and institutional biases in which they are used.

5.2 Practical Implications

For audit firms and users of artificial intelligence in accounting and auditing, the model suggests that AI governance frameworks accompanying AI adoption should explicitly separate two categories of control. Firstly, controls designed to sustain human professional scepticism toward AI outputs, and secondly, controls designed to test the AI system itself for bias. Standard-setters including the International Auditing and Assurance Standards Board and the Public Company Accounting Oversight Board have begun to acknowledge automation bias as a risk category, but neither body has yet articulated a comparably developed framework for algorithmic bias testing specific to accounting estimates.

5.3 Regulatory and Standard-Setting Implications

For standard-setters such as the IAASB and national regulators, the model implies that existing guidance on automation bias and overreliance (Bukowski-Kruszyna, 2024) addresses only the human pathway. Algorithmic bias testing has developed primarily within financial services fair lending regulation rather than within auditing or financial reporting standards specifically. There is an identifiable need for auditing standards to incorporate bias testing procedures for AI derived accounting estimates as a distinct category of audit evidence, alongside the human-oversight provisions already emerging in broader AI regulation such as the EU AI Act.

6. CONCLUSION AND FUTURE RESEARCH AGENDA

This paper has argued that the accounting profession's engagement with Artificial Intelligence has, to date, been shaped on an incomplete view that AI is a correction to human cognitive bias. Drawing on behavioural accounting, automation bias, and algorithmic fairness literatures, this paper proposes the Dual-Bias Model, which treats human cognitive bias and algorithmic bias as two distinct but interacting sources of biases in AI mediated accounting judgment that converge to compound material misstatement risk in ways neither pathway produces alone.

This paper is conceptual rather than empirical and as a result is subject to certain limitations. The Dual-Bias Model and its propositions are gotten from existing literature across behavioural accounting, Human-AI interaction and Algorithmic fairness. They have not been tested against financial

statement or experimental data. The extent of the compounding effect in preposition 3 is therefore an open empirical question rather than an established result.

Several research gaps follow this framework and merit future empirical and methodological attention.

- There is no established, accounting-specific methodology for auditing AI driven accounting estimates for algorithmic bias, in contrast to the fairness-testing protocols developed in lending and credit-scoring contexts.
- Empirical evidence on automation bias specifically within accounting and audit judgment as opposed to aviation, medical, or general human-factors contexts remains limited and would benefit from experimental replication using accounting-specific estimation tasks.
- The auditor liability and professional responsibility implications of relying on a materially biased AI-generated estimate remain unresolved in professional standards and largely untested in practice or case law.
- Cross-cultural and cross-jurisdictional research is needed, given that most AI accounting and audit tools are developed and trained predominantly on data from developed, Western financial markets and then deployed globally.
- The moderating role of auditor experience and digital-skill confidence, identified in this paper's Proposition 4, warrants direct empirical testing to distinguish automation bias from AI aversion as distinct organizational risks requiring different interventions.

As AI derived estimates become further embedded in financial reporting, the capacity to interrogate not only the humans who rely on these tools, but the tools themselves, will be central to sustaining the reliability of financial statements.

REFERENCES

Adeyemi, T. O., & Ojogbede, A. O. (2026). Artificial intelligence-enabled accounting information systems and fraud detection in Nigeria's financial services sector: The moderating role of natural language processing [Preprint]. arXiv.

BAHAR, D. S. (2026). AUDITOR COGNITIVE BIAS IN RISK ASSESSMENT: A REVIEW OF COGNITIVE PSYCHOLOGY THEORY. SSRN.

BUKOWSKI-KRUSZYNA, K. (2024, MAY 27). OVER-RELIANCE ON AUTOMATION: A CAUTIONARY TALE FROM PLATO. ICAEW INSIGHTS.

Camilli, R., Cristofaro, M., Hristov, I., & Sargiacomo, M. (2025). Cognitive biases in accounting judgment and decision making: A review, a typology, and a future research agenda. Accounting Forum, 49(5), 1038–1067.

Growth Market Reports. (2025). IFRS 9 expected credit loss modeling platforms market research report [Market research report].

Hedback, A. (2025). Identifying going concern audit opinions using supervised machine learning. Intelligent Systems in Accounting, Finance and Management.

ILLINOIS CPA SOCIETY. (2026). WHAT'S REALLY BEHIND AI AVERSION IN AUDITING? INSIGHT MAGAZINE.

Yoshua Bengio, International AI Safety Report Expert Panel. (2026). International AI safety report [Independent expert report]. arXiv.

Jones, S., & Free, C. (2026). What accountants need to know about artificial intelligence and machine learning: A review and call for future research. Journal of Accounting Literature, 48(1), 133–148.

Kozodoi, N., Jacob, F., & Lessmann, S. (2022). Fairness in credit scoring: Assessment, implementation and profit implications. European Journal of Operational Research, 297(3), 1083–1094.

Laux, J. (2025). Automation bias in the AI Act: On the legal implications of attempting to de-bias human oversight of AI. arXiv.

Meneses Cerón, L. Á., Fernández Rengifo, E., Pino Terán, V. E., Albán Lara, M. A., Vega Sánchez, R. C., & Fuentes Sánchez, J. J. (2026). Artificial intelligence and machine learning in the transformation of forensic accounting: A systematic literature review. Journal of Risk and Financial Management, 19(7), Article 519.

Murikah, W., Nthenge, J. K., & Musyoka, F. M. (2024). Bias and ethics of AI systems applied in auditing: A systematic review. Scientific African, 25, Article e02281.

Noreen, R., Sajid, M., & Amjad, F. (2026). Algorithmic bias and fairness in AI credit scoring: Evidence, mechanisms, and governance responses. Journal of Management Science Research Review, 5(2), 1546–1558.

NYENRODE BUSINESS UNIVERSITEIT. (2026). AI IN AUDITING: BALANCING QUALITY, PROFESSIONAL SCEPTICISM, AND RESPONSIBILITY.

MS Kulkarni (2025). The role of robo-advisors in behavioural finance, shaping investment decisions. (2025). Cogent Business & Management, 12(1), Article 2571403.

UNIQUS. (2025, SEPTEMBER 25). EXPECTED CREDIT LOSSES UNDER IFRS 9.