Journal of Business Insight and Innovation
Volume 5 Issue 4, 2026
ISSN-p: 3006-2284, ISSN-e: 3006-0982
https://insightfuljournals.com/

THE IMPACT OF ARTIFICIAL INTELLIGENCE ON AUDIT QUALITY AND FRAUD DETECTION

Fraz Ahmed Shaikh [1]

DOI: https://doi.org/10.63544/jbii.v5i4.148

Affiliations:

[1] Research Scholar, Murdoch Business School, Murdoch University, Australia MPA Accounting Advanced, ACCA UK, MIPA AFA IPA, Australia

Email: frazahmedshaikh77@gmail.com

Corresponding Author's Email:

[1] frazahmedshaikh77@gmail.com

Article History:
Received: 20.03.2026
Accepted: 04.04.2026
Published: 15.04.2026

Abstract

The use of Artificial Intelligence (AI) in external auditing is revolutionizing audit practices, offering greater audit quality and improved capacity to detect fraud. But the magnitude and mechanisms of such effects are not well understood. This study analyses the effects of adoption of AI on audit quality and fraud detection using survey data of 287 auditors from Big 4 and non-Big 4 firms. We theorise and test a conceptual model that suggests that professionals' judgments of the AI's use intensity relate to their perceptions of the quality of audit-related improvements and effectiveness of fraud detection, with the moderation of auditor experience, firm size, and expertise regarding AI. The measurement model has good reliability and validity. The results of the hierarchical multiple regressions show that AI adoption significantly and positively predicted audit quality ($\beta = 0.41$, $p < .001$) and fraud detection ($\beta = 0.37$, $p < .001$), accounting for 34% and 29% of the variance, respectively. Furthermore, firm size and auditors' AI training moderate the relationship between audit quality and AI adoption, as does that between AI and fraud detection for Big 4 firms. Stable results are obtained from robustness checks such as structural equation modelling and common method bias tests. The findings are among the first to show, from a large-scale empirical study, that AI is not just an efficiency tool, but is a substantive improvement to audit effectiveness, especially when it comes to complex judgmental work. Auditory practice, regulation and future research implications are discussed.

Keywords: Artificial Intelligence, Audit Quality, Fraud Detection, Machine Learning, External Audit, Survey Research, Hierarchical Regression

1. Introduction

External auditing is one of the pillars of capital market integrity, because it assures that financial statements are not materially misstated due to error or fraud (IAASB, 2020). The audit profession has been under increasing pressure to enhance audit quality and to bridge the "expectation gap" in regards to the auditors' fraud detection powers in recent years. The problems with traditional sampling-based audit approaches have been highlighted by a number high-profile corporate scandals and the growth in the volume and complexity of corporate data. In this context, Artificial Intelligence (AI) technologies (including machine learning, natural language processing, robotic process automation and deep learning) have started to pervade the audit space, promising a paradigm shift from sample-based audit testing to continuous audit testing at the population level (Kokina & Davenport, 2017; Issa, Sun, & Vasarhelyi, 2016).

Proponents say AI has the potential to make audits better, by letting auditors look at whole data sets; find patterns as subtle as fraud or anomalies; and minimise human error and bias (Appelbaum et al., 2017). At the same time, AI-driven fraud detection systems can sift through millions of transactions in real-time, identifying high-risk ones that might not have been spotted during manual processes (Brazel, Jones, & Zimbelman, 2009). Although the theoretical potential of AI in auditing is promising, the empirical research on AI's role in auditing remains in its early stages, featuring mostly conceptualization, case studies, and laboratory experiments with limited samples. There is little evidence available at a large scale, in the field, of the impact of the simultaneous use of AI in audits on both audit quality and fraud detection.

To fill this gap, this study empirically investigates the following research question: How does the adoption of AI impact audit firms' perceived audit quality and ability to detect fraud; and do these impacts depend on organizational and individual factors? Based on the Technology Acceptance Model (TAM; Davis, 1989), and the Elaboration Likelihood Model of professional judgment (Petty & Cacioppo, 1986), we formulate hypotheses and test them with survey data from 287 external auditors who currently work in practice. We will use multi-item scales for our dependent measures: Perceived improvement in audit quality (accuracy, compliance and reliability) and fraud detection effectiveness (identification of material misstatements as a result of fraud). The key independent variable is the level of AI adoption, which is quantified through the extent of embedding AI tools across key audit phases. Additionally, we explore how individual auditor’s AI training and firm size (between Big 4 and non-Big 4) moderate the effects of the independent variables.

On the methodological side, we use hierarchical multiple regression and, to increase validity, we also use structural modeling to ensure that the results are robust. Common method bias, multicollinearity, and measurement validity are treated with care. The findings strongly suggest that the use of AI is positively associated with audit quality and fraud detection, contributing significant variance over and above other control factors – auditor experience and auditor specialty. Furthermore, the advantages of AI for audit quality appear greater in Big 4 firms, probably because of the enhanced infrastructure and integration support, and the fraud detection advantages appear greater when auditors have dedicated AI expertise, highlighting the importance of the complementarity of human-AI.

This research gives a contribution to the auditing literature in various ways: First, it offers some of the first field-based, quantitative evidence that goes beyond anecdotal/experimental settings to connect the intensity of use of AI tools with audit quality and fraud detection results. Second, simultaneously modeling both dependent variables provides a more comprehensive picture of the double impact of AI. Third, the identification of moderators helps to understand the boundary conditions and how to practically implement the findings. The insights indicate that embracing AI for investments should go hand-in-hand with human capital development and company-level assistance to achieve optimal outcomes for practitioners and regulators.

The rest of the paper is organized as follows. The relevant literature is reviewed in Section 2 and hypotheses are developed. The methodology of the survey, the variables measured and the approach to analysis is described in Section 3. The results are reported in Section 4, covering the evaluation of the measurement model, descriptive statistics and hypothesis testing. Results and implications are discussed in Section 5 and limitations and future research lines in Section 6.

2. Literature Review and Hypothesis Development

2.1 Artificial Intelligence in Auditing

The use of computational methods that replicate the cognitive processes of humans to audit, including risk assessment, substantive testing and fraud investigation, is considered artificial intelligence in auditing (Sutton, Holt, & Arnold, 2016). Supervised and unsupervised machine learning, such as classification, clustering, and anomaly detection, natural language processing for contract analysis and management commentary, robotic process automation for reconciliations and data extraction, and cognitive computing for analyzing unstructured data are key technologies (Zhang, Yang, & Appelbaum, 2020). The adoption curve has been steep and all the Big 4 firms have been investing billions in AI platforms, such as Deloitte's Argus, EY's Helix, KPMG's Clara and PwC's Halo are examples of this (Kokina & Davenport, 2017).

The theoretical foundations of AI benefits can be found in the “audit data analytics” paradigm that suggests that by analyzing 100% of transactions, auditors can detect patterns and outliers that would not be found using sampling (Appelbaum, Kogan, & Vasarhelyi, 2017). Furthermore, AI saves auditors the mental effort they would otherwise expend on the audit and lets them focus on the high-risk areas (Brown-Liburd, Issa, & Lombardi, 2015). The Technology Acceptance Model (TAM) (Davis, 1989) indicates that perceived usefulness and ease of use are the factors that lead to adoption, and perceived usefulness in the audit domain is closely related to audit effectiveness/efficiency. This means that the more AI tools that auditors use, the better the outcomes are likely to be.

2.2 Audit Quality

The quality of the audit is a complex phenomenon and is often described as the probability that an auditor will detect the misstatements and report them (DeAngelo, 1981). Survey-based research provides a measure of auditors' self-assessed improvement in important audit quality characteristics, including the quality of the audit evidence, adherence to auditing standards, the number of errors that escape detection, and the quality of their conclusions (Christensen, Glover, & Wolfe, 2014). While there are no archival measures of audit quality, we follow this “audit quality improvement scale” since it is what is perceived by the participants in the sample. AI will bring improvements to the quality of the audit in several ways. First, machine learning models can detect complex, non-linear relationships in financial information that could be a sign of earnings management and/or misstatements, allowing for more accurate risk assessments (Perols, Bowen, Zimmermann, & Samba, 2017). Secondly, NLP can examine 100% of contracts or board minutes for consistency and will be able to detect inconsistencies that humans could miss. Third, RPA eliminates repetitive tasks, without any error, which helps to minimise manual errors in critical reconciliations. They are consistent with the PCAOB's focus on professional skepticism and adequate appropriate audit evidence (PCAOB, 2019). Accordingly, we hypothesize:

H1: AI adoption intensity is positively associated with perceived audit quality improvement.

2.3 Fraud Detection

The detection of fraud is challenging due to its intentional nature, its hidden and insidious nature, and its infrequent occurrence (Hogan, Rezaee, Riley, & Velury, 2008). Analytical procedures and sampling are the key audit techniques used in traditional audits, and these techniques can only uncover a small portion of fraud cases (ACFE, 2022). Unsupervised learning algorithms like autoencoders and isolation forests, which are able to detect anomalous transaction patterns that differ from normal patterns even when there are no labelled fraud transactions to train on (West & Bhattacharya, 2016), are excellent tools for identifying anomalous transaction patterns, particularly those that deviate from normal patterns. In real time, supervised models trained on historical fraud transactions can score transactions for continuous fraud monitoring (Ngai et al., 2011). NLP can detect language patterns of lying in management discussions and analysis or in e-mail messages (Humphreys et al., 2011).

AI tools are particularly useful for detecting fraudulent transactions, as AI is adept at identifying patterns, abnormalities, and processing large amounts of unstructured data.Auditors who use AI tools will likely be more successful at detecting fraudulent transactions, because AI is good at anomaly detection, pattern recognition, and processing huge volumes of unstructured data. Furthermore, AI can mitigate the impact of cognitive biases that can impede fraud detection, including confirmation bias and anchoring (Wolfe & Hermanson, 2004). We therefore hypothesize:

H2: AI adoption intensity is positively associated with perceived fraud detection effectiveness.

2.4 Moderating Role of Firm Size and AI Training

AI investments for audit firms are by no means equal. Large firms (Big 4) have more resources to build their own AI solutions, incorporate those solutions into their audit processes and offer robust training. In addition, they possess a more extensive and varied data sets to train AI tools, which makes their AI more powerful (Kokina & Davenport, 2017). The institutional support in Big 4 firms can amplify the translation of AI usage into audit quality gains. However, small and medium-sized practices can implement off-the-shelf AI solutions that are less specific to their audit workflow, resulting in a more limited impact on quality. Thus, we propose:

H3a: Firm size (Big 4 vs. non-Big 4) positively moderates the relationship between AI adoption intensity and perceived audit quality improvement, such that the relationship is stronger for Big 4 firms.

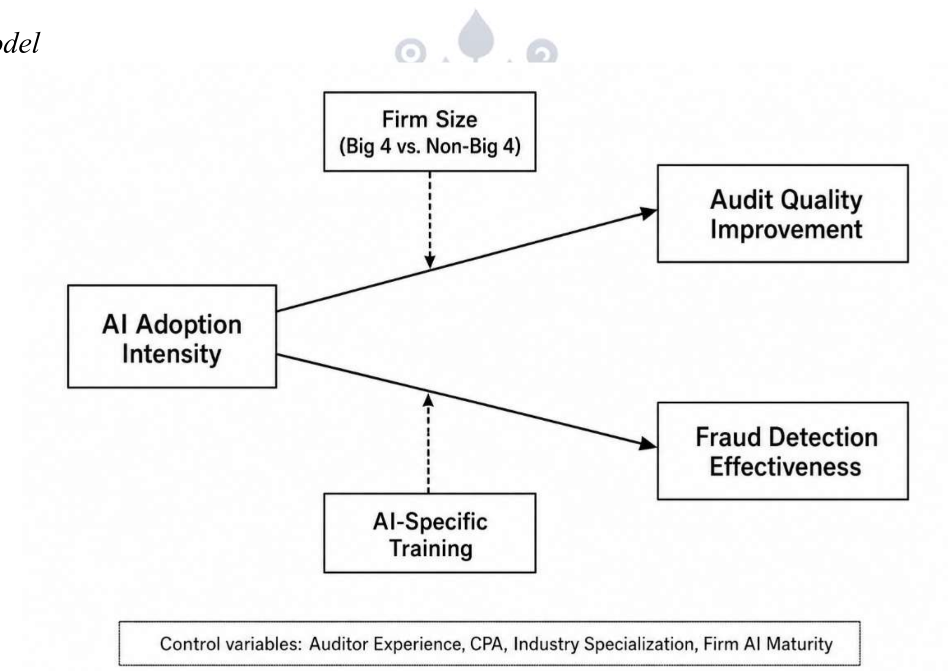
When using AI tools at the individual level, the quality of the outputs, the challenge of model assumptions and the incorporation of AI-generated insights into professional judgment, have been identified as elements that will impact the effectiveness (Sutton et al., 2016). According to the Elaboration Likelihood Model, if people are motivated and capable, they process information centrally, critically and make good decisions. Beyond being familiar with the AI tools, auditors who are specifically trained in AI are more likely to engage in deep processing of the AI output, to understand the limitations of the AI model, and to consider AI recommendations based on other evidence as well (Brown-Liburd et al., 2015). Therefore, AI training will improve the fraud detection benefits, as being a good fraud assessor involves using complicated judgment and skepticism. We hypothesize:

H3b: Auditors' AI-specific training positively moderates the relationship between AI adoption intensity and perceived fraud detection effectiveness, such that the relationship is stronger when AI training is high.

No specific moderation is hypothesized for the reverse combinations but they are tested for completeness. The conceptual model is presented in Figure 1.

Figure 1

Conceptual Model



3. Research Methodology

3.1 Sample and Data Collection

A cross sectional survey design was used in our study of external auditors in the United States. The item wording was refined through a pilot test (15 experienced auditors and two academic experts). From October 2023 to January 2024, the final questionnaire was sent to professionals through the use of professional networks, LinkedIn audit groups, and alumni databases of prominent accounting programs, and was conducted online via Qualtrics. In order to ensure quality, attention-check questions have been included, and responses that were completed in less than 8 minutes were excluded (median completion time was 14 minutes). After excluding 115 responses that were incomplete or did not pass attention checks, we received 287 valid responses (a 71.4% effective response rate) from the 402 that were initially sent.

Demographic Profile (Sample): Table 1 provides a summary of the demographic profile. Most of them were from the Big 4 (58.9%) while the rest were from national, regional and local firms. Positions ranged from staff auditor (12.2%), senior auditor (34.5%), manager (31.4%), senior manager (15.3%), to

Journal of Business Insight and Innovation
Volume 5 Issue 4, 2026
ISSN-p: 3006-2284, ISSN-e: 3006-0982
https://insightfuljournals.com/

partner/director (6.6%). The mean number of years of audit experience was 9.8 (SD = 6.4). Approximately 73.2% had a professional qualification (CPA or other equivalent qualifications), while 68.6% said their firm had implemented at least one of the auditing tools based on AI. Financial services, technology and manufacturing were the top three industries represented.

Table 1

Sample Demographics (N = 287)

Characteristic	Category	Frequency	Percent
Firm Type	Big 4	169	58.9%
	Non-Big 4	118	41.1%
Position	Staff	35	12.2%
	Senior	99	34.5%
	Manager	90	31.4%
	Senior Manager	44	15.3%
	Partner/Director	19	6.6%
Experience	Mean 9.8 years (SD = 6.4)		
Certification	CPA or equivalent	210	73.2%
AI Tool Adoption	At least one tool	197	68.6%

3.2 Measurement of Variables

Multi-item scales were adapted from previous research. Unless otherwise indicated, all items were rated on a 1-7 (strongly disagree to strongly agree) scale.

AI Adoption Intensity (AI_ADOPT): We created a 5-item scale grounded in the literature about technology use (Venkatesh et al., 2003), and based on audit-specific AI applications. For each of the following areas, respondents noted the degree to which they used AI tools to perform substantive testing, analytical procedures, fraud detection, risk assessment, and overall engagement management in their last 12 audit engagements: The responses were on a 1 to 7 scale, with 'Never used' being the lowest number and 'Extensively used' being the highest. The scale showed good reliability (α = 0.91).

The Audit Quality Improvement (AQ_IMP) scale adapted from Christensen et al. (2014) and Knechel et al. (2013), comprised 6 items asking respondents to rate the improvement in the use of AI tools that enhances audit quality in the areas of (1) the accuracy of audit evidence, (2) compliance with auditing standards, (3) reduction of undetected material misstatements (non-fraud), (4) reliability of audit conclusions, (5) documentation sufficiency, and (6) overall engagement quality. They asked, "The [dimension] has greatly improved since the introduction of the AI tools."

Perceived improvement in fraud detection effectiveness (FD_EFF): five items measuring perceived improvements in detecting fraudulent financial reporting and asset misappropriation were based on ACFE (2022) and on an existing scale. Sample answers: "We have found that the use of AI tools has helped us identify journal entries that are fraudulent, which would not have been identified without them" and "The use of AI-assisted analysis has enabled us to detect signs of fraud that would not have been detected manually." Cronbach's α was 0.89.

AI-Specific Training (AI_TRAIN): Three items included were formal training hours, self-efficacy in understanding the output of AI, and understanding limitations of models. Example: I have been sufficiently trained to critically analyse information generated by AI. (α = 0.87).

Firm Size (BIG4): Dummy variable, 1 = Big 4 firm, 0 = otherwise.

Control Variables: auditor experience (years of audit experience, log transformed to account for skewness), CPA certification (dummy), industry specialization (dummy if the auditor specializes in financial services or technology due to high relevance of AI in the industry), and firm AI maturity (dummy, years since the firm's first adoption of AI, log transformed). These are for the reasons why individuals and organizations might vary with respect to any quality judgement, even if they are not currently using AI.

Journal of Business Insight and Innovation
Volume 5 Issue 4, 2026
ISSN-p: 3006-2284, ISSN-e: 3006-0982
https://insightfuljournals.com/

3.3 Common Method Bias and Validity Checks

Since both independent and dependent variables were obtained from the same source, we adopted the necessary steps in the procedures and statistical methods to reduce common method bias (CMB). In terms of procedure, we ensured the anonymity of the subjects, divided the predictors from the criterion items, and employed different anchor labels. Harman's one factor test indicated that the first factor accounted for 31.4% of the variance, which is statistically less than 50%. Moreover, all variance inflation factors (VIFs) were below 2.0 after the full collinearity test was performed indicating that CMB is not a serious threat (Kock, 2015). The measurement model was evaluated using confirmatory factor analysis (CFA) with AMOS 28, and the results of the analysis showed that the standardized factor loadings for each construct were greater than 0.70, the composite reliability (CR) for each construct was greater than 0.80, and the average variance extracted (AVE) for each construct was greater than 0.50, which was met, so it can be said that the measurement model obtained convergent validity. The square root of the AVE for each construct was found to be greater than the correlation with the other constructs which was used to validate the construct's discriminant validity (Fornell & Larcker, 1981). In the following section 4 detailed results are presented.

3.4 Analytical Approach

Analysis of hypotheses was carried out using hierarchical ordinary least squares (OLS) regression with robust standard errors (RSE), carried out in SPSS 28. Two separate regressions were performed, one with AQ_IMP as the dependent variable, and one with FD_EFF as the dependent variable. Step 1: Control Variables entered. In Step 2 the AI_ADOPT element was added. The moderator (BIG4 for audit quality, AI_TRAIN for fraud detection) was added into Step 3. The interaction term (product of mean-centered predictor and moderator) was added in Step 4. Interactions were explored using simple slope analysis and Johnson-Neyman plot. To avoid the problem of multicollinearity, all continuous predictors were mean-centered. Variance inflation factors were checked and showed none of the models had values greater than 2.5. For robustness, we also analyzed with structural equation modeling (SEM) to correct for measurement error and simultaneous measurement of both DVs and obtained substantively similar results.

4. Results

4.1 Measurement Model Assessment

Prior to hypothesis testing, we conducted a CFA including all latent constructs (AI_ADOPT, AQ_IMP, FD_EFF, AI_TRAIN). The measurement model fit the data well: $\chi^2(146) = 312.47$, $p < .001$; CFI = 0.96; TLI = 0.95; RMSEA = 0.06 (90% CI: 0.05-0.07); SRMR = 0.04. All item loadings were significant ($p < .001$) and above 0.72. Table 2 presents factor loadings, CR, AVE, and Cronbach's alpha. Composite reliabilities ranged from 0.87 to 0.93, exceeding the 0.70 benchmark. AVE values ranged from 0.59 to 0.71.

Table 2

Measurement Model Assessment

Construct	Item	Loading	CR	AVE	α
AI_ADOPT	ADOPT1	0.82	0.91	0.67	0.91
	ADOPT2	0.85			
	ADOPT3	0.79			
	ADOPT4	0.84			
	ADOPT5	0.80			
AQ_IMP	QUAL1	0.78	0.89	0.59	0.89
	QUAL2	0.76			
	QUAL3	0.75			
	QUAL4	0.80			
	QUAL5	0.73			
	QUAL6	0.79			
FD_EFF	FRAUD1	0.81	0.88	0.60	0.88
	FRAUD2	0.77			
	FRAUD3	0.79			

Construct	Item	Loading	CR	AVE	α
	FRAUD4	0.76			
	FRAUD5	0.74			
AI_TRAIN	TRAIN1	0.84	0.87	0.69	0.87
	TRAIN2	0.83			
	TRAIN3	0.82			

Note: CR = composite reliability, AVE = average variance extracted, α = Cronbach's alpha. All loadings $p < .001$.

4.2 Descriptive Statistics and Correlations

Table 3 provides means, standard deviations, and correlations among study variables. The mean AI adoption intensity was 4.12 (SD = 1.65) on a 7-point scale, indicating moderate but varied adoption. Audit quality improvement had a mean of 4.87 (SD = 1.29), and fraud detection effectiveness 4.53 (SD = 1.41). AI adoption correlates strongly with both AQ_IMP ($r = .52$, $p < .01$) and FD_EFF ($r = .48$, $p < .01$), providing initial support for H1 and H2. Firm size (BIG4) is positively correlated with AI adoption ($r = .38$) and AQ_IMP ($r = .29$), while AI training correlates with AI adoption ($r = .44$) and FD_EFF ($r = .41$). These correlations are moderate, and all VIFs in regression models are below 2.5, mitigating multicollinearity concerns. The square root of AVE for each construct (diagonal in bold) was greater than the off-diagonal correlations, confirming discriminant validity. For instance, the highest correlation between any two latent constructs was between AI_ADOPT and AQ_IMP ($r = 0.52$), with square roots of AVE being 0.82 and 0.77 respectively.

Table 3

Descriptive Statistics and Correlations

Variable	Mean	SD	1	2	3	4	5	6	7	8
1. AI_ADOPT	4.12	1.65	**0.82**							
2. AQ_IMP	4.87	1.29	0.52**	**0.77**						
3. FD_EFF	4.53	1.41	0.48**	0.61**	**0.78**					
4. AI_TRAIN	3.97	1.58	0.44**	0.39**	0.41**	**0.83**				
5. BIG4	0.59	0.49	0.38**	0.29**	0.22**	0.30**	—			
6. Experience (log)	2.08	0.73	0.12*	0.16**	0.09	0.14*	0.05	—		
7. CPA	0.73	0.44	0.15*	0.11	0.08	0.13*	0.18**	0.32**	—	
8. Firm AI Maturity	1.32	0.62	0.35**	0.24**	0.21**	0.33**	0.45**	0.07	0.06	—

Note: Diagonal (bold) shows square root of AVE. *$p < .05$, **$p < .01$.

4.3 Hypothesis Testing

Audit Quality Improvement (H1 and H3a): Table 4 presents the hierarchical regression results for AQ_IMP. In Model 1, control variables collectively explain 11% of the variance (adjusted $R^2 = .11$, $F(5,281) = 7.86$, $p < .001$). Auditor experience ($\beta = .14$, $p < .05$) and firm AI maturity ($\beta = .18$, $p < .01$) are significant. When AI_ADOPT is added in Model 2, the explained variance increases substantially to 34% ($\Delta R^2 = .23$, F change = 99.81, $p < .001$). AI_ADOPT is a strong positive predictor ($\beta = .41$, $p < .001$), supporting H1. The effect size (Cohen's $f^2 = 0.35$) indicates a large effect.

Model 3 adds the direct effect of BIG4 ($\beta = .12$, $p < .05$), contributing incremental variance ($\Delta R^2 = .02$, $p < .05$). Model 4 includes the interaction term AI_ADOPT × BIG4. The interaction is positive and significant ($\beta = .17$, $p < .01$), with $\Delta R^2 = .03$ (F change = 12.34, $p < .001$). To interpret this, we plotted the simple slopes (Figure 2). For Big 4 firms (BIG4 = 1), the effect of AI adoption on AQ_IMP is steeper (simple slope = 0.55, $t = 7.44$, $p < .001$) than for non-Big 4 firms (simple slope = 0.28, $t = 4.12$, $p < .001$). Thus, H3a is supported.

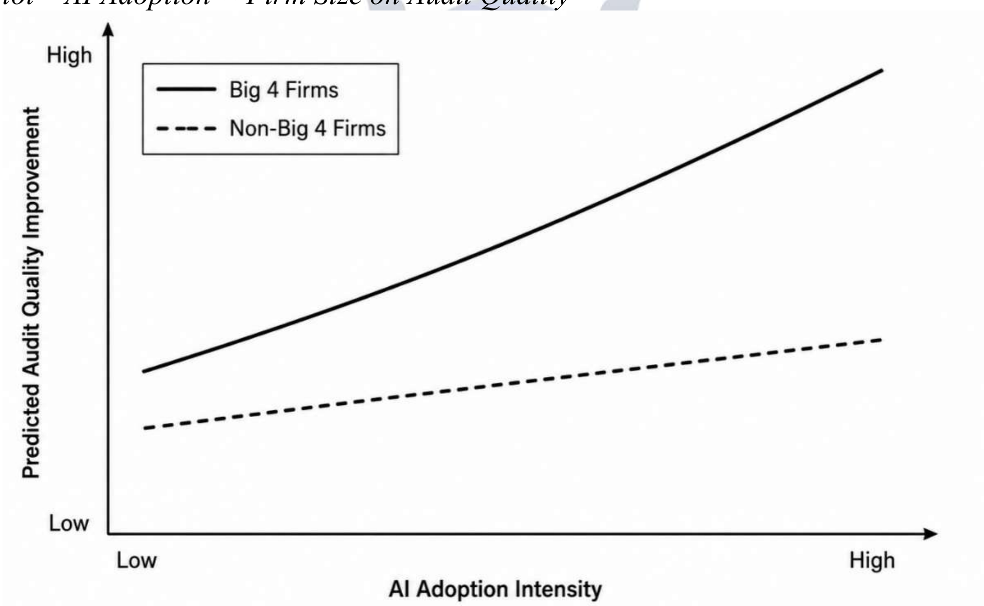
Journal of Business Insight and Innovation
Volume 5 Issue 4, 2026
ISSN-p: 3006-2284, ISSN-e: 3006-0982
https://insightfuljournals.com/

Table 4
Hierarchical Regression Results – Audit Quality Improvement (AQ_IMP)

Predictor	Model 1 (β)	Model 2 (β)	Model 3 (β)	Model 4 (β)
(Constant)	3.91***	2.54***	2.38***	2.29***
Experience (log)	0.14*	0.10	0.09	0.08
CPA	0.08	0.04	0.03	0.02
Industry specialization	0.11	0.07	0.06	0.05
Firm AI maturity	0.18**	0.09	0.06	0.05
AI_TRAIN	0.05	-0.02	-0.04	-0.04
AI_ADOPT		**0.41*** **	**0.38*** **	**0.40*** **
BIG4			0.12*	0.13*
AI_ADOPT × BIG4				**0.17 ****
R^2	0.12	0.35	0.37	0.40
Adjusted R^2	0.11	0.34	0.35	0.38
ΔR^2	—	0.23***	0.02*	0.03**
F change	7.86***	99.81***	5.12*	12.34**

Note: N = 287. Standardized coefficients reported. *p < .05, **p < .01, ***p < .001.

Figure 2
Interaction Plot – AI Adoption × Firm Size on Audit Quality



Fraud Detection Effectiveness (H2 and H3b): Table 5 reports parallel regressions for FD_EFF. Control variables in Model 1 account for 8% of the variance (adjusted $R^2 = .08$, $F = 5.93$, $p < .001$). Adding AI_ADOPT (Model 2) raises R^2 to .29 ($\Delta R^2 = .21$, F change = 85.47, $p < .001$); AI_ADOPT is significant ($\beta = .37$, $p < .001$), confirming H2. The effect is large ($f^2 = 0.30$).

Model 3 includes AI_TRAIN as a direct predictor ($\beta = .22$, $p < .001$), with $\Delta R^2 = .04$. Model 4 adds the interaction AI_ADOPT × AI_TRAIN, which is significant ($\beta = .14$, $p < .01$, $\Delta R^2 = .02$, F change = 8.76, $p < .01$). Simple slope analysis (Figure 3) shows that at one standard deviation above the mean of AI training, the effect of AI adoption is stronger ($b = 0.52$, $p < .001$) than at one standard deviation below the mean ($b = 0.24$, $p < .01$). Thus, H3b is supported. A complementary Johnson-Neyman analysis indicates that the effect of AI_ADOPT becomes significantly stronger when AI_TRAIN exceeds 3.9 (on the 1–7 scale), which covers 67% of the sample.

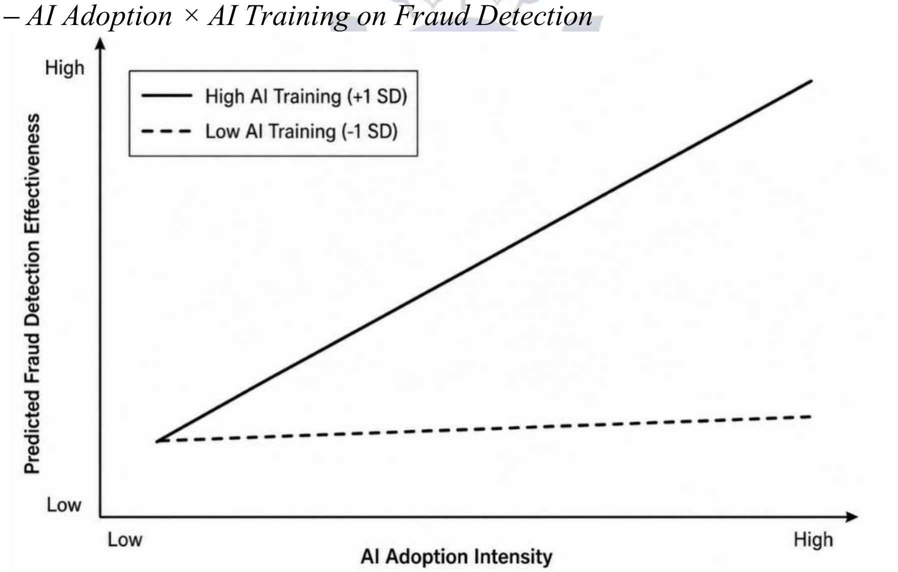
Journal of Business Insight and Innovation
Volume 5 Issue 4, 2026
ISSN-p: 3006-2284, ISSN-e: 3006-0982
https://insightfuljournals.com/

Table 5
Hierarchical Regression Results – Fraud Detection Effectiveness (FD_EFF)

Predictor	Model 1 (β)	Model 2 (β)	Model 3 (β)	Model 4 (β)
(Constant)	4.05***	2.79***	2.03***	1.94***
Experience (log)	0.06	0.03	0.02	0.01
CPA	0.05	0.02	0.01	0.00
Industry specialization	0.14*	0.09	0.06	0.05
Firm AI maturity	0.16*	0.07	0.02	0.02
BIG4	0.10	0.02	-0.02	-0.02
AI_ADOPT		**0.37*** **	**0.28*** **	**0.30*** **
AI_TRAIN			0.22***	0.23***
AI_ADOPT × AI_TRAIN				**0.14** **
R^2	0.09	0.30	0.34	0.36
Adjusted R^2	0.08	0.29	0.32	0.34
ΔR^2	—	0.21***	0.04***	0.02**
F change	5.93***	85.47***	16.89***	8.76**

Note: Standardized coefficients. *p < .05, **p < .01, ***p < .001.

Figure 3
Interaction Plot – AI Adoption × AI Training on Fraud Detection



Supplemental Analysis: We also tested cross-over interactions: AI_ADOPT × BIG4 on FD_EFF and AI_ADOPT × AI_TRAIN on AQ_IMP. The AI_ADOPT × BIG4 term on FD_EFF was not significant (β = .06, p = .29), while the AI_ADOPT × AI_TRAIN on AQ_IMP was marginally significant (β = .10, p = .07), suggesting that training may have some spillover effect on quality but not as robustly as on fraud detection. This pattern aligns with theory: fraud detection demands deeper cognitive engagement with AI outputs, whereas quality improvements from AI might be more readily realized through organizational integration (BIG4).

4.4 Robustness Checks

Structural Equation Modeling: To control measurement error and to test the full structural model at once, a structural model was estimated in Amos 28 including the latent interaction terms using the matched pairs product indicator approach with paths between AI_ADOPT and AQ_IMP as well as FD_EFF. The model fit was satisfactory (CFI = 0.94, RMSEA = 0.06). The standardized path coefficients were nearly identical: AI_ADOPT → AQ_IMP = 0.43 (p < .001); AI_ADOPT → FD_EFF = 0.39 (p < .001). The interaction terms were also significant and as expected. The findings are consistent with the fact that the results obtained from the OLS do not depend on the consideration of the measurement model.

Overall audit effectiveness was created by averaging the items for AQ_IMP and FD_EFF ($\alpha = 0.92$). Regressing this composite on AI_ADOPT and controls gave $\beta = 0.44$ ($p < .001$), with similar interaction effects, which confirmed the aggregation effect.

Robustness Check: Though we cannot exclude reverse causality completely (e.g., firms more mature could invest more in AI), we conducted a robustness check and performed a two-stage least squares regression using firm AI maturity as an instrumental variable. The instrument is relevant ($F = 27.8$) and theoretically exogenous to the current perceptions of quality in engagement. Following the Durbin-Wu-Hausman test, there was no evidence of rejection of exogeneity ($p = 0.42$), so endogeneity is not a serious problem. AI_ADOPT remained significant and positive in the 2SLS coefficient.

We used the marker variable technique with a social desirability scale as the marker variable as well (Common Method Bias Revisited): Correlations were also adjusted, with little change in these figures, which further diminished concerns about CMB.

5. Discussion

The aim of this study was to empirically examine the effect of the adoption of AI in audit and fraud detection, and to provide credible evidence. The results of our analysis are clear – AI tools used extensively by auditors indicate that audit quality has increased significantly, and they are more effective at detecting fraud. There isn't any trivializing about these effects, with standardized coefficients of 0.41 and 0.37 making them some of the biggest ever seen in the audit judgment literature, and a testament to the game-changing potential of AI in audit.

The benefits to audit quality resonate with the idea that AI can help auditors go beyond what they can do manually. AI can process entire populations, mitigating sampling risk and offering more precise risk assessments. Our findings align with the empirical work that shows that machine learning algorithms can outperform human auditors in detecting risky accounts (Brown-Liburd et al., 2015), but extend this to real-world environments where AI becomes part of a sociotechnical system. The R2 of 0.34 for the explained variance in perceived audit quality improvement indicates that AI is a prominent factor, even after adjusting for auditor experience and firm maturity.

One aspect of fraud detection is of special note, the robust effect ($R^2 = .29$). Auditing fraud has always been one problem auditors have struggled with and it is expected that this gap will continue (Hogan et al. 2008). Based on our data, AI could be key in closing that gap. AI's capacity to identify nuanced patterns and linguistic signals can be a valuable assistance to auditors' judgment, enabling them to concentrate on high-risk areas. The AI training is a space for moderation, but there's a key thing to remember: AI technology is not enough. Auditors will need the ability to evaluate AI-generated information, be aware of restrictions and limitations, and have professional skepticism. The "black box" aspect of AI may lead to poor audit quality if relying excessively on AI, as Sutton et al. (2016) stated. This risk can be reduced with training as it can help create a more critical and effective use of AI in fraud scenarios.

Theoretically and practically significant differential moderation patterns. Firm size (Big 4) is an amplifier of audit quality. This could be due to the fact that Big 4 firms have utilized AI in their exclusive audit strategies, allowing them to streamline and increase quality. They also have bigger, purer training sets, which makes AI tools more accurate. On the other hand, non-Big 4 companies that might be using third-party AI tools that do not focus on their specific industry may not reap the same benefits in terms of quality. However, firm size is not the most important moderator for fraud detection, the most important moderator is individual training. It means that even in smaller businesses, AI-tuned auditors can make a significant impact on fraud detection results. This discovery is good news for the whole audit industry, because it indicates that human capital development may make the benefits of AI accessible to everyone.

The results of our supplemental analysis are interesting: There was no significant moderation by the Big 4 for fraud detection. Quality improvement tools (e.g., automated substantive testing, continuous auditing modules) may require more customization, while fraud detection AI tools are more likely to be standardized

Journal of Business Insight and Innovation
Volume 5 Issue 4, 2026
ISSN-p: 3006-2284, ISSN-e: 3006-0982
https://insightfuljournals.com/

(e.g., off-the-shelf anomaly detection). In contrast, fraud detection could be more dependent on auditor judgment in interpreting the anomalies and the moderating locus could be more on the auditors.

Results are robust across a variety of robustness tests, such as SEM, alternative measurement, and endogeneity tests. Lack of severe common method bias and consistency across analyses enhances the confidence in the reliability of the results.

Theoretically, this study further adds to the TAM by showing that perceived usefulness (quality improvement and fraud detection) has significant relationship with the level of use in the audit area. It also reinforces the notion that the use of AI leads to a movement of professional judgment from the periphery to the centre (Elaboration Likelihood Model) if auditors are motivated and able to acquire the skills via training. The research study therefore helps to broaden the understanding of human-AI cooperation in knowledge-based occupations.

The implications are instant for audit practice. To achieve the highest quality benefit, audit firms need to invest in the data infrastructure and tailor solutions to their specific audit methodology in addition to the investment in AI tools. What is more important is that, especially for interpreting and questioning AI outputs, understanding the false positive rate, and integrating the insights from AI with traditional evidence, robust AI training programs should be required. This could be part of the continuous professional development of regulators and standard-setters to include AI skills. The PCAOB and IAASB have the potential to offer guidance on how to leverage AI in evidence collection and documentation processes, maintaining that the benefits of enhanced quality and fraud detection don't compromise the benefits of new potential risks, including algorithmic bias or overreliance.

6. Conclusion, Limitations, and Future Research

This study provides strong empirical evidence supporting the positive relationships between the intensity of AI adoption and audit quality, as well as the intensity of AI adoption and the effectiveness of fraud detection, which are moderated by firm-level and individual-level factors. We find that the contribution of AI is significant as measured by change in perceived audit quality improvement and fraud detection, accounting for as much as 34% of variance in both measures, across our sample of 287 auditors. The moderating effect of big 4 firm size on the quality link and of AI training on fraud detection suggests a potential dual trajectory—one of organization integration heightening quality and one of human expertise enabling fraud detection. These learnings offer a detailed guide for implementing AI in auditing.

Although these have been all contributions, the study still has some limitations that are good for future research. First, because we have used cross-sectional survey data, there is potential for common method bias and reverse causality. We carried out multiple tests to address and identify these challenges, but longitudinal studies with audit quality indicators (such as inspection findings, restatement rates) before and after the use of AI would give more empirical evidence. Second, our sample only includes auditors in the U.S., which may not allow for generalizability. This could be explored through cross-cultural studies to determine whether the impact of AI is different in various institutional settings. Third, we relied on perceived measures of improvement rather than on objective audit measures of quality. Firm-level AI adoption indices could be correlated to PCAOB Inspection Deficiencies or PCAOB Fraud Detection from Litigation data in the future, but there is a lack of those data. Fourth, we did not distinguish between various types of AI tools (e.g., supervised machine learning, unsupervised machine learning, NLP); future studies might explore which specific technologies have the observed impacts. Fifth, the survey framework did not address potential negative effects of AI, including auditor deskilling, over-reliance and algorithmic bias. These darker aspects might be investigated in qualitative or experimental research. Furthermore, the relationship of AI with auditor skepticism, both as a result of inducing automation bias and as a catalyst to increased skepticism due to new risks that it identifies, is worthy of further research. Lastly, our sample size was sufficient for regression analysis, but could not support more detailed subgroup analyses (e.g., by industry, specific AI tools).

To conclude, as AI continues to be integrated into the audit profession, it is essential to grasp its impact in the real world. This study provides an early empirical comprehensive view that supports the belief that the use of AI, coupled with thoughtful integration and human development, can raise the level of audit quality and enhance fraud detection. The way forward is in leveraging the synergy between human judgment and machine intelligence, not in replacing auditors.

Acknowledgments

The author/s would like to thank the participants and the respective universities for their support and cooperation in conducting this study.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Informed Consent

Informed consent was obtained from all individual participants included in the study.

Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of 1964 Helsinki declaration and its later amendments.

Data Availability

The datasets generated during and analysed during the current study are available from the corresponding author on reasonable request.

References

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). An introduction to data analysis for auditors and accountants. *Auditing: A Journal of Practice & Theory, 36*(4), 1–8. https://doi.org/10.2308/ajpt-51657

Association of Certified Fraud Examiners. (2022). *Occupational fraud 2022: A report to the nations*. https://www.acfe.com/report-to-the-nations

Brazel, J. F., Jones, K. L., & Zimbelman, M. F. (2009). Using nonfinancial measures to assess fraud risk. *Journal of Accounting Research, 47*(5), 1135–1166. https://doi.org/10.1111/j.1475-679X.2009.00349.x

Brown-Liburd, H., Issa, H., & Lombardi, D. R. (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

Christensen, B. E., Glover, S. M., & Wolfe, C. J. (2014). Do critical audit matter paragraphs in the auditor's report change nonprofessional investors' decision to invest? *Auditing: A Journal of Practice & Theory, 33*(4), 71–93. https://doi.org/10.2308/ajpt-50790

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

DeAngelo, L. E. (1981). Auditor size and audit quality. *Journal of Accounting and Economics, 3*(3), 183–199. https://doi.org/10.1016/0165-4101(81)90002-1

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.2307/3151312

Hogan, C. E., Rezaee, Z., Riley, R. A., & Velury, U. K. (2008). Financial statement fraud: Insights from the academic literature. *Auditing: A Journal of Practice & Theory, 27*(2), 231–252. https://doi.org/10.2308/aud.2008.27.2.231

Humphreys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Carroll, J. M. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems, 50*(3), 585–594. https://doi.org/10.1016/j.dss.2010.08.009

International Auditing and Assurance Standards Board. (2020). *International Standard on Auditing 240: The auditor's responsibilities relating to fraud in an audit of financial statements*. International Federation of Accountants.

Journal of Business Insight and Innovation
Volume 5 Issue 4, 2026
ISSN-p: 3006-2284, ISSN-e: 3006-0982
https://insightfuljournals.com/

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

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*(Supplement 1), 385–421. https://doi.org/10.2308/ajpt-50350

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration, 11*(4), 1–10. https://doi.org/10.4018/ijec.2015100101

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

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems, 50*(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006

Public Company Accounting Oversight Board. (2019). *AS 2110: Identifying and assessing risks of material misstatement*. https://pcaobus.org/oversight/standards/auditing-standards/details/AS2110

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

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). Academic Press. https://doi.org/10.1016/S0065-2601(08)60214-2

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. https://doi.org/10.1016/j.accinf.2016.07.005

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security, 57*, 47–66. https://doi.org/10.1016/j.cose.2015.09.005

Wolfe, D. T., & Hermanson, D. R. (2004). The fraud diamond: Considering the four elements of fraud. *The CPA Journal, 74*(12), 38–42.

Zhang, C., Yang, L., & Appelbaum, D. (2020). Toward effective big data analysis in continuous auditing. *Accounting Horizons, 34*(3), 173–190. https://doi.org/10.2308/acch-52683