

Does Artificial Intelligence Improve Audit?**Reconciling Divergent Empirical Evidence on Audit Quality, Efficiency, and the Audit Workforce**Author: Jeanpaul Ziad Daccache¹*Affiliation:*

1. *Business School, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*
2. *CIRAME Research Center, Holy Spirit University of Kaslik, P.O. Box 446, Jounieh, Lebanon*

Abstract:

Audit firms have spent more than a decade investing in artificial intelligence (AI), yet the academic evidence on whether those investments improve the audit remains openly contradictory. This study asks why such a divergence exists and what the accumulated evidence, read as a whole, actually supports. This study conducts a structured evidence synthesis of seventeen empirical studies of AI and audit data analytics in external auditing published between 2015 and 2026, coding each study on five dimensions: evidence type, unit of analysis, technology construct, outcome domain, and direction of the reported effect. Vote-counting is then performed within strata rather than across the full corpus, so that studies measuring different things are never pooled. Three findings emerge. First, every one of the seven studies that measures a realized audit outcome — restatements, going-concern accuracy, misstatement detection, inventory count error, audit fees, or task completion time — reports a favorable effect, with magnitudes ranging from a 5.0% reduction in restatement likelihood to a fall in inventory count time from 681 hours to 19 hours. Second, all three studies that measure human judgment report adverse or mixed effects, because auditors discount machine-generated evidence and reviewers penalize procedures that appear less effortful. Third, the workforce divergence is not a genuine empirical contradiction but an artefact of the unit of analysis and of the outcome definition: firm-level headcount of accounting staff and office-level counts of auditor positions are different quantities that can legitimately move in opposite directions.

Keywords:

Branding, Consumer Demand, Brand Equity, Brand Credibility, Brand Trust, Willingness to Pay, Price Sensitivity

Introduction:

The external audit exists to reduce information asymmetry between the parties who own a company and the parties who run it. For most of its modern history it has done so with a comparatively simple technology: a sample of transactions, a set of documented procedures, and the professional judgment of a qualified human being. That technology is now changing faster than at any point since the audit became a statutory requirement. Artificial intelligence, robotic process automation, and audit data analytics are no longer pilot projects confined to innovation departments; they are embedded in the standard working papers of the largest firms.

The scale of the shift is measurable. Fedyk et al. (2022) document that the share of AI workers at the thirty-six largest United States audit firms rose from 0.08% of the workforce in 2010 to 0.39% in 2018, a fivefold increase over less than a decade. The United Kingdom Financial Reporting Council (FRC, 2017) found that all six of the largest United Kingdom audit firms were investing heavily in audit data analytics capability and that every one of them cited audit quality as the primary justification. Standard setters have moved with, rather than ahead of, that shift: the International Auditing and Assurance Standards Board launched a Technology Quality Management initiative in June 2025 and convened global roundtables engaging more than 240 stakeholders across six continents (IAASB, 2025).

What has not kept pace is a settled empirical answer to the obvious question: does any of this make the audit better? The literature is not merely incomplete on this point. It is openly contradictory, and the contradiction sits between two of the most prominent studies in the field. Fedyk et al. (2022), analyzing more than 310,000 employee resumes, report that AI investment is followed by a reduction in the accounting workforce of 3.6% after three years and 7.1% after four, concentrated entirely among junior staff. Law and Shen (2025), exploiting the staggered hiring of AI employees across United States audit offices, report the opposite sign: offices that hire AI employees show a 4.3% increase in the number of auditor jobs, again concentrated among junior and mid-level staff. Two careful studies, drawing on the same country, the same industry, and broadly similar data infrastructure, disagree about the direction of the effect.

A parallel tension runs between method families. Archival work consistently finds that AI investment tracks improvements in measured audit outcomes. Experimental work, by contrast, finds that auditors systematically discount evidence when it comes from a machine (Commerford et al., 2022) and that external reviewers judge analytics-based procedures to be of lower quality precisely because they appear to require less effort (Emett et al., 2023). Meanwhile, the survey and interview literature reports that actual adoption of advanced analytics on real engagements remains limited (Eilifsen et al., 2020; Krieger et al., 2021).

Sawaya, Khalil, Jabbour Al Maalouf, and Kaspard (2026) demonstrated that not only AI does enhance Audit quality and efficiency, but in fact it revolutionized the Audit industry, enabling entire organization redesign (Subrahmanyam & Saber, 2027).

This is the research gap the present study addresses. The field has accumulated a substantial body of empirical evidence but has not systematically asked why that evidence points in different directions. Existing reviews of technology in auditing tend to be narrative and thematic, cataloguing applications and challenges rather than interrogating the structure of the disagreement. The consequence is that practitioners, regulators, and audit committees are presented with a literature that appears to support whichever conclusion they arrive with.

Accordingly, this study addresses three research questions:

RQ1. What is the direction and magnitude of the effects of AI and audit data analytics on audit quality and audit efficiency reported in the empirical literature?

RQ2. Do those effects differ systematically by evidence type and by unit of analysis?

RQ3. What accounts for the divergent workforce findings reported by Fedyk et al. (2022) and Law and Shen (2025), and is that divergence a genuine empirical contradiction?

The objectives that follow from these questions are to assemble a defined corpus of empirical studies, to code that corpus on dimensions capable of explaining disagreement, to perform effect-direction analysis within strata rather than across the whole corpus, and to decompose the workforce divergence into its measurable sources. The practical significance is straightforward: audit committees are being asked to accept fee structures and staffing models justified by technology investment, and regulators are drafting standards for a technology whose measured benefit varies by two orders of magnitude depending on what is being measured and at what level and while transparency and objectivity may increase (Saber & Sassine, 2022) but ethical implications may be severe. The theoretical significance is that auditing offers an unusually clean setting in which to test whether AI substitutes for or complements high-skilled professional labor, because the output is standardized, the errors are observable ex post through restatements, and the labor force is well documented.

Literature Review:

2.1 From computer-assisted techniques to audit data analytics

The conceptual foundations of technology in auditing were laid well before the current wave of AI. Kogan et al. (2014) designed and evaluated a continuous data-level auditing system, demonstrating that audit testing need not be periodic. Gray and Debreceeny (2014) proposed a taxonomy to guide research on data mining for fraud detection in financial statement audits. A cluster of papers in *Accounting Horizons* then framed the opportunity and its limits: Cao et al. (2015) considered big data analytics in financial statement audits, Yoon et al. (2015) argued that big data can serve as complementary audit evidence, Alles (2015) examined the drivers and obstacles to the profession's evolution, and Brown-Liburd et al. (2015) raised the behavioral implications for audit judgment. Appelbaum et al. (2017) subsequently set out a research agenda for big data and analytics in the modern audit engagement, and Gepp et al. (2018) surveyed the gap between the techniques available and the techniques actually used, while many argue that big data analytics has significant impact on firm performance (Subrahmanyam et al., 2024). A consistent theme across this foundational work is that the technical potential was never seriously in doubt; what was in doubt was whether the profession would use it, and whether doing so would change the assurance actually delivered. Salijeni et al. (2019) documented growing reliance on analytics while voicing explicit skepticism about whether these changes had any substantial effect on the nature of the audit or on audit quality. That skepticism has proved durable.

2.2 Archival evidence on realized audit outcomes

The most important development of the last five years has been the arrival of large-scale archival evidence. Fedyk et al. (2022) constructed a firm-level measure of AI human capital following Babina et al. (2020), matching more than 310,000 resumes to Audit Analytics data for thirty-six audit firms over 2010–2019. Using restatements as the audit-quality proxy — the proxy that DeFond and Zhang (2014) and Knechel et al. (2013) identify as the most robust and universally applicable — they report that a one-standard-deviation increase in recent AI investment is

associated with a 5.0% reduction in restatement likelihood, a 1.4% reduction in material restatements, a 1.9% reduction in accrual and revenue-recognition restatements, and a 0.3% reduction in restatement-related SEC investigations. Crucially, they show that the effect is attributable to the auditor's AI and not the client's, and that it survives controls for general software engineering and data analysis skills.

Law and Shen (2025) approach the same question at the office rather than the firm level, treating the staggered arrival of AI employees at individual audit offices as the treatment. They also find quality gains, reporting more accurate going-concern and internal control opinions. Their labor result, however, runs in the opposite direction to that of Fedyk et al. (2022), and their interviews with eleven seasoned audit partners record that none believed AI had replaced or would replace human auditors. Related archival work on detection technologies is uniformly positive: Perols et al. (2017) show that data analytics improve fraud prediction, and Brown et al. (2020) demonstrate that topic-based textual analysis detects financial misreporting.

2.3 Experimental evidence on human judgment

The experimental literature tells a markedly less comfortable story. Commerford et al. (2022) find that auditors who receive contradictory evidence from their firm's AI system, rather than from a human specialist, propose smaller adjustments to management's complex estimates — a manifestation of algorithm aversion — and that the effect is strongest where management's inputs are relatively objective. Emmett et al. (2023) identify a complementary mechanism on the review side: external reviewers rely on an effort heuristic, judging data-and-analytics procedures as lower in quality than traditional procedures because they appear less effortful, even when the level of assurance is held constant. Their second experiment shows the effect can be attenuated by a theory-based intervention, which matters because it locates the problem in the evaluator rather than in the procedure. Rose et al. (2017) had earlier shown that big data visualizations shift auditors' evaluation of evidence and their professional judgments in ways that are not uniformly beneficial. Libby and Witz (2024) extend the question to independence conflicts and audit firm liability. Subrahmanyam and Beainy (2027) analyzed the role of smart decision support systems and its positive impact on human judgment.

The newest layer concerns generative AI. Jezierski et al. (2025) conducted a preregistered randomized field experiment with one hundred professional auditors at a global audit firm, in which participants completed a complex inventory valuation task with or without access to a firm-specific large language model. Access improved output quality and reduced completion time, with the largest gains among active users, but did not improve adherence to internal audit guidelines. Akpan and Ogunade (2026) tested four large language models on their ability to predict critical audit matters from management discussion and analysis disclosures across 141 firm-year observations, and Li and Vasarhelyi (2024) compared methodologies for applying such models in accounting.

2.4 Adoption, ethics, and the theoretical frame

Evidence on actual use tempers all of the above. Eilifsen et al. (2020) surveyed 216 engagement partners and managers and examined 109 real audit engagements in Norway, concluding that audit data analytics use was relatively limited and that advanced analytics were rare, with greater use on clients running integrated enterprise resource planning systems and on newly tendered engagements. Krieger et al. (2021) developed a process theory of why advanced analytics are not adopted. Cooper et al. (2019) documented the uneven spread of robotic process automation across public accounting. Austin et al. (2021) interviewed matched dyads of company managers and their audit partners alongside regulators, surfacing three conflicts: tensions over audit fees,

frustration at the absence of regulation specific to analytics, and regulator concern that using analytics to deliver business insight to clients may impair independence. Munoko et al. (2020) set out the ethical implications of AI in auditing, and Christ et al. (2021) showed through a design-science study that drone-enabled inventory procedures deliver very large efficiency and accuracy gains while first-mover reluctance impedes adoption.

Theoretically, the audit is a response to the agency problem articulated by Jensen and Meckling (1976): a monitoring technology that constrains the divergence of interest between principals and agents. Frey and Osborne (2017) place auditing among the occupations most exposed to computerization, which frames the labor question. Taken together, the prior literature establishes that the technology works, that auditors and reviewers do not consistently trust it, that adoption is partial, and that its labor consequences are disputed. What it does not establish is why these strands disagree, which is the contribution attempted here.

2.5 AI and technology's impact on different sectors and financial assets

Kaspard, Kamel, Khalil, and Beainy (2026) examined the positive impact of AI on sustainability oriented indices while (Khalife et al., 2024) conducted a study according to which AI and Blockchain technology significantly enhance the performance of companies, especially SMEs. In a data-driven culture (Azoury et al., 2024), AI has allowed even intangible assets to be audit, even investor sentiments (Khalife et al., 2025) and non-financial metrics (Kaspard et al., 2025). Finally, while AI's impact on financial results may vary (Athari et al., 2026), AI's impact on financial reporting quality is significant and the ability to accurately price shares is better than ever (Khalife et al., 2023), however business experience and entrepreneurial knowledge remain a must to properly control, evaluate and interpret AI to achieve better outcomes (Chaanine, Mhanna, Verdie, & Beainy, 2023).

Methodology:

3.1 Research design

This study uses a structured evidence synthesis, a design appropriate where the object of inquiry is not a new phenomenon but an unresolved disagreement within an existing body of findings. The unit of observation is the individual empirical study; the unit of analysis is the study-outcome pair, since a single study may report effects in several outcome domains. The design proceeds in four stages: corpus construction, coding, within-stratum effect-direction analysis, and decomposition of the workforce divergence.

3.2 Corpus construction and inclusion criteria

Candidate studies were identified from the reference networks of the two anchor archival papers (Fedyk et al., 2022; Law and Shen, 2025) and from forward citation tracing of those papers, supplemented by targeted searches of the leading accounting and auditing journals. Four inclusion criteria were applied. A study had to (i) generate primary empirical evidence, so conceptual essays, taxonomies, and narrative reviews were excluded and used only as background in Section 2; (ii) concern external auditing or financial statement assurance, so

studies of internal audit and of tax practice were excluded; (iii) examine artificial intelligence, machine learning, audit data analytics, robotic process automation, drone-enabled procedures, or large language models, rather than information technology in general; and (iv) be published between 2015 and 2026 in a peer-reviewed outlet, with one preregistered working paper retained because it is the only randomized field experiment on generative AI in audit currently available.

Applying these criteria produced a final corpus of seventeen studies. The corpus is purposive rather than exhaustive: it is weighted towards high-impact studies capable of shifting the field's conclusions, and it is not a substitute for a full PRISMA-compliant systematic review. This is a deliberate trade-off, discussed in Section 5.

3.3 Coding scheme and variable definitions

Each study was coded on five dimensions, defined as follows.

Evidence type. Four mutually exclusive categories: archival (observational analysis of firm, office, or filing data), experimental and design science (laboratory experiments, field experiments, and design-science artefact evaluation), computational benchmark (evaluation of model performance against a criterion), and qualitative or survey (interviews, questionnaires, and regulatory thematic reviews).

Unit of analysis. The level at which the effect is measured: audit firm, audit office, audit engagement, individual auditor or task, or filing.

Technology construct. What the study actually operationalizes, distinguishing AI human capital, applied analytics, automation, and generative models, since these are frequently conflated.

Outcome domain. Five domains: audit quality and detection effectiveness; efficiency and cost; workforce and skills; auditor or reviewer judgment; and adoption intensity.

Effect direction. Coded as favorable, adverse, null, or mixed, relative to the outcome as the original authors define it. A reduction in restatements is favorable; a reduction in the size of proposed audit adjustments in the presence of contradictory AI evidence is adverse, because it indicates suppressed professional skepticism.

3.4 Analytic procedure

Effect-direction analysis, sometimes described as vote counting, is used rather than meta-analytic pooling. Pooling is inappropriate here because the corpus reports effects in incommensurable units — percentage changes in restatement probability, hours of count time, adjustment magnitudes in an experimental instrument — and because a common effect size cannot be constructed across archival and experimental designs without heroic assumptions. The critical methodological safeguard is that vote counting is performed strictly within outcome domains and, where relevant, within units of analysis. Pooling a restatement result with an algorithm-aversion result would produce an average that describes nothing.

For RQ3, the workforce divergence is decomposed by comparing the two anchor studies on four attributes on which they differ and which can each independently generate a sign reversal: the unit at which treatment is measured, the population counted in the dependent variable, the direction of the comparison implied by the research design, and the time horizon over which the effect is estimated. The decomposition is diagnostic rather than statistical; it establishes whether the two estimates are the same quantity, which is a logical prior to asking whether the estimates conflict.

All coding was performed by the author against the published text of each study, and all reported magnitudes are taken directly from the original papers rather than from secondary summaries. The coded corpus is reported in full in Table 1 and is available as a supplementary dataset, so that any reader may re-code the corpus and reproduce the counts in Tables 2 and 3.

Results:

4.1 Composition of the corpus

The seventeen studies divide into four archival studies, five experimental and design-science studies, two computational benchmarks, and six qualitative or survey studies. Table 1 presents the full coded corpus.

Study	Evidence type	Unit	Technology	Outcome domain	Direction
Fedyk et al. (2022)	Archival	Audit firm	AI human capital	Quality; Efficiency; Workforce	+ / + / -
Law & Shen (2025)	Archival	Audit office	AI human capital	Quality; Workforce; Skills	+ / + / +
Perols et al. (2017)	Archival	Firm-year	Predictive analytics	Detection effectiveness	+
Brown et al. (2020)	Archival	Filing	NLP / topic models	Detection effectiveness	+
Christ et al. (2021)	Design science	Engagement task	Drones counting software +	Efficiency; Effectiveness	+ / +

Study	Evidence type	Unit	Technology	Outcome domain	Direction
Jezierski et al. (2025)	Field experiment	Auditor task	Firm-specific LLM	Quality; Efficiency; Compliance	+ / + / 0
Commerford et al. (2022)	Experiment	Auditor judgment	AI evidence source	Judgment	–
Emett et al. (2023)	Experiment	Reviewer judgment	D&A procedures	Judgment (perceived quality)	–
Rose et al. (2017)	Experiment	Auditor judgment	Big data visualization	Judgment	Mixed
Akpan & Ogunade (2026)	Benchmark	Filing	Large language models	Risk identification	+
Li & Vasarhelyi (2024)	Benchmark	Task	Large language models	Feasibility	Mixed
Eilifsen et al. (2020)	Survey / interview	Engagement	Audit data analytics	Adoption intensity	Limited
Austin et al. (2021)	Interview	Auditor–client dyad	Data analytics	Adoption; Independence	Contested
Salijeni et al. (2019)	Interview	Audit firm	Big data technology	Adoption intensity	Limited
Krieger et al. (2021)	Interview / process	Audit firm	Advanced analytics	Adoption intensity	Limited
Cooper et al. (2019)	Survey	Audit firm	Robotic process automation	Adoption intensity	Partial
FRC (2017)	Regulatory review	Audit firm	Audit data analytics	Adoption intensity	Investing

Table 1: Coded corpus of seventeen empirical studies of AI and analytics in external auditing, 2015–2026. "+" denotes an effect favorable to the outcome as defined by the original authors, "–" an adverse effect, and "0" a null effect.

4.2 RQ1: Direction and magnitude of effects on quality and efficiency

Seven studies in the corpus measure a realized audit outcome, meaning an outcome observed in archival records, in a benchmark against a known criterion, or in the physical result of an audit

procedure. All seven report a favorable effect. The direction is unanimous; the magnitudes, presented in Table 2, are not.

Study	Reported effect	Magnitude	Level
Fedyk et al. (2022)	Restatement likelihood, per 1 SD rise in three-year AI share	−5.0%	Firm
Fedyk et al. (2022)	Material restatement likelihood	−1.4%	Firm
Fedyk et al. (2022)	Accrual / revenue restatement likelihood	−1.9%	Firm
Fedyk et al. (2022)	Restatement-related SEC investigation	−0.3%	Firm
Fedyk et al. (2022)	Audit fees after one, two, three years	−0.9% / −1.5% / −2.1%	Firm
Law & Shen (2025)	Going-concern and internal control opinion accuracy	Improved	Office
Christ et al. (2021)	Inventory count time, drones vs manual	681 h → 19 h	Task
Christ et al. (2021)	Inventory count error rate	0.15% → 0.03%	Task
Jezierski et al. (2025)	Documentation quality and completion time, LLM access	Both improved	Task
Jezierski et al. (2025)	Adherence to internal audit guidelines	No improvement	Task

Table 2: Reported effect magnitudes on realized audit outcomes. All magnitudes are taken directly from the original studies.

Two observations follow. First, the effect sizes scale inversely with the level of aggregation. At task level the improvements are enormous — a 97% reduction in inventory count time and an 80% reduction in count error rate. At firm level they are measurable but modest: a 5.0% relative reduction in restatement likelihood and a fee reduction that does not exceed 2.1% even after three years. This pattern is exactly what one would expect if AI substantially improves a limited subset of audit procedures whose contribution to the overall audit opinion is itself limited, and it cautions strongly against citing task-level demonstrations as evidence of firm-level transformation.

Second, the efficiency evidence is more equivocal than it appears. A reduction in audit fees is favorable to the client and is interpreted by Fedyk et al. (2022) as evidence of efficiency, but Austin et al. (2021) document that auditors resist exactly this interpretation, arguing that innovation costs justify higher rather than lower fees. The same observed quantity therefore carries opposite normative valence depending on which party is asked.

4.3 RQ2: Effects by evidence type and unit of analysis

Stratifying the corpus by what is being measured produces the central result of this study, summarized in Table 3. Where the outcome is a realized audit outcome, seven of seven studies are favorable, contributing eight favorable study-outcome pairs and no adverse or null ones. Where the outcome is human judgment or perception, none of the three studies is favorable: two report adverse effects and one reports mixed effects. Where the outcome is compliance with internal guidelines, the single relevant study reports a null effect. Where the outcome is adoption intensity, all six studies report partial, limited, or contested uptake.

Outcome domain	Distinct studies	Study-outcome pairs	Favourable	Adverse	Null / mixed	Unanimity
Realized audit outcomes	7	8	8	0	0	Unanimous, favorable
Human judgment / perception	3	3	0	2	1	Unanimous, unfavorable
Guideline compliance	1	1	0	0	1	Null
Workforce and skills	2	2	1	1	0	Divergent
Adoption intensity	6	6	0	0	6	Unanimous, constrained
Method feasibility	1	1	0	0	1	Mixed
Total	17	21	9	3	9	—

Table 3: Effect-direction analysis within outcome domains. A study reporting effects in several domains contributes one study-outcome pair per domain, so the seventeen studies generate twenty-one pairs. Fedyk et al. (2022) contribute two pairs to realized audit outcomes (quality and fees) and one to workforce.

The disagreement in the literature is therefore not random. It falls almost perfectly along the boundary between what the technology does and what people do with what the technology produces. No study in the corpus finds that AI degrades a measured audit outcome, and no study finds that auditors and reviewers respond to AI-generated evidence in the way that the technology's effectiveness would warrant. Commerford et al. (2022) show auditors proposing smaller adjustments when the contradictory evidence carries a machine label; Emmett et al. (2023) show reviewers penalizing the procedure for looking easy. Both effects reduce realized audit quality through a channel that no improvement in model accuracy can address.

4.4 RQ3: Decomposing the workforce divergence

The apparent contradiction between Fedyk et al. (2022) and Law and Shen (2025) dissolves under decomposition. Table 4 compares the two studies on the four attributes capable of generating a sign reversal.

Attribute	Fedyk et al. (2022)	Law & Shen (2025)
Treatment level	Firm-wide share of AI workers	Office-level arrival of AI employees
Dependent variable	Total accounting (audit and tax) employees at the firm	Number of auditor positions at the office
Comparison	Cross-sectional, across firms with differing AI intensity	Staggered adoption, treated vs not-yet-treated offices
Horizon	Three and four years ahead	Contemporaneous with adoption
Reported result	−3.6% at three years; −7.1% at four years	+4.3% auditor jobs
Junior-level result	−5.7% at three years; −11.8% at four years	+4.7% junior; +3.3% mid-level

Table 4: Decomposition of the divergent workforce findings of the two anchor archival studies.

The two studies do not estimate the same quantity. Fedyk et al. (2022) measure the total accounting workforce of a firm, which includes tax practice and back-office accounting roles, at a horizon of three to four years, and they explicitly report that the effect is absent in years one and two. Law and Shen (2025) measure auditor positions at offices in the period contemporaneous with AI arrival. A firm that concentrates AI capability in a handful of centralized hubs — which is precisely what Fedyk et al. (2022) document, with AI workers clustered in New York, California, Washington D.C., Illinois, and Texas — can simultaneously staff up the offices where AI is being deployed and reduce total accounting headcount elsewhere over a longer horizon. The office-level and firm-level estimates are not merely different; they are not even expected to share a sign.

This reading is supported by evidence internal to both studies. Fedyk et al. (2022) report that their quality results weaken substantially when AI is measured at office rather than firm level, confirming that the level of measurement is doing real work. Law and Shen (2025) report that the eleven partners they interviewed uniformly rejected the displacement thesis, while Fedyk et al. (2022) report that their seventeen partners described labor effects as the removal of lower-level tasks rather than of people. Both sets of interviews are consistent with a single underlying process: tasks are automated, junior roles are redefined, hiring at deployment sites rises in the short run, and aggregate junior headcount falls over a multi-year horizon.

Discussion and Conclusion:

This study set out to determine what the empirical literature actually supports about artificial intelligence in external auditing, and to establish whether its most visible disagreements are real. Three conclusions follow.

First, the evidence that AI improves measured audit outcomes is stronger than the field's reputation for disagreement suggests. Seven of seven studies measuring a realized outcome report improvement, across restatements, going-concern accuracy, misstatement detection, inventory count accuracy, and task completion time, and across four independent research designs. That is an unusually consistent result for an accounting literature. The qualification is one of magnitude rather than direction: the firm-level effects are modest, and the spectacular figures that circulate in professional discourse — a count time falling from 681 hours to 19 — are task-level results that do not translate proportionally into audit opinions.

Second, the binding constraint on realized benefit is human and institutional, not technical. The corpus contains no study in which the technology underperforms and no study in which humans respond to it optimally. Auditors discount machine evidence (Commerford et al., 2022); reviewers penalize efficient procedures (Emett et al., 2023); generative AI improves the quality of documentation without improving adherence to the firm's own guidelines (Jezierski et al., 2025); and adoption on real engagements remains limited a decade after the tools became available (Eilifsen et al., 2020; Krieger et al., 2021). Investment in model capability therefore has a low expected marginal return relative to investment in evaluation frameworks, reviewer training, and inspection practices. The intervention tested by Emmett et al. (2023), which removed reviewers' reliance on the effort heuristic, is more consequential for audit quality than another increment of model accuracy.

Third, the workforce controversy is largely a measurement artefact. Fedyk et al. (2022) and Law and Shen (2025) estimate different quantities at different levels over different horizons, and the reconciliation offered in Section 4.4 — centralized AI capability, local staffing increases at deployment sites, and longer-horizon contraction of aggregate junior headcount — is consistent with the archival results and with the partner interviews in both papers. The profession should stop treating this as an open question about whether AI displaces auditors and start treating it as a question about which auditors, at which sites, on what timescale.

Three implications follow for practice and regulation. For audit firms, the evidence favors reallocating investment from tool development towards the reception of tool output, since the profession's bottleneck lies in evaluation rather than capability. For regulators, the absence of analytics-specific guidance that Austin et al. (2021) document as a source of confusion is now being addressed through the IAASB's technology program and its proposed revisions to the audit evidence standards (IAASB, 2025); the evidence here suggests that guidance which legitimizes less effortful but equally assuring procedures would have a direct quality benefit, by removing

the incentive to prefer procedures that merely look rigorous. For audit committees, fee reductions attributable to AI are real but small, and a firm proposing substantial fee changes on technology grounds is making a claim the literature does not yet support at that magnitude.

Several limitations bound these conclusions. The corpus is purposive rather than exhaustive and is weighted towards high-impact journals, which may understate null results held in the file drawer. Effect-direction analysis discards information about effect size and gives a small study the same vote as a large one; it also cannot distinguish a marginally significant result from a decisive one. The evidence base is heavily concentrated in the United States and, for the survey literature, in northern Europe, so generalization to emerging markets and to smaller audit firms is not warranted. The archival studies establish association under controls and fixed effects rather than clean causal identification. Finally, one corpus member is a working paper rather than a peer-reviewed article, and the generative AI evidence is necessarily preliminary given how recently these tools entered practice.

Future research should pursue four directions. The first is direct replication of both anchor workforce studies using a common estimate which would convert the reconciliation proposed here from a diagnostic argument into a tested one. The second is field evidence on interventions that correct algorithm aversion and the effort heuristic in live engagement reviews and regulatory inspections, since that is where the literature locates the binding constraint. The third is evidence from outside the United States, and particularly from jurisdictions where audit markets are less concentrated and enterprise systems less mature, where the enabling conditions identified by Eilifsen et al. (2020) are largely absent. The fourth is longitudinal study of generative AI specifically, distinguishing its effects from those of the earlier analytics wave with which it is routinely and unhelpfully conflated. Until that work exists, the defensible position is the one this synthesis supports: artificial intelligence demonstrably improves what auditors can measure, and does not yet reliably change what auditors conclude.

References

- Akpan, M., & Ogunade, K. M. (2026). Generative artificial intelligence in audit practice: Enhancing risk assessment. *Current Issues in Auditing*. <https://doi.org/10.2308/CIIA-2025-034>
- Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review*, 40, 100402. <https://doi.org/10.1016/j.cosrev.2021.100402>
- Alles, M. G. (2015). Drivers of the use and facilitators and obstacles of the evolution of big data by the audit profession. *Accounting Horizons*, 29(2), 439–449. <https://doi.org/10.2308/acch-51067>
- American Institute of Certified Public Accountants. (2017). Guide to audit data analytics: An overview. AICPA.
- 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>
- Athari, S. A., Sassine, M. E., Agyemang, E. T., Kirikkaleli, D., & Saliba, C. (2026). Nonlinear Dynamics of AI Investment and Economic Growth in Germany: Evidence from a NARDL Approach. *Economies*, 14(1), 31.
- Austin, A. A., Carpenter, T. D., Christ, M. H., & Nielson, C. S. (2021). The data analytics journey: Interactions among auditors, managers, regulation, and technology. *Contemporary Accounting Research*, 38(3), 1888–1924. <https://doi.org/10.1111/1911-3846.12680>
- Azoury, N., Subrahmanyam, S., & Sarkis, N. (2024). The Influence of a Data-Driven Culture on Product Development and Organizational Success through the Use of Business Analytics. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 15(2), 123-134.
- Babina, T., Fedyk, A., He, A., & Hodson, J. (2020). Artificial intelligence, firm growth, and industry concentration [Working paper]. SSRN. <https://ssrn.com/abstract=3651052>
- Brown, N. C., Crowley, R. M., & Elliott, W. B. (2020). What are you saying? Using topic to detect financial misreporting. *Journal of Accounting Research*, 58(1), 237–291. <https://doi.org/10.1111/1475-679X.12294>
- 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>
- Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting Horizons*, 29(2), 423–429. <https://doi.org/10.2308/acch-51068>
- Chaanine, N., Mhanna, R. Y., Verdie, J. F., & Beainy, R. (2023). Leveraging business experience and knowledge to boost entrepreneurial achievement. *International Journal of Work Innovation*, 4(2), 104-115.
- Christ, M. H., Emett, S. A., Summers, S. L., & Wood, D. A. (2021). Prepare for takeoff: Improving asset measurement and audit quality with drone-enabled inventory audit procedures. *Review of Accounting Studies*, 26(4), 1323–1343. <https://doi.org/10.1007/s11142-020-09574-5>
- 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>

- Cooper, L. A., Holderness, D. K., Jr., Sorensen, T. L., & Wood, D. A. (2019). Robotic process automation in public accounting. *Accounting Horizons*, 33(4), 15–35. <https://doi.org/10.2308/acch-52466>
- DeFond, M. L., & Zhang, J. (2014). A review of archival auditing research. *Journal of Accounting and Economics*, 58(2–3), 275–326. <https://doi.org/10.1016/j.jacceco.2014.09.002>
- Eilifsen, A., Kinserdal, F., Messier, W. F., Jr., & McKee, T. E. (2020). An exploratory study into the use of audit data analytics on audit engagements. *Accounting Horizons*, 34(4), 75–103. <https://doi.org/10.2308/HORIZONS-19-121>
- Emett, S. A., Kaplan, S. E., Mauldin, E. G., & Pickerd, J. S. (2023). Auditing with data and analytics: External reviewers' judgments of audit quality and effort. *Contemporary Accounting Research*, 40(4), 2314–2339. <https://doi.org/10.1111/1911-3846.12894>
- 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>
- Financial Reporting Council. (2017). Audit quality thematic review: The use of data analytics in the audit of financial statements. FRC.
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>
- Gepp, A., Linnenluecke, M. K., O'Neill, T. J., & Smith, T. (2018). Big data techniques in auditing research and practice: Current trends and future opportunities. *Journal of Accounting Literature*, 40, 102–115. <https://doi.org/10.1016/j.acclit.2017.05.003>
- Gray, G. L., & Debreceeny, R. S. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems*, 15(4), 357–380. <https://doi.org/10.1016/j.accinf.2014.05.006>
- International Auditing and Assurance Standards Board. (2025). Technology. IAASB. <https://www.iaasb.org/focus-areas/technology>
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305–360. [https://doi.org/10.1016/0304-405X\(76\)90026-X](https://doi.org/10.1016/0304-405X(76)90026-X)
- Jezierski, M., Kaltenpoth, S., Müller, O., & Weißenberger, B. E. (2025). The effect of large language models on audit quality: Causal evidence from a randomized field experiment [Working paper]. SSRN. <https://ssrn.com/abstract=6095166>
- Kaspard, J. A., Khalil, F. C., & Kamel, C. A. (2025). Factors contributing to non-financial performance of companies. In *Sustainable Economic Development: Utilizing Non-Financial Performance, Sustainability Reporting, and Corporate Governance* (pp. 413–442). Cham: Springer Nature Switzerland.
- Kaspard, J., Kamel, C., Khalil, F., & Beainy, R. (2026). Financial Performance, Risk, and Market Integration of Sustainability-Oriented Equity Indices: Implications for the Sustainability Transition (2010–2025). *Risks*, 14(5), 99.
- Khalife, D., Subrahmanyam, S., & Farah, A. (2024). A Sustainable Circular Business Model to Improve the Performance of Small and Medium-sized Enterprises Using Blockchain Technology. *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, 15(2), 240–250.

- Khalife, D., Yammine, J., El Bazi, T., Zaki, C., & Jabbour Al Maalouf, N. (2025). Analyzing the impact of investor sentiment on S&P 500 prices using deep learning models. *Journal of Financial Reporting and Accounting*, 23(4), 1365-1386.
- Khalife, D., Yammine, J., Rahal, S., & Freiha, S. (2023). Pricing Asian and Barrier Options Using a Combined Heston Model and Monte Carlo Simulation Approach with Artificial Intelligence. *Mathematical Modelling of Engineering Problems*, 10(5), 1690.
- Knechel, W. R., Krishnan, G. V., Pevzner, M., Shefchik, L. B., & Velury, U. K. (2013). Audit quality: Insights from the academic literature. *Auditing: A Journal of Practice & Theory*, 32(Suppl. 1), 385–421. <https://doi.org/10.2308/ajpt-50350>
- Kogan, A., Alles, M. G., Vasarhelyi, M. A., & Wu, J. (2014). Design and evaluation of a continuous data level auditing system. *Auditing: A Journal of Practice & Theory*, 33(4), 221–245. <https://doi.org/10.2308/ajpt-50844>
- Krieger, F., Drews, P., & Velte, P. (2021). Explaining the (non-)adoption of advanced data analytics in auditing: A process theory. *International Journal of Accounting Information Systems*, 41, 100511. <https://doi.org/10.1016/j.accinf.2021.100511>
- Law, K. K. F., & Shen, M. (2025). How does artificial intelligence shape audit firms? *Management Science*, 71(5), 3641–3666. <https://doi.org/10.1287/mnsc.2022.04040>
- Li, H., & Vasarhelyi, M. A. (2024). Applying large language models in accounting: A comparative analysis of different methodologies and off-the-shelf examples. *Journal of Emerging Technologies in Accounting*, 21(2), 133–152. <https://doi.org/10.2308/JETA-2023-065>
- Libby, R., & Witz, P. D. (2024). Can AI reduce independence conflicts on audit firm liability? *Contemporary Accounting Research*, 41(2), 1346–1375. <https://doi.org/10.1111/1911-3846.12943>
- Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics*, 167(2), 209–234. <https://doi.org/10.1007/s10551-019-04407-1>
- Perols, J. L., Bowen, R. M., Zimmermann, C., & Samba, B. (2017). Finding needles in a haystack: Using data analytics to improve fraud prediction. *The Accounting Review*, 92(2), 221–245. <https://doi.org/10.2308/accr-51562>
- Rose, A. M., Rose, J. M., Sanderson, K. A., & Thibodeau, J. C. (2017). When should audit firms introduce analyses of big data into the audit process? *Journal of Information Systems*, 31(3), 81–99. <https://doi.org/10.2308/isys-51837>
- Saber, A. M., & Sassine, M. E. (2022). Reflections on internal control and internal audit roles for ensuring financial transparency toward Lebanese NGOs financial donors. *EuroMed Journal of Management*, 4(3), 186-198.
- Salijeni, G., Samsonova-Taddei, A., & Turley, S. (2019). Big data and changes in audit technology: Contemplating a research agenda. *Accounting and Business Research*, 49(1), 95–119. <https://doi.org/10.1080/00014788.2018.1459458>
- Sawaya, C., Khalil, F. C., Jabbour Al Maalouf, N., & Kaspard, J. (2026). Revolutionizing audit practice in emerging economies: the role of Artificial Intelligence in enhancing the audit of consolidated financial statements. *Cogent Business & Management*, 13(1), 2625974.
- Subrahmanyam, S., & Beainy, R. H. (2027). Smart Decision Support for Uncertainty and Disruptions in Industry 4.0 Operations. In *Smart Operational Research Solutions in the Era of Industry 4.0 and Sustainability* (pp. 167-198). IGI Global Scientific Publishing.

- Subrahmanyam, S., & Saber, A. (2027). Role of Competitive Intelligence in Organizational Redesign. In *Navigating Mergers, Acquisitions, and Organizational Redesign* (pp. 169-204). IGI Global Scientific Publishing.
- Subrahmanyam, S., Aishwaryalaxmi, N. S., Khalife, D., Shaikh, I. A. K., Faldu, R., & Asthana, N. (2024, April). Impact of knowledge management and big data analytics capabilities on firm performance. In *2024 Ninth International Conference on Science Technology Engineering and Mathematics (ICONSTEM)* (pp. 1-5). IEEE.
- Yoon, K., Hoogduin, L., & Zhang, L. (2015). Big data as complementary audit evidence. *Accounting Horizons*, 29(2), 431–438. <https://doi.org/10.2308/acch-51076>