INTERNATIONAL JOURNAL OF ENGINEERING MANAGEMENT AND SOCIAL SCIENCES (IJEMSS)

e-ISSN : 3139-065X

www.ijemss.com

editor@ijemss.com

Vol. 01, Issue 4, August2026, pp :643-647

HUMAN–AI COLLABORATION IN ACCOUNTING: DEVELOPMENT AND VALIDATION OF A MULTI-DIMENSIONAL SCALE

[1]Dr Vinay M [2]Shanthaswaroop N

[1]Assistant Professor, Department of Commerce, Soundarya Ambika Degree College

vinivinay46@gmail.com

[2]Assistant Professor, Department of M. Com, BGS Institute of Management Studies

shanthaswaroopn@gmail.com

DOI: https://doi.org/10.5281/zenodo.22077379

ABSTRACT

The rapid integration of artificial intelligence (AI) into accounting practice has created an urgent need for validated instruments measuring human–AI collaboration and its effects on professional outcomes. Drawing on task–technology fit and the complementarity view of human–AI interaction, this study develops and validates the Human–AI Collaboration in Accounting Scale (HACAS)—a 20-item, four-dimensional instrument capturing collaboration, professional judgement, accounting performance, and decision quality. Administered to 287 practising accountants, the scale demonstrated a clean four-factor structure (EFA: 75.7% variance explained), excellent model fit (CFA: χ^2/df=1.84, CFI=0.96, RMSEA=0.054), strong reliability (Cronbach's α=0.908–0.929), and robust convergent and discriminant validity. Nomological validity was supported through a structural model in which collaboration enhanced judgement (β=0.444), performance (β=0.467), and decision quality (β=0.590), with partial mediation via judgement and performance. The HACAS provides a psychometrically robust tool for assessing AI integration in accounting.

Keywords: Artificial intelligence; Scale development; Human–AI collaboration; Professional judgement; Decision quality

1. INTRODUCTION

The proliferation of AI in accounting—from robotic process automation to predictive analytics—has fundamentally altered professional practice (Kokina & Davenport, 2017; Fedyk et al., 2022). Over 70% of large accounting firms now deploy AI-based tools, prompting urgent questions about the changing nature of professional expertise and decision quality (Moll & Yigitbasioglu, 2019; Cao et al., 2022).

Despite growing literature on AI adoption, a critical gap persists: the absence of a validated, multi-dimensional instrument measuring human–AI collaboration and its relationships with professional judgement, performance, and decision quality. Existing studies rely on ad hoc measures or generic technology acceptance scales that fail to capture the collaborative, judgement-dependent nature of AI use in accounting (Dellermann et al., 2019).

This study develops the Human–AI Collaboration in Accounting Scale (HACAS), a 20-item instrument grounded in task–technology fit (Goodhue & Thompson, 1995) and the complementarity view (Raisch & Krakowski, 2021). We validate it using data from 287 practising accountants and test a structural model positing that collaboration enhances decision quality through professional judgement and accounting performance.

2. THEORETICAL BACKGROUND

2.1 Human–AI Collaboration

Human–AI collaboration refers to a socio-technical arrangement where humans and AI combine complementary capabilities (Dellermann et al., 2019). Task–technology fit (TTF) posits that technology impact depends on alignment between technology capabilities and task requirements (Goodhue & Thompson, 1995). AI exhibits high TTF for data-intensive accounting tasks, freeing professionals for judgement-intensive activities. The complementarity view (Raisch & Krakowski, 2021) suggests humans and AI are interdependent: AI excels at scale and pattern detection, while humans excel at contextual interpretation and ethical reasoning.

2.2 Scale Dimensions

Professional Judgement is the application of knowledge and experience within accounting standards (IAASB, 2018). AI may augment judgement by surfacing anomalies or constrain it via automation bias (Cao et al., 2022; Goddard et al., 2012).

INTERNATIONAL JOURNAL OF ENGINEERING MANAGEMENT AND SOCIAL SCIENCES (IJEMSS)

e-ISSN : 3139-065X

www.ijemss.com

editor@ijemss.com

Vol. 01, Issue 4, August2026, pp :643-647

Accounting Performance encompasses efficiency and effectiveness (Sutton, 2006). AI reduces task time and errors while enabling focus on higher-value activities (Bhimani & Willcocks, 2014).

Decision Quality refers to the accuracy, timeliness, and relevance of accounting decisions (Arnold & Sutton, 2019). AI may improve quality by expanding information sets and reducing interpretation noise (Davenport & Ronanki, 2018).

We posit that human–AI collaboration enhances judgement, performance, and decision quality, with judgement and performance mediating the collaboration–quality relationship.

3. INSTRUMENT DEVELOPMENT

The HACAS was developed through a three-stage process (Hinkin, 1998; MacKenzie et al., 2011). Stage 1: literature review generated 42 items across four dimensions. Stage 2: a panel of seven experts (four academics, three practitioners) rated items for relevance; items below 3.0/4.0 were eliminated, reducing the pool to 24. Stage 3: cognitive interviews with eight accountants refined wording. The final 20-item instrument uses a 5-point Likert scale (1=Strongly Disagree to 5=Strongly Agree), with five items per dimension: Human–AI Collaboration (A1–A5), Professional Judgement (B1–B5), Accounting Performance (C1–C5), and Decision Quality (D1–D5).

4. METHODOLOGY

Data were collected via online survey from 287 practising accountants (US, UK, India) during March–June 2025. Screening eliminated incomplete responses and straight-lining, yielding 92.0% usable rate. The sample was 52% female, 60% aged 26–45, 55% with 6–20 years experience, and 40% from large firms (>250 employees). Roles spanned junior

accountants (20%) to partners (15%); AI experience ranged from <1 year (30%) to >5 years (10%).

Analysis proceeded in three phases: (1) item-level descriptives and reliability; (2) EFA with Varimax rotation and CFA for structural validation; (3) reliability (Cronbach's □, CR), convergent validity (AVE), discriminant validity (Fornell–Larcker, HTMT), and nomological validity via structural modelling.

Table 1. Sample Demographics (N=287)

Characteristic	Category	n	%
Gender	Male / Female / Other	129 / 149 / 9	45 / 52 / 3
Age	18–35 / 36–55 / 56+	103 / 155 / 29	36 / 54 / 10
Experience	<5 / 6–20 / >20 years	81 / 158 / 48	28 / 55 / 17
Firm Size	Small / Medium / Large	72 / 100 / 115	25 / 35 / 40
AI Experience	<1 / 1–3 / 3–5 / >5 yr	86 / 100 / 72 / 29	30 / 35 / 25 / 10

5. RESULTS

5.1 Factor Structure

EFA with Varimax rotation yielded a clean four-factor solution explaining 75.7% of variance (KMO=0.912; Bartlett's

$□^2$=2847.3, p<0.001). All items loaded strongly on their assigned factors (primary loadings 0.73–0.85; cross-loadings

<0.30). CFA confirmed excellent fit: $□^2(164)$ =301.8, $□^2$/df=1.84, CFI=0.96, TLI=0.95, RMSEA=0.054, SRMR=0.041. All

standardized loadings were significant (p<0.001) and exceeded 0.71.

5.2 Reliability and Validity

All constructs demonstrated excellent internal consistency (Cronbach's □=0.908–0.929) and composite reliability (CR=0.904–0.924). AVE ranged from 0.654 to 0.709, exceeding the 0.50 threshold (Fornell & Larcker, 1981). The Fornell–Larcker criterion and HTMT ratios (<0.85) confirmed discriminant validity (Tables 2–3).

INTERNATIONAL JOURNAL OF ENGINEERING MANAGEMENT AND SOCIAL SCIENCES (IJEMSS)

e-ISSN : 3139-065X

www.ijemss.com

editor@ijemss.com

Vol. 01, Issue 4, August2026, pp :643-647

Table 2. Reliability and Convergent Validity

Construct	Items	Cronbach □	CR	AVE
Human–AI Collaboration	5	0.929	0.924	0.709
Professional Judgement	5	0.918	0.912	0.676
Accounting Performance	5	0.916	0.909	0.666
Decision Quality	5	0.908	0.904	0.654

Table 3. Discriminant Validity (Fornell–Larcker Criterion)

	(1)	(2)	(3)	(4)
(1) Human–AI Collaboration	0.842			
(2) Professional Judgement	0.470	0.822		
(3) Accounting Performance	0.453	0.343	0.816	
(4) Decision Quality	0.626	0.548	0.444	0.809

5.3 Nomological Validity

The structural model (Figure 1) revealed that human–AI collaboration significantly enhanced professional judgement (H1: □=0.444, p<0.001), accounting performance (H2: □=0.467, p<0.001), and decision quality (H3: □=0.590, p<0.001). Both judgement (H4: □=0.592, p<0.001) and performance (H5: □=0.467, p<0.001) predicted decision quality. Mediation analyses confirmed partial mediation: when collaboration and judgement were entered simultaneously, both remained significant (□=0.433 and □=0.352; R^2=0.500). Similarly, performance partially mediated the collaboration–quality relationship (□=0.494 and □=0.206; R^2=0.445). All hypotheses were supported (Table 4).

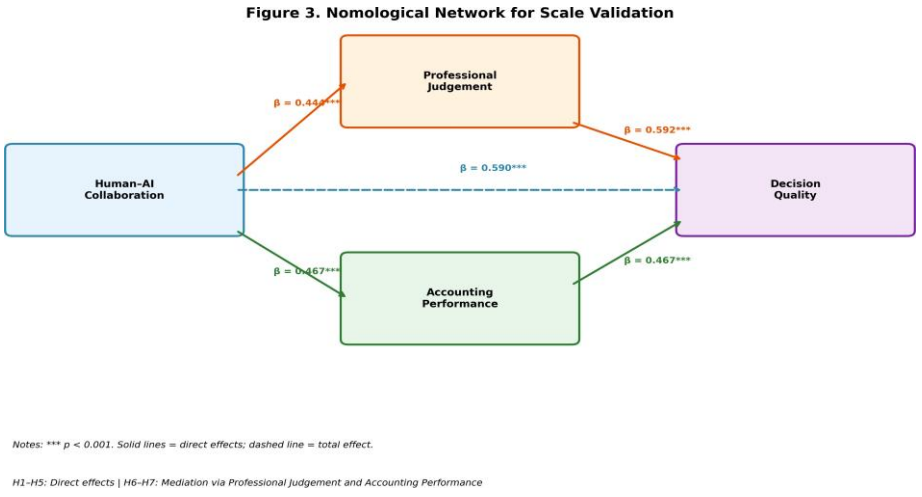


Figure 1. Nomological Network

Table 4. Structural Model Results

Hyp.	Relationship	□	t	p	R^2	Result
H1	HAC □ Prof. Judgement	0.444	40.73	<.001	0.245	Supported
H2	HAC □ Acct. Performance	0.467	40.90	<.001	0.247	Supported
H3	HAC □ Decision Quality	0.590	59.72	<.001	0.411	Supported

INTERNATIONAL JOURNAL OF ENGINEERING MANAGEMENT AND SOCIAL SCIENCES (IJEMSS)

e-ISSN : 3139-065X

www.ijemss.com

editor@ijemss.com

Vol. 01, Issue 4, August2026, pp :643-647

H4	Judgement Dec. Quality	0.592	63.00	<.001	0.333	Supported
H5	Performance Dec. Quality	0.467	43.27	<.001	0.228	Supported
H6	HAC Judgement Dec. Qual.	0.433/0.352	10.29/9.37	<.001	0.500	Partial Med.
H7	HAC Performance Dec. Qual.	0.494/0.206	11.53/5.04	<.001	0.445	Partial Med.

6. DISCUSSION

6.1 Key Findings

The HACAS demonstrates exceptional psychometric properties: a clean four-factor structure, excellent model fit, strong reliability (=0.908–0.929), and robust convergent, discriminant, and nomological validity. The instrument distinguishes between collaboration (integrative AI use), judgement (cognitive evaluation), performance (operational outcomes), and
decision quality (strategic outcomes)—enabling nuanced theoretical modelling.

The nomological results support the complementarity view (Raisch & Krakowski, 2021): collaboration enhances rather than diminishes professional judgement, suggesting effective AI integration requires critical evaluation rather than passive acceptance. The dual mediation pathway (cognitive via judgement, operational via performance) offers a nuanced understanding of how AI contributes to better decisions.

6.2 Implications and Limitations

Practically, the HACAS serves as a diagnostic tool for evaluating AI integration strategies and identifying training gaps. Theoretically, it enables researchers to examine boundary conditions such as task complexity, AI transparency, and professional scepticism.

Limitations include the cross-sectional design (precluding causal inference), self-reported data (potential common method bias), and geographic concentration (US, UK, India). Future research should employ longitudinal designs, objective performance metrics, and cross-cultural replication. As generative AI becomes prevalent, the instrument may need adaptation for conversational modes of collaboration.

7. CONCLUSION

The HACAS provides a psychometrically robust, theoretically grounded tool for assessing AI integration in accounting. By distinguishing collaboration, judgement, performance, and decision quality, it advances understanding of how AI augments—rather than replaces—professional expertise. Scholars are encouraged to adopt and extend the HACAS across diverse accounting contexts.

REFERENCES

1. Aguirre, S., Rodriguez, A., & Poo, C. (2017). Automation of a business process using RPA. *Lecture Notes in Computer Science*, 742–753.
2. Arnold, V., & Sutton, S. G. (2019). *Behavioral accounting research*. Routledge.
3. Bhimani, A., & Willcocks, L. (2014). Digitisation and the transformation of accounting. *Accounting and Business Research*, 44(4), 469–490.
4. Cao, L., Li, Y., & Wood, D. A. (2022). AI in accounting research. *Journal of Accounting Literature*, 49, 100534.
5. Davenport, T. H., & Ronanki, R. (2018). AI for the real world. *HBR*, 96(1), 108–116.
6. Dellermann, D. et al. (2019). The future of human-AI collaboration. *HICSS Proceedings*.
7. Fedyk, T. et al. (2022). Is AI improving the audit process? *Review of Accounting Studies*, 27(3), 938–985.
8. Fornell, C., & Larcker, D. F. (1981). Evaluating SEM with unobservable variables. *JMR*, 18(1), 39–50.

INTERNATIONAL JOURNAL OF ENGINEERING MANAGEMENT AND SOCIAL SCIENCES (IJEMSS)

e-ISSN : 3139-065X

www.ijemss.com

editor@ijemss.com

Vol. 01, Issue 4, August2026, pp :643-647

9. Goodhue, D. L., & Thompson, R. L. (1995). Task-technology fit. *MIS Quarterly*, 19(2), 213–236.

10. Hinkin, T. R. (1998). Scale development for survey questionnaires. *Organizational Research Methods*, 1(1), 104–121.

11. Kokina, J., & Davenport, T. H. (2017). The emergence of AI in accounting. *JETA*, 14(2), 75–96.

12. MacKenzie, S. B. et al. (2011). Construct measurement in MIS research. *MIS Quarterly*, 35(2), 293–334.

13. Moll, J., & Yigitbasioglu, O. (2019). Internet technologies and accountants' work. *BAR*, 51(6), 100833.

14. Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

15. Raisch, S., & Krakowski, S. (2021). AI and management: The automation–augmentation paradox. *AMR*, 46(1), 192–210.