TANMEAH BASSMAH PRESS
International Journal of Digital Accounting and Fintech Sustainability (IJDAFS). Volume 3, Issue 1, 2026
Journal Homepage: https://journals.tanmeah.com/index.php/ijdafs
ISSN: 3008-1505 (print), 3008-1513 (online)
PP. 350-373

Generative Artificial Intelligence in Accounting and Auditing: A Bibliometric and Thematic Review

Nehad Ibrahim Ineizeh[1]*, Muntasir Khaled Hamdan[2], Zakaria Mahmoud Jammal[2], Fayez Lotfi Dweikat[3]

[1]*Faculty of Economics and Muamalat, Universiti Sains Islam Malaysia (USIM), Nilai, Malaysia.*
[2]*Faculty of Administrative and Economic Sciences, AL - Quds Open University, Jenin Palestine.*
[3]*Technical and Vocational Education, Alummah University College, Jerusalem, Palestine.*

C H R O N I C L E
Article history:
Received: March, 07, 2026
Received in revised format: April, 21, 2026
Accepted: June 27, 2026
Available online: June 30, 2026

***Keywords**:*

large language models; ChatGPT; generative artificial intelligence; auditing; financial reporting; bibliometric analysis

A B S T R A C T

This paper aims to map the knowledge structure of the research on the generative artificial intelligence and the large language models in accounting and auditing, and to synthesise the previous findings into an integrative framework and a research agenda. A bibliometric analysis was combined with a thematic content analysis. The records were retrieved from OpenAlex and Crossref for the period from January 2022 to April 2026 and screened following the PRISMA 2020 logic, and the final corpus consists from 321 peer-reviewed documents. The science mapping was conducted through a concept co-occurrence network and the Louvain clustering. The research grew in an exponential manner, where 75.4% of the corpus was published in 2025 and 2026. The United States and China dominate the production, while only 16.5% of the documents were internationally co-authored. The citations are concentrated, since the ten most cited documents received 47.7% of the citations. Four clusters were identified, namely the reporting, disclosure and markets, the education, acceptance and the profession, the judgment and model reliability, and the auditing and assurance. The evidence moved from the benchmarking of the chatbots on the examinations toward the evidence on the audit quality, the disclosure and the market efficiency, but the causal and longitudinal evidence is still limited. The study proposes an enablers, mechanisms and outcomes framework and a research agenda for the researchers, the regulators, the audit firms and the educators, especially in the emerging economies.

***JEL Classification:** M41, M42, M15, O33, G14*

1. Introduction

The accounting and auditing professions were always affected by the technological changes, and these changes include the spreadsheets, the enterprise resource planning systems, the data analytics and the robotic process automation. However, the release of ChatGPT in November 2022 is considered different from the

* Corresponding author.
E-mail address: n.ineizeh@usim.edu.my
https://doi.org/10.70568/ijdafs.3.1.5

previous technologies, because the generative artificial intelligence (GenAI) produces a new text, a new code and a new analysis in response to the instructions written in a natural language. In other words, the GenAI tools are able to perform the tasks that were considered before as a knowledge work, such as the drafting of the disclosures, the interpretation of the accounting standards and the preparation of the audit working papers (Dwivedi et al., 2023; Vasarhelyi et al., 2023). Accordingly, the GenAI was described in the accounting literature as a technology that may disrupt the accounting and auditing industries (Eulerich et al., 2024). The earlier wave of the artificial intelligence (AI) in accounting was focused on the machine learning models that classify, predict and detect anomalies, and these models were used mainly by the audit firms and the large companies (Kokina & Davenport, 2017; Lehner et al., 2022). In contrast, the GenAI is available for the individual accountant, the student and the small firm with a low cost, and it interacts with the user in a conversational way. Moreover, the GenAI is able to process the unstructured textual information, which represents a large part of the accounting information, such as the annual reports, the notes to the financial statements, the audit reports and the sustainability reports (de Kok, 2025; Huang et al., 2023). It should be noted that the same features that make the GenAI attractive also create new risks, and these risks include the hallucination of the incorrect information, the bias, the lack of the explainability and the leakage of the confidential client data (Bender et al., 2021; Ji et al., 2023; PCAOB, 2024).

As a result of these developments, the number of the studies on the GenAI in accounting and auditing increased rapidly within a short period. The early studies were mostly commentaries and experiments that tested the performance of ChatGPT on the accounting assessment questions and the professional examinations (Eulerich et al., 2024; Wood et al., 2023). Later, the studies started to examine the use of the GenAI in the internal audit functions, the external audit firms, the financial reporting, the accounting education and the accounting research (de Kok, 2025; Emett et al., 2025; Li & Goel, 2026). However, this literature is fragmented between the different journals, the different disciplines and the different methods, and it includes the studies from the accounting journals, the information systems journals, the education journals and the computer science conference. Therefore, there is a need for a review that maps the structure of this literature and synthesises the main findings in a systematic manner.

A few reviews were already published on this topic. Dong et al. (2024) conducted a scoping review of the ChatGPT research in accounting and finance, and they identified three themes that are the applications of ChatGPT, the use of ChatGPT as a research tool, and the implications for the professionals and the organisations. Moreover, Anica-Popa et al. (2024) reviewed 103 papers in order to identify the competencies that are required from the accounting and audit professionals in the GenAI context, and Li et al. (2025) reviewed the impact of the LLMs on the accounting practices and the future application scenarios. Nevertheless, these reviews covered the literature until 2024 or early 2025, and they did not combine the bibliometric indicators with the thematic synthesis. It is worth mentioning that more than three quarters of the documents in the current corpus were published in 2025 and 2026, which means that the previous reviews could not capture the most recent empirical evidence, such as the evidence on the audit quality, the disclosure practices and the market efficiency (Bertomeu et al., 2025; Blankespoor et al., 2026; Du et al., 2026). Based on the above, the current study aims to answer the following research questions:

RQ1: What are the publication trends and the performance indicators (sources, countries, authors and citations) of the research on the GenAI in accounting and auditing?

RQ2: What is the intellectual and conceptual structure of this research, as reflected in the concept co-occurrence network?

RQ3: What are the main themes and findings of the previous studies, and how the evidence developed over the time?

RQ4: What are the research gaps, and what is the future research agenda that can be derived from the literature?

The study contributes to the literature in three ways. Firstly, the study provides an updated bibliometric map of the GenAI research in accounting and auditing until April 2026, and the map is based on two open databases, which increases the transparency and the reproducibility of the review (Hendricks et al., 2020; Priem et al., 2022). Secondly, the study combines the data-driven clusters with a qualitative synthesis of the

key studies, and in this way the study explains not only what is studied but also what is known. Thirdly, the study proposes an integrative framework of the enablers, the mechanisms and the outcomes of the GenAI use in accounting, and it derives a research agenda that is relevant for the digital accounting and the fintech sustainability research, especially in the emerging economies.

2. Literature Review

2.1 Generative Artificial Intelligence in Accounting

The artificial intelligence in accounting is not a new subject, and the previous literature discussed the use of the expert systems, the machine learning and the natural language processing in the accounting and auditing tasks for several decades. In the auditing field, Kokina and Davenport (2017) explained that the automation will change the audit work by taking over the repetitive and structured tasks, and they argued that the auditors will need to focus more on the judgment and the communication. Moreover, Moll and Yigitbasioglu (2019) discussed how the internet-related technologies, including the AI and the cloud computing, shape the work of the accountants and change the boundaries of the accounting profession. In the same direction, Lehner et al. (2022) highlighted the ethical challenges of the AI-based decision making in accounting and auditing, such as the objectivity, the privacy, the transparency and the accountability. The GenAI differs from the previous AI applications in several aspects. The GenAI models, and especially the LLMs, are trained on a very large amount of the textual data, and they generate a new content by predicting the next word in a sequence (Dwivedi et al., 2023). As for the accounting context, the earlier language models such as FinBERT were used mainly for the classification of the financial text, for example the classification of the sentiment of the analyst reports (Huang et al., 2023). In contrast, the generative models such as GPT-4 are able to answer the questions, to summarise the documents, to write the reports and to perform the calculations through the external tools, and all of these tasks can be done without a specific programming skills from the user (de Kok, 2025; Vasarhelyi et al., 2023). Therefore, the GenAI moved the AI from the back office of the large firms to the desk of every accountant, auditor and student.

2.2 Theoretical lenses used in the literature

The previous studies used several theoretical lenses in order to explain the adoption and the consequences of the GenAI in accounting. Firstly, the technology acceptance perspective, which includes the technology acceptance model (Davis, 1989) and the unified theory of acceptance and use of technology (Venkatesh et al., 2003), is the most used lens in the studies that examine the intention of the accountants, the auditors and the students to use the GenAI (Aziki, 2025; Jabin, 2025; Al-Okaily, 2025). Secondly, the human and automation interaction perspective explains how the users rely on the algorithms, and it includes the algorithm aversion (Dietvorst et al., 2015), the algorithm appreciation (Logg et al., 2019) and the automation bias and complacency (Parasuraman & Manzey, 2010). This perspective is important for the auditing, because the previous evidence showed that the auditors may rely less on the advice that comes from the AI compared with the advice that comes from a human specialist (Commerford et al., 2022). Thirdly, the trust in the AI is considered as a central concept, since the trust determines whether the user accepts the output of the system or verifies it (Glikson & Woolley, 2020). Fourthly, the cognitive offloading concept explains how the users transfer a part of the cognitive work to the external tools, which may improve the performance in the short term but may reduce the learning in the long term (Risko & Gilbert, 2016). Finally, a critical stream of the literature used the critical and decolonial perspectives in order to question whose knowledge is reproduced by the LLMs, and this stream argued that the GenAI may reinforce the Western-centric accounting knowledge in the developing countries (Muzanenhamo & Power, 2024). These lenses are used in the current study as a background for the interpretation of the themes and for the development of the integrative framework.

3. Methodology

3.1 Research design

The current study adopts a mixed review design that combines the bibliometric analysis with the thematic content analysis. The bibliometric analysis is suitable for the mapping of a large body of the literature in an

objective way, and it includes the performance analysis and the science mapping (Donthu et al., 2021; Zupic & Čater, 2015). However, the bibliometric indicators alone do not explain the content of the findings, and for this reason the bibliometric results were complemented by a thematic synthesis of the key studies (Braun & Clarke, 2006). The reporting of the search and the screening steps followed the PRISMA 2020 statement (Page et al., 2021).

3.2 Data sources and search strategy

The data were collected from two open scholarly databases, which are OpenAlex and Crossref. OpenAlex is an open index of the scholarly works that covers the journals, the conferences and the books, and it provides the information about the authors, the institutions, the countries and the citations (Priem et al., 2022). Crossref is the main registration agency of the digital object identifiers (DOIs), and it provides the metadata deposited by the publishers (Hendricks et al., 2020). The use of the open databases was preferred because it allows other researchers to replicate the search without a subscription, and it reduces the bias toward the journals that are indexed in the commercial databases only. In addition, the coverage of OpenAlex for the recent publications from the emerging economies is wider than the commercial databases. Table 1 summarises the search protocol. In OpenAlex, the search was applied on the titles of the documents using a Boolean string that combines the GenAI terms with the accounting terms. In Crossref, the search engine does not support the Boolean search on the titles, and therefore 25 relevance-ranked queries were run and the first 2,000 results of each query were retrieved, which resulted in 30,996 unique records. Then, an automated filter was applied on the titles of the Crossref records, where the title should contain at least one GenAI term and at least one accounting term. The search covered the period from January 2022 to April 2026, and the data were retrieved on 5 November 2026.

Table 1. Search protocol

Item	Description
Databases	OpenAlex (Priem et al., 2022) and Crossref (Hendricks et al., 2020)
GenAI terms	“generative AI”, “generative artificial intelligence”, GenAI, ChatGPT, GPT-4, “large language model(s)”, LLM
Accounting terms	accounting, accountant(s), auditing, audit, auditor(s), “financial reporting”, plus disclosure, assurance, tax and related terms in the Crossref title filter
Field searched	Title (OpenAlex); relevance-ranked full metadata followed by title filter (Crossref)
Period	January 2022 to April 2026
Document types	Journal articles, reviews, conference papers and book chapters
Excluded types	Preprints (e.g., SSRN, arXiv, Zenodo), peer-review reports, datasets and books
Snowballing	Backward snowballing of the key reviews (Dong et al., 2024; Li et al., 2025)

3.3 Screening and eligibility

Figure 1 presents the flow of the records. The search identified 673 records in OpenAlex and 756 records in Crossref, and the total is 1,429 records. Firstly, 512 records were removed because they are preprints or non-research documents, such as the peer-review reports and the datasets. Secondly, 233 duplicate records between the two databases were removed, and 684 records remained for the screening of the titles and the abstracts. Thirdly, the records were screened according to the inclusion criterion, where the study should examine the GenAI or the LLMs in an accounting, auditing, financial reporting, tax or accounting education context. It is important to note that the word “audit” is used widely in the computer science literature to describe the testing of the AI systems for the bias or the safety, and these studies were excluded because they do not relate to the accounting audit. Similarly, the studies on the payment fraud and the banking fraud were excluded, together with the clinical and the engineering audits. As a result, 366 records were excluded, including five title duplicates, and 318 records were retained. Finally, three highly relevant studies that were

cited by the key reviews and were not captured by the title search were added through the backward snowballing (Bertomeu et al., 2025; de Villiers et al., 2024; Gu et al., 2024). Accordingly, the final corpus consists from 321 documents.

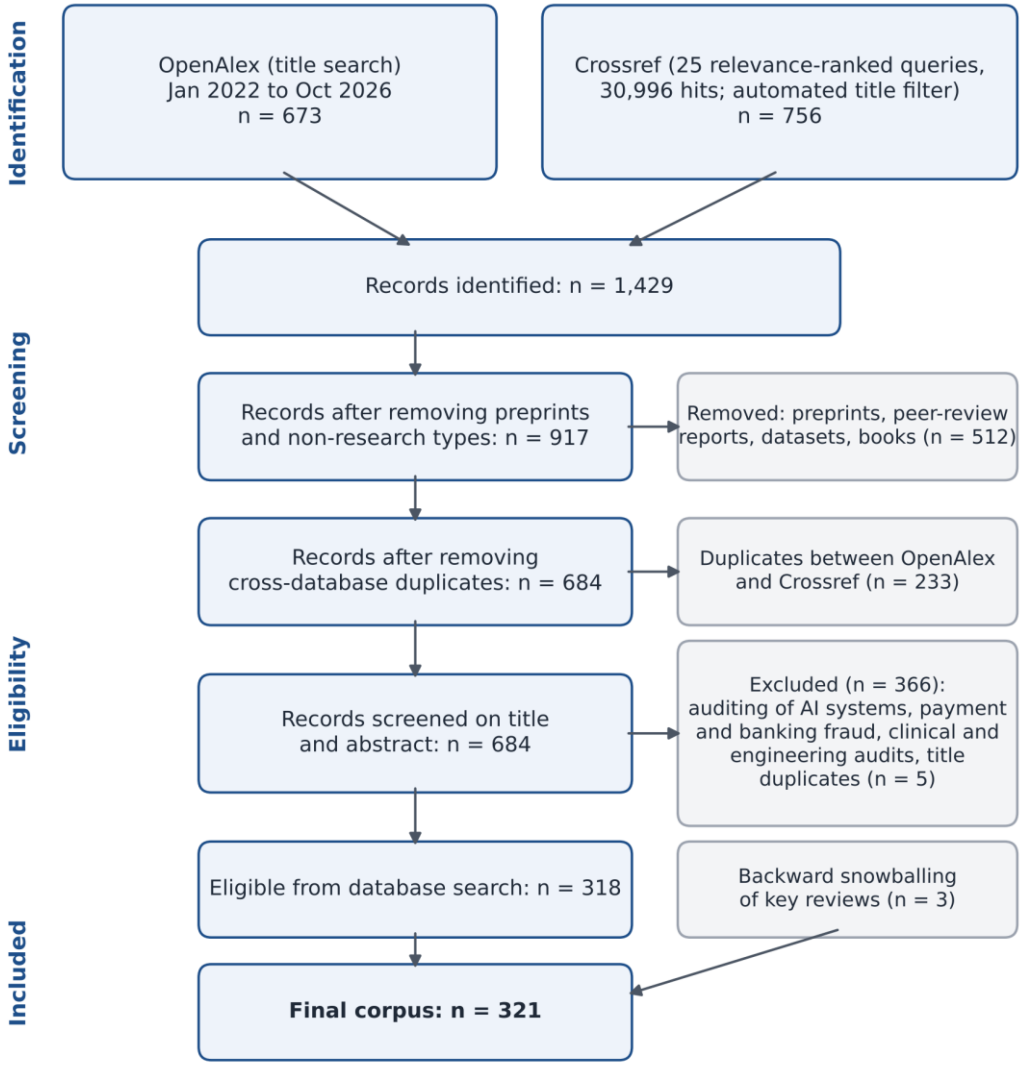


Figure 1. Flow diagram of the identification, screening and inclusion of the records (adapted from PRISMA 2020).

3.4 Data analysis

The bibliometric metadata of the final corpus were enriched through the OpenAlex database by using the DOIs, and 311 documents out of 321 were matched. The performance analysis included the annual production, the document types, the most productive sources, the countries of the authors, the international collaboration and the citations. The citation counts reflect the status of OpenAlex on the retrieval date. The country analysis used the full counting, where a document is counted once for each country that appears in its affiliations. As for the science mapping, the current study used a concept co-occurrence analysis instead of the author keywords, because many of the documents in the corpus do not have the author keywords in the database records, and the available keywords are not standardised. Therefore, a controlled vocabulary of

35 concepts was developed from the reading of the literature, and each concept was defined with a set of the synonyms and the spelling variants. The vocabulary was applied on the titles and the abstracts of the documents, and the abstracts were available for 270 documents. A co-occurrence network was then built for the 32 substantive concepts, where the three technology concepts (GenAI, ChatGPT and LLMs) were excluded because they appear in almost all the documents. The links with a co-occurrence frequency below four were removed, and the Louvain community detection algorithm was used in order to identify the clusters. The analysis was conducted in Python using the NetworkX library, and it follows the same logic of the co-word analysis that is implemented in the common tools such as VOSviewer and bibliometrix (Aria & Cuccurullo, 2017; van Eck & Waltman, 2010). Regarding the thematic analysis, each document was assigned to one primary theme based on its title, and then the assignment was checked manually for the most cited documents. Five primary themes were identified, which are the capability benchmarking, the auditing and assurance, the reporting, disclosure and markets, the education and academic integrity, and the adoption, profession and governance. After that, the most cited and the most recent empirical studies in each theme were read in more detail in order to extract the research questions, the methods and the main findings.

4. Bibliometric Results

4.1 Publication trend and document types

Figure 2 shows the annual production of the documents. The production started with only two documents dated 2022 and 21 documents in 2023, and then it increased to 56 documents in 2024 and 109 documents in 2025. In the first ten months of 2026, 133 documents were published, which means that the production of 2026 already exceeded the production of 2025. In other words, 75.4% of the corpus was published in 2025 and 2026, and this confirms that the GenAI research in accounting is a new and fast-growing research field. It should be noted that the two documents dated 2022 appeared before the public release of ChatGPT, and their dating reflects the metadata deposited by the publishers. Regarding the document types, the journal articles represent the majority with 262 documents, followed by 34 conference papers and 25 book chapters.

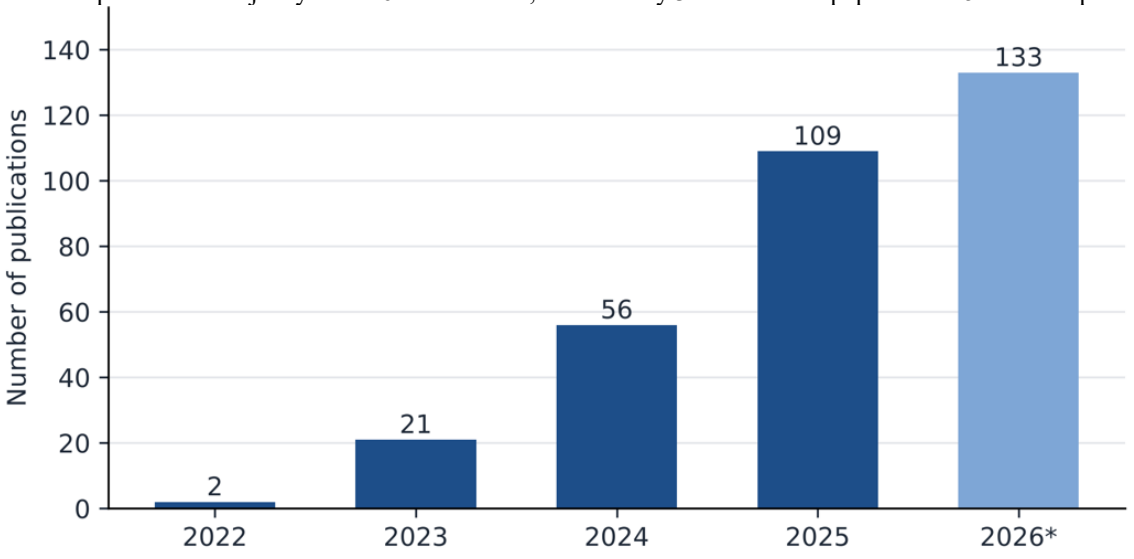


Figure 2. Annual production of the documents on the GenAI in accounting and auditing (n = 321).

4.2 Most productive sources

The documents were published in a large number of the sources, and this indicates that the research field is still dispersed. The most productive source is the Journal of Emerging Technologies in Accounting with nine documents, followed by the Journal of Accounting Education with seven documents, and then Current Issues in Auditing and the International Journal of Accounting Information Systems with four documents

for each one. Moreover, Accounting Horizons, Issues in Accounting Education, Accounting Education and Higher Education, Skills and Work-based Learning published three documents for each one. It is worth mentioning that the most productive sources are the specialised journals of the accounting information systems and the accounting education, while the top general accounting journals such as the Journal of Accounting Research, the Journal of Accounting and Economics and the Review of Accounting Studies published a small number of the documents, but these documents are among the most cited ones (Bertomeu et al., 2025; Blankespoor et al., 2026; Eulerich et al., 2024).

4.3 Countries and collaboration

Out of the 311 documents that were matched in OpenAlex, 255 documents have the information about the country of the authors. Figure 3 shows that the United States is the most productive country with 69 documents (27.1%), followed by China with 54 documents (21.2%), Indonesia with 26 documents, the United Kingdom with 11 documents and Jordan with 10 documents. The presence of Indonesia and Jordan among the top five countries shows that the researchers from the emerging economies are active in this field, and their studies focus mostly on the adoption of the GenAI by the students and the professionals. However, only 42 documents (16.5%) were written by the authors from more than one country, and the average number of the authors per document is 3.03, while 90 documents were written by a single author. This means that the international collaboration in the field is still weak.

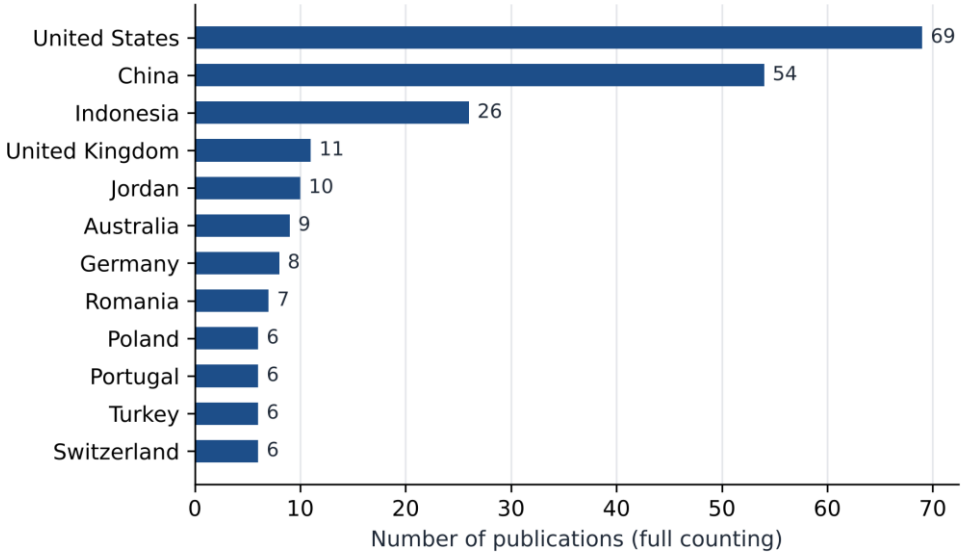


Figure 3. The most productive countries (full counting, n = 255 documents with country information).

4.4 Citation analysis

The 311 documents received 2,671 citations in total, with a mean of 8.59 citations per document and a median of zero. Moreover, 161 documents (51.8%) were not cited yet, and the h-index of the corpus is 26. The citations are highly concentrated, since the ten most cited documents received 1,273 citations, which represent 47.7% of the total citations. Table 2 presents the ten most cited documents. The list shows that the most influential studies are the early studies that tested the capabilities of ChatGPT (Eulerich et al., 2024; Wood et al., 2023), the studies that provided a methodological guidance for the use of the LLMs in the accounting research (de Kok, 2025), the conceptual studies on the sustainability reporting and the auditing (de Villiers et al., 2024; Gu et al., 2024), and the scoping review of Dong et al. (2024). In addition, the study of Bertomeu et al. (2025) is the only archival study with a quasi-natural experiment in the top ten list.

Table 2. The ten most cited documents in the corpus

No.	Document	Source	Type of study	Citations
1	de Villiers et al. (2024)	Sustainability Accounting, Management and Policy Journal	Conceptual framework	183
2	Eulerich et al. (2024)	Review of Accounting Studies	Benchmark experiment	179
3	de Kok (2025)	Management Science	Methodological guidance	175
4	Dong et al. (2024)	International Journal of Accounting Information Systems	Scoping review	157
5	Wood et al. (2023)	Issues in Accounting Education	Benchmark experiment	144
6	Gu et al. (2024)	International Journal of Accounting Information Systems	Design and demonstration	119
7	Alshurafat et al. (2024)	Journal of Financial Reporting and Accounting	Survey (PLS-SEM)	83
8	Zhao and Wang (2024)	Journal of Corporate Accounting and Finance	Conceptual	81
9	Cheng et al. (2024)	Issues in Accounting Education	Benchmark experiment	78
10	Bertomeu et al. (2025)	Journal of Accounting and Economics	Quasi-natural experiment	74

4.5 Conceptual structure

Figure 4 presents the concept co-occurrence network. The most frequent concepts in the corpus are the accounting education (123 documents), the efficiency and productivity (98 documents), the benchmarking of the performance (94 documents) and the financial reporting (84 documents), followed by the external audit and the disclosure and textual analysis (57 documents for each one). On the other side, some concepts that are central for the accounting profession appear in a small number of the documents, such as the professional skepticism (5 documents), the small and medium enterprises (3 documents), the public sector (8 documents) and the continuous auditing (10 documents).

The Louvain algorithm identified four clusters. Cluster 1 (reporting, disclosure and markets) is the largest cluster, and it includes the financial reporting, the disclosure and textual analysis, the capital markets and analysts, the sustainability and ESG reporting, the fraud detection, the tax, and the governance and regulation. Cluster 2 (education, acceptance and the profession) includes the accounting education, the technology acceptance, the skills and competencies, the ethics, the academic integrity, the trust and the management accounting. Cluster 3 (judgment and model reliability) includes the professional judgment, the hallucination and reliability, the retrieval and fine-tuning, the human and AI collaboration, and the prompt engineering. Cluster 4 (auditing and assurance) includes the external audit, the internal audit, the IT audit and controls, the public sector audit, the professional skepticism, and the assessment and examinations. It is interesting that the professional judgment was grouped with the model reliability and the human and AI collaboration rather than with the auditing cluster, and this indicates that the literature discusses the judgment mainly as a question of whether the output of the model can be trusted.

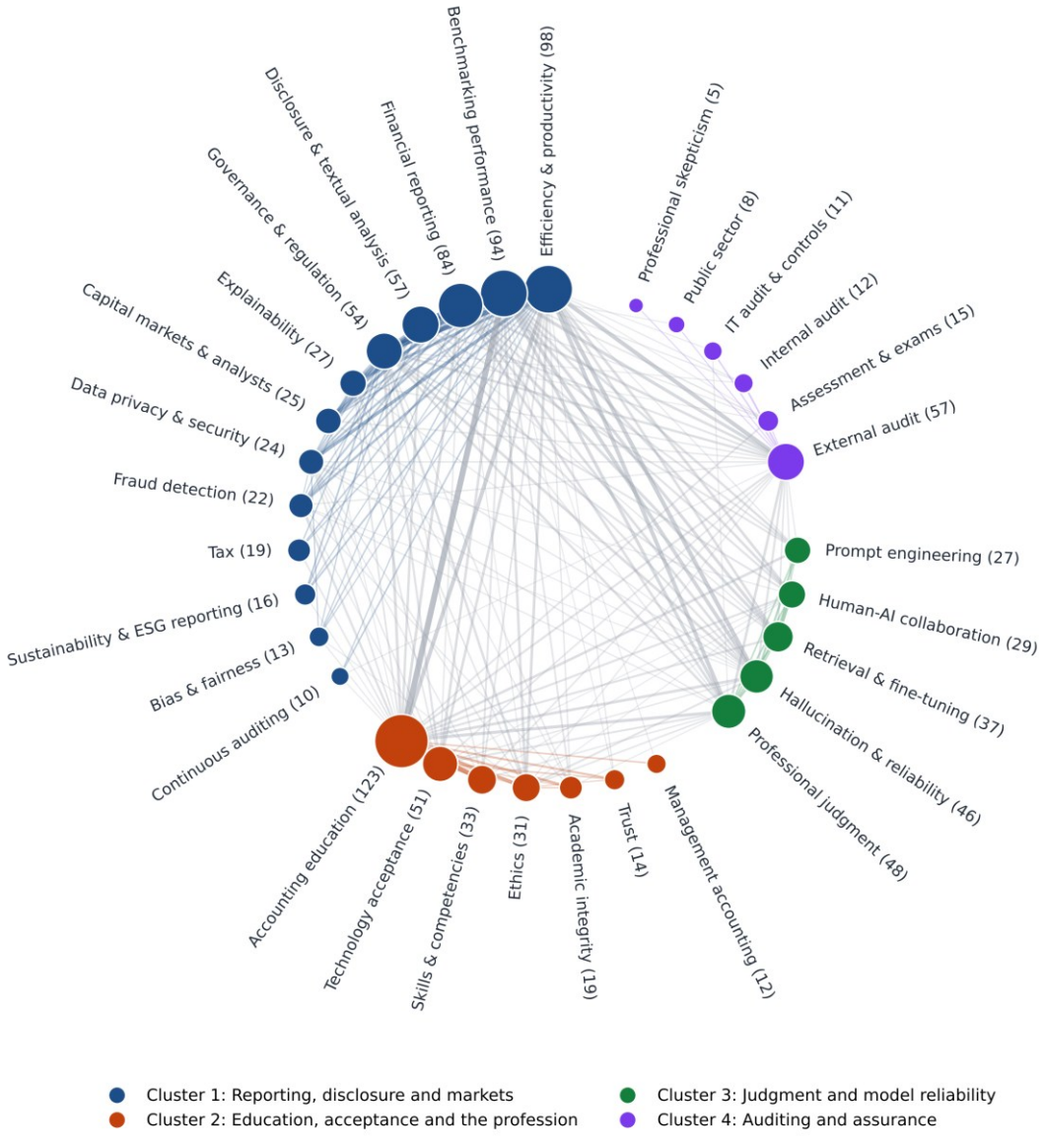


Figure 4. Concept co-occurrence network (node size = document frequency, shown in brackets; links with fewer than four co-occurrences removed; colours = Louvain clusters).

4.6 Thematic evolution

Figure 5 shows the distribution of the five primary themes over the time. The education and academic integrity theme (79 documents) and the auditing and assurance theme (78 documents) are the largest themes, followed by the reporting, disclosure and markets theme (71 documents), the adoption, profession and governance theme (68 documents) and the capability benchmarking theme (25 documents). In 2023, the literature was dominated by the commentaries on the profession and the early tests of the capabilities. Later, in 2025 and 2026, the auditing and assurance theme and the education theme grew faster than the other themes, and the auditing theme alone included 36 documents in the first months of 2026. This development

shows that the field moved from the question of "what the GenAI can do" toward the question of "how the GenAI is used and with what consequences".

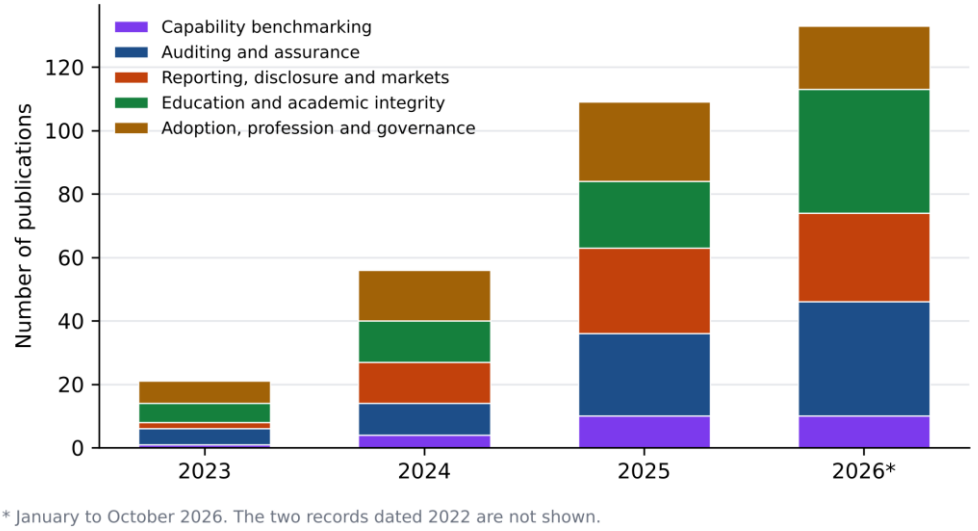


Figure 5. Evolution of the primary themes, 2023 to 2026.

5. Thematic Synthesis of the Literature

5.1 Theme 1: Capability benchmarking of the GenAI on accounting tasks

The first stream of the literature tested how well the GenAI models perform on the accounting tasks, and these studies were the earliest and the most cited studies in the field. Wood et al. (2023) compared the performance of ChatGPT with the performance of the students on 28,085 questions from the accounting assessments and the test banks, which were collected from 186 institutions in 14 countries. The study found that ChatGPT, as of January 2023, provided correct answers for 56.5 percent of the questions and partially correct answers for an additional 9.4 percent, while the students outperformed ChatGPT when the point values of the questions were considered. In the same direction, Cheng et al. (2024) evaluated ChatGPT 3.5 and 4 on seven educational accounting cases, and they found that ChatGPT performed better on the tasks that require the explanation and the application of the rules, but it performed poorly on the tasks that require the preparation of the financial statements and the journal entries.

However, the performance of the models improved quickly with the newer versions and the better prompting techniques. Eulerich et al. (2024) examined whether ChatGPT can pass the major accounting certification examinations, including the CPA, CMA, CIA and EA examinations. The study found that GPT-3.5 could not pass any examination, with an average score of 53.1%, but the upgrade to GPT-4 increased the scores by 16.5% on average, the 10-shot training added 6.6%, and the use of the reasoning and acting tools such as the calculator added 8.9%. As a result, the model with all the enhancements passed all the sections of the examinations with an average score of 85.1%. Similar tests were conducted on the national examinations, such as the Portuguese chartered accountant examination (Albuquerque & Gomes dos Santos, 2024) and the certified internal auditor examination (Żytniewski & Pankowska, 2026), and on the domain benchmarks that are specific for a country, such as the Romanian tax and accounting regulations (Olariu et al., 2026).

Moreover, a group of the studies moved from the multiple-choice questions toward the more realistic tasks that require the judgment. Henry et al. (2025) applied ChatGPT-4 on the Deloitte Trueblood case questions, and they asked the model to provide the support from the relevant FASB standard and to answer the same questions several times in order to test the consistency. Leitner-Hanetseder et al. (2026) used 480 prompts across four task categories, six GPT models, two prompting techniques and two accounting

standards (IFRS and Austrian GAAP) in order to analyse the performance of ChatGPT in the financial accounting judgment tasks. In addition, Albuquerque and Gomes dos Santos (2025) found that ChatGPT recognises the key ethical issues in the accounting dilemmas, but its answers lack the reliability when the professional judgment is required, although the answers are persuasive and well structured. Based on the above, it can be concluded that the benchmarking literature shows a clear improvement in the capabilities of the GenAI on the structured and knowledge-based tasks, but the reliability on the judgment-based tasks is still limited. It should be noted that most of the benchmark studies used the public versions of the models, and the results of these studies become outdated quickly with the release of the new models. Therefore, the value of the future benchmark studies depends on their ability to develop a stable evaluation framework rather than to report a single score at a single point of the time.

5.2 Theme 2: Auditing and assurance

The auditing and assurance theme is one of the two largest themes, and it covers the internal audit, the external audit, the IT audit and the public sector audit. In the internal audit, Eulerich and Wood (2025) demonstrated with specific examples how the GenAI can be used in all the stages of the internal audit process, from the planning to the reporting. Moreover, Emett et al. (2025) described how the internal audit function of Uniper, a large multinational energy company, integrated ChatGPT in the audit preparation, the fieldwork and the reporting tasks, and the function estimated efficiency gains ranging from 50 to 80 percent for various processes based on the initial tests. These two studies are important because they provide an evidence from the practice, but it should be noted that the efficiency gains were estimated by the function itself and were not measured through a controlled design.

In the external audit, the studies focused on the design of the GenAI tools and on the perceptions of the auditors. Gu et al. (2024) proposed the concept of the AI co-piloted auditing, where the auditors and the foundation models collaborate, and they used the chain-of-thought prompting in order to show that the model can identify the transactions according to the instructions of the auditor. Other studies designed the LLM-based tools for the verification of the compliance of the financial reports with the accounting standards (Berger et al., 2023), for the continuous auditing of a governmental payroll system (Li et al., 2026), and for the prediction of the audit opinions by combining the textual information of the MD&A with the financial data (Lu et al., 2025). In the public sector, Pereira et al. (2025) introduced a system that integrates the LLMs in the case analysis of the Brazilian Federal Court of Accounts.

Regarding the perceptions of the auditors, Fotoh and Mugwira (2025) surveyed the auditors from the non-Big Four firms and asked the experienced audit partners to evaluate the responses of ChatGPT on the audit scenarios without knowing the source of the responses. The study found that the LLMs perform efficiently the routine tasks, such as the preparation of the basic working papers, but they struggle to produce the comprehensive audit-specific reports. Moreover, Li and Goel (2026) interviewed 37 Big 4 audit professionals, and they found that the GenAI supports various audit tasks, but the overreliance without verification, the lack of the transparency, the confidentiality risks and the inability to exercise the skepticism may affect the audit quality in a negative way. In addition, Booker et al. (2026) found that the GenAI changes the socialisation of the early-career auditors, since it supports the independent learning and the confidence, but it may not contribute to the real acquisition of the skills.

The professional skepticism is a central concept in the auditing, but it is examined in a small number of the studies. Grey and Li (2026) found that ChatGPT was most skeptical when it conducted the preliminary assessments and the substantive analytical testing, while it exhibited lower levels of the skepticism in the going concern assessments and when it was asked to act as a junior auditor. In the Indonesian context, Amarullah et al. (2026) surveyed 102 auditors and found that the perceived transparency, the explainability and the social influence increase the trust of the auditors in the LLMs, and Harahap and Fahmi (2026) found that the data governance has a significant positive effect on the readiness of 248 external auditors for the GenAI-based audits ($\beta = 0.47$, $p < 0.001$), and that the AI literacy partially mediates this relationship. Finally, the most recent studies started to use the archival data in order to examine the consequences of the GenAI on the audit outcomes. Du et al. (2026) used the accessibility of ChatGPT as a quasi-natural experiment in

China in order to examine its effect on the audit quality. Moreover, Zhang (2026) constructed a textual indicator of the GenAI application for the Chinese listed firms during the period from 2015 to 2024, and the study found that the application of the GenAI by the client firms increases the audit fees, which is explained by the higher audit risk premium. These results are consistent with the earlier evidence of Estep et al. (2024), who found that the effect of the AI use by the auditor depends on whether the client company also uses the AI. In other words, the GenAI is not only a tool of the auditor, but it is also a source of a new risk in the client environment, and this dual role is not studied enough until now.

5.3 Theme 3: Financial reporting, disclosure and capital markets

The third theme examines the use of the GenAI in the production and the processing of the accounting information. On the side of the preparers, Blankespoor et al. (2026) examined the earnings press releases, the conference call remarks, the risk factors, the MD&As and the IPO filings until 2024, and they found a statistically significant usage of the GenAI in all the five types of the disclosures, with up to 4.5% of the new text written by the GenAI in 2024. Moreover, the study found that the filings with a higher usage of the GenAI have different linguistic properties. As for the sustainability reporting, de Villiers et al. (2024) developed a conceptual framework on how the AI text generation affects the sustainability reporting, the assurance and the use of the reports, and they warned from the risk that the GenAI may facilitate the greenwashing. In the same context, Brune and Hinsberger (2026) conducted an online experiment with 96 participants in Germany, and they found that the sustainability reports co-created with the AI were perceived as credible as the human-written reports, while the fully automated reports were rated significantly lower.

On the side of the users, the GenAI changes the way in which the investors and the analysts process the information. Bertomeu et al. (2025) used the ban of ChatGPT in Italy as a natural experiment, and they found that the ban coincided with a lower AI usage by the Italian financial analysts and fewer earnings forecasts compared with the foreign analysts that cover the same firms. Moreover, the forecast accuracy declined and the bid-ask spreads widened, which indicates a lower market efficiency. In addition, the working papers of Kim et al. (2023, 2024) showed that the summaries of ChatGPT are much shorter than the original disclosures and explain the market reactions better, and that GPT-4 with a chain-of-thought prompt predicted the direction of the future earnings with an accuracy of 60.35%, compared with 52.71% for the analysts' consensus forecasts one month after the earnings release. Although these working papers were not included in the corpus because they are preprints, they are highly influential in the literature.

A third group of the studies examined the GenAI as a tool for the extraction and the analysis of the financial information. Balsiger et al. (2024) assessed the ability of ChatGPT-4 and Bard to extract the data from the tables of the annual financial reports in the PDF format, and Han et al. (2024) found that the LLMs have limitations in the comprehension of the financial domain knowledge and the mathematical calculations when they analyse the XBRL reports, and they proposed the use of the external tools in order to solve these problems. Moreover, Bronzini et al. (2024) used the LLMs in order to derive the structured insights from the sustainability reports, and Wu et al. (2025) compared the scores of three LLMs with the scores of the human experts for the sustainability reports of the European football clubs.

As for the accounting research, de Kok (2025) argued that the generative LLMs can solve any textual analysis task that was solvable by the non-generative methods, and he provided a framework for the selection of the model, the prompt engineering and the construct validity. In a case study on the detection of the non-answers in the earnings conference calls, the GPT-based method achieved 96% accuracy and reduced the error rate by 70% compared with an established method from the earlier literature. However, d'Arcy et al. (2026) compared GPT-4o with the topic modelling and the human research assistants in the classification of the analyst reports, and they found that GPT-4o has a lower reliability than the human assistants because of the risk of the topic over-classification. Similarly, Jang and Wu (2026) applied the LLMs on the non-English textual analysis of the MD&A sentiment. These studies show that the LLMs are powerful research tools, but the validity of the measures derived from them should be tested and reported carefully.

5.4 Theme 4: Accounting education and academic integrity

The accounting education theme is the largest theme in the corpus, and this is not surprising since the students were among the first users of ChatGPT. The studies on the students' adoption showed that the use of the GenAI is already widespread. Zhou and Luo (2025) surveyed 259 accounting students and 126 non-accounting students in a public university in the United States, and they found that 80% of the students use the AI at least about once a week, primarily for the help with the difficult concepts. Moreover, the accounting students were less likely than the other students to accept the AI outputs without questioning their accuracy. Sundkvist and Kulset (2024) found that the perceived usefulness of ChatGPT increases with the enthusiasm of the students, the ease of use, the trust and the friends, and several studies from Jordan, Poland and Indonesia applied the extended TAM and UTAUT models in order to explain the intention of the students to use ChatGPT (Al-Okaily, 2025; Ziemba et al., 2024).

On the other side, the academic integrity represents the main concern of the educators. Alshurafat et al. (2024) applied the fraud triangle theory on 279 accounting students in the Jordanian public universities, and they found that the opportunity, the rationalisation and the pressure are all significant determinants of the misuse of ChatGPT for the academic dishonesty. Moreover, Al-Hajaya (2026) interviewed 17 accounting instructors in the UK universities, and the study discussed the threats of the GenAI to the academic integrity and the strategies to overcome them. In addition, Dosumu et al. (2026) used a collaborative autoethnography in order to reflect on the marking of an open-book online accounting examination in January 2023, and Abeysekera (2024) examined whether ChatGPT can act as a student agent in the multiple-choice assessments of the financial accounting courses.

Regarding the learning outcomes, the evidence is mixed. Calderon et al. (2023) found that the financial accounting students modified their answers in a beneficial way after the use of a GenAI tool, and that the students with more prior knowledge benefited more from the experience. Tharapos et al. (2025) evaluated the performance of ChatGPT on 18 assessments in six courses, and they found that ChatGPT is effective in the lower and middle order learning tasks according to the revised Bloom's taxonomy, but it has a limited utility in the higher order tasks. Moreover, Blondeel et al. (2025) examined the effects of the GenAI on the learning in an accounting data analytics course, and Xiong and Bao (2025) proposed the integration of the GenAI in the digital role-playing games in order to increase the engagement of the students. It is worth mentioning that the education theme is also connected with the professional judgment, since Sugahara and Kano (2026) used an experiment with the accounting majors in order to examine how the GenAI advice, compared with the advice of a human auditor, influences the ethical judgment.

5.5 Theme 5: Adoption, the accounting profession and governance

The fifth theme includes the studies on the adoption of the GenAI by the accounting professionals, the future of the profession, and the ethical and governance issues. Regarding the adoption, Jabin (2025) collected the data from 450 accountants in Bangladesh, and the study found that the performance expectancy, the effort expectancy, the social influence and the reliability perception affect the intention to use the GenAI. In Morocco, Aziki (2025) surveyed 262 chartered accountants and found that the effort expectancy ($\beta = 0.538$) and the social influence ($\beta = 0.498$) are the primary determinants of the GenAI acceptability, while the model explains 40.4% of the variance. Moreover, Toumeh (2024) conducted 13 interviews with the accounting professionals in Jordan in order to explore the potential integration of the LLMs in the accounting practices, and Aku et al. (2026) interviewed 18 tax professionals and found that the GenAI is used as a reference and a co-pilot tool for the tax research and the compliance, rather than to automate the tax judgments.

As for the future of the profession, the early commentaries discussed the potential applications and challenges of ChatGPT in the accounting field (Vasarhelyi et al., 2023; Zhao & Wang, 2024), and Li and Vasarhelyi (2024) compared the different methods of the LLM use, including the user interface, the application programming interface and the robotic process automation. Moreover, Anica-Popa et al. (2024) identified the knowledge, the skills and the competencies that the accounting and audit professionals need in the GenAI context. These studies agree that the routine tasks will be automated, and that the value of the

accountant will depend more on the judgment, the communication and the ability to verify the outputs of the AI. This expectation is consistent with the general evidence on the productivity effects of the GenAI in the knowledge work (Brynjolfsson et al., 2025; Noy & Zhang, 2023). Regarding the ethics and the governance, the studies raised several concerns. Leong and Sung (2024) found that the LLMs assign the gender labels to the accounting job titles in different ways, and that the job titles classified as male are associated with a higher salary, which shows that the GenAI may reproduce the gender stereotypes in the profession. Moreover, Muzanenhamo and Power (2024) argued from a critical perspective that ChatGPT may amplify the epistemic injustice in the African contexts, because it reproduces the Western-centric accounting knowledge. At the regulatory level, the PCAOB (2024) reported that the audit firms use the GenAI mainly for the administrative and research tasks, and that the firms emphasise the need for a strong human supervision and the protection of the data privacy. These issues are cross-cutting, since the concepts of the governance, the ethics, the privacy and the bias appear in all the four clusters of the co-occurrence network.

5.6 Summary of the themes

Table 3 summarizes the five themes, the dominant methods and the main findings. It can be seen that the methods differ between the themes, where the benchmarking theme relies on the experiments with the models, the education and adoption themes rely on the surveys and the structural equation modelling, and the auditing theme relies on the case studies, the interviews and the design science. In contrast, the archival studies that examine the economic consequences of the GenAI are concentrated in the reporting and markets theme and in the recent auditing studies.

Table 3. Summary of the themes, methods and main findings

Theme (n)	Dominant methods	Main findings	Key studies
Capability benchmarking (25)	Experiments on exams, cases and benchmarks	Fast improvement on structured tasks; passing of certification exams with GPT-4 and tools; weak reliability on judgment tasks	Wood et al. (2023); Eulerich et al. (2024); Cheng et al. (2024); Leitner-Hanetseder et al. (2026)
Auditing and assurance (78)	Case studies, interviews, design science, early archival studies	Efficiency gains in internal audit; co-pilot designs; risks of overreliance, confidentiality and weak skepticism; client GenAI use raises audit risk and fees	Emett et al. (2025); Gu et al. (2024); Li and Goel (2026); Grey and Li (2026); Zhang (2026)
Reporting, disclosure and markets (71)	Archival studies, natural experiments, NLP evaluations	GenAI already used in disclosures; GenAI improves analysts' information processing; LLMs are powerful but need validity checks	Blankespoor et al. (2026); Bertomeu et al. (2025); de Kok (2025); de Villiers et al. (2024)
Education and academic integrity (79)	Surveys (TAM/UTAUT, PLS-SEM), qualitative studies, classroom experiments	Widespread student use; integrity threats explained by the fraud triangle; benefits on lower-order learning, limits on higher-order learning	Zhou and Luo (2025); Alshurafat et al. (2024); Tharapos et al. (2025); Calderon et al. (2023)
Adoption, profession and governance (68)	Surveys, interviews, commentaries, critical studies	Effort expectancy and social influence drive adoption; GenAI used as a co-pilot; bias, epistemic injustice and privacy concerns	Aziki (2025); Jabin (2025); Aku et al. (2026); Leong and Sung (2024); Muzanenhamo and Power (2024)

6. Discussion

6.1 An integrative framework

Based on the bibliometric clusters and the thematic synthesis, the current study proposes an integrative framework that organises the literature into the enablers, the mechanisms and the outcomes of the GenAI use in accounting and auditing (Figure 6). The enablers include the capability of the models and their adaptation to the accounting domain through the retrieval-augmented generation, the fine-tuning and the prompting, the data governance and the AI literacy, the perceptions of the users as explained by the acceptance models, the regulation and the professional guidance, and the redesign of the curriculum and the assessment. These enablers activate five mechanisms, which are the automation of the routine tasks, the augmentation of the professional judgment through the co-pilot and human-in-the-loop designs, the improvement of the information processing of the disclosures, the calibration of the trust and the professional skepticism, and the cognitive offloading.

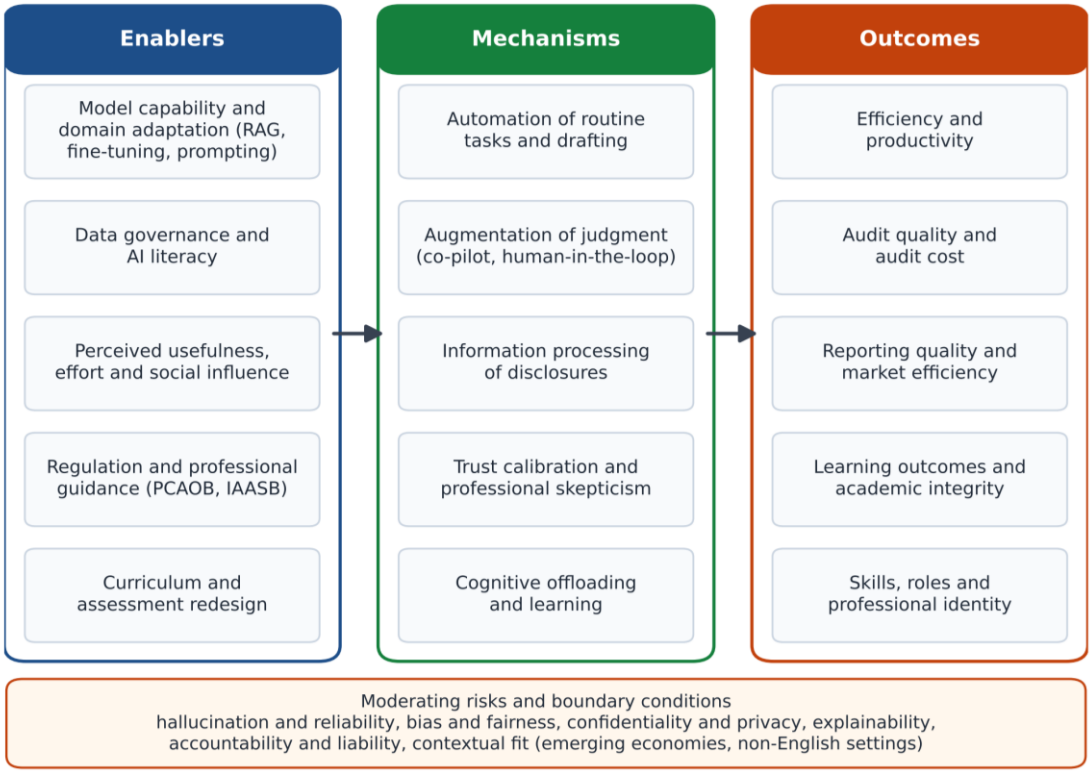


Figure 6. Integrative framework of the GenAI in accounting and auditing.

The mechanisms lead to five groups of the outcomes, namely the efficiency and productivity, the audit quality and the audit cost, the reporting quality and the market efficiency, the learning outcomes and the academic integrity, and the skills and the professional identity. Moreover, the relationships in the framework are moderated by a set of the risks and the boundary conditions, and these include the hallucination, the bias, the confidentiality, the explainability, the accountability and the contextual fit of the models, especially in the emerging economies and the non-English settings. It should be noted that the framework shows that the same mechanism may lead to a positive or a negative outcome. For example, the cognitive offloading may increase the efficiency in the short term, but it may reduce the learning of the students and the junior auditors in the long term (Booker et al., 2026; Risko & Gilbert, 2016). Similarly, the trust in the GenAI may increase

its use, but the excessive trust may weaken the professional skepticism (Amarullah et al., 2026; Parasuraman & Manzey, 2010).

6.2 Research gaps and future research agenda

The analysis of the literature revealed several gaps, and these gaps were translated into a future research agenda in Table 4. The first gap is the lack of the causal and longitudinal evidence. Most of the studies in the corpus are commentaries, cross-sectional surveys or benchmark experiments, while the archival and quasi-experimental studies are few (Bertomeu et al., 2025; Blankespoor et al., 2026; Du et al., 2026). Therefore, the future research should use the natural experiments, such as the staggered access to the GenAI tools inside the audit firms, and the panel data in order to examine the effect of the GenAI on the audit quality, the financial reporting quality and the cost of the capital. The second gap relates to the professional skepticism and the judgment. Although the judgment is one of the most frequent concepts, the professional skepticism appears in only five documents. Accordingly, the future studies should use the behavioural experiments with the practising auditors in order to examine whether the GenAI advice increases or decreases the skepticism, and how the explainability of the output affects the reliance (Commerford et al., 2022; Grey & Li, 2026). The third gap is the narrow geographical and contextual coverage. The international collaboration is weak, and the studies from the emerging economies focus mostly on the adoption intentions rather than on the actual use and its consequences. Moreover, the small and medium enterprises and the public sector are rarely studied, although the GenAI may be most useful for the firms that do not have the resources to employ a large accounting staff.

The fourth gap relates to the sustainability reporting and the assurance. The sustainability concepts appear in 16 documents only, while the new sustainability reporting standards increase the demand for the processing of a large amount of the narrative information. The future research should examine whether the GenAI improves the quality of the sustainability disclosures or facilitates the greenwashing, and how the assurance providers can verify the AI-generated sustainability information (Brune & Hinsberger, 2026; de Villiers et al., 2024). This gap is especially relevant for the research on the fintech sustainability and the digital accounting. The fifth gap is the methodological validity of the use of the LLMs in the accounting research itself, since the measures derived from the LLMs may have a lower reliability than the human coding (d'Arcy et al., 2026; de Kok, 2025). Finally, the governance of the GenAI inside the accounting firms, including the documentation of the AI use in the audit files, the liability and the regulation, needs more empirical studies (PCAOB, 2024).

Table 4. Future research agenda

Research area	Gap identified	Suggested research questions	Suggested methods and theories
Audit quality and audit risk	Few causal studies; dual role of GenAI (auditor tool and client risk)	Does GenAI use by audit firms affect audit quality and audit fees? How does client GenAI use change the audit risk assessment?	Quasi-natural experiments, difference-in-differences, panel data; audit risk model
Professional skepticism and judgment	Skepticism studied in only five documents	How does GenAI advice, and its explainability, influence auditors' skepticism and reliance?	Behavioural experiments with practitioners; algorithm aversion and automation bias
Financial reporting and disclosure	Early evidence on GenAI-written disclosures	Does GenAI-assisted drafting improve or reduce disclosure quality, readability and informativeness?	Textual analysis, archival data; signalling and disclosure theories

Research area	Gap identified	Suggested research questions	Suggested methods and theories
Sustainability reporting and assurance	Only 16 documents on sustainability and ESG	Does GenAI facilitate greenwashing or improve ESG disclosure? How can assurance providers verify AI-generated content?	Mixed methods, experiments with users; legitimacy and stakeholder theories
Education and skill formation	Mixed evidence on learning; integrity concerns	How does GenAI affect higher-order learning and the long-term competence of graduates?	Longitudinal and controlled classroom designs; cognitive offloading and cognitive load theories
Emerging economies, SMEs and public sector	Weak international collaboration; focus on intentions	How do SMEs and public sector entities use GenAI for accounting, and with what effects on compliance and accountability?	Multi-country studies, case studies; TOE framework, institutional theory
Research methods	Validity of LLM-derived measures	How reliable and replicable are LLM-based measures compared with human coding?	Validation studies, replication; construct validity frameworks
Governance and regulation	Limited empirical evidence on firm-level AI governance	How do firms document, supervise and take responsibility for GenAI use in reporting and audit?	Interviews, surveys of firms and regulators; accountability theory

7. Conclusion

The current study reviewed 321 documents on the GenAI in accounting and auditing that were published during the period from January 2022 to April 2026, and it combined the bibliometric analysis with the thematic synthesis. The results showed that the research field grew very fast, where three quarters of the documents were published in 2025 and 2026, and that the production is dominated by the United States and China, while the international collaboration is weak. The concept co-occurrence analysis revealed four clusters, which are the reporting, disclosure and markets, the education, acceptance and the profession, the judgment and model reliability, and the auditing and assurance. Moreover, the thematic synthesis showed that the literature moved from the testing of the capabilities of ChatGPT toward the study of its use in the practice and its consequences for the audit quality, the disclosures and the market efficiency.

7.1 Implications

The study has several implications. For the researchers, the study provides a map of the field and a research agenda that directs the future studies toward the causal designs, the professional skepticism, the sustainability reporting and the emerging economies. For the audit firms and the accounting departments, the findings suggest that the GenAI provides efficiency gains in the routine tasks, but the firms need to establish clear policies for the verification of the outputs, the protection of the confidential data and the documentation of the AI use. For the standard setters and the regulators, the findings indicate that the existing standards need guidance on the use of the GenAI in the audit evidence and in the preparation of the disclosures. For the accounting educators, the findings show that the assessment should be redesigned in order to focus on the higher order skills and on the critical evaluation of the AI outputs, rather than on the tasks that the GenAI can perform easily.

7.2 Limitations

This paper has some limitations. Firstly, the search was based on two open databases, and although these databases have a wide coverage, some relevant documents that are indexed in the commercial databases only may be missed. Moreover, the search in OpenAlex was conducted on the titles, and the Crossref search

depended on the relevance ranking of the search engine, which means that some studies that do not mention the GenAI or the accounting terms in their titles may be missed. Secondly, the preprints were excluded in order to focus on the peer-reviewed literature, but some influential working papers in this field are still preprints. Thirdly, the citation counts change quickly in a new field, and the recent documents had less time to receive the citations. Finally, the concept vocabulary and the assignment of the themes involve a degree of the subjective judgment, although the vocabulary was applied in a transparent and replicable way. The future reviews can address these limitations by using additional databases and by updating the analysis as the field develops.

References

Abeysekera, I. (2024). ChatGPT and academia on accounting assessments. Journal of Open Innovation Technology Market and Complexity, 10(1), 100213. https://doi.org/10.1016/j.joitmc.2024.100213

Aku, H., Parker, C., & Stone, C. W. (2026). Applications of generative artificial intelligence in tax practices: Evidence from tax professionals. Accounting Horizons. Advance online publication. https://doi.org/10.2308/horizons-2025-200

Al-Hajaya, K. (2026). Academic integrity is under fire in the generative AI age: Insights from accounting educators to overcome challenges, threats and ethical concerns. Higher Education Skills and Work-based Learning, 16(1), 237–255. https://doi.org/10.1108/heswbl-02-2025-0060

Al-Okaily, M. (2025). ChatGPT as an educational resource for accounting students: Expanding the classical TAM model. Education and Information Technologies, 30(12), 16671–16685. https://doi.org/10.1007/s10639-025-13391-1

Albuquerque, F., & Gomes dos Santos, P. (2024). Can ChatGPT be a certified accountant? Assessing the responses of ChatGPT for the professional access exam in Portugal. Administrative Sciences, 14(7), 152. https://doi.org/10.3390/admsci14070152

Albuquerque, F., & Gomes dos Santos, P. (2025). Can ChatGPT be a knowledgeable and accurate ethical accountant? Assessing the overall usefulness of ChatGPT's answers to ethical dilemmas. Journal of Accounting Education, 73, 101000. https://doi.org/10.1016/j.jaccedu.2025.101000

Alshurafat, H., Al Shbail, M. O., Hamdan, A., Al-Dmour, A., & Ensour, W. (2024). Factors affecting accounting students' misuse of ChatGPT: An application of the fraud triangle theory. Journal of Financial Reporting and Accounting, 22(2), 274–288. https://doi.org/10.1108/jfra-04-2023-0182

Amarullah, K., Ritchi, H., & Mubarrok, A. Z. (2026). When trust sharpens skepticism: Large language models in Indonesian audit practice. Dinasti International Journal of Economics Finance & Accounting, 7(3), 2382–2400. https://doi.org/10.38035/dijefa.v7i3.7181

Anica-Popa, I., Vrîncianu, M., Anica-Popa, L.-E., Cișmașu, I.-D., & Tudor, C. G. (2024). Framework for integrating generative AI in developing competencies for accounting and audit professionals. Electronics, 13(13), 2621. https://doi.org/10.3390/electronics13132621

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics, 11(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Aziki, A. (2025). Empirical investigation of generative artificial intelligence acceptability determinants among Moroccan chartered accountants: A structural equation modeling approach. International Journal of Research in Economics and Finance, 2(11), 1–17. https://doi.org/10.71420/ijref.v2i11.210

Balsiger, D., Dimmler, H.-R., Egger-Horstmann, S., & Hanne, T. (2024). Assessing large language models used for extracting table information from annual financial reports. Computers, 13(10), 257. https://doi.org/10.3390/computers13100257

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

Berger, A., Hillebrand, L., Leonhard, D., Deußer, T., De Oliveira, T. B. F., Dilmaghani, T., Khaled, M. B., Kliem, B., Loitz, R., Bauckhage, C., & Sifa, R. (2023). Towards automated regulatory compliance verification in financial auditing with large language models. In 2023 IEEE International Conference on Big Data (BigData) (pp. 4626–4635). IEEE. https://doi.org/10.1109/bigdata59044.2023.10386518

Bertomeu, J., Lin, Y., Liu, Y., & Ni, Z. (2025). The impact of generative AI on information processing: Evidence from the ban of ChatGPT in Italy. Journal of Accounting and Economics, 80(1), 101782. https://doi.org/10.1016/j.jacceco.2025.101782

Blankespoor, E., deHaan, E., & Li, Q. (2026). Generative AI in financial reporting. Journal of Accounting Research, 64(3), 1189–1232. https://doi.org/10.1111/1475-679x.70050

Blondeel, E., Bullock, T., Gaskin, J. E., Schuetzler, R. M., Serre, R., Steffen, J., Wells, T., & Wood, D. A. (2025). The effects of generative artificial intelligence (GenAI) on learning in an accounting data analytics course. Journal of Accounting Education, 72, 100987. https://doi.org/10.1016/j.jaccedu.2025.100987

Booker, D. D., Pelzer, J. R. E., & Wu, Y.-J. (2026). From potential to practice: How GenAI shapes early-career auditors' work and socialization. Accounting Horizons. Advance online publication. https://doi.org/10.2308/horizons-2025-192

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bronzini, M., Nicolini, C., Lepri, B., Passerini, A., & Staiano, J. (2024). Glitter or gold? Deriving structured insights from sustainability reports via large language models. EPJ Data Science, 13(1). https://doi.org/10.1140/epjds/s13688-024-00481-2

Brune, N., & Hinsberger, B. (2026). How stakeholders perceive generative AI in sustainability reports: Assistance or interference? In Proceedings of the 59th Hawaii International Conference on System Sciences. https://doi.org/10.24251/hicss.2026.104

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at work. The Quarterly Journal of Economics, 140(2), 889–942. https://doi.org/10.1093/qje/qjae044

Calderon, T. G., Gao, L., & Cardoso, R. L. (2023). Generative artificial intelligence in the classroom: A financial accounting experience. In Advances in accounting education: Teaching and curriculum innovations (Vol. 27, pp. 125–144). Emerald Publishing. https://doi.org/10.1108/s1085-462220230000027006

Cheng, X., Dunn, R. T., Holt, T. P., Inger, K. K., Jenkins, J. G., Jones, J. P., Long, J. H., Loraas, T. M., Mathis, M. E., Stanley, J. D., & Wood, D. A. (2024). Artificial intelligence's capabilities, limitations, and impact on accounting education: Investigating ChatGPT's performance on educational accounting cases. Issues in Accounting Education, 39(2), 23–47. https://doi.org/10.2308/issues-2023-032

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

d'Arcy, A., Haas, C. S., & Wallner, N. (2026). Large language models as research assistants: The risk of topic-overclassification in financial disclosure research. The British Accounting Review. Advance online publication. https://doi.org/10.1016/j.bar.2026.101986

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

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

Dong, 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, 100715. https://doi.org/10.1016/j.accinf.2024.100715

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. https://doi.org/10.1016/j.jbusres.2021.04.070

Dosumu, O. E., Porumb, V., Stafford, A., & Zimmer, A. (2026). In the wake of ChatGPT: Early reflections on marking open-book online accounting assessments. Accounting Education, 35(4), 425–456. https://doi.org/10.1080/09639284.2025.2487487

Du, X., Zhang, X., & Lai, S. (2026). ChatGPT accessibility and audit quality: Evidence from a quasi-natural experiment in China. The British Accounting Review. Advance online publication. https://doi.org/10.1016/j.bar.2026.101950

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Emett, S. A., Eulerich, M., Lipinski, E., Prien, N., & Wood, D. A. (2025). Leveraging ChatGPT for enhancing the internal audit process: A real-world example from Uniper, a large multinational company. Accounting Horizons, 39(2), 125–135. https://doi.org/10.2308/horizons-2023-111

Estep, C., Griffith, E. E., & MacKenzie, N. L. (2024). How do financial executives respond to the use of artificial intelligence in financial reporting and auditing? Review of Accounting Studies, 29(3), 2798–2831. https://doi.org/10.1007/s11142-023-09771-y

Eulerich, M., & Wood, D. A. (2025). A demonstration of how ChatGPT and generative AI can be used in the internal auditing process. Journal of Emerging Technologies in Accounting, 22(2), 47–77. https://doi.org/10.2308/jeta-2023-041

Eulerich, M., Sanatizadeh, A., Vakilzadeh, H., & Wood, D. A. (2024). Is it all hype? ChatGPT's performance and disruptive potential in the accounting and auditing industries. Review of

Accounting Studies, 29(3), 2318–2349. https://doi.org/10.1007/s11142-024-09833-9

Fotoh, L. E., & 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

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. Academy of Management Annals, 14(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Grey, L., & Li, Y. (2026). Can ChatGPT be skeptical? A case-based evaluation of generative AI in financial statement auditing. Accounting Research Journal. Advance online publication. https://doi.org/10.1108/arj-12-2025-0527

Gu, H., Schreyer, M., Moffitt, K. C., & 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

Han, S., Kang, H., Jin, B., Liu, X.-Y., & Yang, S. Y. (2024). XBRL agent: Leveraging large language models for financial report analysis. In Proceedings of the 5th ACM International Conference on AI in Finance (pp. 856–864). ACM. https://doi.org/10.1145/3677052.3698614

Harahap, R. U., & Fahmi, M. (2026). Data governance and auditor readiness for generative AI-based audits: A survey study of Indonesian external auditors with AI literacy as a mediator. Ilomata International Journal of Tax and Accounting, 7(4), 1–10. https://doi.org/10.61194/ijtc.v7i4.2497

Hendricks, G., Tkaczyk, D., Lin, J., & Feeney, P. (2020). Crossref: The sustainable source of community-owned scholarly metadata. Quantitative Science Studies, 1(1), 414–427. https://doi.org/10.1162/qss_a_00022

Henry, T. F., Rosenthal, D. A., & Weitz, R. R. (2025). ChatGPT-4: Can it handle real-world accounting cases? Journal of Emerging Technologies in Accounting, 22(2), 97–108. https://doi.org/10.2308/jeta-2024-028

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. Contemporary Accounting Research, 40(2), 806–841. https://doi.org/10.1111/1911-3846.12832

Jabin, S. (2025). The extended unified theory of adoption and use of generative artificial intelligence in accounting profession: Evidence from emerging economy. Jahangirnagar University Journal of Business Research, 25(01), 93–112. https://doi.org/10.53461/jujbr.v25i01.92

Jang, J., & Wu, X. (2026). Non-English textual analysis with large language models: Analysts' use of MD&A sentiment in earnings forecasting. Journal of Contemporary Accounting & Economics, 22(1), 100524. https://doi.org/10.1016/j.jcae.2025.100524

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38. https://doi.org/10.1145/3571730

Kim, A. G., Muhn, M., & Nikolaev, V. V. (2023). Bloated disclosures: Can ChatGPT help investors process information? (arXiv:2306.10224) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2306.10224

Kim, A. G., Muhn, M., & Nikolaev, V. V. (2024). Financial statement analysis with large language models (arXiv:2407.17866) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2407.17866

de Kok, T. (2025). ChatGPT for textual analysis? How to use generative LLMs in accounting research. Management Science, 71(9), 7888–7906. https://doi.org/10.1287/mnsc.2023.03253

Kokina, J., & Davenport, T. H. (2017). The emergence of artificial intelligence: How automation is changing auditing. Journal of Emerging Technologies in Accounting, 14(1), 115–122. https://doi.org/10.2308/jeta-51730

Lehner, O. M., Ittonen, K., Silvola, H., Ström, E., & Wührleitner, A. (2022). Artificial intelligence based decision-making in accounting and auditing: Ethical challenges and normative thinking. Accounting, Auditing & Accountability Journal, 35(9), 109–135. https://doi.org/10.1108/AAAJ-09-2020-4934

Leitner-Hanetseder, S., Perkhofer, L., Frenkenberger, S., & Eisl, C. (2026). ChatGPT as digital accounting assistant: Evaluating output performance in financial accounting tasks. International Journal of Accounting and Information Management, 34(2), 330–361. https://doi.org/10.1108/ijaim-01-2025-0017

Leong, K., & Sung, A. (2024). Gender stereotypes in artificial intelligence within the accounting profession using large language models. Humanities and Social Sciences Communications, 11(1). https://doi.org/10.1057/s41599-024-03660-8

Li, H., & Vasarhelyi, M. A. (2024). Applying large language models in accounting: A comparative analysis of different methodologies and off-the-shelf examples. Journal of Emerging Technologies in Accounting, 21(2), 133–152. https://doi.org/10.2308/jeta-2023-065

Li, H., de Freitas, M. M., Lee, H., & Vasarhelyi, M. A. (2026). Enhancing continuous auditing with large language models: AI-assisted real-time accounting information cross-verification. International Journal of Accounting Information Systems. Advance online publication. https://doi.org/10.1016/j.accinf.2026.100786

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. Advance online publication. https://doi.org/10.1108/jal-12-2024-0357

Li, Y., & Goel, S. (2026). Generative artificial intelligence in the Big 4: Auditor adoption and its implications for audit quality. Journal of Information Systems. Advance online publication. https://doi.org/10.2308/isys-2024-069

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Lu, L. Y., Hao, J.-X., & Tang, X. (2025). Dual-model synergy for audit opinion prediction: A collaborative LLM agent framework approach. International Review of Economics & Finance, 104, 104642. https://doi.org/10.1016/j.iref.2025.104642

Moll, J., & Yigitbasioglu, O. (2019). The role of internet-related technologies in shaping the work of accountants: New directions for accounting research. The British Accounting Review, 51(6), 100833. https://doi.org/10.1016/j.bar.2019.04.002

Muzanenhamo, P., & Power, S. B. (2024). ChatGPT and accounting in African contexts: Amplifying epistemic injustice. Critical Perspectives on Accounting, 99, 102735. https://doi.org/10.1016/j.cpa.2024.102735

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586

Olariu, M.-E., Buinceanu, V.-G., Simionescu, C., Dospinescu, O., Georgescu, R., Tudor, C., Iftene, A., &

Bores, A.-M. (2026). RO-FIN-LLM: A benchmark with LLM-as-a-Judge and human evaluators for Romanian tax and accounting. Systems, 14(3), 244. https://doi.org/10.3390/systems14030244

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., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

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

Pereira, J., Assumpcao, A., Trecenti, J. A. Z., Airosa, L., Lente, C., Cléto, J., Dobins, G., Nogueira, R. F., Mitchell, L., & Lotufo, R. (2025). INACIA: Integrating large language models in Brazilian audit courts: Opportunities and challenges. Digital Government Research and Practice, 6(1), 1–20. https://doi.org/10.1145/3652951

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts (arXiv:2205.01833) [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2205.01833

Public Company Accounting Oversight Board. (2024). Spotlight: Staff update on outreach activities related to the integration of generative artificial intelligence in audits and financial reporting. PCAOB. https://pcaobus.org/documents/generative-ai-spotlight.pdf

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Sugahara, S., & Kano, K. (2026). The impact of ChatGPT's advice on professional judgment related with accounting and the role of accounting education. Journal of Business Ethics. Advance online publication. https://doi.org/10.1007/s10551-026-06365-x

Sundkvist, C. H., & Kulset, E. M. (2024). Teaching accounting in the era of ChatGPT: The student perspective. Journal of Accounting Education, 69, 100932. https://doi.org/10.1016/j.jaccedu.2024.100932

Tharapos, M., Lau, K. H., Peszynski, K. J., Nguyen, L. A., Magdziarz, S., Morton, E., Borg, V., & Duan, S. X. (2025). Generative AI in accounting education: Evaluating ChatGPT's role in assessment and skill development. Accounting and Finance, 65(3), 3231–3244. https://doi.org/10.1111/acfi.70051

Toumeh, A. A. (2024). Assessing the potential integration of large language models in accounting practices: Evidence from an emerging economy. Future Business Journal, 10(1). https://doi.org/10.1186/s43093-024-00368-8

Vasarhelyi, M. A., Moffitt, K. C., Stewart, T. R., & Sunderland, D. (2023). Large language models: An emerging technology in accounting. Journal of Emerging Technologies in Accounting, 20(2), 1–10. https://doi.org/10.2308/jeta-2023-047

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

de Villiers, C., Dimes, R., & Molinari, M. (2024). How will AI text generation and processing impact sustainability reporting? Critical analysis, a conceptual framework and avenues for future research.

Sustainability Accounting, Management and Policy Journal, 15(1), 96–118. https://doi.org/10.1108/sampj-02-2023-0097

Wood, D. A., Achhpilia, M. P., Adams, M. T., Aghazadeh, S., Akinyele, K., Akpan, M., Allee, K. D., Allen, A. M., Almer, E. D., Ames, D., Arity, V., Barr-Pulliam, D. D., Basoglu, K. A., Belnap, A., Bentley, J. W., Berg, T., Berglund, N. R., Berry, E., Bhandari, A., . . . Zoet, E. (2023). The ChatGPT artificial intelligence chatbot: How well does it answer accounting assessment questions? Issues in Accounting Education, 38(4), 81–108. https://doi.org/10.2308/issues-2023-013

Wu, Y., Hu, P., & Wang, D. D. (2025). The AI annotator: Large language models' potential in scoring sustainability reports. Systems, 13(10), 899. https://doi.org/10.3390/systems13100899

Xiong, F., & Bao, J. (2025). Next-generation gamification in accounting education: Digital role-play games (D-RPGs) with generative AI (GAI). Journal of Emerging Technologies in Accounting, 22(2), 171–180. https://doi.org/10.2308/jeta-2024-012

Zhang, Y. (2026). Technology empowerment or audit burden? A study on the impact of generative AI on corporate audit costs. Advances in Economics and Management Research, 18(1), 1–5. https://doi.org/10.56028/aemr.18.1.1-5.2026

Zhao, J., & Wang, X. (2024). Unleashing efficiency and insights: Exploring the potential applications and challenges of ChatGPT in accounting. Journal of Corporate Accounting & Finance, 35(1), 269–276. https://doi.org/10.1002/jcaf.22663

Zhou, A., & Luo, Y. (2025). Exploring the impact of generative AI on student learning in accounting. Journal of Accounting Education, 72, 100982. https://doi.org/10.1016/j.jaccedu.2025.100982

Ziemba, E. W., Maruszewska, E. W., Grabara, D., & Renik, K. (2024). Acceptance and use of ChatGPT among accounting and finance higher education students. In Lecture notes in networks and systems (pp. 185–202). Springer. https://doi.org/10.1007/978-3-031-66761-9_16

Zupic, I., & Čater, T. (2015). Bibliometric methods in management and organization. Organizational Research Methods, 18(3), 429–472. https://doi.org/10.1177/1094428114562629

Żytniewski, M., & Pankowska, M. B. (2026). Evaluating large language models on the certified internal auditor exam. Applied Sciences, 16(16), 8106. https://doi.org/10.3390/app16168106