Financial Services Agency

Discussion Paper Series

Study on the Potential Utilization of Generative AI in Audit Practices and Challenges Related to Audit Quality

NOMA Mikiharu

DP 2025-3
July 2025
(English version released in December 2025)

Financial Research Center (FSA Institute)
Financial Services Agency
Government of Japan
3-2-1 Kasumigaseki, Chiyoda-ku, Tokyo 100-8967, Japan

The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Financial Services Agency or the FSA Institute.

Study on the Potential Utilization of Generative AI in Audit Practices and Challenges Related to Audit Quality

NOMA Mikiharu*

Abstract

Artificial intelligence (AI), including generative AI, is advancing at a rapid pace and is affecting multiple facets of our society. The adoption, application, and exploration of generative AI is becoming increasingly widespread in the audit industry. The utilization of generative AI and other technology in auditing is expected to contribute to improving audit quality and alleviate the shortage of audit resources by enhancing efficiency. However, concerns remain about potential issues that could negatively impact auditors' performance and compromise audit quality, such as over-reliance on AI, AI's "black box" nature, information leakage, and hallucinations.

Numerous discussions are underway among audit firms, industry associations, audit regulators, and international organizations regarding the potential benefits and challenges of using generative AI in auditing. In academia, prior studies have explored themes such as the potential use of generative AI and possible effects on audit firms' recruitment strategies and practices. For example, Law and Shen (2025) examine whether AI can replace human auditors or complement them, and Fedyk et al. (2022) survey the impact of AI on audit quality and efficiency.

This study examines the potential utilization of generative AI in audit practices and associated challenges in terms of audit quality, based on these prior studies and by observing developments in the actual utilization practices of generative AI as well as discussions among and initiatives taken by audit regulators and international organizations.

Keywords: generative AI; auditing; and audit quality.

---

* Professor, Graduate School of Business Administration, Hitotsubashi University (Special Research Fellow, Financial Research Center, Financial Services Agency)
I received valuable comments from many officials at the Financial Services Agency, including Professor YANAGAWA Noriyuki (Director of the Financial Research Center), Professor YOSHINO Naoyuki and Mr. OHGO Naoki, who are Chief Advisors of the Center. I would like to express my gratitude to them. The views expressed in the paper are those of the, and do not necessarily reflect the formal views of the Financial Services Agency or the Financial Research Center.

1. Introduction

1.1 Objective

This paper examines the potential opportunities and challenges to audit quality associated with the use of generative AI in the audit industry. In recent years, regulatory authorities, international organizations, audit firms, and industry associations have engaged in vigorous discussions on the use of AI in auditing. Meanwhile, prior research on AI and generative AI in auditing has been gradually accumulating in the academic field. For example, Igou et al. (2023) surveyed six international accounting journals published between 2011 and 2020 containing keywords related to new technologies, discussing the digitalized future of the accounting industry. They reported that papers on analytics were the most numerous, with many papers discussing machine learning and deep learning alongside blockchain.

This paper reviews relevant reports and papers on the impact of AI utilization on auditing. It examines the effects of AI on auditing and financial reporting, regulatory trends, changes in the auditor's role, and future prospects. The paper aims to clarify the multifaceted impact of AI on auditing and financial reporting and to derive implications for future auditing practices.

The utilization of AI in audit work has the potential benefits, such as improving audit quality and alleviating the shortage of audit personnel. On the other hand, concerns exist regarding potential impacts on audit quality, including risks of over-reliance on AI, black-boxing, and, particularly with generative AI, hallucinations and information leakage. This paper discusses the potential opportunities and challenges for sound utilization of generative AI in the audit industry, based on prior research and current practices regarding these issues.

1.2 Conventional AI and Generative AI

This paper adopts the definitions of traditional AI and generative AI provided by the Financial Services Agency (2025), which outlined initial points for promoting the sound use of AI in the financial sector. The Financial Services Agency (2025) defines conventional AI as systems that learn "characteristics and trends" by feeding data to AI in advance (including rule-based models, such as chatbots, where complex rules are created from data to build and operate the model) to obtain responses to input data. Generative AI, on the other hand, is defined as models with large parameters, such as large language models (LLMs), that learn from internet data and content (unstructured data like text and images) to generate new outputs (text, images, audio, video, etc.).

The Financial Services Agency (2025) presents survey results on AI utilization among financial institutions, including audit firms. These results highlight differences in usage patterns between traditional AI and generative AI. In the financial sector, traditional AI use cases predominantly involve "optical character recognition" (OCR) for documents, customer service operations, information retrieval, and marketing. In contrast, common use cases for generative AI include "text summarization," "proofreading, correction, and evaluation of texts," "translation," and "draft creation." Regarding the key points of this paper—"audit operations" and "accounting/bookkeeping-related tasks"—the adoption of traditional AI is more advanced than that of generative AI. Furthermore, for both generative AI and traditional AI, adoption is more advanced in "accounting and finance-related tasks" than in "audit operations." However, the Financial Services Agency (2025) survey does not clarify in which specific audit tasks financial institutions utilize AI. That is, it remains unclear whether audited companies use AI for audits conducted by audit firms or for internal audit purposes.

This paper generally distinguishes between traditional AI and generative AI in its discussion. The term "AI" is used when referring to both traditional and generative AI collectively. However, the term "AI" is also used when prior research does not necessarily distinguish between the two. Furthermore, while this paper aims to discuss the impact of generative AI on audit quality and related areas, it will also introduce relevant content from prior research discussing big data and data analysis in auditing that pertains to generative AI and auditing.

2. Overview and Status of AI Utilization in Auditing

The rapid advancement of AI technology is significantly impacting various sectors, including the financial industry. AI, particularly generative AI and large language models (LLMs), have diverse applications within financial markets. IOSCO (2025) notes that AI is widely used in capital markets for applications such as robo-advisors, algorithmic trading, investment research, and sentiment analysis. These technologies excel at analyzing vast amounts of data, identifying patterns, and making predictions, thereby enhancing decision-making processes in the financial industry. Specifically, AI has the potential to increase market efficiency and provide more personalized financial services to investors. For example, robo-advisors can automatically manage portfolios tailored to individual investors' risk tolerance and investment goals, while algorithmic trading enables capturing minute market fluctuations to generate profits.

In auditing, AI also holds potential to streamline tasks such as data analysis, risk assessment, and anomaly detection (CAQ, 2024; KPMG, 2024). CAQ (2024) states that AI can automate audit procedures, optimize sample sizes, and improve the evaluation of audit evidence, enabling auditors to focus on deeper analysis and judgment. AI could allow auditors to perform their duties more efficiently and effectively. For instance, AI can analyze large volumes of transaction data to detect anomalous patterns, helping auditors identify high-risk transactions requiring focused investigation. KPMG (2024) highlights the potential of generative AI to streamline and enhance the processes of creating, analyzing, and distributing financial reports, citing examples such as automatically generating financial statements, identifying risk factors, and visualizing financial data. Generative AI holds the potential to enhance the quality and transparency of financial reporting. For instance, it can draft financial statements based on historical reports and industry best practices, accelerating the reporting process.

AI is expected to transform financial reporting and auditing from the digital era to the AI era, bringing smarter, integrated information flows, improved risk identification and response, and enhanced anomaly detection capabilities (CAQ, 2024; KPMG, 2024). Below, I introduce perspectives on AI from international organizations, audit firms, and industry associations, along with the impact AI has on auditing. I also provide an overview of prior research examining the potential for utilizing AI in auditing and its effects on audit quality.

2.1 Generative AI and Accounting-Related Exams

First, to demonstrate the potential for utilizing generative AI in audit practices, I introduce prior research on its response capabilities regarding bar exams and accounting-related national examinations.

OpenAI has reported that its AI possesses the capability to pass highly challenging exams such as the U.S. bar exam and medical licensing exams. These exams feature many essay-type questions and contain fewer arithmetic elements compared to the CPA exam. The CPA exam includes subjects with significant arithmetic components, such as bookkeeping, financial accounting, management accounting, and taxation.

Wood et al. (2023) compared ChatGPT-3.5's responses to those of students using accounting exam data collected from 186 educational institutions across 14 countries by a team of 327 researchers. Wood et al. (2023) evaluated how accurately ChatGPT-3.5, developed by OpenAI, could answer accounting questions. Using 28,085 accounting textbooks and test questions gathered from educational institutions, they compared ChatGPT-3.5's performance against student scores. Specifically, they compared performance across seven topics: Accounting Information Systems, Analysis and Technology, Auditing, Financial Accounting, Management Accounting, Other, and Taxation. The results revealed three key points: First, overall, students outperformed ChatGPT-3.5. While ChatGPT-3.5's average score ranged from 47.4% to 56.5%, the students' average score was 76.8%. Second, GPT performed poorly on financial accounting, management accounting, and taxation questions involving arithmetic elements, but performed well on accounting information systems and auditing questions, which contain more conceptual elements. Third, while ChatGPT-3.5 struggled with short-answer questions and work-through problems, it performed well on true/false and multiple-choice questions.

Cheng et al. (2024) examined the extent to which ChatGPT-4 can answer accounting questions. They analyzed the AI's question-answering capabilities using ChatGPT-4 on seven accounting education cases. The results revealed three key points: First, generative AI performed well on tasks involving concept explanation, application of regulations and rules, and ethical judgments using existing frameworks. Second, generative AI performed poorly on tasks involving journal entries, financial statement preparation, and software usage. Third, ChatGPT-4 achieved an average score improvement of 22% compared to ChatGPT-3.5, reaching a level closer to student responses.

Furthermore, Eulerich et al. (2023) analyzed whether generative AI can pass accounting certification exams. Specifically, they verified whether generative AI could pass major U.S. accounting certification exams such as the CPA, CMA, CIA, and EA (enrolled agent). First, using ChatGPT-3.5 as the generative AI resulted in not passing any exam. Next, by employing prompt engineering—specifically, providing ChatGPT-4 with additional training using a small number of concrete examples and a mechanism for iterative reasoning and action using external resources—it passed all exams. Based on these results, Eulerich et al. (2023) argue that generative AI poses a threat to the auditing industry.

Furthermore, it has been reported that Deep Dean, an accounting-specialized AI, achieved a perfect score on the short-answer questions of Japan's Certified Public Accountant (CPA) exam. Specifically, by implementing additional training aligned with legal frameworks and accounting standards, it achieved a 100% accuracy rate across all subjects: Financial Accounting Theory, Management Accounting Theory, Auditing Theory, and Corporate Law. This result signifies that as AI evolves, it can reach a passing standard even on short-answer questions.

These results indicate that generative AI has reached a level capable of passing accounting-related qualification exams. This suggests the potential for utilizing generative AI in auditing. However, even if it reaches the level required to pass qualification exams, this does not directly imply generative AI will be useful in actual auditing practice. Therefore, the following section introduces prior research examining the potential for utilizing generative AI in auditing practice.

2.2 Potential for the Utilization of Generative AI in Audit Practices

How are new technologies like AI being utilized in audit practices? This point has also been discussed in several prior studies.

CAQ (2024) and KPMG (2024) state that AI can automate audit and financial reporting processes, enhance analytical capabilities, and provide higher-quality information. AI is thus considered to have the potential to improve the efficiency and effectiveness of audit and financial reporting. First, CAQ (2024) emphasizes that AI enhances the transparency, reliability, and relevance of financial reporting. It argues that AI can enhance the reliability of financial reporting and provide more useful information for stakeholders. For example, AI is thought to help verify the consistency of financial reporting data and identify potential errors or fraud. KPMG (2024) points out that AI enables real-time auditing, improved predictability, and a proactive approach, suggesting it could support continuous auditing, early risk warnings, and predictive analytics for fraud. AI enhances the efficiency and effectiveness of auditing while increasing the reliability of financial reporting. Specifically, it is argued that AI can continuously monitor the effectiveness of a company's internal control systems and provide early warnings of potential risks.

Fotoh and Lorentzon (2023) examine how shifting the paradigm from traditional auditing to digital auditing impacts the audit expectation gap. After discussing audit digitization, they argue that digital technology has the potential to strengthen internal controls and promote fraud prevention and detection, and that it can narrow expectation gaps on critical issues such as internal controls and fraud prevention and detection.

For specific uses of digital technology or AI in audit procedures, the IFIAR (2025) survey provides useful reference points. As part of its annual survey, IFIAR (2025) presents information provided by member bodies of the International Forum of Independent Audit Regulators (IFIAR), showing the frequency with which the use of Automated Tools and Techniques (ATT) are observed across various audit areas within the jurisdictions of IFIAR member authorities. The survey results highlight a trend toward increasing adoption of ATT in auditing.

For example, in 2024, 88% of IFIAR member authorities reported frequently observing the use of ATT in verifying journal entries (86% in 2023), and 55% reported frequent use in risk assessment (49% in 2023). Additionally, 43% of member bodies (33% in 2023) frequently observed ATT use in verifying internal controls, and 55% (51% in 2023) frequently observed its use in substantive audit procedures testing accounts such as revenue and inventory. While this survey was not limited to AI, it demonstrates ATT's use across various audit domains.

The Japanese Institute of Certified Public Accountants (2024) and the Financial Services Agency (2025) also discuss aspects of AI adoption within the audit industry. The Japanese Institute of Certified Public Accountants (2024) argues that AI use is advancing in simple tasks and auxiliary roles within audit work. Specifically, it anticipates increased AI use in: understanding the company and its environment during the audit planning phase; assessing risk evaluation; evaluating the status of internal controls during the internal control evaluation phase; performing substantive procedures such as corroborating evidence, witnessing, confirming, analytical procedures, and journal voucher testing; verifying disclosures; and conducting overall review during the audit opinion expression phase.

The Financial Services Agency (2025) reports that traditional AI is being utilized as a tool for anomaly detection in transaction and journal data, identifying and assessing fraud risks, searching internal corporate information such as audit standards and audit manuals, and converting documents into text (OCR). Regarding generative AI, it states that the introduction and consideration of tools to assist auditors, such as summarizing and translating documents, proofreading, and editing, is progressing.

Previous studies have discussed the potential for utilizing generative AI in audit practices. One characteristic of these prior studies is that they document the prompts input into the generative AI and the resulting outputs in their papers.

Gu et al. (2024) propose the concept of AI copilot auditing and attempt to present a framework for improving audit work using GPT-4. Using Chain-of-Thought prompting—a method that breaks down tasks into distinct thought processes—they divide audit tasks into six steps: 1) Task description, 2) Action description, 3) Input data description, 4) Output data description, 5) Input and output examples (for additional learning), and 6) Data input and execution. They then provide prompts for each step. They then illustrate its applicability to audit tasks such as financial ratio analysis, post-application review, and journal entry testing. Specifically, they report that the six-step process can detect anomalies in journal entries.

Fohr et al. (2023) focused on the EU requirement for ESG reports to comply with the EU Taxonomy and investigated ChatGPT's usefulness in the audit process. Applying the created prompt to actual companies yielded responses with a certain degree of validity.

Eulerich and Wood (2025) qualitatively organized the potential use of ChatGPT in internal auditing. They then provided specific prompts and responses for four internal audit processes: 1) developing a risk-based audit plan, 2) preparing for the audit, 3) conducting the audit, and 4) reporting.

Additionally, Emett (2025) details how the internal audit department at Uniper, a German energy company, utilizes ChatGPT across three stages—1) audit preparation, 2) execution, and 3) reporting—for Q&A, research, text analysis, and content creation, providing concrete prompt examples.

2.3 Impact of Generative AI on Auditing

AI holds significant potential to impact the fields of auditing and financial reporting. Its adoption is expected to play a crucial role in shaping the future of auditing and financial reporting. Empirical research on the effects of generative AI on audit quality, auditors, and audit firms is gradually accumulating.

First, regarding the potential impact of AI on audit firm recruitment, I introduce the research by Law and Shen (2025). Law and Shen (2025) analyze the debate over whether AI replaces or complements auditors' duties by focusing on the hiring of AI employees[1] at audit firms across the United States. Specifically, they use the staggered hiring of AI employees at audit firms across different U.S. regions as a proxy for AI adoption at regional audit firms.

The analysis utilizes résumé data from Revelio Labs and job posting data from Burning Glass. The résumé data identifies when auditors and AI employees were hired at each audit firm and the primary tasks performed by employees in each job category. Additionally, Burning Glass job posting data from 2010 to 2019 is used. The final main sample consists of panel data comprising 4,417 audit firm-years based on 648 audit firms across 163 audit firms from 2011 to 2019. The résumé sample size is 407,000.

[1] Law and Shen (2025) define AI employees as including workers who perform tasks such as data analysis.

Prior to hypothesis testing, Law and Shen present several interesting descriptive statistics regarding AI talent hiring. First, they report that AI job postings have consistently increased, doubling from 2010 to 2019. Furthermore, Big4 firms[2] hired AI employees much earlier than non-Big4 firms, with Big4 firms hiring approximately three times as many AI employees as non-Big4 firms in 2019.

Law and Shen (2025) further conduct empirical research on the characteristics of audit firms adopting AI employees, the impact of AI adoption on audit headcount, its effect on internal controls from the perspective of going concern errors, and its influence on restatements and audit fees.

First, regarding audit firms hiring AI employees, they identified three points. First point is that, audit firms belonging to Big4 firms or receiving higher audit fees started hiring AI employees earlier than other firms. Second, firms with higher employee turnover rates lagged in AI employee adoption. This result suggests that audit firms have less incentive to use AI when they need to hire new staff more frequently to replace departing employees. And third, it indicates that audit firms located in regions with higher education levels, larger populations, and younger demographics tend to hire AI employees sooner. The linear probability model reports that audit firms belonging to the Big4 are more likely to adopt AI employees.

Second, the study examines whether hiring AI employees increases the number of auditors. It reveals that hiring AI employees increases the number of auditors by 4.3%. However, it confirms that hiring AI employees does not increase the number of tax advisors or consultants, and while it increases the number of junior and middle-level auditors, there is no change in the number of senior auditors. The junior-to-senior ratio also increases. Furthermore, the study examined whether the skills required of auditors change after an audit firm hires AI employees. It defined soft skills as cognitive skills, efficiency skills, creativity skills, writing skills, social skills, customer service skills, management skills, people management skills, and project management skills. Conversely, it defined hard skills as general software skills, business systems skills, databases skills, data skills, machine learning skills, AI skills, automation skills, and RPA skills. Analysis confirmed that hiring AI employees requires not only cognitive skills like problem-solving, decision-making, and analysis, but also efficiency skills such as time management, task prioritization, and goal setting, along with customer service soft skills. Conversely, it revealed that while the need for general software skills has increased, other hard skills are not demanded. In other words, hiring AI employees resulted in a demand for auditors possessing soft skills, and as AI employees increased, so did auditors.

Third, it categorizes going concern errors into two types and analyzes material weaknesses in internal control. First, it defines a Type 1 error as a situation where a client receives a going concern opinion but does not go bankrupt within the next 12 months. Conversely, a Type 2 error is defined as a situation where a client does not receive a going concern opinion but does go bankrupt within the

[2] Refers to Deloitte, PwC, EY, and KPMG.

next 12 months. Furthermore, it defines an error concerning a significant internal control deficiency as a situation where a significant deficiency in internal control was predicted, but the client was actually judged to have effective internal control. The analysis reveals that hiring AI employees reduces both Type 1 errors concerning going concern and errors concerning significant internal control deficiencies.

Fourth, the study reports that hiring AI employees reduces the probability of non-reliance accounting adjustments (where the auditee cannot rely on previously issued financial statements, related audit reports, or completed interim reviews), mentions of 8-K Item 4.02,[3] and SEC investigations involving restatements. It also indicates no relationship between AI employee adoption and audit fees.

Law and Shen (2025) report these empirical findings and also present results from interviews with 11 partners, many of whom are affiliated with the Big4. First, they confirmed that the centralized approach to AI investment by audit firms combines strategic oversight through country-level decision-making with the flexibility to adapt and customize solutions to local firm needs, ensuring the effective and efficient integration of AI into audit practices. Second, they clarified that while AI tools are developed centrally, their successful implementation depends on decision-making and adaptation at the local firm level. Partners also responded that the composition of client companies has only a limited impact on investments in AI and human capital, and that AI does not significantly influence client selection. This is because audit firms generally adhere to consistent criteria when selecting clients, regardless of their level of involvement with AI technology. However, within each audit firm, the development of personnel who can effectively leverage AI in audit work is key. There is an accelerating trend to develop personnel capable of taking full responsibility regarding AI, and these individuals require intensive training and skill enhancement in AI and related technologies. Furthermore, interviews reveal that AI usage has not reduced audit staff numbers, that AI improves audit quality, and that despite efficiency gains from AI, audit fees do not decrease significantly. Reasons cited for stable fees include high initial investment costs for AI, ongoing compliance with audit standards, and increased fee demands.

Additionally, Fedyk et al. (2022) analyzed a unique dataset of over 310,000 detailed individual résumés from the top 36 U.S. audit firms to explore how AI impacts audit quality and efficiency, identifying AI employee hiring patterns within audit firms. Their analysis confirmed that AI employees are predominantly male, relatively young, and hold technical degrees. AI capabilities are concentrated within firms, focused on a small number of teams and regions. The study also examined the impact of AI investment on audit quality. Analysis revealed that a one-standard-deviation change in recent AI

[3] Section 4.02 of Form 8-K is part of the Form 8-K report filed with the U.S. SEC. It requires disclosure when circumstances arise that prevent reliance on previously published financial statements, related audit reports, or completed interim reviews.

investment reduces the likelihood of audit report reissuance by 5.0%. This result suggests AI contributes to improved audit quality.

Analysis of AI investment's impact on audit fees revealed that AI investment results in a 0.9% reduction in audit fees. This finding indicates that AI may enhance the efficiency of the audit process, leading to cost savings.

Furthermore, the analysis examines the impact of AI investment on employment. While AI investment could ultimately lead to the replacement of human auditors, it argues that this impact will take several years to materialize. Specifically, it reports that after AI investment, the number of accounting employees decreased by 3.6% after three years and by 7.1% after four years. These empirical findings are also supported by detailed interviews with 17 audit partners.

Fedyk et al. (2022) indicate that while AI improves the audit process, positively impacting both quality and efficiency, it may also have long-term effects on the labor market, such as a reduction in auditors.

The utilization of AI in auditing also presents risks and challenges, as discussed in "3. Issues Regarding the Use of AI in Auditing." Regulators, auditors, auditees, and other stakeholders need to collaborate to promote the appropriate use of AI and mitigate its potential adverse effects. The academic field must also focus on the ethical aspects of AI, the professional development of auditors, and the effectiveness of regulatory frameworks. Furthermore, research investigating the long-term impact of AI on auditing and financial reporting, and developing best practices for AI use, is crucial.

3. Issues Regarding the Use of AI in Auditing

Thus far, I have reviewed the potential for utilizing generative AI in auditing and its impact on audit quality, as discussed by regulators and in prior research. Below, I organize the issues and challenges that may arise when utilizing AI in audit practices.

3.1 Hallucination

The use of AI also carries risks to investor protection, market integrity, and financial stability (IOSCO, 2025). IOSCO (2025) points out challenges such as the black-box nature of AI models, inconsistency in outputs, bias, and lack of explainability. These risks can be amplified by inappropriate use or misuse of AI. For example, the black-box nature of AI models (where their internal operations are unclear) makes it difficult to understand AI decision-making processes and obscures accountability. Bias in AI models, reflecting biases present in training data, can lead to unfair outcomes for specific groups and raise ethical concerns.

CAQ (2024) and KPMG (2024) emphasize that data quality issues (inaccurate, incomplete, or outdated data) can negatively impact the reliability and accuracy of AI models. High-quality data is essential for training and validating AI models. Inappropriate data can degrade AI model performance and lead to erroneous conclusions. For example, an AI model trained on inaccurate data may make incorrect predictions, leading to erroneous audit judgments. Furthermore, model governance, ethical use, and the misuse of AI technology are also critical considerations (IOSCO, 2025). AI models must be developed and deployed in accordance with principles of fairness, transparency, and accountability. This is essential to ensure AI usage occurs within an ethical framework.

Beyond this, when utilizing generative AI in auditing, risks associated with hallucination become a key concern. This paper defines hallucination in the context of generative AI for auditing as the phenomenon where AI generates information that is not based on facts or is incorrect, yet presents it as if it were true. While hallucination is a risk associated with AI utilization beyond auditing, auditors are specifically required to conduct audits with professional skepticism and are expected to ensure audit quality and reliability. Therefore, if AI-generated hallucinations occur, they could lead to inappropriate judgments or erroneous conclusions, potentially resulting in serious consequences that undermine the very reliability of the audit.

The Financial Services Agency (2025), which clarified the actual state of AI usage among financial institutions, confirmed that these institutions face "hallucination" and "low response accuracy" as new challenges related to generative AI. Specifically, regarding new challenges related to generative AI, "hallucination" was cited most frequently by respondents (approximately 90%), followed by "low response accuracy" (approximately 50%).

Particularly in auditing, generative AI may misinterpret provided data or context, leading to erroneous conclusions regarding audit risk assessments or recommending inappropriate audit procedures. For example, AI might judge "this transaction is low in risk" when it actually suggests a significant fraud indicator. Furthermore, this could lead to errors in audit planning, collection of inadequate audit evidence, or overlooking material misstatements, potentially lowering audit quality.

Furthermore, generative AI may generate provisions that do not actually exist or that differ in content, while appearing to quote from specific documents or regulations. This risk is particularly heightened in the interpretation of laws and accounting standards. Additionally, errors in applying the laws and accounting standards forming the basis for audit opinions could lead to material defects in those opinions. Consequently, if audits are conducted based on erroneous information, material misstatements or fraud could be overlooked, potentially diminishing audit quality. Furthermore, should errors be discovered in audit reports or if AI-generated hallucinations cause harm to clients, audit firms could face regulatory scrutiny and potentially be held legally or ethically liable. If the utilization of generative AI leads to reduced accuracy in responses or diminished explainability of the output process, it could undermine market and stakeholder trust in audit firms and, by extension, in the audit process itself. Moreover, addressing the risks associated with generative AI utilization—such as the need for additional work to detect and correct hallucinations—cannot be ruled out as potentially reducing the overall efficiency of the audit process.

3.2 Development Costs, Audit Fees, and Audit Quality

Developing AI in house requires significant investment by audit firms. Whether the investment costs associated with AI development can be reflected in audit fees is also a point of contention.

The investment required to develop generative AI for auditing is not insignificant. For the Big4, one approach could be to develop generative AI globally that can be utilized for auditing and then promote its use across local firms. Smaller audit firms, however, face the issue of how to bear the development costs of generative AI given financial constraints.

The Financial Services Agency (2025) also presents three key issues surrounding AI investment. First, for both traditional AI and generative AI, predicting the return on investment (ROI) in advance is difficult, requiring significant time for internal consensus building. Second, due to the rapid pace of technological evolution in AI, even if an AI system is introduced, it may become obsolete in the short term. Third, while AI's effects may be limited in the short term, continuous learning can improve accuracy, potentially altering effectiveness over the medium to long term. These findings suggest that AI investment outcomes are difficult to predict, making it crucial to base decisions not solely on short-term perspectives but on a medium-to-long-term vision.

Regarding the issue of how to reflect the development and costs of new technologies like AI in audit fees, research by Austin et al. (2021) is useful. Austin et al. (2021) conducted interviews with managers and partner auditors at audited companies, as well as regulatory authorities, attempting to clarify how these stakeholders interact with data analysis and the impact this has on the adoption of data analysis in financial reporting.

Austin et al. (2021) argue that tensions have emerged between auditors and auditees regarding investments in data analysis tools and audit fees. Two of the interviewed auditors mentioned these tensions, with a common theme being that while audit firms incur costs related to investments in data analysis and technology, auditees expect these data analyses to reduce their audit fees.

Austin et al. (2021) argue that this stems from auditees believing audit fees are calculated purely based on audit hours, while audit firms fail to recognize the need to pass on R&D costs to clients. Furthermore, they state that auditee managers expect audit fees to decrease as data analysis enhances audit efficiency, particularly when combined with their belief that their own data analysis work promotes efficient auditing.

Additionally, Austin et al. (2021) introduce auditors' claims that the use of data and analytical tools by audit firms increases audit procedures. Specifically, by leveraging data analysis, auditors are more likely to identify exceptions and anomalies than with traditional audit procedures. Consequently, the likelihood of the "surprise-free" audit preferred by auditees[4] decreases, and auditors must understand why exceptions occur, leading to budget overruns.

Austin et al. (2021) also report that some auditee managers expect audit hours and fees to remain unchanged, despite predictions that data analysis will improve audit efficiency. This is because audit teams will allocate time to other critical audit activities that previously received insufficient attention.

However, Austin et al. (2021) reports that, overall, there is tension between auditors and auditees regarding audit fees. Auditors, focused on recouping investments in data analysis, expect higher fees. Conversely, auditees consider only the aspect of audit efficiency and reduced audit time from data analysis, expecting lower audit fees.

For large audit firms like the Big4, which invest in AI and can leverage it for audits, the key issue discussed in Austin et al. (2021) is how to reflect AI development costs in audit fees. Conversely, for smaller audit firms, investing in AI is not easily feasible. If the audit quality of large firms utilizing AI improves while smaller firms fail to advance their use of generative AI and thus see no improvement in their audit quality, this could lead to increased oligopolization by the large firms.

Currently, the number of listed companies audited by small and medium-sized audit firms is increasing. If audit quality improves at large audit firms capable of utilizing AI, while that of small and medium-sized firms unable to sufficiently invest in or utilize AI remains unchanged, this could lead to a disparity in audit quality driven by AI adoption.

3.3 Regulation

Regulations also impact audits utilizing data analysis, including AI, in various ways. Austin et al. (2021) derived three insights from interviews regarding the relationship between regulation and data-driven auditing.

First, the absence of clear regulations concerning data analysis is hindering its advancement. Auditee managers and auditors believe there is no clear, definitive guidance on the feasibility or permissible purposes of using data analysis in the financial reporting process. Both auditors and auditee managers feel that while standards assuming data analysis in audits are necessary, the lack of regulations is slowing its adoption.

Austin et al. (2021) argue that existing auditing standards were not developed with data analysis in mind and thus do not explicitly permit auditors to rely on data analysis for audit procedures. In other words, they further state that the standards even appear to discourage auditors from using data analysis. This indicates that both auditors and managers of audited entities believe auditing standards that incorporate data analysis are necessary.

[4] A "no-surprise" audit refers to an audit that is as expected for the audited company and contains no surprises.

Furthermore, Austin et al. (2021) reveal from auditor interviews that the absence of data analysis standards leads to instances of duplicated audit procedures. One auditor reported modifying procedures to perform one set of traditional audit tests for PCAOB inspections and a separate set of high-value-added data analysis tests. The auditors interviewed by Austin et al. (2021) stated this change was due to uncertainty about how inspectors would interpret procedures based on data analysis during PCAOB inspections.

Second, the absence of clear regulations on data analysis means auditors determine the scope of data analysis-based auditing based on inspection outcomes and informal feedback from regulators.

Third, the increasing use of AI, the rapidly evolving technology, and the diverse nature of data analysis make it challenging to develop useful regulations for leveraging data analysis in audits. Some managers at audited companies stated that, even when regulations exist, the ambiguity surrounding data analysis regulations makes it unclear whether their data analysis procedures meet the standards. Austin et al. (2021) found from interviews with regulators[5] that regulators are considering how to modify regulations to align with useful regulations for stakeholders and evolving data analysis techniques. However, they have not yet developed regulations that meet the expectations of auditors and auditees.

3.4 Approach to Auditees

Not only audit firms but also auditees are utilizing AI and data analysis across various accounting and auditing domains. For example, Tang et al. (2017) interviewed internal audit department heads from six for-profit companies and six non-profit organizations. They found that these departments consider the use of data analysis in internal audit work a high priority and intend to increase its use going forward. However, Tang et al.'s (2017) findings are from eight years ago, so progress in AI use is likely to be more advanced today.

The increasing use of AI by auditees may also change the relationship between auditors and auditees in terms of actual audit work.

Salijeni et al. (2019) discuss recent developments in audit technology based on interviews with 22 individuals possessing extensive experience in evaluating the development, implementation, and impact of big data and data analytics in auditing. Their findings indicate several changes in the auditor-auditee relationship.

First, there are changes in audit-related communication and enhanced value added. Big data and data analytics enable auditors to communicate audit findings in ways that are more understandable and perceived as adding value by the audited entity. They particularly highlight the usefulness of visualizing the thought processes behind the auditor's judgments regarding risk assessment.

---

[5] Austin et al. (2021) conducted six interviews related to regulators. The composition included the former Chairman of the U.S. FASB, the AICPA, a former PCAOB member, a CAQ member, and two anonymous individuals.

In addition, big data and data analysis demonstrate usefulness in resolving disagreements between auditors and auditees. They are particularly helpful in resolving differences of opinion between auditors and auditees in areas involving various judgments.

Furthermore, auditees are increasingly interested in developing systems that utilize analytics for decision making. While audit firms make significant investments in big data and data analytics, auditors frequently find opportunities to assist auditees in utilizing data for other business functions, particularly internal audit. Conversely, auditees are often reluctant to expose their ERP systems[6] to big data and data analytics due to concerns about data security and the potential use of their proprietary information for benchmarking, which could compromise their competitive advantage.

4. Countermeasures for Addressing Challenges in Ensuring Audit Quality in the Use of AI by Audit Firms

As discussed, various issues exist regarding AI utilization in auditing. Below, I discuss countermeasures.

4.1 Hallucination

To mitigate hallucination risks in auditing, the following countermeasures can be considered.

First, given the current accuracy of generative AI, it is essential that analysis results generated by such AI are always ultimately verified by human auditors to confirm their accuracy and validity. AI remains a tool; humans must retain responsibility for judgments concerning audit opinions and related matters. Interviews with several audit firms during the preparation of this paper confirmed that when generative AI is utilized in their audit work, human auditors re-verify the results based on their professional judgment.

Furthermore, hallucinations can be mitigated by providing clear and specific prompts to generative AI. Ambiguous instructions can cause AI to misinterpret information. Fine tuning the AI and using limited datasets to avoid overloading it with unnecessary information are also necessary. Data sources include publicly available data like auditing standards, the audit firm’s proprietary data, and data from the audited entity. When utilizing data containing personal information from the audited entity or others, careful consideration of privacy protection and information leakage risks is necessary. Fine tuning AI with audit-specific data and training it solely on accurate information regarding auditing standards and internal controls can reduce the risk of hallucinations.

Additionally, ensuring the transparency and explainability of generative AI is crucial. By clarifying the data and reasoning process underlying how generative AI arrives at its conclusions, signs of hallucination can be identified.

[6] ERP stands for “Enterprise Resource Planning,” referring to a system that centrally manages business data.

Furthermore, comparing and validating multiple generative AI models can be effective. Utilizing several different AI models or tools and cross-checking results is thought to enable the dispersion of hallucination risk.

4.2 Auditors in the Era of Generative AI

CPAB (2024) argues that auditors must verify that AI tools comply with auditing standards and regulatory requirements. Auditors must ensure AI tools are fit for audit purposes and do not compromise audit quality. For example, auditors must evaluate whether AI tools provide sufficient and appropriate audit evidence as required by auditing standards.

The introduction of AI will require auditors to not only fulfill their traditional roles but also to appropriately utilize AI tools and evaluate and judge their results (CAQ, 2024; IFIAR, 2025). CAQ (2024) states that auditors must confirm that AI tools are appropriate for the audit's purpose and scope. This means auditors must ensure AI tools are used appropriately to enhance audit quality. For example, auditors must evaluate whether AI tools provide sufficient and appropriate audit evidence as required by auditing standards.

IFIAR (2025), in its report on the use of technology in auditing, states that while AI provides powerful capabilities to audit teams, auditors are required to oversee its use and evaluate and judge its results. Auditors must develop the ability to effectively utilize AI tools and critically evaluate their outputs. However, this implies that AI implementation does not diminish the importance of auditors' professional judgment but rather enhances it. Auditors require advanced analytical skills and critical thinking to interpret results of AI tools and determine their suitability for audit purposes.

Auditors must manage the risks associated with AI use and adhere to ethical considerations (KPMG, 2024). KPMG (2024) emphasizes the importance of ensuring AI models to possess transparency, explainability, and fairness. To uphold the ethical aspects of auditing, auditors must verify that AI tools are used in accordance with ethical principles. For example, auditors must ensure AI tools are free from bias and guarantee the fairness of their results.

Additionally, auditors need to engage in continuous learning and professional development to keep pace with advances in AI technology (IFIAR, 2023). IFIAR (2023) notes that auditors need to participate in training and education programs to enhance their knowledge and skills regarding AI. This is crucial to enabling auditors to respond to the latest advancements in AI. For instance, auditors must acquire the technical skills necessary to effectively use AI tools and interpret their results.

Needless to say, improving auditors' digital literacy, including AI, is necessary. Appelbaum et al. (2021) argue that with the digitalization of auditing, auditors must learn about visualization, classification and regression analysis, association analysis, text mining, process mining, neural networks and deep learning, computer-assisted audit techniques, blockchain and smart contracts, and robotics. Appelbaum et al. (2021) also proposed changes to the Certified Public Accountant (CPA) exam to facilitate the application of data analysis in auditing.

In the United States, to address the advancing digitalization facing the audit industry and the changing roles required of auditors, the American Institute of Certified Public Accountants (AICPA) introduced a new CPA exam starting in 2024. Under this new exam system, called CPA Evolution, the structure of the exam subjects was first changed. It consists of a core set of required subjects and a discipline set of elective subjects. The core subjects comprise three courses: Auditing and Attestation (AUD), Financial Accounting and Reporting (FAR), and Taxation and Regulation (REG). The discipline subjects are Business Analysis and Reporting (BAR), Information Systems and Controls (ISC), and Tax Compliance and Planning (TCP). ISC assesses knowledge related to IT and data governance, assurance or advisory services concerning business processes, internal control testing, and information systems security. While AI itself is not an exam subject, it can be evaluated as an exam subject aligned with the digitalization of auditing.

4.3 Regulatory Authority Responses

Regulatory authorities are also closely monitoring the development and impact of AI, conducting assessments to understand the situation and consider necessary responses.[7] IOSCO has published a report on the risks and challenges associated with AI use and is soliciting feedback. This report emphasizes the importance of international regulatory cooperation and information sharing.

While coordinating globally, national regulators need to develop appropriate regulatory frameworks to manage AI risks and maximize its potential benefits. This is considered essential in ensuring AI usage is based on principles of transparency, fairness, and accountability.

For example, the U.S. audit oversight body, the PCAOB, is conducting outreach activities regarding the integration of generative AI in auditing and financial reporting, gathering feedback from stakeholders. The PCAOB is developing guidance for integrating AI into auditing standards and the audit oversight framework. This is essential for maintaining audit quality and reliability.

The UK's FRC has investigated AI's impact on actuarial practice, providing guidance and reports on model usage (FRC, 2019; FRC, 2023; FRC, 2024). The FRC has established guidance issued to promote the appropriate use of AI in actuarial practice and manage associated risks. This is crucial for fostering the ethical use of AI in actuarial work.

The Canadian CPAB (2024) provides best practices for auditors to effectively utilize AI tools in auditing, focusing on enhancing audit quality. CPAB also oversees AI usage to ensure audit quality and reliability. These are considered important initiatives contributing to promoting the responsible use of AI in auditing.

---

[7] Austin et al. (2021) report differing views between listed and unlisted companies regarding regulations on data analysis. Compared to listed companies, unlisted companies express less concern about formal regulations. However, the reasons for this divergence in regulatory views between listed and unlisted companies remain unclear.

The Financial Services Agency also strongly supports efforts for the sound utilization of AI within the financial industry, including the audit sector. Anticipating future constructive dialogue with the financial industry, it has published a discussion paper (Financial Services Agency, 2025) summarizing AI use cases and initial issues within financial institutions. The paper references the "risk of not taking actions," where financial institutions risk being left behind by technological innovation, making it difficult to provide high-quality financial services over the medium-to-long term.

AI technologies, including generative AI, remain in their early stages. Their future development is expected to further enhance the sophistication and efficiency of auditing and financial reporting (CAQ, 2024; KPMG, 2024). It is crucial for regulators to keep pace with the timely advancement of AI and other technological innovations.

International regulatory and standard-setting bodies are collaborating to assess AI's impact on auditing and financial reporting and develop necessary guidance and standards (IOSCO, 2025). IOSCO (2025) notes that international cooperation is essential to establish a consistent regulatory framework for AI use. This is vital to addressing AI's global impact. For example, international regulators should collaborate to establish common principles for the ethical use of AI and set international standards for its application.[8]

5. Conclusion

This paper discussed the potential benefits and audit quality challenges associated with the utilization of generative AI. Various perspectives on the potential benefits of generative AI were presented in prior research, reports from international organizations and regulators, and audit firms.

While the introduction of generative AI offers diverse benefits—such as streamlining audit procedures, deepening data analysis, and refining risk assessments—it is also clear that its implementation and operation inherently involve various challenges. Risks associated with AI generating erroneous information (hallucination), data privacy and security, and the "black box" problem of AI's opaque decision-making processes all carry the potential to threaten the reliability and independence of audits. Regarding the use of AI in auditing, including generative AI, some reports highlight the importance of auditors' expertise and suggest considering appropriate scope of use and quality control methods, given the risks that could impact audit quality (Japan Institute of Certified Public Accountants, 2024).

On the other hand, generative AI has the potential to be a powerful tool enabling auditors to efficiently analyze more data within limited time and resources, allowing them to focus on high-risk areas. For instance, the language processing and pattern recognition capabilities of generative AI hold immense potential for enhancing audit quality and coverage in areas such as anomaly detection,

[8] For ethical issues surrounding the use of new technologies, including AI, in auditing, see Munoko et al. (2020).

automating contract reviews, and evaluating the appropriateness of financial statement notes. Some prior research has demonstrated the potential for utilizing generative AI in audit practices even at this stage by presenting outputs generated from prompts entered into the AI.

Therefore, when introducing generative AI into auditing, it is essential to establish a robust governance framework and internal controls to appropriately assess these risks and mitigate them to an acceptable level. Specific measures include review procedures for AI outputs, data quality management, and efforts to ensure algorithm transparency. Furthermore, to advance the utilization of generative AI, it is considered necessary not only to establish guidelines for its use within audit firms but also to consider disclosure standards regarding the transparency of AI audit tools.

Furthermore, regarding AI investment within audit firms, key issues include how audit fees are set and support for the introduction and utilization of AI by small and medium-sized audit firms. Traditionally, audit fees have been determined based on the number of auditors, the time spent, and the hourly rate. Therefore, the introduction of AI into auditing raises the question of how costs associated with AI investment should be reflected in audit fees. This also prompts the need to reconsider the very nature of audit fees, including whether they should fundamentally be determined by the number of auditors and their hourly rates.

Assuming AI investment improves audit quality, disparities in the scale and quality of AI investment could emerge between large audit firms (capable of substantial AI investment) and other firms (i.e., small and medium-sized audit firms), potentially widening the gap in audit quality. The audit industry as a whole, including regulators, must discuss support for AI adoption and utilization by small and medium-sized audit firms, as well as responses to widening audit quality disparities.

Generative AI holds the potential to transform auditing, but its adoption should proceed with careful consideration of its impact on audit quality. By balancing the benefits and risks brought by generative AI, audit quality and efficiency could be enhanced, which will contribute to the enhancement of the reliability of capital markets.

References

Appelbaum D., D. S. Showalter, T. Sun and M. A. Vasarhelyi (2021), “A Framework for Auditor Data Literacy: A Normative Position,” *Accounting Horizons*, Vol.35 No.2, pp.5-25.

Austin, A. A., T. D. Carpenter, M. H. Christ, and C. S. Nielson (2021). “The Data Analytics Journey: Interactions Among Auditors, Managers, Regulation, and Technology,” *Contemporary Accounting Research*, Vol. 38 No. 3, pp. 1888-1924.

Canadian Public Accountability Board (2024). *The Use of Artificial Intelligence in the Audit –*

Balancing Innovation and Risk.

Center for Audit Quality (2024). *Auditing in the Age of Generative AI*.

Cheng X., R. Dunn, T. Holt, K. Inger, J. G. Jenkins, J. Jones, J. H. Long, T. Loraas, M. Mathis, J. Stanley, and D. A. Wood (2024). "Artificial Intelligence's Capabilities, Limitations, and Impact on Accounting Education: Investigating ChatGPT's Performance on Educational Accounting Cases," *Issues in Accounting Education,* Vol.39 No.2, pp.23-47.

Emett, S, M. Eulerich, E. Lipinski, N. Prien, Nicolo, D. A. Wood (2025). "Leveraging ChatGPT for Enhancing the Internal Audit Process—A Real-World Example from Uniper, a Large Multinational Company," *Accounting Horizons*, Vol.39 No.2, pp.125-135.

Eulerich, M., A. Sanatizadeh, H. Vakilzadeh, and B. Wood (2023). "Can Artificial Intelligence Pass Accounting Certification Exams? ChatGPT: CPA, CMA, CIA, and EA?," Working Paper, University of Duisburg-Essen.

Eulerich, M. and D. A. Wood (2025). "A Demonstration of How ChatGPT and Generative AI Can be Used in the Internal Auditing Process," *Journal of Emerging Technologies in Accounting*, pp.1-31.

Fedyk, A., J. Hodson, N. Khimich, and T. Fedyk (2022). "Is Artificial Intelligence Improving the Audit Process?", *Review of Accounting Studies*, Vol.27 No.3, pp.938-985.

Financial Services Agency (2025). "AI Discussion Paper (Version 1.0) - Preliminary Discussion Points for Promoting the Sound Utilization of AI in the Financial Sector."

Financial Reporting Council (2019). *Artificial Intelligence and Corporate Reporting: How Does It Measure Up?*.

Financial Reporting Council (2023). *The Use of Artificial Intelligence and Machine Learning s in UK Actuarial Work*.

Financial Reporting Council (2024). *Technical Actuarial Guidance: Models*.

Föhr T. L., M. Schreyer, T. A. Juppe, and K. Marten (2023). "Assuring Sustainable Futures: Auditing Sustainability Reports using AI Foundation Models," Working Paper University of Ulm.

Fotoh L. E. and J. I. Lorentzon (2023), "Audit Digitalization and Its Consequences on the Audit Expectation Gap: A Critical Perspective," *Accounting Horizons*, Vol. 37 No.1, pp.43-69.

Gu, H., M. Schreyer, K. Moffitt, and M. Vasarhelyi (2024). "Artificial Intelligence Co-piloted Auditing," *International Journal of Accounting Information Systems*, Vol. 54.

Igou, A., D. J. Power, S. Brosnan, and C. Heavin (2023). "Digital Futures for Accountants," *Journal of Emerging Technologies in Accounting*, Vol. 20 Issue 1, pp. 39-57.

International Forum of Independent Audit Regulators (2023). *Use of Technology in Audits-Observations, Risks and Further Evolution*.

International Forum of Independent Audit Regulators (2025). *Use of Technology in Audits—Observations, Risks, and Further Evolution*.

International Organization of Securities Commissions (2025). *Artificial Intelligence in Capital Markets Use Cases, Risks, and Challenges*.

Japan Institute of Certified Public Accountants (2024). “The Study into the Use of AI in Audits, Technology Committee,” Technology Committee Research Report No. 11.

KPMG. (2024). *AI in Financial Reporting and Audit: Navigating the New Era*.

Law, K. K. F. and M. Shen (2025). “How Does Artificial Intelligence Shape Audit Firms?”, *Management Science*, Vol. 71 No. 5, pp. 3641-3666.

Munoko I., H. L. Brown-Liburd and M. Vasarhelyi (2020). “The Ethical Implications of Using Artificial Intelligence in Auditing,” *Journal of Business Ethics*, Vol.167 No.2, pp.209-234.

National Institute of Standards and Technology (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*, NIST Trustworthy and Responsible AI, NIST AI 600-1.

Public Company Accounting Oversight Board. (2024). “Staff Update on Outreach Activities Related to the Integration of Generative Artificial Intelligence in Audits and Financial Reporting.”

Salijeni, G., A. Samsonova-Taddei, and S. Turley (2019). “Big Data and Changes in Audit Technology: Contemplating a Research Agenda,” *Accounting and Business Research*, Vol. 49 No. 1, pp. 95-119.

Tang, F., C. S. Norman, and V. P. Vendrzyk (2017). “Exploring Perceptions of Data Analytics in the Internal Audit Function,” *Behaviour & Information Technology*, Vol. 36 No. 11, pp. 1125-1136.

Wood, D.A., M.P. Achhpilia, M.T. Adams,... and E. Zoet (2023). “The ChatGPT Artificial Intelligence Chatbot: How Well Does It Answer Accounting Assessment Questions?”, *Issues in Accounting Education*, Vol.38 No.4, pp.81-108.

Financial Research Center (FSA Institute)
Financial Services Agency
Government of Japan

3-2-1 Kasumigaseki, Chiyoda-ku, Tokyo 100-8967, Japan

TEL：03-3506-6000
URL：https://www.fsa.go.jp/frtc/english/index.html