JEBA Vol.1 No.2 (2025) – 4929 – pp. 26-38

Journal of Entrepreneurship and Business Administration

journal homepage: http://journals.lagh-univ.dz/index.php/jeba

Calibrated Skepticism: A Conceptual Framework of Auditor Reliance on Artificial Intelligence

Ammar BENNOUAR [a,*]

[a] Faculty of Economics and Commercial Sciences, and Management, University of Tiaret - Algeria

a r t i c l e i n f o

Article history:
Received:
17 June 2025
Accepted:
15 September 2025
Published:
01 October 2025

Keywords:
Foreign bank entry
Domestic bank performance
Professional skepticism
Artificial intelligence
Audit judgement
Automation bias
Algorithm aversion
Audit quality
Conceptual framework

JEL Classification:
M42, M41, O33, D91

a b s t r a c t

Purpose: Artificial intelligence (AI) is rapidly entering the financial statement audit, yet the empirical literature reports two contradictory reliance failures: auditors sometimes defer uncritically to AI output (automation bias) and sometimes discount reliable AI evidence after observing its errors (algorithm aversion). This paper develops a theory that explains when each failure arises and how professional skepticism can be preserved in AI-enabled audits.

Design/methodology/approach: The paper adopts a conceptual research design combining problematization, an integrative review of auditing, judgement and decision-making, and human–automation research, theory synthesis and model building. Professional skepticism serves as the domain theory, while trust-in-automation and algorithm-aversion research serve as method theories.

Findings: The paper introduces the construct of calibrated skepticism, defined as the alignment of an auditor's reliance on AI output with that output's actual reliability for a given assertion, sustained through skeptical judgement and action. A framework specifies AI-system, auditor and task–context antecedents of calibrated skepticism, eight testable propositions and a typology of four reliance stances: calibrated partnership, blind deference, reflexive rejection and ritualistic compliance.

Practical implications: The framework informs standard setters revising audit-evidence requirements, audit firms designing AI governance and review procedures, regulators planning inspections and educators developing AI literacy in accounting curricula.

Originality/value: The paper reframes over- and under-reliance on AI as two symmetrical failures of professional skepticism rather than as separate phenomena, and offers a research agenda for experimental, archival and qualitative audit research.

* Corresponding author.
E-mail addresses: A. Bennouar a.bennouar@univ-tiaret.dz
How to cite: Bennouar, A. (2025). Calibrated Skepticism: A Conceptual Framework of Auditor Reliance on Artificial Intelligence. *Journal of Entrepreneurship and Business Administration, 1*(2), 26-38.
DOI: https://doi.org/10.34118/jeba.v1i2.4929
This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).
ISSN: 3088 - 8980

1. Introduction

Artificial intelligence (AI) is transforming how audit evidence is obtained and evaluated. Machine-learning models now score journal entries for anomalies, natural-language tools read contracts and minutes, and predictive models produce independent estimates of complex accounts such as expected credit losses or fair values (Kokina & Davenport, 2017; Issa et al., 2016; Appelbaum et al., 2017). Archival evidence suggests that AI investment by audit firms is associated with fewer restatements and lower fees, indicating potential gains in both effectiveness and efficiency (Fedyk et al., 2022). Standard setters have responded: the Public Company Accounting Oversight Board (PCAOB) amended its audit-evidence and risk-response standards to address technology-assisted analysis for fiscal years beginning on or after 15 December 2025 (PCAOB, 2024), and the International Auditing and Assurance Standards Board (IAASB) has re-exposed a revised ISA 500, Audit Evidence, which addresses information generated by automated tools (IAASB, 2026).

These developments place professional skepticism, the cornerstone of audit quality (IAASB, 2009; Nelson, 2009), under new strain. Yet empirical research on how auditors treat AI output paints a contradictory picture. On the one hand, research on automation bias shows that decision-makers supported by automated aids tend to reduce their own information search and accept system recommendations even when they are wrong (Skitka et al., 1999; Parasuraman & Manzey, 2010). On the other hand, experimental audit research shows that auditors may rely less on evidence from an AI system than on identical evidence from a human specialist, particularly when estimates are complex and the AI's output contradicts management (Commerford et al., 2022). The broader literature on algorithm aversion likewise shows that people abandon algorithms after seeing them err, even when the algorithm outperforms human judgement (Dietvorst et al., 2015).

Existing studies tend to examine these phenomena separately and to treat reliance as inherently desirable or undesirable. This paper argues that neither over-reliance nor under-reliance is the core problem; the problem is miscalibration, a gap between how much auditors rely on AI output and how reliable that output actually is for the assertion at hand. Both failures are, at root, failures of professional skepticism: automation bias reflects insufficient questioning of the machine, while algorithm aversion reflects questioning that is directed at the source of evidence rather than at its quality.

Accordingly, the paper addresses the following research question: Under what conditions do auditors calibrate their reliance on AI-generated audit evidence to its actual reliability, and how does such calibration affect audit quality? To answer it, the paper develops the construct of calibrated skepticism, builds a conceptual framework of its antecedents and consequences, formulates eight propositions and derives a typology of auditor reliance stances. The paper makes three contributions. First, it integrates two literatures that have developed in parallel, automation bias and algorithm aversion, within the established professional skepticism model of Nelson (2009). Second, it extends skepticism theory from human-generated to machine-generated evidence, a shift that current standards are only beginning to address. Third, it provides a structured agenda that can guide empirical research and standard-setting.

The remainder of the paper is organised as follows. Section 2 describes the conceptual methodology. Section 3 reviews the theoretical foundations. Section 4 problematises current assumptions and defines calibrated skepticism. Section 5 presents the framework and propositions, and Section 6 develops the typology. Section 7 discusses implications and a research agenda, and Section 8 concludes.

2. Research methodology

Conceptual papers make a contribution by bridging existing theories in novel ways, linking work across disciplines and broadening the scope of thinking, rather than by testing hypotheses with data (MacInnis, 2011; Gilson & Goldberg, 2015). Their methodological rigour lies in a transparent research design, an explicit choice of theories and a clear logic for deriving new constructs and relationships (Jaakkola, 2020). Following Jaakkola (2020), Table 1 positions the present study among four established designs for conceptual articles.

Table 1
Conceptual research designs and the approach adopted in this study

Design	Aim	Typical output	Use in this paper
Theory synthesis	Integrate fragmented findings into a higher-order understanding	New construct or integrative view	Primary: integrates skepticism, automation-bias and algorithm-aversion research into calibrated skepticism
Theory adaptation	Revise a domain theory using a method theory	Extended or re-scoped theory	Secondary: extends Nelson's (2009) skepticism model to machine-generated evidence
Typology	Classify a phenomenon along theoretically derived dimensions	Ideal types with predicted outcomes	Applied: four reliance stances (Section 6)
Model	Specify antecedents, mechanisms and outcomes	Nomological network and propositions	Applied: framework and propositions P1–P8 (Section 5)

Source: Adapted from Jaakkola (2020) and MacInnis (2011).

The study proceeds in five stages (Figure 1). First, it uses problematisation to identify and challenge an implicit assumption of the literature, namely that more reliance on AI is either good or bad in itself (Alvesson & Sandberg, 2011). Second, it conducts an integrative literature review (Snyder, 2019; Webster & Watson, 2002) of three bodies of work: auditing research on professional skepticism and technology, judgement and decision-making research on algorithm aversion and appreciation, and human-factors research on trust in automation. The review covered peer-reviewed journals in accounting, management, psychology and information systems together with current IAASB and PCAOB pronouncements. Third, it engages in theory synthesis, using professional skepticism as the domain theory and trust-in-automation research as the method theory (Lee & See, 2004; Jaakkola, 2020). Fourth, it builds a model with explicit propositions, a form that facilitates subsequent empirical testing (Cornelissen, 2017; Whetten, 1989). Fifth, it derives a typology whose types are ideal configurations of theoretically meaningful dimensions (Doty & Glick, 1994). Each stage is evaluated against the criteria of originality, utility and parsimony proposed for theoretical contributions (Corley & Gioia, 2011;

Whetten, 1989).

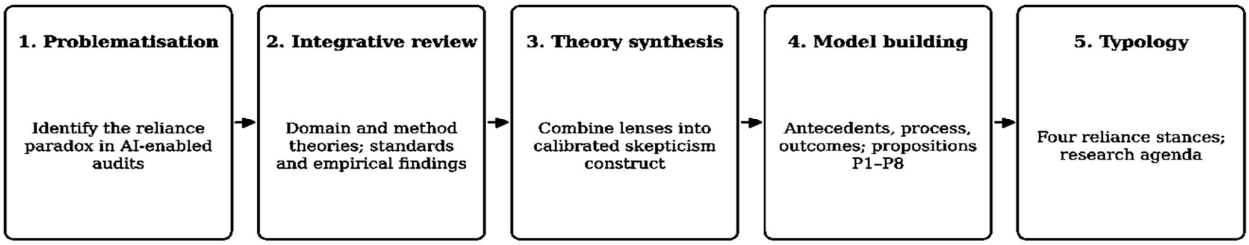


Fig. 1. Conceptual research design. Source: Authors' elaboration based on Jaakkola (2020) and Alvesson and Sandberg (2011).

3. Theoretical foundations

3.1. *Professional skepticism*

International standards define professional skepticism as an attitude that includes a questioning mind, alertness to conditions that may indicate misstatement and a critical assessment of audit evidence (IAASB, 2009). Nelson (2009) models skepticism as a sequence in which skeptical judgement, the recognition that a problem may exist, must be translated into skeptical action, such as gathering additional evidence or challenging management. Both stages are shaped by auditors' knowledge, traits and incentives. Hurtt (2010) identifies six trait characteristics of skepticism, including a questioning mind, suspension of judgement and self-esteem, and subsequent research synthesises how trait, state and contextual factors jointly drive skeptical behaviour (Hurtt et al., 2013). A key insight from this literature is that skepticism is costly: it consumes time and can create conflict with clients, so it is vulnerable to budget pressure and to incentives that favour verification of management's numbers (Griffith et al., 2015).

3.2. *Automation bias and complacency*

Human-factors research distinguishes the appropriate use of automation from its misuse and disuse (Parasuraman & Riley, 1997). Misuse includes automation bias, the tendency to use automated cues as a heuristic replacement for vigilant information search, leading to errors of commission (following incorrect advice) and omission (failing to notice problems the system does not flag) (Skitka et al., 1999; Goddard et al., 2012). Complacency and automation bias are more pronounced when workload is high, when the system has a strong track record and when the operator lacks insight into how the system reaches its outputs (Parasuraman & Manzey, 2010). In auditing, these conditions are common: budgets are tight, AI tools are promoted as firm-approved and many models are opaque (Brown-Liburd et al., 2015).

3.3. *Algorithm aversion and appreciation*

A parallel literature documents the opposite tendency. Dietvorst et al. (2015) show that people lose confidence in algorithms more quickly than in humans after observing the same error, a phenomenon termed algorithm aversion. Aversion is reduced when users can modify the algorithm's output, even slightly (Dietvorst et al., 2018). Conversely, Logg et al. (2019) report algorithm appreciation in many numerical estimation tasks, especially among non-experts, and

find that experts are more likely to discount algorithmic advice. Reviews conclude that reactions to algorithms depend on expertise, task characteristics, decision autonomy and the visibility of errors (Burton et al., 2020; Glikson & Woolley, 2020). In auditing, Commerford et al. (2022) find that auditors propose smaller adjustments to management's complex estimates when contradictory evidence comes from an AI system rather than from a human specialist, consistent with aversion in a domain where auditors view themselves as experts.

3.4. *Trust in automation as an integrating lens*

Lee and See (2004) argue that the central design goal for human–automation interaction is not high trust but appropriate trust, that is, trust calibrated to the capabilities of the automation. Calibration is high when trust matches capability, while over-trust produces misuse and distrust produces disuse. Trust is built on information about the system's performance, process and purpose, which parallels the ability, integrity and benevolence dimensions of interpersonal trust (Mayer et al., 1995). This lens is particularly useful because it treats over- and under-reliance as two sides of the same calibration problem. Table 2 compares the four lenses and shows what each contributes to the synthesis.

Table 2
Comparison of theoretical lenses

Lens	Core mechanism	Predicted reliance failure	Contribution to the framework	Key sources
Professional skepticism	Questioning mind; judgement → action	Insufficient skepticism under pressure	Domain theory; defines desired audit behaviour	Nelson (2009); Hurtt (2010)
Automation bias	Heuristic substitution of system cues for vigilance	Over-reliance (misuse)	Explains blind deference under workload and opacity	Parasuraman & Manzey (2010); Skitka et al. (1999)
Algorithm aversion	Asymmetric penalty for machine errors; expert self-confidence	Under-reliance (disuse)	Explains reflexive rejection after salient errors	Dietvorst et al. (2015); Commerford et al. (2022)
Trust in automation	Calibration of trust to system capability	Both, depending on calibration	Method theory; integrates both failures	Lee & See (2004); Glikson & Woolley (2020)

4. Problematising reliance: from more or less trust to calibrated skepticism

Much of the discussion of AI in auditing rests on an implicit assumption that the main risk lies in a single direction. Practitioner and regulatory discourse tends to emphasise over-reliance, warning that auditors may accept AI output uncritically, whereas firm investment cases emphasise under-use of technology and its forgone efficiency gains. Following Alvesson and Sandberg (2011), this paper challenges the assumption that reliance can be evaluated independently of evidence quality. Under ISA 500 and AS 1105, evidence is persuasive in proportion to its relevance and reliability, not its source. A skeptical auditor should therefore rely heavily on AI output when it is reliable for the specific assertion and data population, and rely little on it when it is not. Both blanket trust and blanket distrust violate this principle.

This reasoning also connects to the automation–augmentation paradox identified in the management literature: organisations that automate tasks risk deskilling the humans who must oversee automated systems, while organisations that merely augment human judgement forgo the consistency gains of automation (Raisch & Krakowski, 2021). Paradox theory suggests that such tensions should be managed through ongoing both/and responses rather than resolved through either/or choices (Smith & Lewis, 2011). For auditing, this implies that professional skepticism must operate simultaneously through the AI (using it to question management) and upon the AI (questioning the AI itself).

Building on these arguments, the paper defines calibrated skepticism as the extent to which an auditor's reliance on AI-generated audit evidence corresponds to the actual reliability of that evidence for the assertion under examination, sustained by skeptical judgement about the AI's performance and process and by skeptical action when reliability is in doubt. The construct has three properties. It is assertion-specific, because the same model may be reliable for one account and unreliable for another. It is bidirectional, because miscalibration may take the form of over- or under-reliance. And it is process-based, because calibration requires the auditor to evaluate the AI's inputs, logic and outputs rather than to accept or reject it wholesale. Figure 2 illustrates the construct.

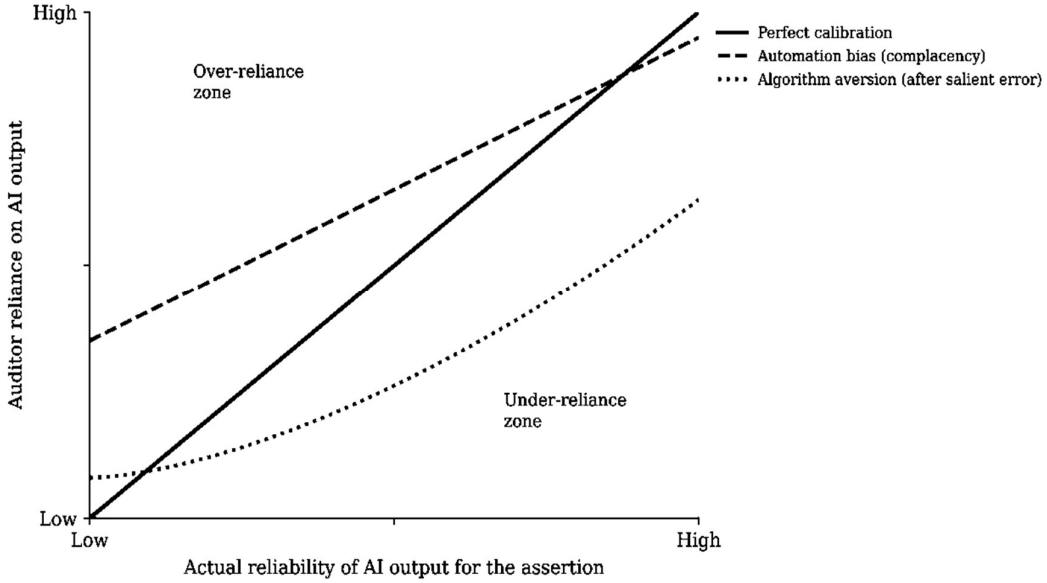


Fig. 2. Calibrated skepticism: reliance relative to the actual reliability of AI output (illustrative)

Note: The curves are conceptual illustrations, not empirical estimates. The diagonal represents perfect calibration; deviations above it represent over-reliance and deviations below it represent under-reliance. Source: Authors' elaboration based on Lee and See (2004).

5. Conceptual framework and propositions

Figure 3 presents the framework. Three groups of antecedents, namely task and context characteristics, AI-system attributes and auditor attributes, shape the degree of calibrated skepticism. Calibrated skepticism determines the auditor's reliance stance, which in turn affects audit quality, understood as both effectiveness (detecting material misstatement) and efficiency (avoiding unnecessary procedures).

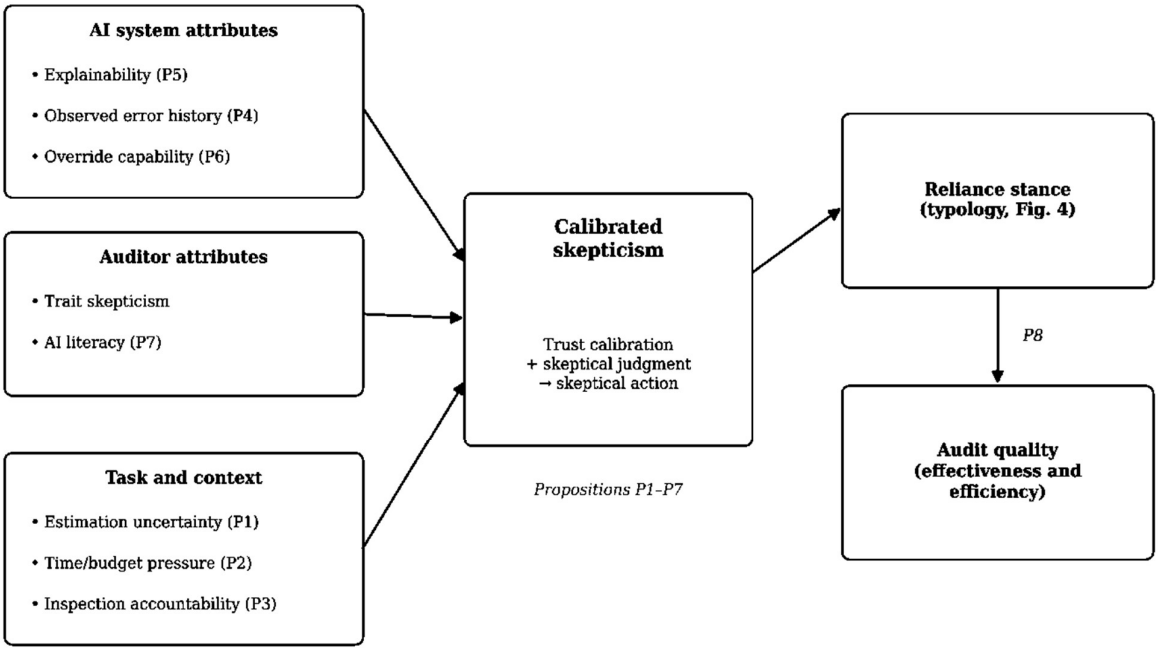


Fig. 3. Conceptual framework of calibrated skepticism in AI-enabled audits. Source: Authors' elaboration.

5.1. *Task and context antecedents*

Estimation uncertainty: Complex estimates involve subjective assumptions in which auditors perceive their own expertise as decisive. Evidence that auditors are reluctant to act on AI-generated contradictory evidence in such tasks (Commerford et al., 2022), together with evidence that experts discount algorithmic advice (Logg et al., 2019), suggests that aversion increases with uncertainty.

> *P1. The higher the estimation uncertainty of an account, the more auditors will under-rely on AI-generated evidence relative to equally reliable evidence from human specialists.*

Time and budget pressure: Automation bias is more likely when cognitive resources are constrained, because system cues offer a low-effort substitute for independent search (Parasuraman & Manzey, 2010). Audit research shows that budget pressure already impairs skeptical action (Nelson, 2009). When pressure coincides with an opaque tool, over-reliance becomes the path of least resistance.

> *P2. Time and budget pressure increase over-reliance on AI output, and this effect is stronger when AI explainability is low.*

Inspection accountability: Auditors anticipate how their judgements will be evaluated by reviewers and regulators. Where inspectors are perceived to scrutinise reliance on AI more severely than reliance on human specialists, auditors may choose evidence that is more defensible rather than more reliable, consistent with institutional pressures that shape audit practice (Griffith et al., 2015; Power, 1997).

> *P3. Perceived inspection risk associated with AI tools increases defensive under-reliance on AI-generated evidence, reducing audit efficiency without improving effectiveness.*

5.2. AI-system antecedents

Observed error history. Because people penalise algorithms more harshly than humans for identical errors (Dietvorst et al., 2015), a single salient AI error is likely to reduce reliance below the level justified by the system's overall accuracy.

> *P4. After observing a salient AI error, auditors reduce reliance on that AI system more than they would after an equivalent error by a human specialist, producing under-reliance.*

Explainability. Trust calibration depends on information about how a system works (Lee & See, 2004). Explanations of the variables driving an anomaly score or estimate allow auditors to judge whether the output is reliable for the specific assertion, which supports skeptical judgement and reduces both forms of miscalibration (Glikson & Woolley, 2020; Shneiderman, 2020).

> *P5. AI explainability improves calibrated skepticism by reducing both over-reliance and under-reliance.*

Override capability. Allowing users to adjust algorithmic output reduces aversion while preserving most of the algorithm's accuracy gains (Dietvorst et al., 2018). In audits, documented override procedures also force auditors to articulate why they accept or reject AI output, an instance of skeptical action.

> *P6. The ability to interrogate and adjust AI output, combined with a documentation requirement for overrides, increases calibrated skepticism.*

5.3. Auditor antecedents

Trait skepticism and AI literacy. Trait skepticism predisposes auditors to question evidence (Hurtt, 2010), but questioning is only productive if the auditor understands what could make AI evidence unreliable, such as unrepresentative training data, data-lineage gaps or model drift. Without such knowledge, skepticism may be directed at the technology per se rather than at the quality of its output.

> *P7. AI literacy moderates the effect of trait skepticism: with high AI literacy, trait skepticism increases calibrated skepticism; with low AI literacy, trait skepticism increases reflexive rejection of AI evidence.*

5.4. Consequences for audit quality

Over-reliance threatens audit effectiveness because misstatements outside the model's scope go undetected, while under-reliance threatens efficiency because auditors duplicate procedures or disregard reliable signals. Archival evidence that AI investment improves audit outcomes (Fedyk et al., 2022) is consistent with gains being realised when tools are used appropriately.

> *P8. Calibrated skepticism is positively associated with audit quality; both over-reliance and under-reliance are negatively associated with it, through reduced effectiveness and reduced efficiency, respectively.*

Table 3 summarises the propositions, their theoretical grounding and the reliance failure each addresses.

Table 3
Summary of propositions

No.	Antecedent	Predicted effect	Failure addressed	Theoretical basis
P1	Estimation uncertainty	↑ under-reliance	Algorithm aversion	Commerford et al. (2022); Logg et al. (2019)
P2	Time/budget pressure × low explainability	↑ over-reliance	Automation bias	Parasuraman & Manzey (2010); Nelson (2009)
P3	Inspection accountability	↑ defensive under-reliance	Algorithm aversion	Griffith et al. (2015); Power (1997)
P4	Salient AI error	↑ under-reliance	Algorithm aversion	Dietvorst et al. (2015)
P5	Explainability	↑ calibration	Both	Lee & See (2004); Shneiderman (2020)
P6	Override capability + documentation	↑ calibration	Both	Dietvorst et al. (2018)
P7	Trait skepticism × AI literacy	↑ calibration / ↑ rejection	Both	Hurtt (2010); Burton et al. (2020)
P8	Calibrated skepticism	↑ audit quality	—	Fedyk et al. (2022); Nelson (2009)

6. A typology of auditor reliance stances

Calibrated skepticism can be decomposed into two dimensions that vary across auditors and engagements: the level of trust in AI output and the level of skeptical engagement with that output, that is, the extent to which the auditor interrogates inputs, logic and results and acts on doubts. Crossing the two dimensions yields four ideal types (Figure 4; Table 4). Consistent with Doty and Glick (1994), the types are not mutually exclusive categories of auditors but theoretical configurations with distinct predicted outcomes; the same auditor may adopt different stances across assertions.

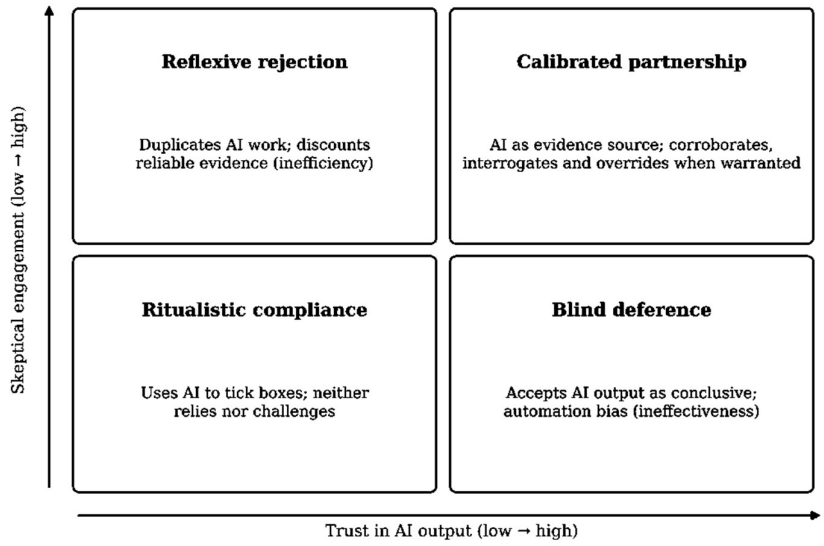


Fig. 4. Typology of auditor reliance stances. Source: Authors' elaboration.

Table 4
Characteristics and predicted outcomes of reliance stances

Stance	Trust	Engagement	Typical behaviour	Effectiveness	Efficiency
Calibrated partnership	High	High	Uses AI as a primary evidence source; corroborates, investigates exceptions, overrides with documented reasons	High	High
Blind deference	High	Low	Accepts AI output as conclusive; limits independent search	Low	Apparently high
Reflexive rejection	Low	High	Re-performs AI work manually; discounts contradictory AI evidence	Moderate	Low
Ritualistic compliance	Low	Low	Runs mandated tools to document use, without relying on or challenging results	Low	Low

Two implications follow. First, blind deference is the most dangerous stance because it appears efficient: audit hours fall while detection risk rises, and the deficiency may only surface after a restatement or inspection finding. Second, ritualistic compliance echoes Power's (1997) critique of audit as a ritual of verification. Firm mandates that require tool use without changing how evidence is evaluated may increase this stance, generating documentation without assurance. Moving auditors towards calibrated partnership therefore requires changes in both trust (through validation and explainability) and engagement (through training, review and incentives).

7. Discussion and research agenda

7.1. Theoretical contributions

The paper contributes to auditing theory in three ways. First, by recasting automation bias and algorithm aversion as symmetrical failures of a single underlying capability, it offers a parsimonious explanation for contradictory empirical findings and shifts the research focus from how much auditors rely on AI to how well their reliance tracks evidence quality. Second, it extends Nelson's (2009) skepticism model by specifying a new object of skeptical judgement: the evidence-generating process of the AI itself. Third, it imports the concept of trust calibration from human-factors research (Lee & See, 2004) into auditing and links it to established audit-quality outcomes, thereby connecting the behavioural and technology streams of audit research.

7.2. Implications for standard setters, firms and educators

For standard setters, the framework suggests that requirements focusing only on the risk of over-reliance may inadvertently encourage defensive under-reliance (P3). Guidance could therefore emphasise that the persuasiveness of evidence depends on its reliability, irrespective of whether it is generated by a person or a tool, and describe how auditors evaluate the reliability of AI-generated information, a topic addressed by the PCAOB (2024) amendments and the IAASB's (2026) re-exposure of ISA 500. For audit firms, P5 and P6 imply that tool design should prioritise explainability and structured override functions, and that review procedures should examine whether reliance was calibrated rather than merely whether a tool was used. For

regulators, inspection approaches that evaluate the appropriateness of reliance in both directions would reduce defensive behaviour. For educators, P7 suggests that skepticism training should be combined with AI literacy, including model validation, data lineage and the limits of machine learning.

7.3. *Research agenda*

The propositions are designed to be empirically testable. Table 5 outlines research questions and suitable methods.

Table 5
Agenda for future research

Theme	Illustrative research questions	Suggested methods
Measuring calibration	How can calibrated skepticism be measured at the assertion level? Can existing skepticism scales be adapted?	Scale development; experiments with known AI accuracy
Task and context (P1–P3)	Does aversion increase with estimation uncertainty? Do inspection regimes induce defensive under-reliance?	Between-subjects experiments; archival analysis of inspection findings
AI-system design (P4–P6)	Which explanation formats improve calibration? Do override requirements reduce aversion after errors?	Experiments; design science; field experiments in firms
Auditor attributes (P7)	Does AI literacy convert trait skepticism into calibrated rather than reflexive skepticism?	Experiments; surveys; training interventions
Audit quality (P8)	Is calibrated reliance associated with fewer restatements and lower audit costs?	Archival studies using firm AI-adoption data
Generative AI	Do large language models, whose outputs are fluent but may be fabricated, alter the balance between over- and under-reliance?	Experiments; interviews; qualitative case studies

7.4. *Limitations*

As a conceptual paper, the study does not provide empirical tests, and its propositions should be treated as theoretically derived expectations. The framework focuses on the individual auditor; team-level and firm-level dynamics, such as review hierarchies and partner incentives, deserve further attention. The typology simplifies continuous dimensions into four ideal types. Finally, because AI technologies evolve rapidly, the relative importance of antecedents may change as tools mature and as regulation develops.

8. Conclusion

AI promises to make audits more thorough and efficient, but only if auditors rely on it to the right degree. This paper argues that the central challenge is not trust or distrust in AI per se but the calibration of reliance to the actual reliability of AI-generated evidence. By introducing calibrated skepticism, specifying its antecedents and consequences in eight propositions and identifying four reliance stances, the paper provides a unified theoretical basis for research,

standard-setting and practice. Preserving professional skepticism in the age of AI means questioning management through the machine and questioning the machine itself.

References

Alvesson, M., & Sandberg, J. (2011). Generating research questions through problematization. *Academy of Management Review, 36*(2), 247–271. https://doi.org/10.5465/amr.2009.0188

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory, 36*(4), 1–27. https://doi.org/10.2308/ajpt-51684

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. https://doi.org/10.2308/acch-51023

Burton, J. W., Stein, M.-K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Commerford, B. P., Dennis, S. A., Joe, J. R., & Ulla, J. W. (2022). Man versus machine: Complex estimates and auditor reliance on artificial intelligence. *Journal of Accounting Research, 60*(1), 171–201. https://doi.org/10.1111/1475-679X.12407

Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: What constitutes a theoretical contribution? *Academy of Management Review, 36*(1), 12–32. https://doi.org/10.5465/amr.2009.0486

Cornelissen, J. (2017). Editor's comments: Developing propositions, a process model, or a typology? Addressing the challenges of writing theory without a boilerplate. *Academy of Management Review, 42*(1), 1–9. https://doi.org/10.5465/amr.2016.0196

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

Doty, D. H., & Glick, W. H. (1994). Typologies as a unique form of theory building: Toward improved understanding and modeling. *Academy of Management Review, 19*(2), 230–251. https://doi.org/10.5465/amr.1994.9410210748

Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies, 27*(3), 938–985. https://doi.org/10.1007/s11142-022-09697-x

Gilson, L. L., & Goldberg, C. B. (2015). Editors' comment: So, what is a conceptual paper? *Group & Organization Management, 40*(2), 127–130. https://doi.org/10.1177/1059601115576425

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association, 19*(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

Griffith, E. E., Hammersley, J. S., & Kadous, K. (2015). Audits of complex estimates as verification of management numbers: How institutional pressures shape practice. *Contemporary Accounting Research, 32*(3), 833–863. https://doi.org/10.1111/1911-3846.12104

Hurtt, R. K. (2010). Development of a scale to measure professional skepticism. *Auditing: A Journal of Practice & Theory, 29*(1), 149–171. https://doi.org/10.2308/aud.2010.29.1.149

Hurtt, R. K., Brown-Liburd, H., Earley, C. E., & Krishnamoorthy, G. (2013). Research on auditor professional skepticism: Literature synthesis and opportunities for future research. *Auditing: A Journal of Practice & Theory, 32*(Supplement 1), 45–97. https://doi.org/10.2308/ajpt-50361

IAASB. (2009). *International Standard on Auditing 200: Overall objectives of the independent auditor and the conduct of an audit in accordance with International Standards on Auditing*. International Federation of Accountants.

IAASB. (2026). *Exposure draft: Proposed International Standard on Auditing 500 (Revised), Audit Evidence*. International Auditing and Assurance Standards Board.

Issa, H., Sun, T., & Vasarhelyi, M. A. (2016). Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation. *Journal of Emerging Technologies in Accounting, 13*(2), 1–20. https://doi.org/10.2308/jeta-10511

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review, 10*(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0

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. https://doi.org/10.2308/jeta-51730

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

MacInnis, D. J. (2011). A framework for conceptual contributions in marketing. *Journal of Marketing, 75*(4), 136–154. https://doi.org/10.1509/jmkg.75.4.136

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

Nelson, M. W. (2009). A model and literature review of professional skepticism in auditing. *Auditing: A Journal of Practice & Theory, 28*(2), 1–34. https://doi.org/10.2308/aud.2009.28.2.1

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39*(2), 230–253. https://doi.org/10.1518/001872097778543886

PCAOB. (2024). *Amendments related to aspects of designing and performing audit procedures that involve technology-assisted analysis of information in electronic form* (PCAOB Release No. 2024-007). Public Company Accounting Oversight Board.

Power, M. (1997). *The audit society: Rituals of verification*. Oxford University Press.

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction, 36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Skitka, L. J., Mosier, K. L., & Burdick, M. (1999). Does automation bias decision-making? *International Journal of Human-Computer Studies, 51*(5), 991–1006. https://doi.org/10.1006/ijhc.1999.0252

Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review, 36*(2), 381–403. https://doi.org/10.5465/amr.2009.0223

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly, 26*(2), xiii–xxiii.

Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review, 14*(4), 490–495. https://doi.org/10.5465/amr.1989.4308371