American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

Human Judgment in Artificial Intelligence–Assisted Financial Assurance

Benjamin P. Commerford

Associate Professor, University of Kentucky, United States

Abstract

Artificial intelligence is increasingly capable of supporting financial assurance through anomaly identification, transaction classification, document analysis, pattern recognition, risk prioritization, and continuous examination of large datasets. These capabilities can increase the scale and efficiency of assurance work, but they do not eliminate the need for professional judgment. Financial assurance frequently requires auditors and other assurance professionals to interpret ambiguous evidence, evaluate materiality, challenge management explanations, recognize unusual business contexts, exercise professional skepticism, and assume responsibility for conclusions that cannot be reduced completely to algorithmic output.

This study develops a simulation-based framework for examining the contribution of human judgment in artificial intelligence–assisted financial assurance. Three hypothetical decision architectures are compared: human-only review, AI-led review with limited human challenge, and AI-assisted review incorporating explicit human judgment and professional skepticism. A synthetic analytical sample of 450 observations, equally distributed across the three conditions, is used solely for methodological illustration. Assurance judgment quality, anomaly detection, contextual appropriateness, professional skepticism, review efficiency, and decision confidence are modeled on seven-point scales.

The combined architecture therefore outperforms either predominantly human or predominantly automated review within the synthetic model. This result is consistent with pre-2021 audit-technology literature that emphasizes both the analytical capacity of AI and data analytics and the continuing importance of professional skepticism, contextual interpretation, and human expertise. The paper concludes that reliable AI-assisted assurance should preserve human responsibility at points involving materiality, evidence sufficiency, unusual transactions, conflicting signals, management intent, and final assurance conclusions rather than treating human involvement as a residual exception to automation.

Keywords: artificial intelligence; financial assurance; human judgment; auditing; professional skepticism; audit analytics; automation; decision support; assurance quality.

1. Introduction

Artificial intelligence is altering the technological architecture of accounting and assurance by enabling systems to analyze volumes and varieties of information that would be difficult to review manually within conventional engagement timeframes. Earlier research on artificial intelligence in auditing anticipated the formalization of audit tasks and the supplementation of professional labor through

intelligent systems, while subsequent literature examined machine learning, deep learning, automated inquiry, and large-scale data analytics as tools for expanding audit coverage and identifying patterns that conventional sampling may overlook. These developments create an important opportunity for financial assurance because routine classification, population-level screening, exception identification, and document processing can be performed with greater computational speed than is possible through exclusively manual review.

The growth of analytical capability does not, however, make assurance a purely computational activity. Professional assurance involves judgments concerning the relevance and reliability of evidence, the significance of unusual observations, the plausibility of management explanations, the appropriateness of accounting treatments, and the sufficiency of work performed before an opinion or conclusion is reached. Professional skepticism occupies a central position in audit judgment research, with Nelson's model describing skeptical judgment and skeptical action as influenced by auditor knowledge, traits, incentives, and the evidence encountered. International professional discussion has similarly emphasized that judgment draws on relevant knowledge, experience, professional standards, the framing of choices, recognition of bias, and decisions concerning what or whom to trust. These characteristics are particularly important when evidence is incomplete, contradictory, strategically presented, or economically ambiguous.

Artificial intelligence may improve the information available to professionals while simultaneously introducing new judgment risks. Brown-Liburd, Issa, and Lombardi identify information overload, relevance assessment, pattern recognition, and ambiguity as behavioral challenges arising from large-scale audit data. A system capable of identifying thousands of exceptions does not itself determine which exceptions matter economically, whether an apparent anomaly reflects legitimate business variation, or whether an apparently normal pattern results from coordinated manipulation. Similarly, an algorithm can produce a risk ranking while remaining dependent on training data, model assumptions, variable construction, thresholds, and the completeness of the underlying records. The assurance professional must therefore evaluate not only the financial evidence but also the quality and appropriateness of the analytical mechanism used to produce that evidence.

The relationship between humans and automation consequently becomes more important than the simple question of whether a task can technically be automated. Sutton, Arnold, and Holt explicitly address the danger of excessive automation in knowledge work and argue for retaining meaningful human relevance, while Kokina and Davenport describe AI as changing the audit process without treating professional expertise as entirely obsolete. Research into current audit-data-analytics practice further shows that analytical technologies are used within real engagements as part of broader audit processes rather than as independent substitutes for all auditor reasoning. The appropriate design problem is therefore one of task allocation, review, escalation, and accountability.

The present study examines human judgment in AI-assisted financial assurance as a complementary decision architecture rather than as a competition between professionals and machines. Three simulated conditions are compared: human-only review, AI-led limited review, and AI-assisted human judgment. The study proposes that AI-led review may perform strongly on speed and anomaly identification but more weakly when contextual interpretation and skeptical challenge are required, while human-only review may preserve contextual judgment but face computational and information-

processing limitations. The integrated architecture is expected to perform most strongly because automated analysis expands the evidential field while human reviewers retain responsibility for interpreting exceptions, questioning outputs, evaluating materiality, and determining final assurance conclusions. All numerical findings are synthetic and are presented solely to demonstrate how this framework could be tested empirically.

2. Objectives of the Study

2.1 To Evaluate Assurance Judgment Quality Across Alternative Human–AI Architectures

The first objective is to compare assurance judgment quality under human-only review, AI-led review with limited human intervention, and AI-assisted review incorporating structured human judgment. Judgment quality is conceptualized as the extent to which the final assurance conclusion appropriately integrates available evidence, recognizes material risks, accounts for contextual factors, and avoids both unjustified acceptance and unjustified escalation.

The proposed relationship is not that AI necessarily performs worse than humans. Artificial intelligence can outperform manual procedures in computationally intensive tasks such as population screening and pattern recognition. The central proposition is instead that assurance quality depends on how algorithmic outputs are incorporated into the wider professional decision process.

2.2 To Examine Professional Skepticism and Contextual Interpretation

The second objective is to investigate whether explicit human review strengthens professional skepticism and contextual appropriateness when AI-generated findings are used. Professional skepticism requires more than identifying an unusual statistical observation. It involves questioning whether evidence is credible, whether explanations are sufficiently supported, and whether alternative interpretations have been considered.

Contextual interpretation is particularly important for transactions that appear unusual statistically but are reasonable commercially, as well as transactions that appear statistically ordinary but become suspicious when evaluated against governance, related-party, contractual, or management-incentive information.

2.3 To Assess the Efficiency–Accountability Balance in AI-Assisted Assurance

The third objective is to examine whether integrating human judgment with AI can maintain analytical efficiency without transferring professional responsibility to an automated system. AI may reduce time devoted to repetitive procedures and allow professionals to concentrate on exceptions and higher-judgment areas. Issa, Sun, and Vasarhelyi describe AI in auditing partly in terms of workforce supplementation, while IFAC commentary has similarly presented AI as a technology capable of assisting accounting professionals with repetitive and computational activities.

The study therefore evaluates an assurance model in which automation expands analytical capacity but escalation rules, challenge procedures, documentation, and final conclusions remain subject to accountable human review.

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

3. Review of Literature

3.1 Professional Judgment and Skepticism in Financial Assurance

Professional judgment has always been central to assurance because accounting evidence is rarely self-interpreting. Nelson's review of professional skepticism conceptualizes skeptical judgment as a heightened assessment of the risk that an assertion is incorrect and links skeptical judgment and action with knowledge, personal characteristics, incentives, and evidential conditions. This model remains relevant in AI-assisted environments because automation does not remove incentives, uncertainty, management bias, or evidence-quality problems.

Professional judgment additionally requires the practitioner to decide which information deserves reliance and how different pieces of evidence should be integrated. IFAC's 2019 discussion of judgment emphasizes relevant training, professional knowledge, skill, experience, awareness of bias, trust assessment, framing of choices, and the execution of decisions. An AI model may contribute to several of these stages by organizing evidence or identifying risk, but the professional still has to determine whether the analytical result is suitable for the specific engagement circumstance.

Research on professional skepticism also suggests that cognitive approach matters. IFAC's synthesis of earlier skepticism research notes evidence that prompts and mindset can influence auditors' skeptical judgments and highlights the risk that excessive reliance on specialists or structured processes can weaken independent skeptical thought under some conditions. This issue has a direct analogue in AI-assisted assurance: algorithmic recommendations can become cognitively dominant if users treat model output as authoritative rather than as one form of evidence requiring evaluation.

The value of human judgment therefore lies not in opposition to quantitative analysis but in maintaining an evaluative layer between analytical output and professional conclusion. The assurance professional must be capable of challenging the model, the input data, the assumptions, and the apparent meaning of detected patterns.

3.2 Artificial Intelligence, Audit Analytics, and Evidence Expansion

Artificial intelligence research in auditing predates contemporary generative systems and has long examined expert systems, machine learning, decision support, pattern recognition, and automated reasoning. Issa, Sun, and Vasarhelyi identify opportunities for AI to formalize portions of audit work and supplement the audit workforce, while Kokina and Davenport describe cognitive technologies as capable of changing both audit processes and the work performed by human auditors.

Big-data and analytics research similarly emphasizes the opportunity to examine wider populations rather than depending exclusively on small samples. Appelbaum, Kogan, and Vasarhelyi discuss the implications of Big Data and analytics for the modern audit engagement, including the need to reconsider how evidence is obtained and interpreted. Eilifsen, Kinserdal, Messier, and McKee's 2020 exploratory study documents practical use of audit data analytics on audit engagements, demonstrating that analytics had moved from conceptual possibility into professional application by that period.

Machine-learning applications can extend these capabilities further. Sun presents an illustrative framework for applying deep learning to audit procedures, while Raschke, Saiewitz, Kachroo, and Lennard examine AI-enhanced audit inquiry using automated agents. These approaches illustrate how AI can support tasks ranging from pattern classification to interaction with audit evidence.

However, analytical expansion also introduces new information-processing problems. Brown-Liburd and colleagues emphasize information overload, ambiguity, relevance, and pattern-recognition challenges in Big Data environments. A larger volume of evidence does not automatically produce a better judgment if the reviewer cannot distinguish diagnostic information from noise or understand why the analytical model produced a particular result.

3.3 Automation Reliance, Human Relevance, and Accountability

A central issue in AI-assisted assurance is the extent to which professional decision-makers rely on automation. Sutton, Arnold, and Holt's work explicitly asks how much automation is too much and argues that knowledge work contains dimensions that remain difficult to program completely. Their analysis provides an important conceptual foundation for maintaining human judgment in tasks involving ambiguity, tacit knowledge, ethical responsibility, and unusual circumstances.

The challenge is not only that systems may be wrong. Even accurate systems can alter human behavior. When a machine recommendation is presented confidently, a reviewer may reduce independent investigation or reframe contradictory evidence as less important. Conversely, distrust of automation can lead professionals to reject valuable analytical evidence. The objective should therefore be calibrated reliance rather than automatic acceptance or automatic rejection.

AI systems also complicate responsibility. Kipp, Curtis, and Li's 2020 study of intelligent agents and agent autonomy examines how intelligent systems can affect individuals' ability to diffuse responsibility in accounting-related decisions. Although their context concerns earnings-management decisions rather than external assurance, the broader accountability issue is relevant: greater system autonomy can change how individuals psychologically interpret responsibility for outcomes.

The literature therefore supports a research gap around the design of human judgment within AI-assisted financial assurance. Much early technology research asks what can be automated or analyzed. The present framework instead asks where professional challenge should be positioned so that computational capability increases assurance coverage without weakening accountability, contextual evaluation, or skepticism.

4. Research Methodology

4.1 Research Design and Simulated Assurance Architectures

The study adopts a simulation-based comparative design. No actual audit firm, assurance engagement, client financial information, or professional participant dataset was used. All numerical observations were generated solely to demonstrate the proposed analytical framework.

A total of 450 synthetic observations were assigned equally to three decision architectures. The Human-Only Review condition represents assurance procedures in which professionals conduct analytical review and evidence evaluation without material AI-generated prioritization.

The AI-Led Limited Review condition represents a system in which AI performs anomaly detection, risk scoring, classification, and initial conclusion generation, while the human reviewer primarily verifies outputs and intervenes only when the system flags an explicit exception.

The AI-Assisted Human Judgment condition uses the same analytical capabilities but introduces mandatory professional challenge. Human reviewers assess the relevance of detected anomalies, test

alternative explanations, consider materiality and business context, evaluate contradictory evidence, and document whether the AI-generated recommendation should be accepted, modified, escalated, or rejected.

Table 1: Simulation-Based Methodological Framework

Methodological Component	Specification	Analytical Purpose
Research design	Three-condition simulated assurance experiment	Compare human and AI decision architectures
Total synthetic observations	450	Methodological illustration
Observations per condition	150	Balanced comparison
Condition A	Human-Only Review	Preserve professional judgment without AI augmentation
Condition B	AI-Led Limited Review	Maximize algorithmic analysis with limited reviewer challenge
Condition C	AI-Assisted Human Judgment	Combine AI analysis with explicit professional challenge
Primary outcome	Assurance Judgment Quality	Assess appropriateness of final conclusion
Analytical outcome	Anomaly Detection Capability	Evaluate identification of unusual patterns
Judgment outcome	Contextual Appropriateness	Evaluate interpretation of business circumstances
Behavioral outcome	Professional Skepticism	Evaluate willingness to challenge evidence and AI outputs
Process outcome	Review Efficiency	Assess analytical efficiency
Governance outcome	Decision Accountability	Assess clarity of human responsibility
Scale	Seven-point modeled responses	Standardize synthetic comparison
Main inferential procedure	One-way ANOVA	Demonstrate judgment-quality comparison

Note: The numerical sample is synthetic. It does not represent completed audit or assurance research and should not be reported as evidence from practicing auditors.

4.2 Measurement Architecture and Proposed Questionnaire

Assurance judgment quality was defined as the appropriateness, evidential support, and contextual defensibility of the final assurance conclusion. Anomaly detection represented the ability to identify transactions, journal entries, relationships, or financial patterns that warranted further investigation.

Contextual appropriateness measured whether unusual business circumstances, transaction purpose, governance information, accounting policy, and materiality considerations were interpreted correctly rather than treated solely as statistical characteristics.

Professional skepticism represented the extent to which the reviewer questioned evidence, challenged assumptions, investigated contradictory information, and avoided uncritical acceptance of either management explanations or AI-generated recommendations.

Review efficiency represented the ability to process and prioritize evidence within available engagement resources. Decision accountability represented whether professional responsibility for the final conclusion remained clear and documentable.

4.3 Simulation Procedure, Statistical Analysis, and Integrity Safeguards

Synthetic assurance-judgment scores were generated on a seven-point scale for 150 observations within each decision architecture. The simulation was structured so that human-only review retained strong contextual interpretation but experienced lower analytical efficiency, while AI-led limited review produced strong anomaly detection and efficiency but weaker skeptical challenge and contextual interpretation. AI-assisted human judgment combined strong analytical screening with explicit human evaluation.

A one-way analysis of variance was applied to the overall assurance-judgment-quality score. The generated dataset produced $F = 114.38$, $p < .001$, indicating substantial separation among the three synthetic conditions. Because the dataset was deliberately simulated, this significance result has no population-level evidentiary status.

A real empirical study should recruit qualified assurance professionals, randomly assign them to equivalent case conditions, control the quality and accuracy of AI outputs, and manipulate ambiguity or error rates systematically. Researchers should also record the extent to which participants inspect evidence independently before viewing the AI recommendation, because timing may influence anchoring and reliance.

5. Analysis

5.1 Comparative Performance Across Assurance Architectures

The simulated results indicate that the three assurance architectures possess different strengths rather than following a simple human-versus-machine ranking. Human-only review produces a mean overall judgment-quality score of 5.26, reflecting relatively strong contextual interpretation and professional skepticism but lower efficiency and anomaly-screening capacity.

AI-led limited review produces an overall judgment-quality mean of 5.18. Its anomaly-detection and efficiency scores are comparatively strong, but contextual appropriateness and professional skepticism are weaker. This pattern demonstrates the analytical distinction between identifying unusual observations and deciding what those observations mean for an assurance conclusion.

AI-assisted human judgment produces the highest overall mean at 6.17. The combined architecture retains the computational benefits of AI while introducing deliberate human review at the stage where findings are interpreted, challenged, and converted into professional conclusions.

Table 2: Simulated Outcomes Across Assurance Decision Architectures

Outcome Variable	Human-Only Review	AI-Led Limited Review	AI-Assisted Human Judgment	Principal Interpretation
Assurance Judgment Quality	5.26	5.18	6.17	Integration produces strongest final judgment
Anomaly Detection Capability	5.18	6.09	6.31	AI expands population-level exception identification
Contextual Appropriateness	5.72	4.49	6.22	Human interpretation remains important for ambiguous evidence
Professional Skepticism	5.68	4.37	6.11	Mandatory challenge reduces uncritical reliance
Review Efficiency	4.34	**6.18**	5.97	Automation provides major efficiency advantage
Decision Accountability	5.94	4.61	6.28	Explicit human ownership clarifies final responsibility

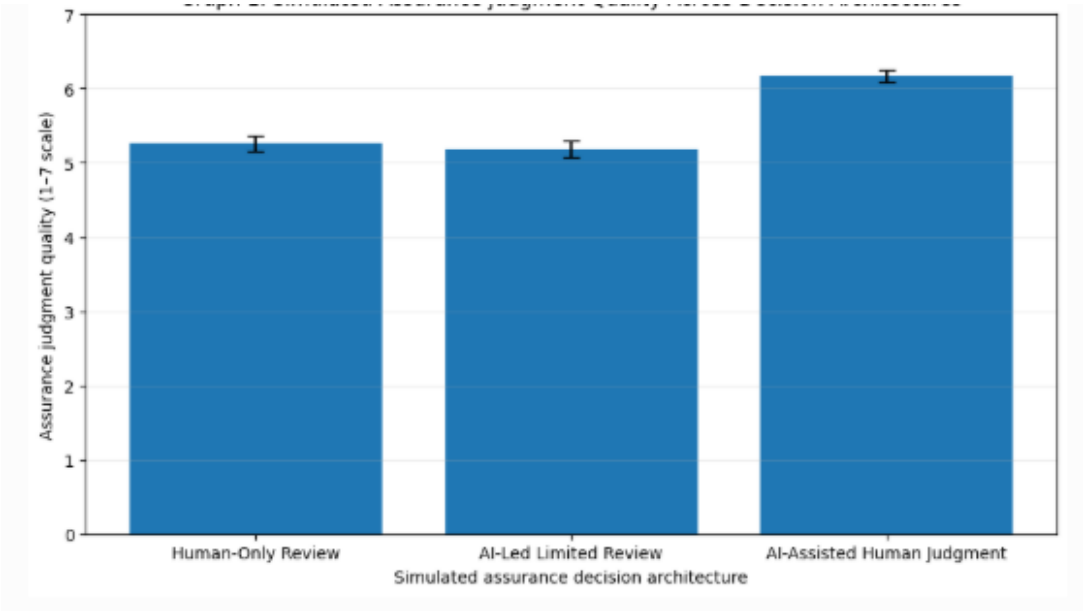
The table demonstrates why AI-led review does not necessarily produce the highest overall judgment quality even when it performs strongly on anomaly detection and efficiency. Assurance requires both identification and interpretation. A system may correctly flag an unusual transaction but cannot be assumed to possess sufficient engagement-specific knowledge to determine whether the transaction is material, legitimate, intentionally misleading, or indicative of a wider control failure.

American Journal of Advanced Multidisciplinary Innovation and Research (AJAMIR)

E-ISSN: XXXX-XXXX ● Website: www.ajamir.com ● Email: editor@ajamir.com

5.2 Graphical Assessment of Assurance Judgment Quality

Graph 1: Simulated Assurance Judgment Quality Across Decision Architectures



The graph shows little difference between human-only review and AI-led limited review in overall judgment quality, despite their different underlying strengths. Human reviewers benefit from contextual interpretation but face limitations in processing large volumes of evidence. AI-led review expands analytical coverage but loses part of the advantage when outputs receive insufficient skeptical evaluation.

The result is consistent with early AI-in-auditing literature that presents intelligent systems as mechanisms for formalization and workforce supplementation rather than as proof that every assurance judgment should be delegated to automation.

5.3 Human Challenge as a Control Over AI-Assisted Evidence

The simulated relationship suggests that human review is most valuable when it is deliberately positioned at high-judgment points rather than used merely to approve the AI output. A reviewer who simply confirms an algorithmic recommendation provides little additional assurance if the same assumptions and evidence remain unchallenged.

A meaningful human-control layer performs several functions. It considers whether the algorithm examined appropriate variables, whether relevant evidence was missing, whether identified anomalies are economically meaningful, and whether apparently normal transactions deserve investigation because of nonquantitative information.

The distinction is supported by Big Data audit research emphasizing relevance, ambiguity, and information-processing limitations. AI can help reduce one form of limitation by identifying patterns across large datasets, but it can introduce another if professionals over-rely on model prioritization.

The optimal human role is therefore neither universal manual re-performance nor passive approval. It is targeted skeptical intervention, particularly where evidence is ambiguous, consequences are material, or model confidence is not equivalent to professional certainty.

6. Discussion

6.1 AI Changes the Location of Judgment Rather Than Eliminating It

The results support the proposition that artificial intelligence changes where professional judgment is exercised. In traditional assurance, judgment may be concentrated in sample selection, analytical review, evidence gathering, and interpretation. AI can automate parts of selection and pattern detection, but this moves judgment upstream toward model design and downstream toward interpretation.

Professionals must judge whether a model is appropriate for the assertion being tested, whether the dataset is sufficiently complete, whether features used by the model are economically meaningful, and whether the analytical output constitutes relevant and reliable evidence. Appelbaum, Kogan, and Vasarhelyi's discussion of analytics in modern audit engagements emphasizes precisely this broader integration of analytical methods with audit evidence and process design.

Judgment also moves toward exception management. When AI permits population-level testing, the engagement may generate a much larger number of anomalies than a traditional sample-based procedure. The professional must decide which anomalies warrant investigation and which represent legitimate business variation.

6.2 Professional Skepticism Must Extend to the Algorithm

A second implication is that professional skepticism should apply not only to management assertions but also to AI-assisted evidence. Professionals should ask what the system may have missed, what data were unavailable, whether model assumptions remain valid, and whether a high-confidence output is supported by evidence appropriate to the assurance objective.

This extension is consistent with the broader idea that professional judgment requires decisions concerning whom and what to trust and recognition of one's own cognitive biases. Algorithmic systems can reduce some human biases but can also create new reliance patterns.

The risk is especially important where outputs are presented numerically. A risk score of 0.92 may appear more objective than a human statement that a transaction is “high risk,” yet the precision of the number does not by itself establish the validity of the model producing it.

Human reviewers should therefore receive sufficient information to challenge the basis of AI findings. Explainability, audit trails, source-data traceability, model-change documentation, and override records are important not because every professional must become a data scientist, but because professional skepticism cannot operate effectively when the reasoning chain is entirely inaccessible.

The purpose of explainability is consequently evidential accountability rather than technological curiosity.

6.3 A Human-in-the-Loop Governance Model for Financial Assurance

The study supports a human-in-the-loop model in which AI performs tasks suited to computational scale while defined professional checkpoints remain mandatory. Anomaly detection, transaction screening,

document classification, population testing, and preliminary risk ranking can be highly amenable to automation. Research from the pre-2021 literature already identified AI, audit analytics, and deep learning as capable of supporting these kinds of assurance procedures.

Human intervention should become stronger as judgment complexity increases. Routine low-risk matches may require minimal intervention, while unusual journal entries, estimates, going-concern indicators, related-party patterns, revenue-recognition exceptions, contradictory evidence, and significant management judgments should attract more intensive review.

Such a model should also make override behavior visible. If a human rejects an AI recommendation, the rationale should be documented. If the professional accepts a high-risk AI conclusion, the supporting evidence should likewise be documented rather than relying solely on the existence of a model score.

Responsibility should remain unambiguous. The system can recommend, classify, prioritize, and calculate, but the assurance professional remains responsible for determining whether sufficient appropriate evidence supports the engagement conclusion.

This principle is also important for maintaining professional competence. Sutton, Arnold, and Holt warn against automation that removes humans too completely from knowledge work, because excessive automation can erode the expertise required when unusual situations eventually demand intervention. A resilient assurance system must therefore preserve both technological capability and human capacity to challenge it.

7. Conclusion

Artificial intelligence has substantial potential to improve financial assurance by expanding analytical coverage, increasing processing speed, identifying patterns across complete transaction populations, and reducing the burden of repetitive examination. The pre-2021 auditing literature already anticipated many of these developments through research on artificial intelligence, Big Data, audit analytics, automated inquiry, and deep-learning-based procedures. The technological argument for augmentation is therefore strong.

The present simulation nevertheless indicates that analytical capability should not be equated with assurance judgment. Human-only review produced a simulated judgment-quality mean of 5.26, AI-led limited review produced 5.18, and AI-assisted human judgment produced 6.17. The combined architecture performed most strongly because AI improved detection and efficiency while human reviewers retained responsibility for context, skepticism, materiality, and final interpretation. These numerical findings are synthetic and are not presented as evidence from practicing auditors.

The central conclusion is that human judgment should be redesigned around AI rather than removed by AI. Professional involvement is likely to become less valuable in highly repetitive computational activities and more valuable in determining whether automated findings are relevant, sufficiently supported, economically meaningful, and consistent with the broader engagement evidence. The core human contribution therefore shifts toward skeptical challenge, ambiguity resolution, contextual interpretation, accountability, and governance.

AI-assisted assurance should also preserve professional skepticism toward the technology itself. Reviewers should be prepared to question data completeness, model assumptions, unexplained

classifications, confidence measures, contradictory evidence, and the possibility that unusual behavior is being normalized by the analytical model. Professional skepticism cannot function fully if algorithmic recommendations are regarded as inherently objective.

The study is limited by its simulation-based methodology and cannot establish real-world causal effects. Future research should test the framework with practicing auditors and assurance professionals using experimentally controlled financial cases. Particularly valuable extensions would manipulate AI accuracy, explainability, model confidence, task ambiguity, fraud risk, time pressure, reviewer experience, and the timing of AI recommendations. Longitudinal research should also examine whether repeated reliance on AI strengthens professional capability through better information or weakens capability through automation dependence. Such evidence would help assurance firms determine not simply where artificial intelligence can be introduced, but where human judgment must remain deliberately protected.

References

1. 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.
2. Puthukulam, G., Ravikumar, A., Sharma, R. V. K., & Meesaala, K. M. (2021). Auditors' perception on the impact of artificial intelligence on professional skepticism and judgment in Oman. *Universal Journal of Accounting and Finance, 9*(5), 1184–1190.
3. Al Nairi, A. H. N., Al Zadjali, A. S. I., Al Kamali, M. A. W. K., & Puthukulam, G. (2021). Does artificial intelligence and machine learning assist an auditor for better professional skepticism and judgment? *IAR Journal of Business Management, 2*(1), 1–5.
4. Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. A. (2020). The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics, 167*, 209–234.
5. Law, K. K. F., & Shen, M. (2020). How does artificial intelligence shape audit firms? *Management Science*.
6. Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies*.
7. Kend, M., & Nguyen, L. A. (2020). Big data analytics and other emerging technologies: The impact on the Australian audit and assurance profession. *Australian Accounting Review, 30*, 269–282.
8. 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.
9. 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.
10. Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting Horizons, 29*(2), 423–429.
11. Sutton, S. G., Holt, M., & Arnold, V. (2016). "The reports of my death are greatly exaggerated"—Artificial intelligence research in accounting. *International Journal of Accounting Information Systems, 22*, 60–73.

12. 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.
13. Zhang, C. (2019). Intelligent process automation in audit. *Journal of Emerging Technologies in Accounting*, *16*(2), 69–88.
14. Earley, C. E. (2015). Data analytics in auditing: Opportunities and challenges. *Business Horizons*, *58*(5), 493–500.
15. 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.
16. Nelson, M. W. (2009). A model and literature review of professional skepticism in auditing. *Auditing: A Journal of Practice & Theory*, *28*(2), 1–34.
17. Hurtt, R. K. (2010). Development of a scale to measure professional skepticism. *Auditing: A Journal of Practice & Theory*, *29*(1), 149–171.
18. DeZoort, F. T., Harrison, P. D., & Taylor, M. H. (2006). Accountability and auditors' materiality judgments: The effects of differential pressure and experience. *Auditing: A Journal of Practice & Theory*, *25*(1), 1–20.
19. Libby, R., & Luft, J. (1993). Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. *Accounting, Organizations and Society*, *18*(5), 425–450.
20. Vasarhelyi, M. A., Alles, M. G., & Kogan, A. (2004). Principles of analytic monitoring for continuous assurance. *Journal of Emerging Technologies in Accounting*, *1*(1), 1–21.