Journal of Finance and Investment Analysis, Vol. 15, No. 4, 2026, 85-111
ISSN: 2241-0988 (print version), 2241-0996(online)
https://doi.org/10.47260/jfia/1545
Scientific Press International Limited

Generative AI and Audit Quality in Jordanian Listed Industrial Companies

Hamza Alaodat[1*] and Ahmad El-Khateeb[1]

Abstract

Generative artificial intelligence (GenAI) may support audit work, but its association with audit quality depends on how auditors use and evaluate its outputs. This study examines task–technology fit, quality-oriented utilization, professional judgment quality, and digital competence among 214 external auditors serving Jordanian listed industrial companies. Respondent-level averaged indicators were analyzed using a reflective measurement model and regression equations corresponding to PROCESS Models 1, 4, and 7. Audit-firm size and professional certification were included as covariates. Reflective discriminant validity was supported by HTMT values of 0.587–0.660. Across 5,000 bootstrap samples, the indirect association through professional judgment was 0.115 (95% CI [0.078, 0.159]). Conditional indirect associations increased from 0.048 at low digital competence to 0.154 at high competence; the index of moderated mediation was 0.082 (95% CI [0.043, 0.126]). The findings are consistent with a professional-judgment pathway and a competence-related boundary condition. They support further evaluation of task suitability, verification practices, and competence development. Cross-sectional self-reports, aggregated indicators, and incomplete independent validation of formative composites limit causal interpretation and generalization.

JEL classification numbers: M42, M15, O33.
Keywords: Generative artificial intelligence, External audit quality, Task-technology fit, Professional judgment, Digital competence, Jordan.

[1] Al-Hussein Bin Talal University, Ma'an, Jordan. * Corresponding author.

Article Info: *Received:* October 26, 2026. *Revised*: November 12, 2026.
Published online: November 17, 2026.

1. Introduction

1.1 Background and Motivation

Generative artificial intelligence is entering external audit workflows, but its implications for audit quality remain uncertain. Interviews with audit professionals identify opportunities for automation alongside concerns about explainability, privacy, reliability, and excessive reliance on AI (Kokina et al., 2025). Regulatory guidance also emphasizes documentation and professional judgment when using AI tools (Financial Reporting Council, 2025). Availability and frequency of use alone do not establish whether a tool is appropriate for a particular audit task or whether its output supports sound judgment. Task-technology fit offers a basis for examining that relationship (Goodhue and Thompson, 1995), while research on AI-co-piloted auditing highlights the need to retain professional oversight (Gu et al., 2024). These issues form part of the wider research agenda on AI in accounting (Stratopoulos and Wang, 2025).

External audits of Jordanian listed industrial companies provide a relevant setting for this question. These companies operate across manufacturing activities with different reporting risks and evidence requirements. Auditors may use GenAI to organize narrative information or develop initial explanations, but they remain responsible for evaluating evidence and supporting their conclusions. The quality of that work matters to investors who rely on audited financial statements. Differences in audit-firm resources and individual competence may also shape how auditors use and verify generated output.

The study combines two theoretical perspectives. Task-technology fit theory explains why technology is more useful when its capabilities match task requirements (Goodhue and Thompson, 1995). Research on reliance on automation distinguishes appropriate use from misuse and disuse (Parasuraman and Riley, 1997). Auditor reliance can also be affected by whether advice comes from an AI system or a human specialist (Commerford et al., 2022). Together, these perspectives motivate a model in which task fit is associated with quality-oriented use, and the relationship between use and judgment quality depends on digital competence.

1.2 Research Problem

The study asks how perceived fit between GenAI capabilities and audit tasks relates to quality-oriented utilization, professional judgment, and engagement-process audit quality. It also examines whether digital competence changes the utilization-to-judgment relationship. Gu et al. (2024) discuss AI-co-piloted auditing, while Commerford et al. (2022) examine auditor reliance on AI in a complex-estimate experiment. Zhao et al. (2026) study the relationship between auditors' digital expertise and audit quality using archival data from China. The present study brings these questions together in a survey of auditors serving Jordanian listed industrial companies.

1.3 Research Questions

RQ1: To what extent do external auditors serving Jordanian industrial companies currently use GenAI tools in their audit tasks?
RQ2: Is GenAI task-technology fit positively associated with quality-oriented GenAI utilization among auditors of listed Jordanian industrial companies?
RQ3: Does auditor professional judgment quality mediate the relationship between quality-oriented GenAI utilization and engagement-process audit quality?
RQ4: Does auditor digital competence moderate the relationship between quality-oriented GenAI utilization and professional judgment quality?
RQ5: Is the conditional indirect association between GenAI utilization and audit quality stronger at higher levels of digital competence?

1.4 Research Objectives

To assess the current level and nature of GenAI use in external audits of listed Jordanian industrial companies.
To examine the association between GenAI task-technology fit and quality-oriented utilization.
To examine professional judgment quality as a partial mediator between quality-oriented utilization and engagement-process audit quality.
To determine whether digital competence moderates the utilization-to-judgment path.
To examine the implications of the findings for Jordanian audit firms, JACPA, and the Jordan Securities Commission within the limits of the study design and measurement approach.

1.5 Significance of the Study

The study applies an integrated task-fit and automation-reliance framework to auditors serving Jordanian listed industrial companies. It distinguishes quality-oriented utilization from frequency of use and examines professional judgment as a possible pathway linking utilization with audit quality. For audit firms, the framework identifies questions for task classification, verification procedures, and competence development. For professional bodies and regulators, it provides conditional process estimates relevant to discussions of GenAI governance. The evidence concerns auditors' perceptions of engagement processes rather than independently observed financial-reporting outcomes.

1.6 Scope and Limitations

The study concerns external financial-statement auditing and GenAI tools used during evidence collection and judgment formation. It excludes the separate task of auditing a client's AI systems (Li and Goel, 2025). The sample comprises external auditors serving Jordanian industrial companies listed on the Amman Stock Exchange, with at least one year elapsed since first professional use of a GenAI tool. Field data were collected from February to May 2026. The cross-sectional design

limits causal interpretation, and the self-report measures are susceptible to common method variance.

2. Theoretical Framework

2.1 Generative AI as an AIS Phenomenon

Accounting information systems (AIS) research has long examined how information technologies interact with accounting tasks, controls, and human decision makers (Sutton et al., 2016; Issa et al., 2016). Foundation models, large language models (LLMs), and GenAI are related but distinct constructs (Bommasani et al., 2021). GenAI refers specifically to the class of models that generate new artifacts — text, code, tables, structured outputs — rather than returning classifications or scores. Unlike conventional analytics that return exceptions or visualizations requiring human interpretation, GenAI can additionally generate linguistic interpretations, propose audit procedures, draft correspondence, and create alternative models of the data (Feuerriegel et al., 2024). Dong et al. (2024) review applications of ChatGPT and related language models, their use as research tools, and their implications for accounting and finance professionals. Their synthesis motivates closer examination of the conditions under which GenAI use relates to professional work. The study focuses on GenAI used by auditors in audit tasks, not on auditing a client's AI system (Li and Goel, 2025).

2.2 Task-Technology Fit Theory and Quality-Oriented Utilization

Task-technology fit theory proposes that technology contributes to performance when its capabilities match task requirements and it is used in practice (Goodhue and Thompson, 1995). In this study, task fit is a theoretically specified antecedent of quality-oriented utilization; the survey does not establish that fit perceptions precede use. Potential applications include summarizing board minutes, comparing contract clauses, organizing narrative evidence, and generating alternative explanations (Gu et al., 2024). Deterministic recalculation, physical inspection, and specialist judgments require different capabilities and safeguards. TTF is specified as a formative composite because its indicators represent distinct dimensions of fit rather than interchangeable manifestations of one underlying trait (Jarvis et al., 2003; Coltman et al., 2008). Quality-oriented utilization concerns verification and analytical coverage rather than use frequency alone.

2.3 Calibrated Automation Reliance, Professional Judgment, and Skepticism

Research on automation distinguishes appropriate reliance from misuse and disuse (Parasuraman and Riley, 1997). Commerford et al. (2022) examine algorithm aversion in audit judgments, showing why auditor responses to AI advice should not be assumed to involve excessive reliance alone. Skitka et al. (1999) examine automation bias, while Lee and See (2004) and Parasuraman and Manzey (2010) explain the importance of calibrated trust and attention. Under ISA 200,

professional judgment involves applying relevant training, knowledge, and experience (IAASB, 2025). Professional skepticism concerns a questioning attitude and critical assessment of evidence (Nelson, 2009; Nolder and Kadous, 2018). Here, judgment quality includes identifying relevant cues, evaluating alternatives, recognizing uncertainty, and corroborating conclusions. Research on cognitive forcing functions provides a rationale for encouraging deliberate evaluation of AI suggestions (Buçinca et al., 2021); it does not directly validate a particular audit-prompting protocol.

2.4 Auditor Digital Competence as a Boundary Condition

Drawing on DigComp 2.2 (Vuorikari et al., 2022) re-specified for the evidentiary demands of external auditing, auditor digital competence encompasses four audit-relevant dimensions: (1) task classification — identifying which audit sub-tasks are and are not appropriate for GenAI; (2) source grounding — tracing a generated proposition to original evidence of sufficient authority and relevance; (3) hallucination awareness — recognizing when generated content contains plausible-sounding but unverifiable claims (Ji et al., 2023); and (4) confidentiality management — understanding data-residency implications of submitting client data to external models. Source grounding is the central dimension: observable behaviors include returning to the original document, checking whether a cited passage exists, and documenting the verification process. Zhao et al. (2026) link archival proxies for auditor digital expertise to audit quality in Chinese listed companies, providing relevant prior evidence for the competence-as-moderator hypothesis, though their construct captures general digitization rather than the GenAI-specific verification behaviors specified here.

2.5 Conceptual Model

The conceptual model contains five constructs. GenAI task-technology fit is specified as a formative antecedent of quality-oriented utilization. Utilization is associated with professional judgment quality and with engagement-process audit quality, both directly and through the proposed judgment pathway. Auditor digital competence is specified as a formative composite that moderates the utilization-to-judgment path. The resulting model represents first-stage moderated mediation, corresponding to PROCESS Model 7 (Hayes, 2018). All constructs are analyzed using individual auditors' survey responses; references to tasks and engagements identify the content of the measures, not separate statistical levels. The hypothesized ordering remains theoretical because all measures were collected in one cross-sectional survey.

3. Prior Literature and Research Gap

3.1 GenAI, Automation Reliance, and Audit Quality

Commerford et al. (2022) investigate auditor reliance on AI advice in a complex-estimate task and provide evidence concerning algorithm aversion. Their experiment concerns how advice sources affect judgment, rather than testing a GenAI task-fit and competence model. Kokina et al. (2025) identify AI adoption challenges through interviews with audit professionals, including explainability, privacy, reliability, and overreliance. Their study addresses AI more broadly and should not be read as evidence that all respondents had deployed GenAI in live audit procedures. Brown-Liburd et al. (2015) discuss how large volumes of information can place demands on auditor judgment. Fotoh and Mugwira (2025) examine implications and ethical considerations of large language models in external audit. These studies motivate attention to verification and competence without establishing the proposed mediation relationships.

3.2 Digital Competence, Expertise, and Audit Quality

Zhao et al. (2026) link auditors' digital expertise to audit quality using archival evidence from Chinese listed companies. Their measures differ from the GenAI-specific competence dimensions considered here. Torroba et al. (2025) examine factors associated with audit professionals' adoption of data analysis and AI in Spain, addressing adoption rather than the proposed judgment pathway. Yoon et al. (2015) discuss Big Data as complementary audit evidence. Zhang et al. (2022) examine explainable AI in auditing, making transparency relevant to the evaluation of generated output. Lombardi et al. (2025) report Delphi-panel expectations about automation, professional skills, ethical concerns, and regulatory influences. Together, this literature supports examining competence while leaving its proposed moderating role open to empirical testing.

3.3 Task-Technology Fit in AIS Research

Goodhue and Thompson (1995) link task-technology fit with individual performance. Applied to audit, this perspective asks whether a tool's capabilities match the evidentiary and analytical requirements of a specific procedure. Gu et al. (2024) discuss AI-co-piloted auditing, and Stratopoulos and Wang (2025) set out a broader AI research agenda. Reviews and research agendas also identify the need to connect technologies with accounting tasks and decisions (Appelbaum et al., 2017). Measurement specification matters in this setting: distinct task and technology characteristics need not be interchangeable or highly correlated. Jarvis et al. (2003) and Coltman et al. (2008) explain why formative composites require assessment different from that used for reflective scales.

3.4 Research Gap and Contribution

The study addresses three related questions. The first concerns whether relationships examined in other jurisdictions also appear among auditors serving Jordanian listed industrial companies. The second concerns the proposed sequence from task fit to quality-oriented use, professional judgment, and perceived audit quality. The third concerns GenAI-specific competence, including source verification, hallucination awareness, and confidentiality management. The contribution lies in examining these relationships together in a conditional process analysis of the case-level data.

4. Hypotheses Development

4.1 H1: GenAI Task-Technology Fit → Quality-Oriented Utilization

A match between task requirements and GenAI capabilities may make appropriate uses easier to identify and reduce the effort required to direct a tool toward analytical coverage and verification (Goodhue and Thompson, 1995). Auditors who perceive stronger fit may therefore report more quality-oriented utilization. Reverse causality is also possible: successful use may lead auditors to rate task fit more favorably. The cross-sectional design cannot distinguish these directions.

H1: GenAI task-technology fit is positively associated with quality-oriented GenAI utilization in external audit.

4.2 H2: Quality-Oriented Utilization → Professional Judgment Quality

Quality-oriented GenAI use may support judgment by expanding the set of cues and hypotheses considered before a conclusion is reached. A model used adversarially — to generate alternative fraud explanations, identify assumptions embedded in management's narrative, flag internal inconsistencies — provides hypothesis-diversification inputs otherwise bounded by individual knowledge and time constraints (Nelson, 2009). The mechanism is behavioral: not the volume of generated content, but the adversarial interrogation of that content through verification and critical comparison with direct evidence (Buçinca et al., 2021).

H2: Quality-oriented GenAI utilization is positively associated with auditor professional judgment quality.

4.3 H3: Professional Judgment Quality → Engagement-Process Audit Quality

External audit quality depends on how auditors interpret evidence and translate it into procedures, escalation decisions, and reporting. Sound professional judgment can support the selection of appropriate procedures and the evaluation of material misstatement risks (IAASB, 2025). Research on auditor reliance on AI also indicates that the source and evaluation of advice can influence judgment (Commerford et al., 2022). These arguments motivate a positive association between professional judgment quality and engagement-process audit quality.

H3: Auditor professional judgment quality is positively associated with engagement-process audit quality.

4.4 H4: Utilization and Audit Quality Controlling for Judgment

GenAI utilization may be associated with engagement-process audit quality through several proposed operational pathways beyond the professional-judgment pathway. These include organizing relevant evidence across unstructured material, allowing time for investigation and review, supporting consistent documentation, and allocating attention to substantive analysis (Gu et al., 2024; Brown-Liburd et al., 2015). These theoretical explanations motivate testing the utilization–audit-quality association while controlling for professional judgment quality; they are not separately identified mechanisms in the present survey.

H4: Quality-oriented GenAI utilization is positively associated with engagement-process audit quality, controlling for professional judgment quality.

4.5 H5: Professional Judgment Quality as Partial Mediator

Many GenAI outputs in audit are advisory rather than dispositive. An auditor must evaluate a generated contract summary, candidate fraud explanation, or proposed procedure against relevant evidence (Gu et al., 2024). Partial mediation is specified to examine both the utilization–audit-quality association conditional on judgment and the indirect statistical association through judgment. This decomposition does not establish temporal ordering or a causal mechanism.

H5: Auditor professional judgment quality partially mediates the positive relationship between quality-oriented GenAI utilization and engagement-process audit quality.

4.6 H6: Digital Competence as Moderator

Digital competence affects the value extracted from the same quality-oriented use episode by changing the quality of inputs to the model and the quality of post-generation evaluation. A competent auditor can classify appropriate tasks, trace generated claims to their evidentiary basis (source grounding), identify hallucinated propositions (Ji et al., 2023), and protect client data. These competencies may strengthen the relationship between quality-oriented use and judgment quality. At high digital competence, the positive utilization-to-judgment relationship should therefore be stronger (Zhao et al., 2026; Lee and See, 2004; Parasuraman and Manzey, 2010).

H6: Auditor digital competence positively moderates the relationship between quality-oriented GenAI utilization and professional judgment quality, such that the positive relationship is stronger at higher levels of digital competence.

4.7 H7: Conditional Indirect Association with Audit Quality

H2, H3, and H6 jointly imply a conditional indirect association. The proposed model links quality-oriented utilization with professional judgment quality at different levels of digital competence, and links judgment quality with engagement-process audit quality. The indirect statistical association is hypothesized to be stronger at high than at low digital competence. This constitutes a test of the integrated TTF–calibrated-reliance framework within the cross-sectional data.

H7: The positive indirect statistical association between quality-oriented GenAI utilization and engagement-process audit quality, through professional judgment quality, is stronger at higher levels of auditor digital competence.

5. Methodology

5.1 Research Design

The study uses a cross-sectional survey design to examine associations between GenAI-related perceptions and self-reported audit processes. Task–technology fit (TTF) is specified as an antecedent of quality-oriented utilization (QOU). Professional judgment quality (PJQ) is the proposed mediator between QOU and engagement-process audit quality (AQ), and auditor digital competence (ADC) moderates the QOU–PJQ path. This first-stage moderated mediation specification corresponds to PROCESS Model 7 (Hayes, 2018). Each observation represents one auditor. The measurement analysis distinguishes formative composites (TTF and ADC) from reflective constructs (QOU, PJQ, and AQ). Temporal ordering follows the theoretical model and is not identified by the cross-sectional design.

5.2 Population, Sample, and Data Collection

The target population comprises JACPA-registered external auditors serving industrial companies listed on the Amman Stock Exchange during 2025–2026. The sampling frame was estimated at approximately 350–400 individual auditors. Purposive eligibility criteria required affiliation with a firm holding a current engagement with a listed industrial company, a university degree in accounting, auditing, or an equivalent field, and at least one year since first professional use of GenAI. The elapsed-time criterion does not require continuous use: an auditor with an early initial exposure may currently use GenAI only occasionally or experimentally. The tools considered include ChatGPT, Microsoft Copilot, Google Gemini, and comparable applications. The unit of observation and inference is the individual auditor, rather than the industrial company or audit firm.

An electronic questionnaire was distributed to 240 eligible auditors between February and May 2026, with two reminders at two-week intervals. Of 225 returned questionnaires, 11 were excluded for incompleteness or straight-line responding, leaving 214 usable cases. The gross response rate was 93.8% (225/240), and the usable-response rate was 89.2% (214/240). For a population benchmark of 380, the conventional sample-size calculation yields approximately 191 cases (Krejcie and Morgan, 1970). This comparison does not convert purposive recruitment into

probability sampling or demonstrate power for the interaction and indirect associations. All 214 analytical cases have complete averaged indicators and model covariates. The certification category "None" denotes absence of a professional certification and is retained as a valid category.

5.3 Measurement Instrument

The questionnaire covers five constructs using a five-point agreement response format. The analytical matrix comprises 44 respondent-level averaged indicators. Each indicator summarizes underlying responses from the same participant; aggregation is within the auditor and does not combine different participants. Fractional values are interpreted as averaged indicators, and the matrix is analyzed as continuous without rounding. Reliability and factor-analytic results apply to these averaged indicators rather than to the underlying subitems. Structural models use operational construct scores linked to the indicators through the linear mappings in Table 2 and Appendix A. Distinguishing the response format, averaged indicators, and construct scores prevents interpreting transformed scores as individual five-category answers.

GenAI task–technology fit (TTF): nine formative averaged indicators cover task complexity, processing of unstructured text, usefulness of linguistic outputs, support for alternative hypotheses, and task-misfit areas such as deterministic recalculation. These dimensions jointly define fit and need not be interchangeable manifestations of a single trait (Jarvis et al., 2003; Coltman et al., 2008).

Quality-oriented GenAI utilization (QOU): eight reflective averaged indicators covering verification-focused use, consideration of alternative explanations, and the intended contribution of GenAI to analytical coverage.

Professional judgment quality (PJQ): ten reflective averaged indicators assessing cue identification, evaluation of alternatives, recognition of uncertainty, and skeptical verification, adapted from Nelson (2009) and Nolder and Kadous (2018).

Auditor digital competence (ADC): eight formative averaged indicators covering task classification (one indicator), source grounding (three indicators), hallucination awareness (two indicators), and confidentiality management (two indicators), informed by Vuorikari et al. (2022).

Engagement-process audit quality (AQ): nine reflective averaged indicators assessing substantive-procedure coverage, evidence documentation, escalation decisions, and review-point calibration.

5.4 Content Validity, CFA, and Reliability

Seven subject-matter experts assessed instrument content: four accounting or accounting information systems researchers and three senior practicing auditors. A pilot involving 25 auditors assessed clarity and indicated an average completion time of 17 minutes. Reflective measurement was evaluated through a three-correlated-factor maximum-likelihood CFA of the 27 averaged indicators for QOU, PJQ, and AQ. All 17 formative indicators were excluded from the CFA. Factor

variances were fixed to one, while cross-loadings and residual correlations were fixed to zero. Covariance maximum likelihood was implemented in Python/SciPy. The model converged with $\chi^2(321) = 318.129$, $\chi^2/df = 0.991$, CFI = 1.000, TLI = 1.001, RMSEA = 0.000, and SRMR = 0.037. The unbounded TLI is slightly above one because χ^2 is below its degrees of freedom. These fit indices describe the reflective covariance structure and do not validate the formative composites.

Reflective internal consistency and convergence are summarized in Table 1. Cronbach's α ranges from 0.912 to 0.932, composite reliability from 0.913 to 0.933, and AVE from 0.567 to 0.592. Standardized CFA loadings range from 0.661 to 0.843 for QOU, 0.707 to 0.809 for PJQ, and 0.647 to 0.854 for AQ. Positive loadings, reliability estimates, and extracted variance support coherence of the reflective averaged indicators. These statistics do not establish the reliability of the underlying subitem-level responses or eliminate shared-method inflation.

Table 1: Measurement characteristics and reflective scale reliability (n = 214)

Construct	Indicators	α	CR	AVE	Mean	SD
TTF	9	—	—	—	3.610	0.450
QOU	8	0.912	0.913	0.567	3.280	0.820
PJQ	10	0.932	0.933	0.581	3.720	0.400
ADC	8	—	—	—	3.390	0.650
AQ	9	0.928	0.929	0.592	3.640	0.400

Note: α, CR, and AVE refer to the reflective averaged indicators. Means and SDs refer to the construct-level scores. Dashes indicate criteria not applicable to formative blocks. Appendix A reports the linear score mappings and compares construct scores with equal-weight indicator means.

Formative measurement is evaluated separately because internal consistency and AVE are not appropriate requirements for indicators that jointly define a composite (Jarvis et al., 2003; Coltman et al., 2008). Table 2 presents the linear coefficients defining the operational TTF and ADC mappings and the VIF for each indicator. The mappings use intercepts of −0.905444 for TTF and −2.969031 for ADC. TTF indicator VIFs range from 1.188 to 1.867, while ADC VIFs range from 1.097 to 1.740, indicating limited within-block collinearity. Coefficients specify the numerical score mappings and are not structural effects; significance tests are not attached to them. Content coverage and limited collinearity support the operational composite specification, but independent global criterion measures are absent, so convergent validity is not established through redundancy analysis. The coefficients and intercepts reported in Table 2 were estimated by ordinary least squares regression of each available formative construct score on its corresponding averaged indicators, yielding linear equations that reproduce the operational scores rather than independently validated formative outer weights.

Table 2: Formative composite score coefficients and indicator collinearity

Indicator	Score coefficient	VIF
TTF1	0.095333	1.188
TTF2	0.079064	1.283
TTF3	0.098738	1.230
TTF4	0.207901	1.298
TTF5	0.117277	1.326
TTF6	0.134514	1.309
TTF7	0.300660	1.575
TTF8	0.184625	1.615
TTF9	0.285462	1.867
ADC1	0.156269	1.097
ADC2	0.242500	1.373
ADC3	0.134264	1.211
ADC4	0.239244	1.306
ADC5	0.257368	1.338
ADC6	0.168367	1.483
ADC7	0.335994	1.571
ADC8	0.580310	1.740

Note: Coefficients are on the original indicator scales. For each composite, multiply each indicator by its coefficient, sum the products, and add the stated intercept. Values are rounded to six decimals; calculations use full precision. VIF assesses collinearity and does not independently establish convergent validity.

Discriminant validity among QOU, PJQ, and AQ was evaluated using the heterotrait–monotrait ratio of absolute indicator correlations, HTMT (Henseler et al., 2015). Table 3 reports point estimates and percentile 95% confidence intervals from 5,000 whole-case bootstrap samples (seed 20261110). HTMT is 0.587 for QOU–PJQ, 0.621 for QOU–AQ, and 0.660 for PJQ–AQ; the largest upper confidence limit is 0.736. All estimates and upper limits are below the conservative 0.85 threshold. The reflective constructs are empirically distinguishable at the averaged-indicator level under this criterion. TTF and ADC are excluded from this reflective-scale test. Discriminant validity does not itself exclude common method bias.

Table 3: Reflective construct discriminant validity

Construct pair	HTMT	LL 95%	UL 95%
QOU–PJQ	0.587	0.494	0.676
QOU–AQ	0.621	0.534	0.700
PJQ–AQ	0.660	0.576	0.736

5.5 Institutional Permission and Ethical Considerations

Permission to conduct the study was obtained from the management of the participating companies.

6. Results and Analysis

6.1 Demographic Profile

Table 4: Demographic Characteristics of the Sample (n = 214)

Variable	Category	n	%
Gender	Male	163	76.2
	Female	51	23.8
Academic Qualification	Bachelor's	98	45.8
	Master's	89	41.6
	Doctoral	27	12.6
Years of Experience	< 5 years	58	27.1
	5–10 years	87	40.7
	> 10 years	69	32.2
Professional Certification	JCPA/CPA	112	52.3
	ACCA	43	20.1
	None	59	27.6
Audit Firm Size	Big 4	76	35.5
	Mid-tier	94	43.9
	Small firm	44	20.6
GenAI Use Frequency	Regular (weekly)	89	41.6
	Occasional (monthly)	73	34.1
	Rare / experimental	52	24.3
Primary GenAI Tool	ChatGPT	97	45.3
	Microsoft Copilot	61	28.5
	Google Gemini	31	14.5
	Other / specialized	25	11.7

The sample comprises 163 men and 51 women. Professional certification comprises JCPA/CPA (112; 52.3%), ACCA (43; 20.1%), and none (59; 27.6%); 155 cases (72.4%) hold either certification. Audit-firm categories comprise Big 4 (76), mid-tier (94), and small firms (44). Rare or experimental current GenAI use is reported

by 52 cases (24.3%), consistent with the elapsed-time eligibility criterion in Section 5.2.

The distribution provides variation across firm categories and certification status, allowing their inclusion as covariates. These marginal counts do not establish representativeness of the Jordanian auditor population. The descriptive use patterns characterize the purposively recruited sample and do not estimate national GenAI adoption rates.

6.2 Descriptive Statistics

Table 5: Descriptive Statistics for Study Constructs (n = 214)

Construct	Mean	SD	Median	Skew	Kurtosis
TTF	3.610	0.450	3.582	0.116	-0.568
QOU	3.280	0.820	3.445	-0.113	-0.877
PJQ	3.720	0.400	3.690	0.349	-0.254
ADC	3.390	0.650	3.430	-0.121	-0.447
AQ	3.640	0.400	3.649	0.096	-0.253

Note: Statistics describe the operational construct-score scales. These scales include linear transformations and should not be interpreted using raw five-point agreement categories.

Mean construct scores are 3.610 for TTF, 3.280 for QOU, 3.720 for PJQ, 3.390 for ADC, and 3.640 for AQ. Appendix A distinguishes these operational scores from equal-weight means of the averaged indicators.

Dispersion is greatest for QOU (SD = 0.820) and smallest for PJQ and AQ (both SD = 0.400) on their operational scales. Such differences are partly determined by score transformations and should not be equated with differences in raw agreement responses. Median, skewness, and kurtosis describe score distributions rather than model residuals. No low, moderate, or high agreement labels are assigned to transformed construct scores.

6.3 Correlation Matrix

Table 6: Pearson Correlation Matrix (n = 214)

Construct	TTF	QOU	PJQ	ADC	AQ
TTF	1.000				
QOU	0.481**	1.000			
PJQ	0.264**	0.543**	1.000		
ADC	0.210**	0.372**	0.439**	1.000	
AQ	0.313**	0.570**	0.614**	0.308**	1.000

Note: ** $p < 0.01$ (two-tailed). Correlations use the construct-level scores defined by the mappings in Section 5.4 and Appendix A.

All construct-score correlations are positive and significant at $p < 0.01$. The strongest is PJQ–AQ ($r = 0.614$), followed by QOU–AQ ($r = 0.570$) and QOU–PJQ ($r = 0.543$). TTF is most closely related to QOU ($r = 0.481$), consistent with its proposed antecedent role. Correlations motivate multivariable analysis but do not identify conditional paths or demonstrate mediation. HTMT addresses whether reflective indicator blocks are distinguishable, rather than whether their construct scores are uncorrelated.

6.4 Hypotheses Testing

The structural analysis uses ordinary least squares equations corresponding to PROCESS Models 1, 4, and 7 (Hayes, 2018), implemented directly in Python/NumPy. Model 7 is the principal conditional process specification: X = QOU, M = PJQ, Y = AQ, and W = ADC. All equations include two firm-size indicators (Big 4 and mid-tier; reference = small firm) and a certification indicator (JCPA/CPA or ACCA = 1; none = 0). QOU and ADC are mean-centered before forming their interaction. Tables distinguish unstandardized B from standardized β. Ordinary OLS standard errors and two-sided t tests are reported for regression coefficients; HC3 standard errors provide a heteroskedasticity-consistent sensitivity check. Entire cases are resampled 5,000 times for indirect-association inference, preserving dependence between the mediator and outcome equations. Bootstrap intervals are percentile 95% intervals, with seed 20261110.

The covariates account for measured firm categories and certification status; they do not directly measure technology access, governance, infrastructure, or engagement complexity. No firm-level identifiers are available for clustered inference. TTF is tested in the separate antecedent equation for H1 and is not an additional serial mediator in Model 7. The framework is therefore evaluated through an antecedent regression and a conditional process model rather than a single serial pathway from TTF through both QOU and PJQ to AQ.

Let C = (Big4, MidTier, Certified), $Xc = QOU - \text{mean}(QOU)$, and $Wc = ADC - \text{mean}(ADC)$. The mediator equation is $PJQ = iM + a_1Xc + a_2Wc + a_3XcWc + \gamma'C + eM$. The outcome equation is $AQ = iY + c'Xc + bPJQ + \delta'C + eY$. The conditional indirect association is $(a_1 + a_3Wc)b$, and its moderated mediation index is a_3b. The outcome table uses uncentered QOU; this changes the intercept without changing b, c′, fitted values, or residuals. Competence probes are fixed at the sample mean and mean ± one SD: 2.740, 3.390, and 4.040. Regression coefficient intervals use the relevant t distribution, while indirect-product intervals use whole-case bootstrap resampling.

Sensitivity analyses examine two alternative specifications. First, separate JCPA/CPA and ACCA indicators replace the pooled certification indicator, retaining none as the reference category. Second, equal-weight means of the averaged indicators replace operational scores for all five constructs. Both conditional process equations use the same covariates within each specification. The first analysis examines dependence on covariate coding; the second examines

dependence on score scaling and formative weighting. An unstandardized index changes with variable scales, so numerical differences across specifications are interpreted with the corresponding transformations.

6.4.1 Task fit and utilization

Table 7: Regression Results — H1 (DV: Quality-Oriented Utilization)

Term	B	SE	β	t	p
Constant	0.198	0.418	—	0.474	0.636
TTF	0.883	0.111	0.484	7.955	< 0.001
Big 4 vs Small	-0.082	0.137	-0.048	-0.598	0.551
Mid-tier vs Small	-0.111	0.133	-0.067	-0.835	0.405
Certified vs None	-0.037	0.111	-0.020	-0.329	0.742
$R^2 = 0.234$	F(4,209)	15.994			< 0.001

After adjustment, task fit is positively associated with utilization (B = 0.883, β = 0.484, $p < 0.001$; $R^2 = 0.234$). This pattern is consistent with H1. The cross-sectional association cannot establish the direction of influence.

The H1 equation explains 23.4% of observed utilization variation. A one-unit difference in the operational TTF score is associated with a 0.883-unit difference in QOU, holding the covariates constant. The standardized β of 0.484 expresses this relationship relative to sample SDs and does not represent a 48.4% improvement. The controls are not individually significant at 5%, which does not establish equivalence between firm categories.

6.4.2 Utilization and judgment quality

Table 8: Regression Results — H2 (DV: Professional Judgment Quality)

Term	B	SE	β	t	p
Constant	2.849	0.114	—	24.999	< 0.001
QOU	0.264	0.028	0.541	9.331	< 0.001
Big 4 vs Small	-0.030	0.064	-0.036	-0.474	0.636
Mid-tier vs Small	0.018	0.062	0.022	0.286	0.776
Certified vs None	0.011	0.052	0.012	0.211	0.833
$R^2 = 0.298$	F(4,209)	22.204			< 0.001

Utilization is positively associated with judgment quality after adjustment (B = 0.264, β = 0.541, $p < 0.001$; $R^2 = 0.298$). This contemporaneous statistical relationship is consistent with H2.

The H2 equation explains 29.8% of judgment-quality variation. Its utilization coefficient is the a path in the unmoderated mediation model. The separate H3 equation describes the judgment–audit-quality relationship before conditioning on utilization. Its judgment coefficient must be distinguished from b in the full outcome equation, which includes utilization and judgment simultaneously.

6.4.3 Judgment and audit quality

Table 9: Regression Results — H3 (DV: Engagement-Process Audit Quality)

Term	B	SE	β	t	p
Constant	1.365	0.212	—	6.439	< 0.001
PJQ	0.618	0.055	0.618	11.334	< 0.001
Big 4 vs Small	0.036	0.060	0.043	0.596	0.552
Mid-tier vs Small	0.013	0.058	0.016	0.228	0.820
Certified vs None	-0.056	0.049	-0.063	-1.163	0.246
$R^2 = 0.383$	F(4,209)	32.410			< 0.001

Judgment quality is positively associated with audit quality after adjustment (B = 0.618, β = 0.618, $p < 0.001$; $R^2 = 0.383$). This is consistent with H3. Audit quality is a self-report construct rather than an independent inspection or misstatement measure.

6.4.4 Mediation through professional judgment

Table 10: Mediation estimates for PROCESS Model 4 (n = 214; 5,000 bootstrap resamples)

Path	B	SE[1]	LL 95%	UL 95%	Inference
Total QOU–AQ	0.279	0.028	0.224	0.333	< 0.001
QOU–AQ conditional	0.164	0.030	0.106	0.222	< 0.001
a QOU–PJQ	0.264	0.028	0.208	0.320	< 0.001
b PJQ–AQ conditional	0.435	0.061	0.316	0.555	< 0.001
Indirect a × b	0.115	0.021	0.078	0.159	CI excludes 0

[1] OLS SE for regression coefficients and bootstrap SE for the indirect product. Both equations use the same covariates.

In Model 4, the utilization–audit-quality coefficient conditional on judgment and controls is B = 0.164. The indirect association is 0.115, with a 95% percentile bootstrap CI of [0.078, 0.159]. The indirect-to-total ratio is 41.2%. These estimates are consistent with partial statistical mediation in H4–H5 and do not establish a causal mechanism.

The total adjusted utilization–audit-quality association is B = 0.279. After conditioning on judgment, the utilization coefficient is 0.164 and the judgment coefficient is 0.435. The product of the utilization–judgment coefficient (0.264) and the conditional judgment coefficient (0.435) is approximately 0.115. Adding this indirect association to the conditional utilization coefficient recovers the total association within rounding. The confidence interval excludes zero, supporting a

statistical indirect pathway. The indirect-to-total ratio describes this model decomposition and is not a causal percentage of audit-quality improvement.

6.4.5 Moderation by digital competence

Table 11: Moderation Analysis — H6 (DV: Professional Judgment Quality)

Term	B	SE	t	p
Constant	3.674	0.059	62.734	< 0.001
QOU centered	0.232	0.028	8.256	< 0.001
ADC centered	0.184	0.035	5.245	< 0.001
QOU × ADC	0.188	0.041	4.549	< 0.001
Big 4 vs Small	-0.029	0.059	-0.487	0.627
Mid-tier vs Small	0.032	0.056	0.571	0.569
Certified vs None	0.006	0.047	0.133	0.895
$R^2 = 0.423$	F(6,207)	25.270		< 0.001

Table 12: Simple Slopes — QOU → PJQ at Three Levels of Digital Competence

ADC level	QOU slope	SE	t	p
Low minus 1 SD	0.110	0.036	3.059	0.003
Mean	0.232	0.028	8.256	< 0.001
High plus 1 SD	0.354	0.042	8.511	< 0.001

The utilization-by-competence coefficient is positive ($B = 0.188$, $p < 0.001$). The interaction increment is $\Delta R^2 = 0.058$; the full judgment equation has $R^2 = 0.423$. Utilization slopes are 0.110, 0.232, and 0.354 at low, mean, and high competence. All three are positive; the low-competence slope has $p = 0.003$. These conditional associations are consistent with H6.

The reduced judgment equation includes centered utilization, centered competence, and the covariates. Adding their interaction raises R^2 from 0.365 to 0.423. At mean competence, a one-unit utilization difference is associated with a 0.232-unit judgment-quality difference. This slope rises to 0.354 at competence one SD above the mean and falls to 0.110 one SD below it. Both slopes remain positive: the difference concerns association strength rather than direction. Mean-centering defines the main-effect interpretation at average competence; it does not create the interaction.

6.4.6 Conditional indirect associations

Table 13: Conditional indirect estimates for QOU and AQ through PJQ (5,000 bootstrap resamples)

ADC level or contrast	Indirect B	Boot SE	LL 95%	UL 95%
Low	0.048	0.017	0.017	0.086
Mean	0.101	0.019	0.068	0.142
High	0.154	0.028	0.104	0.214
High minus low	0.106	0.028	0.056	0.164
Index of moderated mediation	0.082	0.021	0.043	0.126

Note: Percentile bootstrap intervals use 5,000 whole-case resamples. Low and high ADC are mean ± one sample SD. Both Model 7 equations include the same institutional controls.

Model 7 conditional indirect associations are 0.048 at low competence, 0.101 at mean competence, and 0.154 at high competence. The high-minus-low contrast is 0.106, 95% CI [0.056, 0.164]; the index a_3b is 0.082, 95% CI [0.043, 0.126]. The intervals exclude zero and are consistent with H7. With two certification category indicators, the index is 0.082, CI [0.043, 0.127]. With equal-weight indicator means, it is 0.370, CI [0.189, 0.575]. The direction persists across specifications. Unstandardized magnitudes differ because of score scaling and, for the formative constructs, alternative weighting.

The index combines the interaction coefficient (a_3 = 0.188) and the conditional judgment coefficient (b = 0.435). Their product is approximately 0.082, expressing the change in the indirect association for a one-unit ADC difference on its operational scale. The positive high-minus-low contrast is 0.106 (95% CI [0.056, 0.164]). The conditional outcome equation explains 46.2% of perceived audit-quality variation. These results support H7 as an associational hypothesis, without establishing a temporal mechanism.

Table 14: Model 7 audit quality equation with all covariates

Term	B	SE	t	p
Constant	1.497	0.200	7.497	< 0.001
QOU	0.164	0.030	5.543	< 0.001
PJQ	0.435	0.061	7.171	< 0.001
Big 4 vs Small	0.043	0.056	0.767	0.444
Mid-tier vs Small	0.022	0.054	0.413	0.680
Certified vs None	-0.054	0.045	-1.187	0.237
R^2 = 0.462	F(5,208)	35.761		< 0.001

HC3 sensitivity checks preserve the main coefficient-test conclusions. The interaction retains B = 0.188, HC3 SE = 0.044, and $p < 0.001$. In the outcome equation, utilization retains B = 0.164 with HC3 SE = 0.028, and judgment retains B = 0.435 with HC3 SE = 0.060; both remain significant at $p < 0.001$. These checks reduce dependence on constant residual variance but do not correct omitted-variable bias or within-firm dependence. None of the firm-size or certification coefficients is individually significant at 5% in the principal equations. This does not establish that institutional conditions are unimportant.

6.4.7 Summary of hypothesis tests

Table 15: Summary of Hypotheses Testing Results

Hyp.	Relationship	Adjusted statistic	Decision
H1	TTF–QOU	β = 0.484	Consistent[1]
H2	QOU–PJQ	β = 0.541	Consistent[1]
H3	PJQ–AQ	β = 0.618	Consistent[1]
H4	QOU–AQ conditional	B = 0.164	Consistent[1]
H5	Indirect through PJQ	0.115 [0.078, 0.159]	Consistent[1]
H6	QOU × ADC	B = 0.188	Consistent[1]
H7	Index a_3b	0.082 [0.043, 0.126]	Consistent[1]

Note: Decisions describe compatibility with the hypotheses under the stated regression specification and do not establish causality. All indirect-association intervals use 5,000 whole-case bootstrap resamples.

7. Discussion

7.1 Task Fit and Quality Oriented Utilization

Task fit is positively associated with utilization after adjustment, consistent with compatibility between technology capabilities and task requirements (Goodhue and Thompson, 1995). Linguistic tools may support narrative organization and alternative explanations, whereas physical inspection and deterministic calculation impose different evidentiary requirements (Gu et al., 2024). The result supports examining task-specific use rather than treating GenAI adoption as uniform. It does not demonstrate that changing fit perceptions increases utilization: successful use may also change perceived fit. The contribution is an auditor-level association in the Jordanian industrial-audit setting.

7.2 The Professional Judgment Pathway (H2, H3, H5)

The positive indirect association is consistent with professional judgment as a statistical pathway between quality-oriented utilization and perceived audit quality. Generated output becomes relevant to a conclusion only after an auditor evaluates cues, alternatives, and corroborating evidence (Commerford et al., 2022; Nolder and Kadous, 2018). The remaining positive utilization coefficient suggests that the measured judgment construct does not exhaust the relationship. Unmeasured

processes, overlapping questionnaire content, and shared response tendencies are also possible explanations. The analysis does not separately identify time savings, improved evidence coverage, or stronger documentation as causal mechanisms.

7.3 Digital Competence and Conditional Associations

Digital competence is associated with a stronger utilization–judgment relationship and a larger indirect association with audit quality. This is consistent with calibrated reliance: use may support judgment when auditors recognize task suitability, verify sources, identify unsupported output, and manage confidentiality (Lee and See, 2004; Vuorikari et al., 2022). Cognitive forcing research motivates deliberate evaluation of AI advice (Buçinca et al., 2021), while archival evidence on broader auditor digital expertise provides a related comparison (Zhao et al., 2026). These studies address different settings and measures. The present results extend that discussion to a GenAI-specific conditional process model without showing that competence training causes the observed differences. Persistence under alternative coding and indicator-mean scoring supports the direction within these data; unstandardized magnitudes remain scale-dependent.

7.4 Contextual Boundary Conditions in the Jordanian Industrial Setting

The Jordanian industrial-audit setting defines the evidence's scope. Respondents serve listed industrial companies but report individual perceptions, so the results cannot be interpreted as observed changes in those companies' financial-reporting quality. Firm resources, engagement complexity, internal guidance, and regulatory exposure may shape both GenAI use and its evaluation; size categories capture only part of this variation. The 24.3% rare or experimental current users indicate heterogeneous experience without identifying infrastructure limitations as its cause. Replication across subsectors, offices, and jurisdictions would help distinguish transferable relationships from setting-specific conditions.

8. Conclusions and Recommendations

8.1 Conclusions

The study identifies a coherent associational pattern: task fit relates to quality-oriented utilization; utilization relates to professional judgment and perceived audit quality; and the utilization–judgment relationship is stronger at higher digital competence. The positive moderated mediation index persists after covariate adjustment and in the sensitivity specifications. The study contributes an integrated account of task fit, judgment, and competence among auditors serving Jordanian listed industrial companies. Its practical value lies in identifying relationships for further evaluation, rather than demonstrating that adoption or training improves independently observed audit outcomes.

Task-technology fit is positively associated with quality-oriented utilization. Audit firms can use this relationship as a reason to evaluate task suitability explicitly, although the survey does not establish the direction of influence.

Professional judgment quality is positively associated with utilization and perceived audit quality. The reported mediation estimates are consistent with a judgment pathway, but they do not establish a causal mechanism or guarantee improvements from adopting GenAI.

Digital competence is associated with a stronger reported utilization-to-judgment relationship. This supports further evaluation of training in source verification, task selection, confidentiality, and the recognition of unsupported output.

Firm-size categories and certification status have been included in the final equations. Infrastructure, proprietary AI access, governance, engagement characteristics, and firm clustering remain unmeasured. Covariate adjustment therefore does not establish that all institutional explanations have been excluded.

The results support examining controlled and documented GenAI use. They do not compare a policy of prohibition with one of adoption and therefore cannot establish which policy produces better audit outcomes.

8.2 Recommendations for Audit Firms

Assess the suitability of GenAI for each proposed audit procedure before authorizing its use. Distinguish assistance with organizing information or developing alternatives from tasks that require direct observation, deterministic calculation, or specialist judgment.

Evaluate adversarial-prompting protocols as potential workflow controls. Such protocols could document the alternative hypotheses generated, source-verification steps performed, and reasons for rejecting competing explanations when GenAI supports substantive procedures.

Develop training and assessment in task classification, source grounding, hallucination awareness, and confidentiality management. Evaluate whether these activities improve observable audit work rather than relying on training completion alone.

Establish firm-level GenAI audit logs capturing use-case type, frequency, and documentation quality — enabling quality-control reviews to distinguish quality-oriented from efficiency-oriented use.

Assess competence and support needs across firm sizes. Target assistance using demonstrated gaps rather than assuming that firm size reliably identifies competence.

8.3 Recommendations for JACPA and the Jordan Securities Commission

Consider technology-neutral guidance on task suitability, source verification, documentation, and confidentiality. International guidance offers useful examples that would require evaluation in Jordanian audit practice (Financial Reporting Council, 2025; Financial Reporting Council, 2026).

Consider evaluating GenAI-specific digital competence within JACPA continuing professional development activities, using learning outcomes informed by the four theoretically specified competence dimensions examined in this study.

Examine whether access to training, technology, and quality-control support differs across audit-firm categories. Any monitoring program should use evidence of actual process or outcome differences rather than the untested assumption of lower quality in small firms.

Engage with IAASB technology governance workstreams to ensure developing Jordanian standards align with evolving international professional guidance (IAASB, 2025).

8.4 Recommendations for Researchers

Conduct longitudinal panel studies tracking individual auditors before and after GenAI deployment, measuring task-fit perception, quality-oriented utilization, digital competence, and archival audit-quality proxies such as restatements and going-concern accuracy. Such designs can clarify temporal sequencing; causal interpretation also depends on addressing confounding.

Use within-auditor experiments to vary task suitability and output-verification requirements, separating these mechanisms from stable individual differences. Experiments on reliance on AI advice provide a methodological reference point (Commerford et al., 2022).

Extend the framework to other MENA contexts (Saudi Arabia, Egypt, UAE) for cross-jurisdictional comparison of infrastructure, regulatory, and competence moderators.

Develop and validate a standalone Auditor GenAI Digital Competence Scale (AGDCS) directly measuring source-grounding, hallucination-awareness, and task-classification behaviors for cross-study comparability.

8.5 Study Limitations

The cross-sectional design cannot establish temporal sequence, exclude reverse relationships, or identify causal mediation. Self-reported audit quality describes perceived processes rather than inspected files, misstatement outcomes, or regulatory findings. Common measurement sources and overlap between judgment and quality content may inflate relationships. Averaged indicators can conceal heterogeneity among their underlying subitems, whose individual measurement properties were not evaluated. Formative assessment addresses content specification and collinearity, but independent global criterion measures for convergent validity are absent. The mappings in Appendix A define operational scores; they do not constitute prospective validation of a scale. Purposive recruitment limits generalization, while unmeasured engagement characteristics, governance, infrastructure, and within-firm dependence remain possible explanations.

References

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

[2] Bommasani, R., Hudson, D.A., Adeli, E., et al. (2021). On the opportunities and risks of foundation models. arXiv:2108.07258. https://arxiv.org/abs/2108.07258.

[3] Brown-Liburd, H., Issa, H. and Lombardi, D. (2015). Behavioral implications of Big Data's impact on audit judgment and decision making and future research directions. Accounting Horizons 29(2), pp. 451-468. https://doi.org/10.2308/acch-51023.

[4] Buçinca, Z., Malaya, M.B. and Gajos, K.Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proceedings of the ACM on Human-Computer Interaction 5(CSCW1), Article 188, 1–21. https://doi.org/10.1145/3449287.

[5] Coltman, T., Devinney, T.M., Midgley, D.F. and Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. Journal of Business Research 61(12), pp. 1250–1262. https://doi.org/10.1016/j.jbusres.2008.01.013.

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

[7] Dong, M.M., Stratopoulos, T.C. and Wang, V.X. (2024). A scoping review of ChatGPT research in accounting and finance. International Journal of Accounting Information Systems 55, 100715. https://doi.org/10.1016/j.accinf.2024.100715.

[8] Feuerriegel, S., Hartmann, J., Janiesch, C. and Zschech, P. (2024). Generative AI. Business & Information Systems Engineering 66, 111–126. https://doi.org/10.1007/s12599-023-00834-7.

[9] Financial Reporting Council (2025). AI in audit: Illustrative example and documentation guidance. FRC, London. https://www.frc.org.uk/library/standards-codes-policy/audit-assurance-and-ethics/guidance/ai-in-audit/

[10] Financial Reporting Council (2026). Generative and Agentic AI Guidance. FRC, London. https://www.frc.org.uk/library/standards-codes-policy/audit-assurance-and-ethics/guidance/ai-in-audit/

[11] Fotoh, L.E. and Mugwira, T. (2025). Exploring Large Language Models in external audits: Implications and ethical considerations. International Journal of Accounting Information Systems 56, 100748. https://doi.org/10.1016/j.accinf.2025.100748.

[12] Goodhue, D.L. and Thompson, R.L. (1995). Task-technology fit and individual performance.
MIS Quarterly 19(2), pp. 213–236. https://doi.org/10.2307/249689.

[13] Gu, H., Schreyer, M., Moffitt, K. and Vasarhelyi, M.A. (2024). Artificial intelligence co-piloted auditing.
International Journal of Accounting Information Systems 54, 100698. https://doi.org/10.1016/j.accinf.2024.100698.

[14] Hayes, A.F. (2018). Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach, 2nd ed. Guilford Press, New York.

[15] Henseler, J., Ringle, C.M. and Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science 43, pp. 115–135. https://doi.org/10.1007/s11747-014-0403-8.

[16] IAASB (2025). 2025 Handbook of International Quality Management, Auditing, Review, Other Assurance, and Related Services Pronouncements. IAASB, New York. https://www.iaasb.org/publications/2025-handbook-international-quality-management-auditing-review-other-assurance-and-related-services

[17] Issa, H., Sun, T. and 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), pp. 1–20. https://doi.org/10.2308/jeta-10511.

[18] Jarvis, C.B., MacKenzie, S.B. and Podsakoff, P.M. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. Journal of Consumer Research 30(2), pp. 199–218. https://doi.org/10.1086/376806.

[19] Ji, Z., Lee, N., Frieske, R., et al. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys 55(12), Article 248, pp. 1–38. https://doi.org/10.1145/3571730.

[20] Kokina, J., Blanchette, S., Davenport, T.H. and Pachamanova, D. (2025). Challenges and opportunities for artificial intelligence in auditing: Evidence from the field. International Journal of Accounting Information Systems 56, 100734. https://doi.org/10.1016/j.accinf.2025.100734.

[21] Krejcie, R.V. and Morgan, D.W. (1970). Determining sample size for research activities. Educational and Psychological Measurement 30(3), pp. 607–610. https://doi.org/10.1177/001316447003000308.

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

[23] Li, Y. and Goel, S. (2025). Artificial intelligence auditability and auditor readiness for auditing artificial intelligence systems.
International Journal of Accounting Information Systems 56, 100739. https://doi.org/10.1016/j.accinf.2025.100739.

[24] Lombardi, D.R., Kim, M., Sipior, J.C. and Vasarhelyi, M.A. (2025). The increased role of advanced technology and automation in audit: A Delphi study. International Journal of Accounting Information Systems 56, 100733. https://doi.org/10.1016/j.accinf.2025.100733.

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

[26] Nolder, C.J. and Kadous, K. (2018). Grounding the professional skepticism construct in mindset and attitude theory: A way forward. Accounting, Organizations and Society 67, 1–14. https://doi.org/10.1016/j.aos.2018.03.010.

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

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

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

[30] Stratopoulos, T.C. and Wang, V.X. (2025). Artificial intelligence and accounting research: A framework and agenda. International Journal of Accounting Information Systems 56, 100760. https://doi.org/10.1016/j.accinf.2025.100760.

[31] Sutton, S.G., Holt, M. and 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.

[32] Torroba, M., Sánchez, J.R., López, L. and Callejón, A. (2025). Investigating the impacting factors for the audit professionals to adopt data analysis and artificial intelligence: Empirical evidence for Spain. International Journal of Accounting Information Systems 56, 100738. https://doi.org/10.1016/j.accinf.2025.100738.

[33] Vuorikari, R., Kluzer, S. and Punie, Y. (2022). DigComp 2.2: The Digital Competence Framework for Citizens – With New Examples of Knowledge, Skills and Attitudes. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/115376.

[34] Yoon, K., Hoogduin, L. and Zhang, L. (2015). Big Data as complementary audit evidence. Accounting Horizons 29(2), pp. 431–438. https://doi.org/10.2308/acch-51076.

[35] Zhang, C.A., Cho, S. and Vasarhelyi, M.A. (2022). Explainable Artificial Intelligence (XAI) in auditing. International Journal of Accounting Information Systems 46, 100572. https://doi.org/10.1016/j.accinf.2022.100572.

[36] Zhao, T., Duan, W. and Mao, Y. (2026). Auditors' digital expertise and audit quality: Empirical evidence based on China's A-share listed companies. Emerging Markets Finance and Trade 62(2), pp. 420–438. https://doi.org/10.1080/1540496X.2025.2535711.

Appendix A Operational score definitions

Operational scores and equal-weight indicator means are reported separately in Table A1. For reflective constructs, let m denote a participant's equal-weight mean of the relevant averaged indicators. The scale mappings are QOU = −3.834943 + 2.375605m, PJQ = 0.238829 + 1.159612m, and AQ = 0.272770 + 1.122477m. For TTF and ADC, the coefficients in Table 2 and intercepts of −0.905444 and −2.969031, respectively, specify the weighted mappings. Full-precision coefficients reproduce the operational score columns within numerical precision; displayed coefficients are rounded. Primary regressions use operational scales, and alternative scoring uses equal-weight indicator means. Numerical equivalence within the dataset is distinct from independent validation of the weighting scheme.

Table A1: Comparison of operational scores and equal-weight indicator means

Construct	Score mean	Indicator mean	Max absolute difference	Correlation
TTF	3.610	3.003	1.081	0.983
QOU	3.280	2.995	1.241	1.000
PJQ	3.720	3.002	0.862	1.000
ADC	3.390	3.012	1.173	0.981
AQ	3.640	3.000	0.765	1.000