Asian Journal of Accounting and Governance 26(2026) ISSN2180-3838
e-ISSN2716-6060
(http://dx.doi.org/10.17576/AJAG-2026-26-5)

Mapping the Structure of Artificial Intelligence in Accounting

YUXIAO LOU, RUHANITA MAELAH & KAMARUL BARAINI KELIWON

ABSTRACT

Research on artificial intelligence (AI) in accounting has expanded rapidly, yet bibliometric reviews report different corpora and structures. This study examines the field's development, conceptual and intellectual structures, and underrepresented research concerns. A rerunnable Web of Science Core Collection search and screening yielded 610 documents published from 1982 to 2025. Methods included performance, co-word and co-citation analyses, pairwise removal of four frequent terms—artificial intelligence, accounting, auditing, and machine learning—and reference publication year spectroscopy. Two applications of the screening protocol agreed on 93.5% of 3,220 records; cases in which both departed from the recorded decision were adjudicated by the lead author. This study finds that five keyword clusters were organised mainly by AI technique rather than accounting task, while pairwise removals involving artificial intelligence, accounting, and auditing produced the largest structural changes. The cited base combined accounting research, analytical methods, and work on automation and judgment. Generative-AI terms appeared in established accounting applications but had few direct links to earlier technology terms, while keywords for external validation, robustness, and governance remained uncommon. The study contributes a rerunnable search, record-level screening decisions, and a sensitivity-tested interpretation of the field's bibliometric structure.

Keywords: Artificial intelligence; accounting; bibliometric analysis; co-word analysis; science mapping

INTRODUCTION

Expert systems were applied to audit judgment in the 1980s, and Sutton et al. (2016: 60–64) argued that research on artificial intelligence (AI) in accounting had continued under changing techniques and labels. Later work extended from robotic process automation in the Big Four (Cooper et al. 2019: 15) to machine-learning fraud detection (Bao et al. 2020: 199) and the association between audit firms' own AI workforce and audit quality (Fedyk et al. 2022: 938). After 2018, publication volume rose sharply.

Growth of this kind invites mapping, and mapping has happened. The maps differ. Coita et al. (2026: 37) analyse 729 Web of Science articles, Liaras et al. (2024: 1431, 1435) 575, Tocev and Atanasovski (2026: 163, 167) 451, Ben Said et al. (2026: 155, 159) 447, and Epelbaum and Farrell (2026: 292–293) 219 from Scopus. These studies cover substantially overlapping technologies and accounting domains, yet their corpora differ by more than threefold.

The differences are only partly traceable. Epelbaum and Farrell (2026: 292–293) report a field-tagged search string, an execution date, and filters, and classify retrieved records into topic categories; Liaras et al. (2024: 1442) publish their final corpus; Tocev and Atanasovski (2026: 167) report a four-phase screening chain but not complete Boolean syntax. Reporting transparency differs between search design and screening procedures. At the record level, however, disclosure is consistently limited: none of the reviewed studies reports the disposition of each retrieved record. Reported structures also range from three to nine clusters, although cluster counts depend on corpus construction and analytical settings. Readers therefore cannot determine how much of the variation reflects scope, query design, screening judgment, or the dimensions represented by the clusters.

This study maps 610 documents on AI in accounting published between 1982 and 2025. It uses a fully specified Web of Science Core Collection search that reproduced the same record count on a verification rerun, an eligibility boundary requiring substantive AI and accounting components, and a deposited file recording the disposition and decision stage of every exported record. The same written protocol was applied twice using ChatGPT 5.6 Sol in separate sessions. Records for which both applications departed from the recorded decision were adjudicated by the lead author. This makes the resulting boundary inspectable without claiming that it is the uniquely correct corpus. Four questions follow:

- RQ_1 How has research on AI in accounting developed in terms of publication activity, sources, authorship, institutions, countries, and international collaboration?
- RQ_2 What conceptual structure does this literature display, and how sensitive is that structure to the removal of high frequency general terms?
- RQ_3 Which works form the intellectual base of the field, and what disciplinary streams do they represent?
- RQ_4 Which concerns are absent from or weakly represented in the vocabulary this literature uses to describe itself?

The study makes three contributions. First, this paper diagnoses differences in the corpus boundaries, reporting practices, and reported structures of existing maps, while recognising that information omitted by earlier reviews cannot now be recovered.

Second, the complete search expression, filters, and search-development tests are reported, while the disposition and decision stage of every exported record are deposited. Repeated classification quantifies variation in relevance decisions.

Third, the study tests rather than merely reports the mapped structure. It identifies the part of the Author Keyword solution most sensitive to targeted term removal, distinguishes the disciplinary functions of the works behind the cited-reference peaks, examines generative-AI continuity at the link level, and identifies validation and normative concerns that remain weakly represented in the field's self-description.

LITERATURE REVIEW

Twenty-two reviews inform this section: sixteen corpus-level maps compared in Table 1 and six focused reviews of application-specific concerns (Abbas 2026; Belahouaoui & Attak 2024; Han et al. 2023; Kureljusic & Karger 2024; Li et al. 2025; Ranta et al. 2023). The set was assembled during search-term development and through a later review search with no end-date restriction; it is not a census.

Despite differences in scope and method, the reviews broadly agree on three points: publication activity grew quickly from the late 2010s; applications concentrate particularly in auditing; and recurring themes include routine-work automation, prediction and classification, and professional adoption. These points provide the baseline for examining how corpus boundaries and reported structures were constructed.

TABLE 1. Reviews of AI in accounting and related domains

Study	Scope	Database and period	Records analysed	Structure reported
Amani & Fadlalla (2017: 32, 35)	Data mining in accounting	Five source types, no date restriction (records 1989–2014)	209	Framework, 4 categories
Lamboglia et al. (2021: 1233, 1236–1237)	Audit and digital technology	Web of Science and Scopus, to 2019	256	3 topics
Mugwira (2022: 201, 203)	Internet-related technology in auditing	Web of Science, 1990–2019	236	6 research streams
Agustí & Orta-Pérez (2023: 417, 421, 424)	Big data and AI in accounting and auditing	Web of Science, period not reported	247	5 co-word clusters; 7 coupling clusters
Abu Huson et al. (2024: 91)	IT, AI, and blockchain in auditing	Web of Science, 2017–2022	328	4 keyword clusters
Liaras et al. (2024: 1431, 1434–1435)	Machine learning in accounting and finance	Web of Science, 1996–2022	575	6 coupling clusters
Dong et al. (2024: 1, 8)	ChatGPT and LLMs in accounting and finance	SSRN 2022–2024 and Web of Science, to March 2024	116	3 themes
Indrayani et al. (2024: 1, 3–5)	Emerging technology in accounting	Scopus, 1982–June 2024	324	4 clusters
Al-Khoury et al. (2025: 1, 3, 8–9)	Emerging technology in AIS	Scopus, 2011–November 2024	195	4 clusters
Elnakeeb & Elawadly (2025: 1)	Automation and AI in accounting	Scopus, 2001–2024	343	4 conceptual clusters
Tlili (2025: 149, 153)	AI in accounting	Dimensions, May 2015–May 2024	475	3 themes
Ben Said et al. (2026: 155, 159, 165–166)	AI and the accounting profession	Web of Science, 1989–2023	447	Keyword co-occurrence and thematic map
Tocev & Atanasovski (2026: 163, 167)	AI in accounting	Scopus, 1984–2024	451	4 thematic clusters
Epelbaum & Farrell (2026: 292–293, 301–303)	AI and auditing	Scopus, no date restriction (records 1986–2024)	219	4 clusters and themes
Alkhatib et al. (2026: 4–6, 10)	AI in auditing	Scopus, January 2015–September 2025	236	9 topics
Coita et al. (2026: 37, 61)	AI in accounting and auditing	Web of Science, post-2015, end year not reported	729	Keyword co-occurrence, 6 dominant terms

Note: Periods, sample sizes, and structures are as stated by each study. Where sections differed, the methods section was followed.

SCOPE CONSTRUCTION AND RESEARCH TRANSPARENCY

The 16 corpora are not directly comparable because periods, databases, search vocabularies, accounting boundaries, and screening procedures differ. Coita et al. (2026: 37) construct their 729-article sample through sequential database filters; Epelbaum and Farrell (2026: 292–293) use ASReview-assisted title/abstract screening followed by review by the second author; Tocev and Atanasovski (2026: 167) apply predefined criteria and resolve borderline cases by consensus.

Scope is also set at different levels. Some reviews isolate one technology, whereas others group AI with broader emerging technologies such as blockchain, big data, and cloud computing (Amani & Fadlalla 2017: 32, 35; Indrayani et al. 2024: 1, 3–5). Search reporting also varies: Epelbaum and Farrell (2026: 292) provide the full field-tagged Boolean string, search date, and filters, whereas Tlili (2025: 153) reports “accounting AND AI OR artificial intelligence” without parentheses. No common reporting standard is evident.

Screening disclosure is also uneven. Epelbaum and Farrell (2026: 292–293) report ASReview-assisted title/abstract screening validated by the second author and category-level inclusion decisions; Tocev and Atanasovski (2026: 167) report predefined criteria and consensus. Such disclosures explain procedures or retained groups but not the disposition of each retrieved record. The present study therefore deposits the final disposition and decision stage of all 3,336 exported records.

LIMITS OF EXISTING BIBLIOMETRIC MAPS

Cluster counts range from three to nine, but these differences do not necessarily indicate contradiction because solutions depend on analytical settings. More importantly, the labels may represent different dimensions.

Some clusters describe dimensions other than accounting tasks. Indrayani et al. (2024: 1, 11–13) and Al-Khoury et al. (2025: 1, 3) each identify accounting-education and research-methodology clusters, showing that one map can mix substantive, methodological, and setting-related dimensions. Both also interpret their structures through the technology

acceptance model (Al-Khoury et al. 2025: 2–5; Indrayani et al. 2024: 1, 11–13), while acceptance or adoption is itself prominent in their clusters. Interpretation can become circular when a corpus term is recovered as a cluster and then used to explain that structure.

Corpus-level maps mainly describe growth, citation, and thematic structure, whereas focused reviews address application-specific limitations, governance, and professional consequences (Abbas 2026: 9, 13; Han et al. 2023: 5, 11–12; Li et al. 2025: 10). These review types answer different questions. The present study does not evaluate whether AI applications work; it asks how strongly validation and governance concerns appear in the field's Author Keyword vocabulary.

Coverage also differs. Most windows in Table 1 end by 2024, whereas 271 of the 610 records appeared in 2024–2025. Reviews extending into the generative-AI period often describe a sequence from expert systems to machine learning and generative AI (Ben Said et al. 2026: 156, 165–166; Epelbaum & Farrell 2026: 301–303; Tocev & Atanasovski 2026: 163), but cluster membership and term years do not show whether newer and older vocabularies occur together in the same records.

Taken together, prior reviews leave a transparency problem and three analytical issues: corpus boundaries and screening are not consistently inspectable; co-word clusters may mix techniques, accounting applications, research designs, and settings; co-citation may combine works serving different disciplinary functions; and validation or governance concerns may be weakly visible in Author Keywords. This study addresses these issues through record-level screening disclosure, pairwise removal of frequent terms, co-citation and reference publication year spectroscopy (RPYS) with inspection of underlying works, link-level analysis of generative-AI vocabulary, and keyword-occurrence checks for validation and governance concerns.

METHODOLOGY

Bibliometric performance indicators describe publication and citation activity, while science mapping captures relationships among records (Donthu et al. 2021: 287–288; Zupic & Čater 2015: 431). Web of Science Core Collection was selected to retain consistent cited-reference metadata for co-citation and RPYS; accordingly, the corpus represents only its indexed coverage (Mongeon & Paul-Hus 2016: 213). Analyses were conducted using bibliometrix through Biblioshiny 5.3.0 and VOSviewer 1.6.20 (Aria & Cuccurullo 2017: 959; van Eck & Waltman 2010: 523).

SEARCH STRATEGY AND TARGETED SENSITIVITY TESTING

An initial search vocabulary was developed from the prior reviews examined in the Literature Review and refined after consultation with one professor of accounting and two doctoral researchers in accounting. Because several vocabulary and category choices remained judgmental, targeted incremental tests were conducted for disputed choices before the final specification was fixed; these tests were not applied to every query term.

Table 2 shows that four disputed terms retrieved eligible records and were retained, whereas four others did not. Adding Computer Science, Information Systems increased retrieval from 1,785 to 3,336 and yielded 131 provisional inclusions plus 27 manual-review candidates; the category was therefore retained.

TABLE 2. Targeted sensitivity tests of search vocabulary and category scope

Panel A: Disputed search terms

Candidate term	Unique records added	Eligible after screening	Excluded after screening	Decision
robotic process automation	47	33	14	Retained
conversational agent*	6	2	4	Retained
data mining	171	23	148	Retained
text mining	139	17	122	Retained
intelligent decision support system*	2	0	2	Dropped
intelligent decision aid*	0	0	0	Dropped
chatbot*	10	0	10	Dropped
conversational AI	0	0	0	Dropped

Panel B: Addition of Computer Science, Information Systems

Category added	Unique records added	Provisionally included at initial triage	Referred for manual review	Decision
Computer Science, Information Systems	1,551	131	27	Retained

After these tests, the specification was fixed. Web of Science Core Collection Advanced Search was run in the Topic field (titles, abstracts, Author Keywords, and Keywords Plus) on 24 July 2026 and rerun on 27 July with identical record counts. The timespan was 1 January 1956–31 December 2025; no eligible pre-1982 record survived. The final query was: TS=((“artificial intelligence” OR “machine learning” OR “deep learning” OR “neural network*” OR “natural language processing” OR “large language model*” OR LLM* OR “generative AI” OR “GenAI” OR “generative artificial intelligence” OR “ChatGPT” OR “GPT-3*” OR “GPT-4*” OR “generative pre-trained transformer*” OR “expert system*” OR “knowledge-based system*” OR “robotic process automation” OR “conversational agent*” OR “data mining” OR “text mining”) AND (“accounting” OR audit* OR “financial reporting” OR “accounting information system*” OR “taxation” OR “tax” OR “assurance” OR “forensic accounting” OR “management accounting”))

The Topic query returned 24,744 records. Restricting the results to English-language Articles and Review Articles in six Web of Science Categories—Business; Business, Finance; Economics; Information Science & Library Science; Management; and Computer Science, Information Systems—produced 3,336 records, exported as Full Record and Cited References.

No restriction was applied at the level of individual Web of Science indexes. The Core Collection configuration available through the institutional subscription included the Science Citation Index Expanded, Social Sciences Citation Index, Emerging Sources Citation Index, Conference Proceedings Citation Index–Science, Book Citation Index–Social Sciences & Humanities, Book Citation Index–Science, Arts & Humanities Citation Index, and Conference Proceedings Citation Index–Social Science & Humanities.

ELIGIBILITY, SCREENING AND ADJUDICATION

Eligibility criteria. Records had to satisfy both substantive AI and accounting conditions. A named AI technique had to be substantively developed, implemented, evaluated, adopted, or governed, or its effects studied; framing-only mentions and contributions unchanged under substitution by a conventional method were excluded. Accounting likewise had to be substantive: a named task, a principal accounting or reporting input, or an accounting outcome. Reviews were eligible; education-only studies, adjacent technologies without substantive AI, retracted publications, and specified lexical false positives were excluded. The criteria were fixed in writing before screening.

Initial screening. After export, 113 records assigned to 2026 and three exact duplicates were removed, leaving 3,220 records. ChatGPT 5.6 Sol triaged these against the written criteria using titles, abstracts, Author Keywords, source titles, and other metadata: 404 provisional inclusions, 2,233 exclusions, and 583 manual-review cases. All provisional inclusions had author confirmation, and all referred cases were individually assessed, producing 595 inclusions. The 2,233 provisional exclusions were not individually reviewed at this stage, so later checks targeted identified risks within that set.

Targeted checks. Three targeted checks addressed identified error patterns. Re-reading all 80 education-only exclusions added nine records. Gemini 3.1 Pro then applied the protocol twice to 72 exclusions containing both AI and accounting terms but lacking a substantive ground; 70 classifications agreed (97.2%; Cohen's kappa = 0.65), and lead-author adjudication yielded three additions overall. These checks produced 607 records. A related-paper consistency check subsequently identified one erroneous exclusion, raising the corpus to 608. These checks were risk-directed rather than systematic audits of all exclusions.

Stability check and adjudication. To assess stability, ChatGPT 5.6 Sol applied the same protocol twice to all 3,220 records in separate sessions with different record orders. Agreement was 3,010 (93.5%; Cohen's kappa = 0.78), with 210 disagreements. Author adjudication was triggered only when both classifications differed from the recorded decision, identifying 132 records; 68 recorded decisions were upheld and 64 reversed (33 exclusions to inclusions; 31 inclusions to exclusions). The final corpus contained 610 documents and 2,610 exclusions. Repeated classification measured consistency and identified cases for renewed human judgment; it was not an independent validation of screening accuracy.

Secondary Web of Science document-type tags did not independently determine exclusion beyond the initial Article and Review Article filters. Figure 1 summarises the identification, screening, and adjudication sequence in PRISMA-style form (Page et al. 2021).

Author names, affiliations, and Author Keywords were standardised in OpenRefine using general rules followed by record-specific corrections. Case and plurals were standardised, acronyms expanded, redundant system-level affiliations removed when the constituent institution was already listed, and classification codes deleted from keyword fields; the same rules were applied to later-added records. Cited references were retained as exported because many lacked reliable identifiers. Co-citation nodes therefore represent exported entries and may not correspond one-to-one with unique publications.

ANALYTICAL PROCEDURES AND SENSITIVITY TESTS

Biblioshiny supplied publication, source, authorship, affiliation, country, and citation indicators. Institution counts used cleaned affiliations; international co-authorship required affiliations from more than one country. Global and local citation counts were both reported.

Author Keyword co-occurrence and cited-reference co-citation networks were built in VOSviewer using Table 3 settings. Author Keywords were chosen because they reflect authors' own labels. Sensitivity analysis reconstructed the co-word network six times, each time removing one pair from artificial intelligence, accounting, auditing, and machine learning while holding other settings constant; partitions were compared on shared terms using the adjusted Rand index. Generative-AI analysis examined direct links involving large language model, ChatGPT, and generative artificial intelligence. RPYS compared each reference year's citations with the median of the two preceding, current, and two following years, and peak references were inspected before interpretation (Marx et al. 2014: 754).

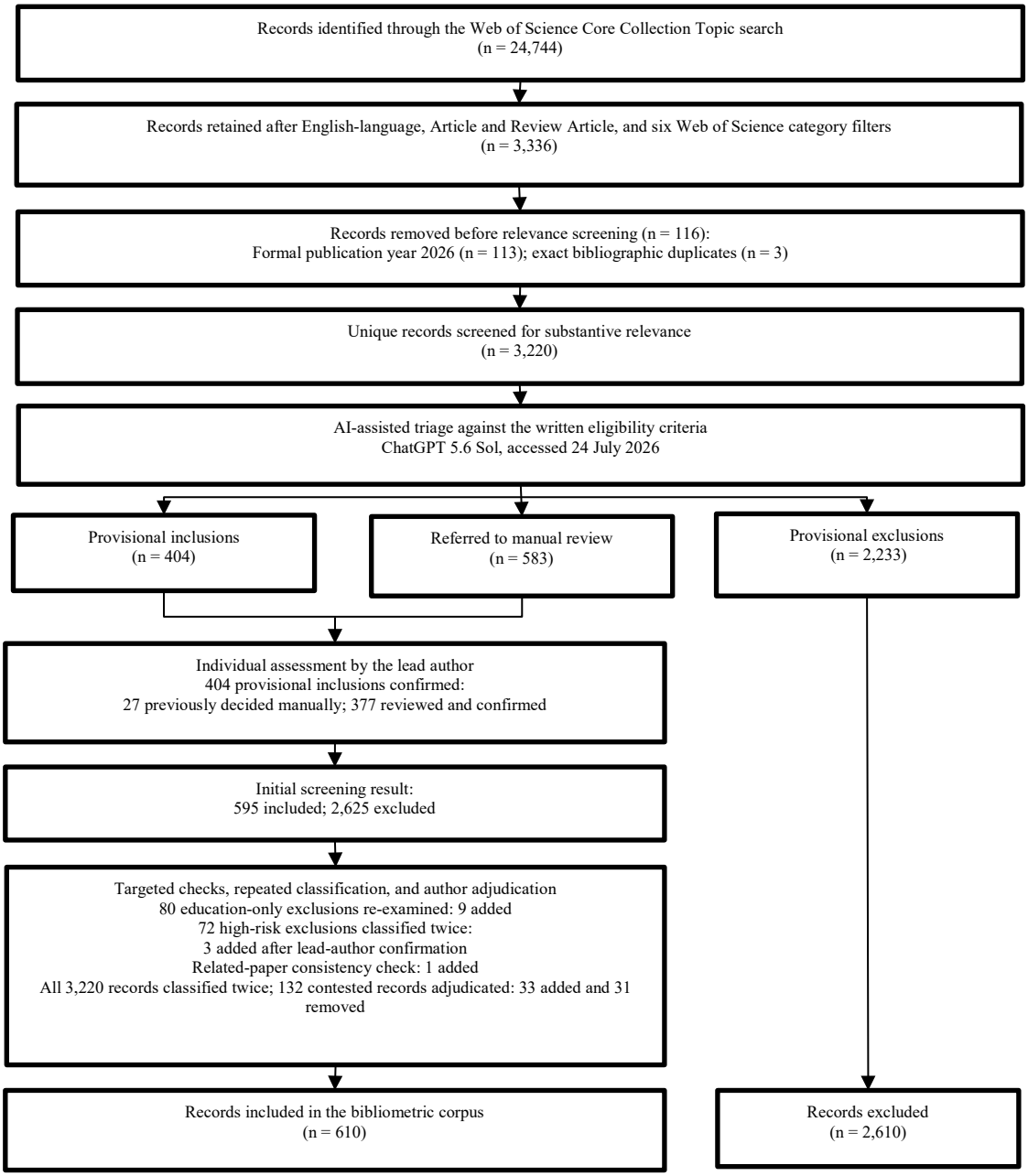


FIGURE 1. PRISMA-style flow of identification, screening, and inclusion
Note: The deposited screening record reports the identifier, final disposition, and decision stage for all 3,336 exported records.

TABLE 3. Network construction settings

	Author Keyword co-occurrence	Cited-reference co-citation
Unit	Author Keywords	Cited references
Counting method	Full	Full
Minimum occurrences or citations	5	15
Normalisation	Association strength	Association strength
Clustering	Smart local moving	Smart local moving
Resolution	1.00	1.00
Minimum cluster size	5	5
Additional link-strength threshold	None	None

Note: The 15-citation threshold was specified in the deposited analysis settings and retained unchanged for the final corpus. Smart local moving follows Waltman and van Eck (2013: 1, 6–8). No additional link-strength threshold was applied; VOSviewer's minimum link-strength control affects visual display only.

The search and screening record and AI-use declaration are available through an anonymised OSF view-only link https://osf.io/xn6hu/overview?view_only=a99c68b059e54e9cae562baf688d60e4.

RESULTS

PUBLICATION, AUTHORSHIP AND COLLABORATION PATTERNS

The corpus contains 610 documents published between 1982 and 2025 in 241 sources by 1,523 authors. Biblioshiny reports 25,502 cited references and a mean of 26 global citations per document. Output remained low for more than two decades. The first year with more than ten documents was 2018; annual output then rose from 23 in 2019 to 167 in 2025, with a fall from 68 in 2022 to 64 in 2023 before the subsequent increase (Figure 2). Two hundred and seventy-one documents, 44.4% of the corpus, appeared in 2024 and 2025. Biblioshiny reports a compound annual growth rate of 12.64%. This rate is dominated by the base year of one document; the annual counts provide the more informative description.

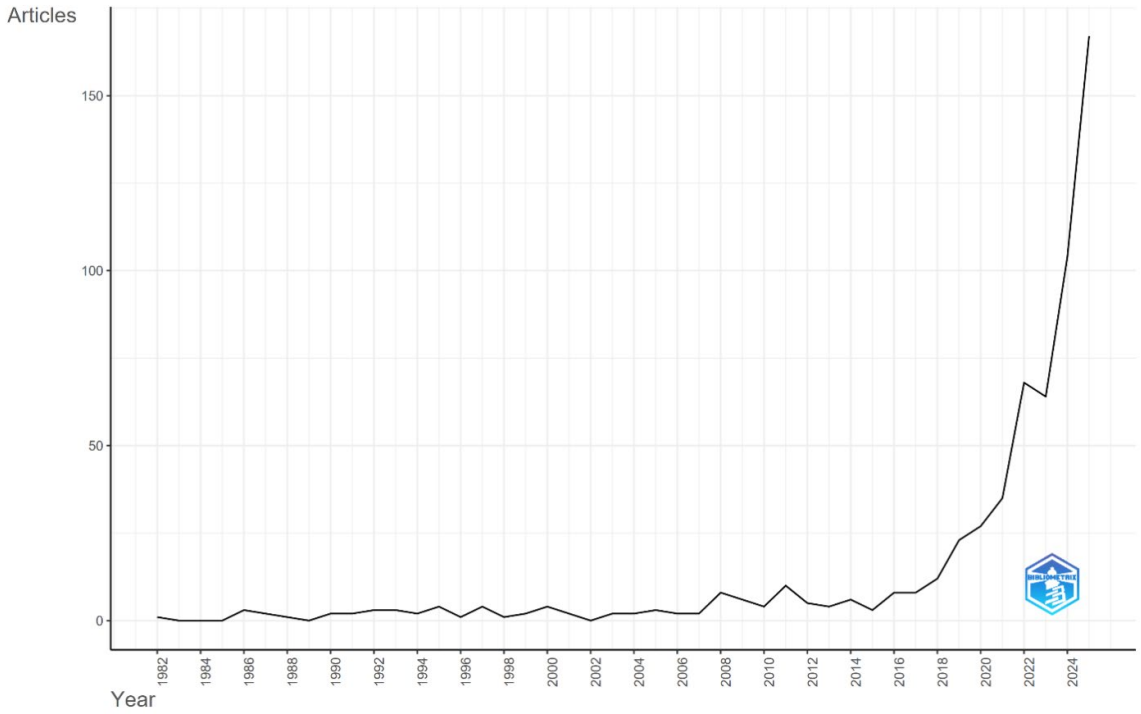


FIGURE 2. Annual number of documents in the corpus, 1982–2025

Eleven sources account for 209 documents under Bradford's law. The *International Journal of Accounting Information Systems* publishes 41 documents and the *Journal of Emerging Technologies in Accounting* publishes 31. *IEEE Access* publishes 29 documents, while *Electronics* publishes 15, showing that a search confined to accounting journals would omit part of the literature.

The corpus averages 2.9 co-authors per document; 106 documents are single-authored and 22.6% involve international co-authorship. Rutgers University is the most frequent affiliation, with 23 documents. When countries are counted across all author affiliations, the United States records 331 occurrences and China records 233, followed by the United Kingdom with 53, Romania with 49, Germany with 41, and Italy with 39. Among publications assigned by corresponding-author country, multiple-country publications account for 13.2% of the United States total and 17.6% of China's total, compared with 55.6% for the United Kingdom and 35.0% for Germany.

CONCEPTUAL STRUCTURE

Eighty-three Author Keywords occur at least five times. VOSviewer partitions them into five clusters, summarised in Table 4.

TABLE 4. Author keyword clusters

Cluster	n	Weighted mean year	Leading terms (occurrences)
1	21	2022.5	deep learning (36), natural language processing (20), bankruptcy prediction (19), data analytics (14), earnings management (11), explainable artificial intelligence (10), accuracy (10)
2	20	2020.5	machine learning (117), neural network (39), fraud (16), fraud detection (16), bankruptcy (10), prediction (10), support vector machine (9), random forest (8), decision tree (7)
3	20	2020.2	data mining (34), robotic process automation (33), large language model (21), ChatGPT (19), expert system (17), text mining (17), financial reporting (14), automation (12), internal control (12)
4	16	2022.3	artificial intelligence (191), auditing (67), accounting (57), digital transformation (13), blockchain (10), ethics (10), accounting profession (9)
5	6	2022.3	big data (22), audit quality (13), going concern opinion (8), audit fee (6), cloud computing (6), auditor (5)

Note: The Author Keyword co-occurrence network was constructed using full counting, a minimum occurrence threshold of five, association-strength normalisation, a resolution of 1.00, and a minimum cluster size of five. The weighted mean year is the mean publication year of the documents supplying each term, weighted by occurrences.

VOSviewer partitioned the 83 Author Keywords into five clusters. Cluster 1 links deep learning and natural language processing with bankruptcy prediction and earnings management. Cluster 2 combines machine learning and classical classifiers with fraud detection and prediction. Cluster 3 groups data mining, robotic process automation, large language models, expert systems, and text-processing techniques with financial reporting and internal control. Cluster 4 contains the general labels of the field and profession, including artificial intelligence, auditing, accounting, and accounting profession. Cluster 5 combines big data and cloud computing with audit-related terms, including audit quality, going-concern opinion, audit fee, and auditor. The first three clusters are more visibly organised around technique families, while accounting applications and outcomes are distributed across the five-cluster solution.

To assess the sensitivity of the partition to high-frequency general terms, the network was rebuilt after removing each of the six possible pairs formed from artificial intelligence, accounting, auditing, and machine learning, while all other settings were held constant. Table 5 compares each resulting partition with the reported network.

TABLE 5. Effect of removing high-frequency terms in pairs

Terms removed	Terms retained	Clusters	Adjusted Rand index
None (reported network)	83	5	1.000
auditing, machine learning	81	6	0.858
accounting, machine learning	81	5	0.829
artificial intelligence, machine learning	81	6	0.742
artificial intelligence, accounting	81	6	0.709
artificial intelligence, auditing	81	7	0.576
accounting, auditing	81	7	0.573

Note: Each row reports a separate reconstruction in which only the listed pair was removed. All other network settings were unchanged. The adjusted Rand index compares that partition with the reported network across the terms retained in both.

Removing pairs of terms drawn from artificial intelligence, accounting, and auditing produced the largest changes, with adjusted Rand indices ranging from 0.573 to 0.709. Removing machine learning together with one of those three terms produced more similar partitions, with adjusted Rand indices ranging from 0.742 to 0.858. Among these targeted removals, the reported partition was most sensitive to pairs drawn from the general domain and profession labels.

Large language model, ChatGPT and generative artificial intelligence are assigned to Cluster 3, together with expert system, data mining, text mining, and natural language processing. The generative-AI terms are recent, with weighted mean publication years between 2024.4 and 2024.6, compared with 2011.1 for expert system, 2017.8 for text mining, 2019.0 for data mining, and 2022.1 for natural language processing. None of the three generative-AI terms co-occurs directly with expert system, data mining, or text mining. ChatGPT and generative artificial intelligence each have one direct link with natural language processing, while large language model has none. Their placement in the same cluster therefore arises mainly through indirect connections with shared accounting-application terms, including financial reporting, automation, and internal control.

INTELLECTUAL STRUCTURE

Ninety-three cited references receive at least fifteen citations within the corpus and form three clusters, summarised in Table 6. Cluster 1 contains work on automation, the profession, ethics, and governance; Cluster 2 contains fraud, misstatement, and textual analysis; and Cluster 3 combines classical distress models with general machine-learning tools.

TABLE 6. Cited-reference co-citation clusters

Cluster	n	Character	Most cited references (local citations)
1	47	Automation, the profession, ethics, and governance	Krejcie & Morgan (1970) 62; Cooper et al. (2019) 54; Issa et al. (2016) 52; Moffitt et al. (2018) 42; Munoko et al. (2020) 41; Sutton et al. (2016) 37; Fedyk et al. (2022) 34; Frey & Osborne (2017) 33
2	31	Fraud, misstatement, and textual analysis	Bao et al. (2020) 45; Dechow et al. (2011) 45; Perols (2011) 39; Hajek & Henriques (2017) 32; Loughran & McDonald (2011) 32; Cecchini et al. (2010) 29; Kirkos et al. (2007) 29
3	15	Classical distress models with general machine-learning tools	Altman (1968) 55; Ohlson (1980) 37; Beaver (1966) 36; Breiman (2001) 33; "Kumar S, 2018, Mol Biol Evol" 28*; Chen & Guestrin (2016) 25; Zmijewski (1984) 21

Note: The cited-reference co-citation network was constructed using full counting, a minimum citation threshold of fifteen, association-strength normalisation, a resolution of 1.00, and a minimum cluster size of five. Local citations are citations received from documents inside the corpus. Cluster median publication years are not reported because they move with the threshold. An asterisk marks a node reported as exported by Web of Science and identified in the text as a reference-matching artefact.

Cluster 3 combines accounting foundations with imported analytical methods. Beaver, Altman, Ohlson, and Zmijewski provide accounting models of financial distress, whereas Breiman and Chen and Guestrin provide general-purpose machine-learning methods. Their placement in the same cluster records their joint citation in later documents, but the cluster alone does not distinguish the disciplinary role of each reference.

Two prominent nodes illustrate why the works behind the network require inspection. Krejcie and Morgan (1970), the most co-cited reference with 62 local citations, provides a general survey sample-size table rather than an argument about AI or accounting. The exported string "Kumar S, 2018, Mol Biol Evol," which receives 28 local citations, results from a cited-reference matching error and is not interpreted as part of the field's intellectual base. These cases show that co-citation prominence records shared citing behaviour and should not, by itself, be interpreted as scientific quality or intellectual origin.

RPYS identifies peaks in 1968, 1980, 1984, 1990, 1995, 2002, 2006, 2008, 2011, and 2017. Table 7 summarises representative works behind these peaks and the functions they perform in the cited base.

TABLE 7. RPYS peak years and representative cited works

Peak year	Representative cited works	Stream represented
1968	Altman (1968); Ball and Brown (1968)	Financial distress and accounting research
1980	Ohlson (1980)	Accounting-based distress prediction
1984	Zmijewski (1984); Breiman et al. (1984)	Accounting model and imported classification method
1990	Tornatzky and Fleischer (1990); Kleinmuntz (1990); Koh (1990); Bonner (1990)	Innovation, judgment, and accounting
1995	Lenard et al. (1995); Cortes and Vapnik (1995)	Auditing application and support-vector method
2002	Chawla et al. (2002)	Class-imbalance method
2006	Kotsiantis et al. (2006)	Machine-learning classification methods
2008	Li (2008)	Accounting disclosure and textual properties
2011	Dechow et al. (2011); Perols (2011); Loughran and McDonald (2011)	Misstatement, fraud, and textual analysis
2017	Frey and Osborne (2017); Hajek and Henriques (2017); Appelbaum et al. (2017)	Automation, machine learning, and auditing

Reading the works behind the peaks identifies three sources of the cited base: accounting research on financial distress, disclosure, earnings quality, and audit; analytical methods imported from statistics and machine learning; and work from labour economics and behavioural science on automation, innovation, and judgment. RPYS identifies years in which cited references are concentrated relative to neighbouring years; it does not, by itself, establish why individual works were cited or how influential they are.

WEAKLY REPRESENTED AUTHOR KEYWORDS

No Author Keyword explicitly denoting external validation, out-of-sample testing, robustness, or generalisation occurs at least five times, although prediction and classification terms are prominent in the network. Accuracy and explainable artificial intelligence each occur ten times, but neither term specifies whether model performance is evaluated beyond the data used for model development or estimation. The normative vocabulary is similarly limited: ethics occurs ten times, whereas governance, trust, automation bias, privacy, fairness, materiality, regulation, accountability, and transparency each occur fewer than five times. These frequencies describe how the included corpus is labelled through Author Keywords; they do not establish that the corresponding issues, practices, or studies are absent.

DISCUSSION

The descriptive results confirm the rapid post-2018 growth reported by prior reviews, but publication volume is evidence of activity rather than technological maturity or realised organisational value. Its role here is mainly to establish the scale and recency of the literature within which the structural findings are interpreted.

The first substantive finding is that the five-cluster co-word solution is organised more by technique families than by accounting tasks. Deep learning and natural language processing, machine learning and classical classifiers, and data mining with automation and text-processing technologies form technique-centred regions, while accounting applications are distributed across them. Because Author Keywords are authors' labels, the map primarily describes how the literature names itself rather than the tasks performed in full texts. Coding accounting tasks directly would test how closely these structures correspond.

The pairwise-removal tests show that the reported partition is not equally sensitive to all four frequent terms. Removing pairs drawn from artificial intelligence, accounting, and auditing produced the largest changes, whereas reconstructions involving machine learning were closer to the reported solution. Because removing frequent or highly connected terms can itself alter community structure, these results indicate dependence of this partition on the general domain/profession labels rather than intrinsic instability of the underlying literature. As these terms also helped define the search boundary, the test is diagnostic of query-induced structure, not a reconstruction of a query-independent field.

A related methodological issue concerns the stability of the corpus boundary. Two applications of the same written screening protocol agreed on 93.5% of 3,220 records, yet 210 decisions differed. Among the 132 cases in which both applications departed from the recorded decision, author adjudication reversed 64. These figures are neither an error-rate estimate nor an independent human replication, but they show that written criteria do not make relevance boundaries perfectly stable. Bibliometric reviews would therefore benefit from reporting the fields used for screening, agreement across repeated assessments and, where possible, record-level dispositions that allow disputed cases to be inspected.

Co-citation and RPYS likewise require functional interpretation. The cited base combines accounting work on distress, disclosure, earnings quality, and audit; methods imported from statistics and machine learning; and labour-economics or behavioural work on automation, innovation, and judgment. These sources can appear together while serving different roles—for example, an accounting model may define the problem while a machine-learning paper supplies the method. The sample-size table and reference-matching artefact among prominent nodes further show that co-citation frequency reflects shared citing behaviour, not scientific quality.

Generative-AI vocabulary shows continuity and separation. Large language model, ChatGPT, and generative artificial intelligence share a cluster and accounting applications with earlier technologies, but none links directly to expert system, data mining, or text mining, and links to natural language processing are limited. The evidence therefore supports continuity in application areas, not a demonstrated technological or intellectual progression. Citation analysis could test whether the separation reflects a break in research practice or simply the short history of generative-AI terminology.

The weak representation of validation and governance vocabulary should also be interpreted cautiously. No Author Keyword relating to external validation, out-of-sample testing, robustness, or generalisation reaches the inclusion threshold, while most governance-related terms also remain below it. Focused reviews nevertheless discuss several of these concerns. Their sparse appearance in the co-word network may therefore reflect differences in terminology, limited substantive attention, or both. Full-text coding would be needed to distinguish whether these practices are absent from the studies themselves or merely absent from the keywords used to describe them.

IMPLICATIONS

Methodologically, the results caution against treating a co-word partition as definitive. Because Author Keywords organise the field more by technique than accounting task, and the domain/profession region is most sensitive to frequent-term removal, cluster interpretations should report this sensitivity. Co-citation and RPYS likewise require inspection of underlying works: jointly cited accounting, analytical, and behavioural or labour studies may serve different functions, so prominence does not imply theoretical equivalence or scientific quality.

For accounting firms and other organisations, the limited direct connection between generative-AI vocabulary and earlier expert system, data mining, and text mining vocabularies suggests that recent publications do not often use the same vocabulary as the earlier research tradition. Evaluation of generative-AI applications may therefore benefit from drawing on both recent studies and earlier evidence concerning model validation, internal controls, human oversight, and process integration. For regulators and standard setters, governance, privacy, fairness, regulation, accountability, and transparency remain weakly represented in the dominant Author Keyword vocabulary. This does not show that research on these matters is absent, but it indicates that relevant evidence may be dispersed across specialised literatures and may not be readily located through general AI-in-accounting terminology. Reviews undertaken for regulatory or standard-setting purposes should therefore use broader search vocabularies and examine specialised work on auditing, taxation, financial reporting, and professional accountability.

LIMITATIONS AND FUTURE RESEARCH

Three limitations qualify the findings. First, the Web of Science-only corpus provides consistent cited-reference metadata but does not cover all AI-in-accounting research; the search vocabulary may also omit unlisted technologies or retrieve lexical false positives. Second, although every included record was read by the lead author, the 2,233 provisional exclusions were not individually reviewed initially. Later targeted checks do not estimate a false-negative rate, and the 610-document corpus should therefore be treated as a transparent rule-based boundary rather than a uniquely correct one. Third, bibliometric measures imperfectly capture substantive content: Author Keywords may omit issues studied in full texts, co-word structures depend on thresholds and general terms, and exported cited-reference strings can contain matching errors, as illustrated by the “Kumar S, 2018, Mol Biol Evol” artefact.

Future research could address these limits through full-text coding of accounting tasks and validation procedures, independent human rescreening, replication of the pairwise-removal test in other corpora, and citation analysis of whether generative-AI studies engage with earlier expert-system and machine-learning research.

CONCLUSION

This study maps 610 Web of Science documents on AI in accounting from 1982 to 2025 using a rerunnable search, substantive eligibility criteria, and a deposited record-level disposition.

Publication activity accelerated after 2018. The Author Keyword structure is organised more by AI techniques than by accounting tasks, and its general domain/profession region is most sensitive to frequent-term removal. Co-citation and RPYS show a cited base spanning accounting research, analytical methods, and work on automation and judgment. Generative-AI terms occupy established accounting applications but connect little with earlier technological vocabularies, while validation, robustness, and most governance terms remain weakly represented.

The contribution is an inspectable corpus boundary and a sensitivity-tested interpretation that distinguishes the disciplinary roles hidden within co-citation structure. The study also demonstrates that repeated application of a written screening protocol can produce different relevance decisions, supporting more explicit reporting of screening fields, agreement, and record-level dispositions.

The findings remain bounded by one database, Author Keyword labels, uncertain relevance judgments, and imperfect cited-reference matching. Full-text coding, independent human screening, and citation-based examination of generative-AI continuity would provide the next tests of the patterns identified here.

ACKNOWLEDGEMENT

The authors would like to acknowledge funding from the Faculty of Economics and Management, Universiti Kebangsaan Malaysia through a Research Initiative Grant (GIP), code EP-2025-026.

REFERENCES

Abbas, K. 2026. Management accounting and artificial intelligence: A comprehensive literature review and recommendations for future research. *The British Accounting Review* 58(1).

Abu Huson, Y., Sierra-García, L. & Garcia-Benau, M.A. 2024. A bibliometric review of information technology, artificial intelligence, and blockchain on auditing. *Total Quality Management & Business Excellence* 35(1–2): 91–113.

Agustí, M.A. & Orta-Pérez, M. 2023. Big data and artificial intelligence in the fields of accounting and auditing: A bibliometric analysis. *Spanish Journal of Finance and Accounting* 52(3): 412–438.

Al-Khoury, A.F., Alastal, A.Y.M., Samara, H., Alslaibi, N.A., Abdulmuhsin, A.A. & Alqudah, M.Z. 2025. The bibliometric landscape of emerging technology in the accounting information systems field. *Cogent Business & Management* 12(1).

Alkhatib, E., Alkhatib, A. & Jarvis, R. 2026. Auditing using artificial intelligence: A systematic literature review with scientometric and topic modeling insights through a socio-technical systems perspective. *Journal of Financial Reporting and Accounting*: 1–28.

Altman, E.I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23(4): 589–609.

Amani, F.A. & Fadlalla, A.M. 2017. Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems* 24: 32–58.

Appelbaum, D., Kogan, A. & Vasarhelyi, M.A. 2017. Big data and analytics in the modern audit engagement: Research needs. *Auditing: A Journal of Practice & Theory* 36(4): 1–27.

Aria, M. & Cuccurullo, C. 2017. bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics* 11(4): 959–975.

Ball, R. & Brown, P. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6(2): 159–178.

Bao, Y., Ke, B., Li, B., Yu, Y.J. & Zhang, J. 2020. Detecting accounting fraud in publicly traded US firms using a machine learning approach. *Journal of Accounting Research* 58(1): 199–235.

Beaver, W.H. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research* 4: 71–111.

Belahouaoui, R. & Attak, E.H. 2024. Digital taxation, artificial intelligence and Tax Administration 3.0: Improving tax compliance behavior – A systematic literature review using textometry (2016–2023). *Accounting Research Journal* 37(2): 172–191.

Ben Said, F., Chakroun, R. & Slim, C. 2026. A bibliometric analysis of the impact of artificial intelligence on the accounting profession. *Fintech and Digital Accounting Review* 2(1): 155–179.

Bonner, S.E. 1990. Experience effects in auditing: The role of task-specific knowledge. *The Accounting Review* 65(1): 72–92.

Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5–32.

Breiman, L., Friedman, J.H., Olshen, R.A. & Stone, C.J. 1984. *Classification and Regression Trees*. Belmont, CA: Wadsworth.

Cecchini, M., Aytug, H., Koehler, G.J. & Pathak, P. 2010. Detecting management fraud in public companies. *Management Science* 56(7): 1146–1160.

Chawla, N.V., Bowyer, K.W., Hall, L.O. & Kegelmeyer, W.P. 2002. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16: 321–357.

Chen, T. & Guestrin, C. 2016. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: 785–794.

Coita, I.F., Filip, L. & Pop, M.V. 2026. Artificial intelligence systems in accounting and auditing: A bibliometric and exploratory analysis. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience* 17(2): 35–65.

Cooper, L.A., Holderness, D.K., Sorensen, T.L. & Wood, D.A. 2019. Robotic process automation in public accounting. *Accounting Horizons* 33(4): 15–35.

Cortes, C. & Vapnik, V. 1995. Support-vector networks. *Machine Learning* 20(3): 273–297.

Dechow, P.M., Ge, W., Larson, C.R. & Sloan, R.G. 2011. Predicting material accounting misstatements. *Contemporary Accounting Research* 28(1): 17–82.

Dong, M.M., Stratopoulos, T.C. & Wang, V.X. 2024. A scoping review of ChatGPT research in accounting and finance. *International Journal of Accounting Information Systems* 55.

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N. & Lim, W.M. 2021. How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research* 133: 285–296.

Elnakeeb, S. & Elawadly, H.S.H. 2025. Automation and artificial intelligence in accounting: A comprehensive bibliometric analysis and future trends. *Journal of Financial Reporting and Accounting* (early access).

Epelbaum, N. & Farrell, P.J. 2026. Artificial intelligence and auditing: A bibliometric study. *Accounting Perspectives* 25(2): 289–311.

Fedyk, A., Hodson, J., Khimich, N. & Fedyk, T. 2022. Is artificial intelligence improving the audit process? *Review of Accounting Studies* 27(3): 938–985.
Frey, C.B. & Osborne, M.A. 2017. The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change* 114: 254–280.
Hajek, P. & Henriques, R. 2017. Mining corporate annual reports for intelligent detection of financial statement fraud – A comparative study of machine learning methods. *Knowledge-Based Systems* 128: 139–152.
Han, H., Shiwakoti, R.K., Jarvis, R., Mordi, C. & Botchie, D. 2023. Accounting and auditing with blockchain technology and artificial intelligence: A literature review. *International Journal of Accounting Information Systems* 48.
Indrayani, Sukoharsono, E.G., Djamhuri, A. & Roekhudin. 2024. Mapping research landscape of emerging technology in the accounting field: A bibliometric analysis. *Cogent Business & Management* 11(1).
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.
Kirkos, E., Spathis, C. & Manolopoulos, Y. 2007. Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications* 32(4): 995–1003.
Kleinmuntz, B. 1990. Why we still use our heads instead of formulas: Toward an integrative approach. *Psychological Bulletin* 107(3): 296–310.
Koh, H.C. 1990. The role of financial ratios in going-concern assessment. *Journal of Business Finance & Accounting* 17(2): 179–199.
Kotsiantis, S.B., Zaharakis, I.D. & Pintelas, P.E. 2006. Machine learning: A review of classification and combining techniques. *Proceedings of World Academy of Science, Engineering and Technology* 12: 284–289.
Krejcie, R.V. & Morgan, D.W. 1970. Determining sample size for research activities. *Educational and Psychological Measurement* 30(3): 607–610.
Kureljusic, M. & Karger, E. 2024. Forecasting in financial accounting with artificial intelligence – A systematic literature review and future research agenda. *Journal of Applied Accounting Research* 25(1): 81–104.
Lamboglia, R., Lavorato, D., Scornavacca, E. & Za, S. 2021. Exploring the relationship between audit and technology: A bibliometric analysis. *Meditari Accountancy Research* 29(5): 1233–1260.
Lenard, M.J., Alam, P. & Madey, G.R. 1995. The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. *Decision Sciences* 26(2): 209–227.
Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45(2–3): 221–247.
Li, W., Liu, W., Deng, M., Liu, X. & Feng, L. 2025. The impact of large language models on accounting and future application scenarios. *Journal of Accounting Literature* 7(2): 78-80.
Liaras, E., Nerantzidis, M. & Alexandridis, A. 2024. Machine learning in accounting and finance research: A literature review. *Review of Quantitative Finance and Accounting* 63(4): 1431–1471.
Loughran, T. & McDonald, B. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance* 66(1): 35–65.
Marx, W., Bornmann, L., Barth, A. & Leydesdorff, L. 2014. Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology* 65(4): 751–764.
Moffitt, K.C., Rozario, A.M. & Vasarhelyi, M.A. 2018. Robotic process automation for auditing. *Journal of Emerging Technologies in Accounting* 15(1): 1–10.
Mongeon, P. & Paul-Hus, A. 2016. The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics* 106(1): 213–228.
Mugwira, T. 2022. Internet related technologies in the auditing profession: A WOS bibliometric review of the past three decades and conceptual structure mapping. *Revista de Contabilidad – Spanish Accounting Review* 25(2): 201–216.
Munoko, I., Brown-Liburd, H.L. & Vasarhelyi, M.A. 2020. The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics* 167(2): 209–234.
Ohlson, J.A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18(1): 109–131.
Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hrobjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P. & Moher, D. 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *British Medical Journal* 372.
Perols, J. 2011. Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice & Theory* 30(2): 19–50.
Ranta, M., Ylinen, M. & Järvenpää, M. 2023. Machine learning in management accounting research: Literature review and pathways for the future. *European Accounting Review* 32(3): 607–636.
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.

Tlili, H. 2025. Bibliometric analysis of artificial intelligence in accounting: Trends and future directions. *International Journal of Advanced and Applied Sciences* 12(8): 149–165.
Tocev, T. & Atanasovski, A. 2026. Artificial intelligence in accounting: A corpus-wide bibliometric review and decade-by-decade thematic evolution. *Zeszyty Teoretyczne Rachunkowości* 50(2): 163–187.
Tornatzky, L.G. & Fleischer, M. 1990. *The Processes of Technological Innovation*. Lexington Books, Lexington, MA.
van Eck, N.J. & Waltman, L. 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84(2): 523–538.
Waltman, L. & van Eck, N.J. 2013. A smart local moving algorithm for large-scale modularity-based community detection. *The European Physical Journal B* 86.
Zmijewski, M.E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* 22: 59–82.
Zupic, I. & Čater, T. 2015. Bibliometric methods in management and organization. *Organizational Research Methods* 18(3): 429–472.

Lou Yuxiao
Faculty of Economics and Management
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor, MALAYSIA.
E-mail: p121136@siswa.ukm.edu.my

Ruhanita Maelah*
Faculty of Economics and Management
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor, MALAYSIA.
E-mail: ruhanita@ukm.edu.my

Kamarul Baraini Keliwon
Faculty of Economics and Management
Universiti Kebangsaan Malaysia
43600 UKM Bangi, Selangor, MALAYSIA.
E-mail: baraini@ukm.edu.my

*Corresponding author