THE HIDDEN RISKS OF GENERATIVE ARTIFICIAL INTELLIGENCE IN ACCOUNTING AND AUDITING: A BIBLIOMETRIC PERSPECTIVE

Anamaria-Georgeta BARBU
Faculty of Economics and Business Administration, West University of Timișoara, Romania
Email: anamaria.barbu99@e-uvt.ro

Abstract: *Generative artificial intelligence has moved quickly from experimental systems to tools used in routine accounting and audit work. Large language models now assist in drafting audit documentation, summarizing contracts, interpreting unexpected variances and supporting elements of risk assessment. These uses promise efficiency gains, yet they also raise questions about evidence quality, professional judgement and the reliability of financial reporting. This paper examines how recent academic and professional literature addresses these issues. Based on a bibliometric analysis of 286 Web of Science publications (October 2025) that link generative AI with accounting, auditing or financial reporting, the study maps the structure of this fast-growing field. Using VOSviewer keyword co-occurrence analysis, frequency counts and close reading of influential works, the paper identifies four recurring groups of risks: persuasive but weakly evidenced AI-generated outputs, pressures on professional judgement including automation bias, data security and confidentiality vulnerabilities, and ethical or reputational biases embedded in training data. The results show a sharp rise in publications after 2023 and a pronounced interdisciplinary profile, with contributions from accounting, computer science, engineering and education. By connecting these risks to core concepts such as verifiability, audit quality and the public interest, the paper argues that generative AI alters the conditions under which professional judgement is formed and evaluated. The conclusion outlines implications for accounting education, organizational AI policies and future research on human–AI interaction and validation of AI-generated evidence. For practitioners, the framework can also serve as a simple checklist when deciding how and where to integrate generative AI into everyday accounting and audit work.*

Keywords: *Generative Artificial Intelligence (GenAI); Accounting; Auditing; Professional Judgement; Accountability; Ethics*

JEL Classification: *M41; M42; O33*

1. Introduction

Generative artificial intelligence, particularly large language models, has expanded rapidly from experimental applications to routine use in accounting and auditing. These systems now assist with drafting audit working papers, summarizing contracts, generating code for data extraction and proposing explanations for unexpected variances. Their appeal lies in efficiency and the ability to process large volumes of structured and unstructured data. Yet these developments also introduce concerns that touch core principles of financial reporting and assurance. AI-generated outputs often appear coherent and professionally crafted while being incomplete, weakly supported by underlying evidence or affected by embedded bias. Such characteristics challenge expectations of reliability, faithful representation and verifiability. Although scholars and professional bodies have begun to address these issues, the literature remains fragmented across studies on automation, ethics, cybersecurity, governance and audit quality. This dispersion makes it difficult to form an integrated understanding of how generative AI reshapes professional judgement and the conditions under which accounting and audit work is carried out.

Against this background, the paper examines the following research question: How does academic and professional literature conceptualize the risks generated by the integration of generative artificial intelligence into accounting and auditing? The objective is to map this emerging field and identify the main risk themes that recur across disciplines. Addressing this question is important because it clarifies how GenAI affects the evaluation of audit evidence, the exercise of judgement and the credibility of financial reporting.

The study proceeds from the assumption that the risks associated with generative AI form identifiable clusters that can be revealed through systematic analysis. A bibliometric approach is appropriate here, as it captures both the quantitative structure of the research field and the dominant conceptual patterns within it. This perspective allows the study to detect areas of convergence and divergence across accounting, computer science, engineering and related domains.

The analysis uses 286 publications indexed in Web of Science as of October 2025, selected for their explicit connection between generative AI, accounting, auditing or financial reporting. VOSviewer is employed to generate keyword co-occurrence maps, supported by frequency counts and close reading of influential contributions. This combination of methods makes it possible to identify both structural patterns and substantive content in the literature.

The paper contributes by documenting the rapid growth and interdisciplinary nature of the GenAI–accounting–audit research domain, by identifying four recurring clusters of risk and by linking these themes to foundational concepts such as verifiability, audit quality and the public interest. The findings offer a clearer framework for evaluating AI-generated evidence, guiding professional judgement and informing organizational AI policies. They also highlight directions for future research on human–AI interaction, validation of AI outputs and the detection of bias in financial reporting and audit contexts.

2. Literature Review

This literature review was developed through a systematic process designed to identify, evaluate and synthesize academic and professional work on generative artificial intelligence in accounting and auditing. The review began by defining a central research question: How does the existing literature conceptualize the risks introduced by the integration of generative AI into accounting and audit practice? Guided by this question, the search strategy focused on keywords such as "generative artificial intelligence," "large language models," "accounting," "auditing," "financial reporting," "automation bias" and "AI governance." These terms and their variants were used across multiple academic databases, with Web of Science serving as the primary source due to its wide coverage of peer-reviewed research.

The search results were screened first by title and abstract, and then by full text, using criteria related to methodological transparency, conceptual relevance and publication quality. The selected studies were then organized and analyzed to identify recurring themes, contradictions and areas requiring consolidation. Three broad and partly overlapping strands emerged from this synthesis: research emphasizing automation and efficiency, work treating GenAI as a decision-support tool and studies focusing on risks, ethics and governance.

The first strand presents GenAI as a tool for automating routine tasks. Publications describe how large language models can draft audit working papers, summarize lengthy documents or assist with data extraction. When these outputs are reviewed by qualified practitioners, authors expect shorter reporting cycles and more time for non-routine analytical work (IFAC, 2023; IAASB, 2024a; IAASB, 2024b). This work highlights operational benefits but pays less attention to how automation interacts with judgement.

The second strand examines GenAI as a form of decision support. Here, the emphasis is on the model's capacity to integrate standards, internal guidance and external sources to propose accounting treatments, identify risks or generate alternative scenarios. These capabilities may strengthen professional judgement, yet their structured and confident presentation increases the risk of over-reliance and automation bias (OECD, 2023).

The third strand concentrates on risks, ethics and governance. These studies highlight concerns about hallucinations, limited transparency and embedded bias, noting that models trained on historical data may reproduce or magnify past patterns in ways that misclassify clients or sectors. Technical analyses describe vulnerabilities such as model inversion and membership inference attacks, which could compromise confidential financial data. Professional bodies emphasize that AI uses reinforces, rather than lessens, ethical obligations relating to integrity, objectivity and confidentiality, and requires clear documentation and strong internal controls (IAASB, 2024a; IAASB, 2024b; IFAC, 2023).

Taken together, these strands depict a research field that is expanding rapidly but remains conceptually fragmented. Automation, judgement, data security and ethics are often examined in isolation and without a shared framework (OECD, 2023; OECD, 2024; World Economic Forum, 2025). This fragmentation confirms the need for a structured synthesis. The present study addresses this by applying a bibliometric analysis of Web of Science publications to identify the main recurring risk themes and arrange them into four risk clusters. These are presented in Section 4 after the methodological outline in Section 3.

3. Methodology

3.1. Data collection

The bibliometric analysis was based on the Web of Science Core Collection. The aim was to capture publications that explicitly connect generative artificial intelligence with accounting, auditing or financial reporting. For this reason, I used a two-step topic search (TS), which required at least one GenAI-related term and one accounting-related term in the title, abstract or keywords:

TS = ("generative artificial intelligence" OR "generative AI" OR "GenAI" OR "ChatGPT" OR "large language models" OR "Gen AI")

AND
TS = (“accounting” OR “audit” OR “financial reporting”).
I did not set any time restrictions in the search. However, all retrieved publications were from the years 2023–2025, which reflects how recent the discussion on GenAI in accounting and auditing is. In a next step, I refined the results only by Web of Science Research Areas. I selected four areas: Business Economics, Computer Science, Engineering and Educational Research. After these filters, the final dataset contained 286 publications. For each item, I exported full records (including author keywords, Keywords Plus and cited references) and basic bibliometric indicators. The 286 documents had together 1,395 citations and an h-index of 19 (Figure 2).

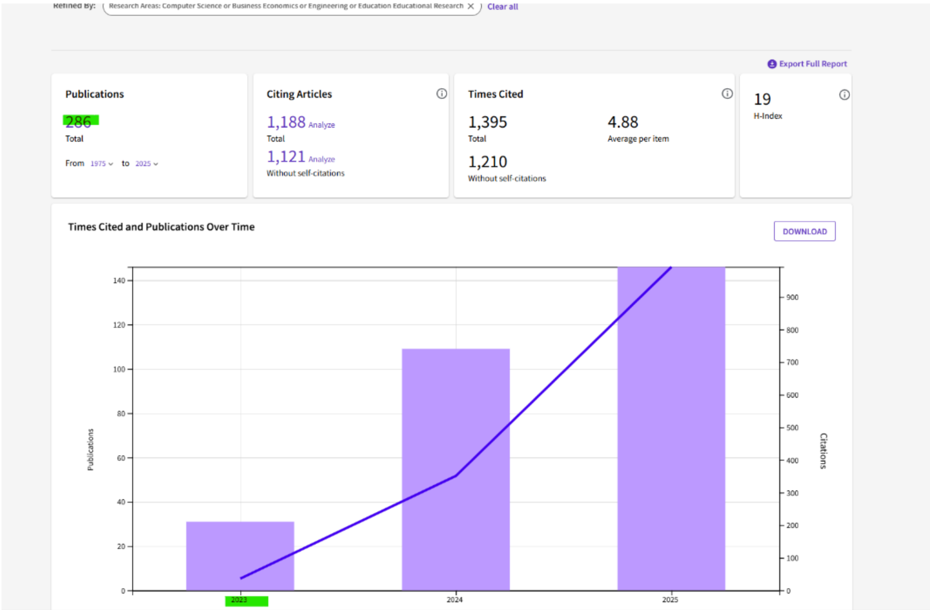

Figure 1: Web of Science overview of the GenAI–accounting publication sample (286 documents, 2023–2025)
Source: Author’s elaboration based on Web of Science data

For the later qualitative interpretation of risk themes, the documents were also sorted by local citation counts and publication date. On this basis, I identified a smaller group of highly cited or conceptually influential papers, which were read in detail and used as reference points when interpreting the clusters from the mapping exercise.

3.2. Bibliometric mapping and clustering

To analyze the conceptual structure of this emerging field, I used VOSviewer to create a keyword co-occurrence map. The Web of Science records for the 286 documents were imported with full bibliographic information. Both author keywords and Keywords Plus were included, because the latter often capture recurring themes that are not visible in individual titles. VOSviewer was then used to construct a network in which nodes represent keywords and links represent their co-occurrence in the same publication (Figure 2). The size of a node reflects how often a keyword appears in the dataset, while the distance and clustering algorithm indicate how strongly keywords are related in the literature. This allowed the identification of groups of terms that frequently appear together when GenAI is discussed in relation to accounting and auditing.

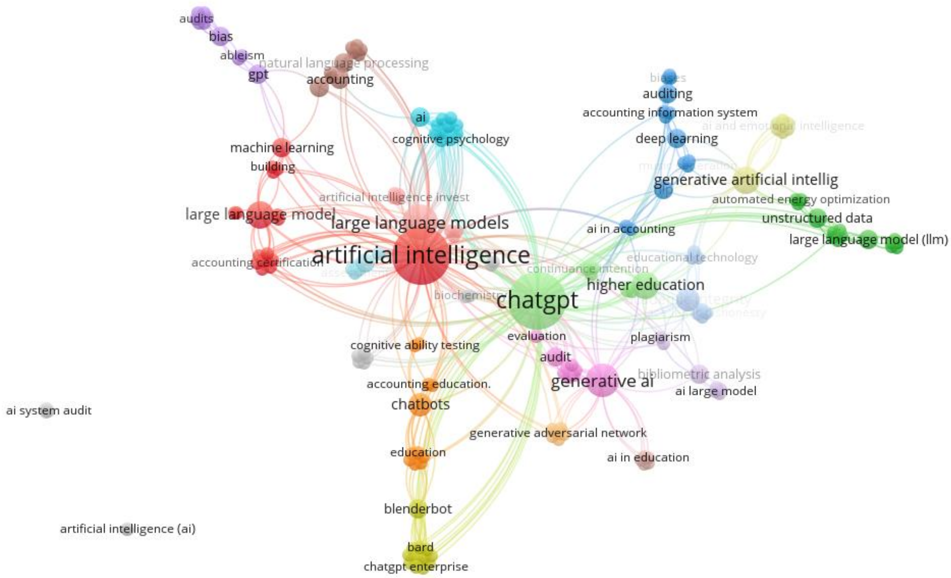

Figure 2: Keyword co-occurrence network for GenAI in accounting and auditing (VOSviewer map)
Source: Author's elaboration based on Web of Science data

To complement the visual network, I prepared simple pivot tables (in Excel) that grouped risk-related keywords into broader families (Figure 3), such as:

- accuracy, hallucinations and reliability,
- judgement, decision-making and automation bias,
- trust, ethics, transparency and quality,
- data protection, privacy and security.

Row Labels	Count of Valu
Bias	6
biases	2
AI bias	2
Algorithmic bias	1
language generation biases	1
Anchoring bias	1
bias mitigation	1
Grand Total	**14**

Row Labels	Count of Value
TRUST	4
patient trust	1
TRUST	1
trust in AI	1
Grand Total	**7**

Row Labels	Count of Value
AI ethics	2
AI Ethics	2
cyberethics	1
Ethics education	1
ethics	1
Ethical AI Adoption	1
LLM Ethics	1
Algorithm Ethics	1
Ethical implications	1
ethical AI in educati	1
Grand Total	**12**

Row Labels	Count of Value
Academic integrity	4
Research integrity	1
Grand Total	**5**

Row Labels	Count of Valu
Cybersecurity	1
cyberethics	1
cyber terrorism	1
cyberattack	1
Grand Total	**4**

Row Labels	Count of Value
Security	2
data security	2
Computer security	1
network security model	1
Security and privacy	1
identity security	1
Cybersecurity	1
network security	1
Grand Total	**10**

Row Labels	Count of Value
TRANSPARENCY	2
TRANSPARENCY	1
Transparency and Enga	1
Grand Total	**4**

Row Labels	Count of Va
QUALITY	4
Qualitative research	3
quality audit	2
QUALITY	1
QUALITIES	1
Quality of experience	1
Qualitative content analysis	1
website quality	1
Grand Total	**14**

Figure 3: Pivot tables summarizing the frequency of risk-related keyword families
Source: Author's elaboration based on Web of Science data

The frequency of these keyword families provided a quantitative indication of which types of risk are most prominent in the current debate.

4. Results

4.1 Overview of the publication landscape

The Web of Science analytics confirm that work on GenAI in accounting and auditing is very recent. The number of papers rises sharply in 2024 and 2025 (Figure 1), and citations follow a similar pattern, in line with the public release and rapid adoption of large language models such as ChatGPT.

With regard to Web of Science subject categories, Business Finance is the most frequent label (77 records). At the same time, a large part of the dataset is classified under several Computer Science categories: Artificial Intelligence, Information Systems, Interdisciplinary Applications etc. Taken together, these labels confirm the interdisciplinary character of the field, where technical discussions of AI are combined with questions about financial reporting, assurance and accounting education.

The keyword analysis shows that the most frequent terms relate to ethics, trust, bias, data security, transparency and quality (Figure 4). This suggests a shift compared to earlier "AI in accounting" discussions, which focused mainly on efficiency and automation. In the current debate, risk, governance and trust are much more prominent.

Based on the VOSviewer network (Figure 3) and the keyword frequency analysis, four thematic risk clusters stand out across the 286 publications. These clusters are not predefined, but emerge from the way authors describe and connect GenAI with accounting and auditing. In the following subsections, I summaries each cluster.

4.2 Illusion of Accuracy and Financial Data Reliability

The papers in this cluster draw attention to something that many practitioners will recognize from their own experiments with GenAI: the model often sounds right when, in fact, it is not. It produces audit-style paragraphs, cross-referenced explanations and neat tables, and at a quick glance everything looks exactly as it should. Only when one starts asking "where does this number come from?" or "what test supports this claim?" do the cracks appear. Similar patterns are documented in the NLP field, where Venkit et al. (2024) audit different types of hallucinations in large language models and show how confidently presented but unfounded outputs can mislead users.

Wang (2025) captures this very well with the term “accounting hallucinations”. He describes AI-generated audit comments that read like the work of an experienced senior, although no underlying procedure or working paper exists. Joshi (2025) reports a similar effect in forecasting and risk analysis: large language models generate detailed projections and risk indicators, but in some cases, these are only loosely connected to the underlying data, especially when the prompt is narrow or vague, a problem he labels “quantitative analysis distortions”. The OECD (2023) also notes that, when key financial information is missing, generative systems tend to “complete” the picture by inventing figures or relationships so that the answer appears coherent.

This behaviour clashes with the basic expectation of verifiability. Information in financial statements and audit files should not only look professional; it should be traceable to evidence. The real danger here is the explanation that sounds entirely reasonable but is wrong, and once it is accepted into the reporting process it can quietly distort indicators and mislead users of financial statements (Wang, 2025; Joshi, 2025; OECD, 2023).

4.3 Erosion of Professional Judgement and Accountability

In the second group of articles I reviewed, the focus shifts from what GenAI can do to what happens to the people who use it. Several authors describe a slow change in working habits. When accountants and auditors turn to GenAI for help with explanations, risk assessments or classifications, the model’s suggestion can quietly become the default answer. Instead of forming their own view first and then comparing it with the tool, some professionals start from what the system proposes and only rarely challenge it.

Choi and Xie (2025) analyze this tendency in a field study on early GenAI use in audit firms. They report signs of automation bias: in some situations, auditors accepted AI-generated assessments even when these differed from their initial judgement. Their results show clear efficiency gains, but also suggest that repeated exposure to AI output can weaken the normal reflex to ask “does this really follow from the evidence we have?”.

If this pattern continues, there is a risk that responsibility becomes blurred. The audit opinion or financial statements are still signed by a human, yet important parts of the reasoning may, in practice, have been shaped by a system that has no ethical duties and no legal accountability. Several authors warn that this is a more profound threat than simple job automation. What is at stake is the habit of independent, critical thinking that has traditionally defined professional judgement in accounting and auditing (Choi and Xie, 2025; Croom, 2025).

4.4 Data Security and Confidentiality Risks

In this cluster, the authors stop talking about “smart tools” in abstract terms and go back to a very old concern in accounting: keeping client information safe. General-purpose accounting systems already store extremely sensitive data about companies, contracts and sometimes identifiable individuals. When the same data are now sent through GenAI applications, very often running on external cloud infrastructure, they leave the protected environment of the firm. The gain in analytical power comes together with a wider circle of systems and people who might, in theory, access or recover that information.

The technical and legal literature does not talk about these risks in abstract terms, but points to a few very concrete issues. Studies on model inversion and membership inference show that, under certain conditions, parts of the data used to train a model or provided through prompts can be reconstructed from what the model outputs. The Hogan Lovells (2024) report illustrates this with practical examples and explains how such attacks could reveal confidential financial information if client data are used to fine-tune GenAI models. Yan et al. (2025) arrive at a similar conclusion for GPT-based applications, showing how so-called “file knowledge leakage” can allow information from documents that should remain separate inside an application environment to be inferred from the model’s behaviour. Other papers note that prompts and training sets may be logged or cached inside the provider’s environment, so that traces of sensitive data remain in the system longer than users expect. All of this sits uncomfortably next to strict rules such as the GDPR and to the traditional duty of confidentiality in the accounting profession. The OECD (2024) argues that, as GenAI is adopted more widely in finance, firms can no longer rely on generic IT policies, but need explicit data governance for AI: clear rules on who may upload which information into GenAI tools, how long it is stored, what encryption and access controls apply, and how data flows are recorded and monitored (Hogan Lovells, 2024; OECD, 2024).

4.5 Ethical and Reputational Biases

In this part of the literature, the focus moves away from what GenAI can do from a purely technical perspective and turns instead to the more discreet effects that its use may have in accounting practice. Because these models are trained on large sets of historical data, they inevitably take over past habits in recognition, valuation and risk assessment. If certain regions, sectors or types of entities have been treated in a particular way over a longer period of time, the model can carry these patterns forward and apply them in new situations. A recommendation that looks neutral at first glance may therefore reproduce older preferences and biases, rather than offering an objective view of the specific case being analyzed. Wilson and Caliskan (2024), for example, show that large language models used in résumé screening can replicate intersectional gender and racial bias, even when the screening procedure is presented as neutral and fair.

A second concern is that many GenAI tools operate in a very opaque way. When an AI system flags a client as high risk or leans towards one accounting treatment, it is often difficult to see why. The IAASB (2024) notes that in an audit this lack of explainability is more than a technical issue, because auditors must be able to justify their conclusions. IFAC (2023) also warns that automation can make processes faster while, at the same time, introducing new vulnerabilities that are harder to detect. For this reason, several reports argue that AI used in finance should be both explainable and open to independent scrutiny, so that its behaviour can be tested against professional values of fairness and transparency (World Economic Forum, 2025; IAASB, 2024). In the end, the standing of the accounting profession depends not only on correct figures, but also on whether stakeholders can trust the methods and algorithms that helped to produce them.

5. Discussion

This discussion expands upon these findings by examining the profound implications of utilizing GenAI on accounting integrity, professional judgement, and ethical governance. Generative AI is changing how accountants do their jobs and what trust, judgement, and accountability mean in the digital age. The real issue is no longer how fast technology changes, but how well the industry uses its power. As algorithms get better at what they do, professionals need to be more careful and aware of what they do.

Accuracy, caution, and moral responsibility have always been the most important things in accounting (IAASB, 2024a; IAASB, 2024b). Smart systems that can copy human expertise without taking responsibility are now testing these foundations. GenAI can speed up analyses, automate reconciliations, and make predictions more accurate. However, it also makes it harder to tell the difference between automation and interpretation. When AI-generated results appear flawless yet lack factual basis, the illusion of accuracy poses greater risk than the error itself (OECD, 2023; Wang, Y. 2025; Joshi, R. 2025; Venkit et al., 2024; Croom, 2025).

The results show that the quality of accounting in the future will depend less on how AI tools improve and more on how strong human oversight and professional skepticism are. Accountants must transition from passive technology users to discerning interpreters of information. They must question, confirm, and challenge algorithmic reasoning. Technical skill alone will no longer be the most important thing that decides who is the most expert. Instead, it will be the ability to know when to have faith in the outcome.

In this new digital world, knowing about ethics is now a technical skill (IAASB, 2024a; IAASB, 2024b; Wilson and Caliskan, 2024). It is now a professional requirement, not just a theoretical ideal, to have clear ways to hold people responsible, to be able to trace AI-generated conclusions, and to be able to see how algorithms work. It is both a legal and moral duty to explain GenAI's financial predictions when they have an impact on markets and people's lives.

To do well in this environment, the profession needs to learn how to be digitally skeptical, which means being accurate while also being aware of ethics. Accountants should think of AI systems as partners that need to be checked all the time, not as experts who can't be questioned. The credibility of the profession will increasingly depend on its ability to evaluate technology: to understand how conclusions are reached, assess the reliability of outputs, and ensure that human judgement remains the primary safeguard of trust (Choi and Xie, 2025; OECD, 2023).

GenAI won't replace accountants, but it will be hard for people who stop thinking critically. The future of accounting will not be those who use technology the most, but those who understand it well, question it smartly, and use it responsibly.

6. Conclusion

This study contributes to the current debate on GenAI in accounting and auditing by offering a bibliometric overview of a very recent body of literature and by organizing the main risk-related concerns into four themes. The analysis shows a rapid growth of publications after 2023 and a strong interdisciplinary character. At the same time, it indicates that the evidence base is still relatively weak and fragmented. Concerns

about accuracy, judgement, data security and ethics are often discussed separately, using different concepts and terms.
The four themes: illusion of accuracy, pressure on professional judgement and accountability, data security and confidentiality risks, and ethical and reputational biases underline that GenAI is not just another efficiency tool. It influences how financial information is produced, how it is justified and how it is perceived by users of financial statements. Keeping financial information reliable therefore requires more than compliance with existing rules. It requires a governance culture in which innovation is aligned with the public interest and where AI use is documented, monitored and open to review.
For practice and regulation, the results suggest that explicit AI policies are needed at firm and system level. These policies should clarify which data may be uploaded into GenAI tools, how outputs are validated, who is responsible for decisions based on AI-generated analyses, and how bias and security risks are addressed. Professional bodies and standard setters, together with organizations such as IFAC, IAASB and OECD, can support this process by providing guidance that links ethical principles such as integrity, objectivity and confidentiality to concrete requirements for AI transparency and oversight.
Future research should move beyond general statements about opportunities and risks and focus on practical frameworks for responsible GenAI use. Possible directions include methods to detect and mitigate algorithmic bias, procedures for validating AI-generated audit or accounting evidence, and designs for training programmes that strengthen critical thinking in a GenAI environment (OECD, 2023; Wang, 2025; Joshi, 2025; Choi and Xie, 2025). Cross-disciplinary collaboration between accountants, data scientists, educators and regulators will be important for turning these ideas into workable tools.
Overall, the findings point to a simple conclusion: GenAI will likely become part of the normal infrastructure of accounting and auditing, but it cannot replace professional judgement. The long-term credibility of the profession will depend less on the sophistication of the technology and more on the integrity, transparency and skepticism with which that technology is used.

7. Acknowledgements
The author thanks Professor Nicolae Bobițan and Professor Diana Dumitrescu for their guidance and support.

References
1. Choi, J.H. and Xie, C. (2025) *Human + AI in Accounting: Early Evidence from the Field.* Stanford Graduate School of Business Research Paper No. 4261; MIT Sloan Research Paper No. 7280-25.
2. Croom, J. (2025) Interactivity and Illusions of Ability: How Using Generative AI Affects Investor Judgments. *Journal of Accounting Research.*
3. Hogan Lovells. (2024). *Model inversion and membership inference: Understanding new AI security risks and mitigating vulnerabilities.* Hogan Lovells Publications.

4. IAASB (2024a) *Elevating Trust in Audit and Assurance: IAASB's Strategy and Work Plan 2024–2027.* New York: International Auditing and Assurance Standards Board.
5. IAASB (2024b) *Technology Position Statement: 8 Guiding Actions.*
6. IFAC (2023) *AI and Intelligent Automation – Opportunities for Professional Accountants in Business.* International Federation of Accountants.
7. Joshi, R. (2025) *Comprehensive Review of AI Hallucinations: Impacts and Mitigation Strategies for Financial and Business Applications.*
8. OECD (2023) *Generative Artificial Intelligence in Finance: Opportunities and Risks.* Paris: OECD Publishing.
9. OECD (2024) *AI, Data Governance and Privacy.* OECD Digital Economy Papers.
10. Venkit, P.N. et al. (2024) An Audit on the Perspectives and Challenges of Hallucinations in NLP. In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024).
11. Wang, Y. (2025) *Financial Reliability and AI Hallucinations: Evaluating the Precision Paradox in Large Language Models for Auditing.* SSRN Electronic Journal.
12. Wilson, K. and Caliskan, A. (2024) *Gender, Race, and Intersectional Bias in Resume Screening via Language Model Retrieval*.
13. World Economic Forum (2025) *Artificial Intelligence in Financial Services.* Geneva: World Economic Forum.
14. Yan, C., Guan, B., Li, Y., Meng, M.H., Wan, L. and Bai, G. (2025) *Understanding and Detecting File Knowledge Leakage in GPT App Ecosystem*. In: *Proceedings of the 22nd International Conference on Computing and Cryptography (ICC '25)*. Association for Computing Machinery.