Explainability and Transparency of AI in Auditing

Vladislav Golubovskij
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT

This thesis investigates the extent of explainability, and transparency of AI systems used in auditing. As AI technology increasingly supports audit procedures the need for transparent and explainable systems has become critical. Using a qualitative approach this study conducted semi-structured interviews with audit company representatives from companies such as Deloitte, EY, and Rabobank. The research applies a coding framework based on Explainable AI (XAI) theory to evaluate six criteria: clarity of explanation, user comprehension, trust, model design, transparency, data provenance, and bias detection.

The findings of the research reveal that while AI systems offer valuable help in the process of audit tasks, their explainability is often limited to surface-level insights. Model design lacks depth for complex or high-risk engagements. Transparency is low, due to limited access to training data and model logic. This study concludes that AI systems are moderately transparent and explainable and are not technically sufficient enough to deliver a stand-alone value without human intervention. Auditors have to use AI as a supportive tool with close oversight, and for now, are not able to fully rely on only AI-generated outputs.

Graduation Committee members:

Dr. Ekaterina Svetlova, University of Twente
Dr. P. Khrennikova, University of Twente

Keywords

AI; Auditing; Big Four; Transparency; Explainability

During the preparation of this work, the author used Perplexity in combination with Google Scholar and Scopus as a research tool to find peer-reviewed academic sources; ChatGPT for grammar checks, and reference list readjustments; Grammarly was used for grammar check-ups, and sentences paraphrasing. After using these tools, the author reviewed and edited the content as needed and took full responsibility for the content of the work

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided properly cited. CC-BY-NC *the original work is*

1. INTRODUCTION

Integrating Artificial Intelligence (AI) into auditing processes is transforming the industry. Big four firms named Deloitte, EY, KPMG, and PwC – are at the forefront of this change. However, this research will focus only on Deloitte, EY and Rabobank. AI technologies, such as machine learning and advanced analytics, enhance auditing efficiency and accuracy by processing vast amounts of data, identifying anomalies, and detecting fraud (Mitan, J., 2024). AI tools can create evidentiary benefits that raise the bar of reasonable assurance provided by issued opinions, thereby fostering stakeholder confidence in audited financial statements and achieving higher audit quality. Moreover, AI tools like Natural Language Processing (NPL) automate data extraction from financial documents, enabling auditors to classify, verify, and analyze information more effectively and efficiently. Mitan, J. (2024). Another key principle of AI in auditing is risk assessment. AI enhances risk assessment by analyzing historical data, identifying patterns, and predicting potential risks or fraud. This allows human auditors to focus on high-risk areas or outlier cases. This allows companies to cut costs and redirect the saved time to another activity (Patel et al., 2023).

However, the increased reliance on AI also raises critical concerns regarding the transparency and explainability of AI tools. The "black-box" nature of many AI systems, where inputs and outputs are known but the decision-making process is "invisible," poses significant challenges for auditing, which requires high levels of accountability and trust (Zhong & Goel, 2024).

Explainability and transparency are two essentials for ensuring that AI-driven decisions are understandable and justifiable. In auditing, where professional skepticism, personal skills, and ethical standards are crucial, the lack of transparency in the AI model can hinder auditors' ability to verify results and maintain trust in the AI model's outputs (Mitan, 2024). One example is KPMG's survey, which highlights transparency as a top concern for AI implementation in financial reporting and auditing. This stems from the lack of disclosure about how AI models are built, trained, and optimized (Mitan, 2024). Similarly, PwC's Responsible AI Toolkit emphasizes the importance of transparent AI systems to ensure ethical and responsible use (Mitan, 2024).

The importance of transparent AI systems is underscored by their ability to build trust, ensure fairness, and comply with regulations. W. Hannah (2024). Explainable AI (XAI) and interpretable models are crucial for providing clear explanations of AI decisions, which is essential for auditors to rely on AI outputs as evidence. W. Hannah (2024).

1.1 Research question

Given the importance of transparency and explainability in AI driven audit processes, this research will focus on answering the following research question:

"To what extent are AI systems transparent and explainable in auditing?"

The adoption of AI systems in auditing must emphasize principles such as interpretability, transparency, and accountability. This framework enables organizations to navigate the complexities of AI adoption while ensuring that AI systems are trustworthy and aligned with established ethical standards.

1.2 Academic and practical relevance

This thesis will examine the primary challenges faced by Deloitte, EY, and Rabobank in adopting AI, with a focus on the critical issues of explainability and transparency. Through qualitative interviews with representatives from the companies and a review of existing literature, this research aims to provide insights into how organizations address these challenges and propose potential solutions on how to enhance transparency and explainability in AI-driven audit processes.

By examining these strategies, this study contributes to the understanding of how AI can be effectively integrated into auditing while maintaining high levels of explainability and transparency, which are fundamental to the auditing field (Zhong, C., & Goel, S., 2024).

2. LITERATURE REVIEW

The integration of Artificial Intelligence (AI) in auditing has transformed the field, offering enhanced efficiency and accuracy. However, this integration also raises critical issues related to explainability and transparency, particularly in complex decision-making processes. This literature review examines the benefits and challenges of auditing firms, with a focus on their experiences and the broader implications of auditing.

2.1 Primary Benefits of AI usage in auditing

The integration of AI in auditing has allowed for numerous benefits. Such as increased efficiency, accuracy, and overall quality of audit processes. These improvements are driven by AI's ability to analyze vast datasets and identify patterns, as well as anomalies that might have been overlooked by the human eye (Datsenko et al., 2024). Moreover, AI enables better task automation. It helps to automate tasks such as data entry and sample selection. Freeing auditors to focus on more complex and strategic aspects of audit. This shift in responsibilities not only improves job satisfaction and expertise among auditors but also allows them to allocate their time more effectively (Mulliqi, 2022). By automating routine tasks, AI enables auditors to concentrate on high-priority tasks that require their judgment and expertise. Enhancing the overall quality of audit Kokina, J., & Davenport, T. H. (2017). AI's real-time data analysis capabilities also play a crucial role in risk identification and management. By detecting anomalies and potential fraud indicators promptly, AI helps prevent financial and reputational damage. This proactive approach enables auditors to target their efforts more effectively, allocating resources to high-risk areas and thereby enhancing the overall effectiveness of risk management (Hu, K, et al., 2021).

Furthermore, AI provides valuable insights by analyzing vast amounts of data, revealing trends and anomalies that inform strategic decisions and help companies stay competitive (Datsenko et al., 2024). In addition to these operational benefits, the adoption of AI significantly contributes to cost savings. By reducing the time and resources spent on audits, firms can allocate their resources more efficiently and effectively. Research indicates that the use of AI is negatively related to audit fees (Kloosterman, 2021). Thanks to AI, organizations in a business environment can achieve additional efficiency gains, which is particularly important in today's fast-paced business environment, where cost management is crucial for maintaining competitiveness. Moreover, AI helps to strengthen compliance management by automatically checking audit findings against regulatory requirements, generating compliance reports, and alerting teams to potential issues. Modern AI auditing systems can

monitor the transaction log in real-time, sending alerts when suspicious activities are identified (Bello, N. O. A., & Olufemi, N. K., 2024).

Despite these substantial benefits, the adoption of AI in auditing is not without challenges. Issues such as AI's lack of transparency and explainability should be addressed.

The adoption of AI in auditing is a critical step towards enhancing audit efficiency, accuracy, and quality. By leveraging AI's analytical capabilities, auditors can improve risk assessment, fraud detection, and decision-making processes, ultimately transforming the auditing profession. As AI technology continues to evolve, its impact on auditing is likely to increase, making it an essential tool for auditors in the future.

2.2 Main risks associated with AI integration in auditing

Recent research highlights significant differences and a lack of consensus regarding the extent to which AI has been adopted in auditing. On the other hand, the integration of AI into auditing practices has been slow, partly because current auditing standards do not mandate the use of these advanced technologies (Fotoh & Lorentzon, 2021). Kokina and Davenport (2017) found that one of the primary barriers to implementing AI in audits is achieving consistency in data formats across clients. Another important feature of AI lies in the explainability of those models. The importance of explainability in AI lies in its role in building trust, ensuring accountability, and enabling regulatory compliance, particularly in high-stakes fields such as auditing (Kokina et al., 2025). According to the work of Kokina et al. (2025), transparency and explainability enable stakeholders to understand how AI systems arrive at decisions, which is crucial for identifying biases, verifying accuracy, and upholding ethical standards.

It is worth noting that the authors document differences in resources between Big 4 and non-Big Four firms. While Big Four firms dedicate substantial resources to auditor training, non-Big Four firms usually depend on off-the-shelf solutions and deal with shortages of IT-skilled auditors, alongside skill gaps in AI tool interpretation. The shortage of IT auditors, combined with the staff's inability to understand AI tool outputs, produces challenges that also include training and expense issues. Challenges related to the use of AI tools by humans lead to additional issues. Fedyk et al. (2022) mention that AI adoption faces its primary challenge due to insufficient training of human resources. Audits face impediments because auditors lack an understanding of the methods used to generate analytic outputs, as well as their unfamiliarity with AI tools. The absence of auditor involvement during tool development creates mistrust, as human auditors struggle to trust the outcomes of the AI model due to its black-box nature (Seethamraju & Hecimovic, 2022; Samiolo et al., 2023).

Considering all AI adoption challenges described in this section, the primary focus of this research will be centered on the explainability and transparency of the AI model in auditing. The concepts of transparency and explainability involve gaining an understanding of how an AI model functions and the decision process behind its outputs. These aspects present a significant challenge when dealing with AI tools that rely on machine learning technology. (Risks of Cognitive Technologies, n.d.). Transparency of AI systems refers to the clarity and openness of how AI systems function, make decisions, and produce outputs. It involves sharing information about datasets, processes and uses with shareholders to ensure understanding and trust (Dittmar, 2024). The explainability concept in AI refers to the ability to understand how an AI model arrives at its outputs or decisions. It focuses on interpreting the relationship between the systems, including the training data, learned components (such as weights and parameters), and the logic behind the prediction. Explainability ensures that stakeholders can understand the reasoning behind the AI-generated results Dittmar (2024).

Technological obstacles are closely related to the human factor, which significantly influences the adoption of AI in auditing. A study by Kokina et al. (2025) revealed that effective solutions often require a combination of heightened human attentiveness and adjustment for clear audit procedures.

2.3 Explainability and transparency evaluation methods

Auditors require transparent and explainable AI systems, along with ML solutions, to integrate them into their auditing procedures. These tools enable auditors to trust AI-based insights while relying on them, as they establish reliable interfaces for explainable AI systems that adhere to regulatory standards. XAI, also known as explainable AI methods, applies to supervised learning. They can be applied to tabular, textual, and image data. XAI techniques can be generally divided into two types: ante hoc techniques and post-hoc techniques. The idea of ante-hoc techniques is to directly adopt ML models that are inherently interpretable, such as decision trees and explainable neural networks.

In contrast, post-hoc techniques can be applied to any machine learning (ML) algorithm because they generate an explanation after the model is trained. When AI systems achieve advanced levels of complexity, they present challenges in demonstrating explainability, as numerous black-box algorithms function as uninterpretable systems. The assessment approaches for audit explainability are investigated through the SHAP and LIME methods, alongside general frameworks, in this section by Zhang et al. (2022).

For instance, Lime explains individual predictions of the model by approximating the complex model with a simpler, interpretable one. This method is beneficial for auditors who need localized explanations for specific transactions or accounts (Zhang et al., 2022). On the other hand, SHAP values are based on cooperative game theory and assign importance scores to features, explaining their contribution to a prediction. SHAP can provide both local-level interpretations (for individual predictions) and global-level insights (feature importance across all instances). Its theoretical foundation ensures consistency and fairness in explanations (Zhang et al., 2022). Both LIME and SHAP have advantages in auditing settings. LIME offers intuitive, human-friendly explanations tailored to individual cases, while SHAP provides robust insights backed by solid mathematical theory. However, limitations exist: LIME's results can vary depending on the choice of perturbed instances, while SHAP requires significant computational resources.

When it comes to transparency, models can be evaluated in several ways. First, Documentation reviews can be used to assess whether comprehensive documentation exists for data provenance, model architecture, training processes, and versioning records. This ensures accountability and traceability throughout the AI lifecycle (Zhang et al., 2022). Bias Detection tools, such as IBM's AI Fairness 360 or Google's Fairness Indicators, help auditors

identify biases in datasets or models that could compromise decision-making integrity (Zhang et al., 2022). Global Sourcing Models help to approximate black-box models using interpretable ones (e.g., decision trees), enabling auditors to understand how the model functions overall.

2.4 Explainable and Transparent AI in auditing

To understand and explain the degree of explainability and transparency of AI models used in auditing, Explainable AI (XAI) theory was employed. Based on the paper by Zhang et al. (2022). A qualitative framework was developed to assess the explainability and transparency of AI models through interviews with auditors. This framework focused on gathering insights about auditors' perceptions, experiences, and expectations regarding AI systems. Key dimensions of the framework were determined from the work of Zhang et al. (2022).

For transparency assessment, Data provenance, Model design, and Bias detection have to be evaluated. Firstly, it is essential to determine if the origin of the training data is disclosed (e.g., sources, collection methods), as analysis by Balasubramaniam et al. (2023) revealed that the importance of data training disclosure is considered an integral part of transparency and is highly valued by nearly all organizations. Secondly, all potential biases in data collection or labeling must be acknowledged, and clear guidelines must be established on the steps to address them (*A Call for Transparency and Responsibility in Artificial Intelligence*, 2023). Additionally, the AI system must align with audit evidence standards and ethical guidelines. These standards require auditors to obtain sufficient information and evidence in order to provide a reasonable basis for their opinion when performing audit procedures.

When it comes to explaining the system, it can be assessed based on three different criteria, one of which is the clarity of explanation, which describes the degree to which the explanation is context-specific to an AI model's outputs, making them understandable and actionable for auditors (Kroeger et al., 2022). The second criterion, which is also part of the explainability assessment framework, is user comprehension – described as the degree of explanation provided to help auditors make informed decisions. Usually, it is written in plain language to minimize misunderstandings (avoidance of technical jargon). Additionally, explanations of AI system outputs must highlight key factors influencing the decision made (e.g., "Your loan was denied due to low income") Shelf, (2024). Last but not least, on the list of criteria is Trust. Trust in an AI model can be seen as the variable that has to enhance auditors' confidence in the AI system. Trust is cultivated when explanations of the AI model align with auditors' "mental models." Surveys indicate that 54% of accountants believe that explainability enhances professional skepticism (ACCA, 2020).

Based on theoretical insights, it is evident that explainability and transparency of AI are integral and crucial in effective AI adoption in auditing. In the empirical section of this research, the degree of explainability and transparency in AI adoption within organizations will be examined while utilizing a developed framework based on XAI theory and transparency criteria

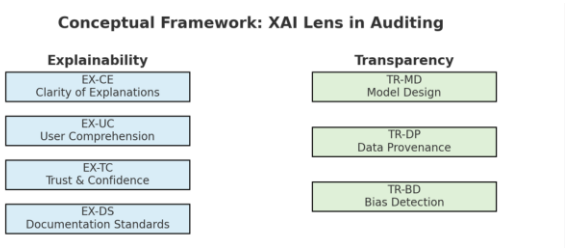

Figure 1.

3. METHODOLOGY

3.1 Research design

The research design of this study is qualitative, focusing on exploring organizational approaches to transparency and explainability in AI-driven audit processes. Qualitative research is particularly suited for studying real-world settings, providing rich insights into the research context (Yin, 2011). Data collection involves collecting primary data through semi structured interviews with representatives from various companies. This approach enables the gathering of detailed, firsthand information about strategies and challenges related to AI transparency and explainability. This approach is similar to studies that have used expert interviews to identify stakeholder specific requirements for explainable AI (XAI) in auditing (Zhong, C., & Goel, S., 2024).

An inductive approach was chosen, where the data collected leads to the emergence of themes and concepts (Yin, 2011). Initially, raw data from interviews will be collected. Subsequently, frequent, dominant, or significant themes will be identified, ultimately informing the development of recommendations for enhancing transparency and explainability in AI adoption within the industry. (Thomas, 2006). This approach aligns well with the research aim of understanding organization-specific experiences and deriving broader insights applicable to the auditing industry.

3.2 Interviews

3.2.1 Sampling approach

For the interviews, a sample of 5 company representatives was taken from the organization's population. Due to the limited sample size and the assumption that auditors from the company themselves would be most relevant to understanding the degree of explainability and transparency of AI models, a purposive sampling approach was chosen. Purposive sampling is a widely used non-probability sampling method in qualitative research; it is beneficial since it allows researchers to focus on specific individuals who are most likely to provide rich and relevant information (Palinkas et al., 2013). Additionally, this method is handy for in-depth exploration of a phenomenon, as it allows for direct access to information cases. For example, this approach helps select participants with specific expertise or experience in the AI auditing field, such as EY, Deloitte, and Rabobank employees who directly interact with AI systems daily. This approach enhances the credibility of the study by ensuring that data collection is focused on participants capable of providing valuable insights. Campbell et al. (2020).

To gather valuable and expert knowledge, a combination of a homogeneous and expert purposive sample was selected from the population. Where homogeneous sampling is a purposive sampling technique that focuses on selecting participants who share specific, predefined characteristics. By concentrating on commonalities among participants, homogeneous sampling minimizes variability, allowing them to focus on detailed analysis

and robust insights that are relevant to the research question Jager et al., (2017)

The primary goal of the interview was to gather expert knowledge and opinions on the degree of explainability and transparency of AI in auditing. A combination of both sampling methods was used. To ensure that interview participants were indeed experts in the field, specific criteria were established, which are outlined in Table 1 below. **Table 1. Criteria for selecting interview participants**

Criteria	Application Example
Job role and function	Select only people who are directly involved in audit activities
Experience with AI in auditing	Choose representatives which had at least 1-year hands on experience with AI
Involvement in AI implementation	Include those who contribute to evaluating AI system performance

3.2.2 Data collection

The primary aim of this research is to evaluate the level of transparency and explainability of AI models employed in auditing practices. To gain a deeper insight into this, semi structured interviews were conducted. Enabling a balance between guided questioning and open dialogue. This method provided the flexibility to explore individual perspectives while maintaining alignment with the core research topics. *Appendix A* shows the complete interview guide. Some examples of relevant interviews. 1. How clear are the explanations provided by the AI system when it flags anomalies or risks? 2. Do you find these explanations actionable for your audit tasks? If not, what improvements would you suggest? 3. How do the explanations provided by the AI system align with audit documentation standards? 4. Does the AI system's ability to explain its decisions affect your trust in its outputs? Why or why not? 5. Do you feel that the AI system is open about how it operates (e.g., its algorithms, data sources, and decision-making processes)? 6. What kind of information about the AI system would help you feel more confident in using it during audits?

3.3.3 Data analysis

After the interviews, transcribed data had to be analyzed. For that purpose, the coding table was chosen. Since coding tables are a foundation tool in qualitative research, they offer several key benefits. First of all, coding tables allow for the systematic organization of qualitative data, such as interview transcripts. By categorizing data into clear codes and themes, the coding table enabled better sorting and comparison of data from the interviews. This systematic methodology not only streamlines the analysis process but also ensures that all relevant data points are considered, supporting a comprehensive exploration of the research question. Sharp, C. A. (2003).

Additionally, utilizing a coding table enabled better pattern identification, similarity tracking, and the highlighting of discrepancies within data sources. The coding table enabled the identification of dominant themes related to the explainability and transparency of AI models and explored the relationship between them. This process not only aids in synthesizing findings but also provides a clear rationale for how themes and interpretations were derived, contributing to a more nuanced and insightful analysis (Sharp, C. A., 2003).

For each interview response, corresponding codes are assigned to the six criteria (three for transparency and explainability. Also, shorthand code notations are used (e.g., TR-DP for transparency data provenance, EX-CE for explainability and clarity of explanations, etc.). Each question from the interview is mapped to one or more criteria. For example, if a question asks, "How does the AI system provide the explanations?" It would be possible to map the response to both clarity of explanations and user comprehension. Also, for each of the data points, "assessment" criteria are assigned. It is a three-level criterion: "Negative," "Neutral," and "Positive." This is done for better topic understanding, data evaluation, and easier finding interpretation. The use of a coding table enables a more comprehensive summary of responses for each criterion across interviews, highlighting common challenges mentioned by respondents. An example of a coding sheet is presented in *Appendix B*.

4. RESULTS

The following section discusses the results obtained through the interviews. The aim of conducting interviews with audit experts was to determine the degree of explainability and transparency of AI systems used by auditing companies. The results will be presented in the coding sheet see *Appendix B* and interpreted in the following paragraphs.

The findings from this analysis offer insight into the degree of explainability and transparency of AI systems used in auditing based on criteria derived from the relevant theories.

4.1 Explainability

Explainability refers to the degree to which an AI system can articulate the reasoning behind its outputs in a way that is understandable to human users. It involves presenting interpretable justifications, data triggers, or decision pathways that allow auditors to grasp why and how a specific output was generated. This aligns with literature from Zhang et al. (2022), which emphasizes explainability as a requirement for trust, user comprehension, and professional judgment.

4.1.1 Clarity of explanations (EX-CE)

The majority of participants agreed that clarity of explanations of AI models was the main reason for usability in their audit work. Simple and repetitive tasks, such as reviewing account balances, were the primary reason most participants agreed that the explanations about AI models were sophisticated enough for their usability in audit work. AI effectively handles repetitive and straightforward tasks, such as reviewing account balances and identifying anomalies in reports. These models, trained explicitly for auditing purposes, also provide guidance on what auditors should check. However, P1 emphasized that when it comes to interpreting key data or complex information, AI does not provide enough context and may omit important details. P3 explained how AI highlighted "heavy cash usage" by mentioning the specific data trigger that led to the alert, such as "a $10,000 cash deposit made twice in one day." This ensured that the bank paid immediate attention to the matter. P3 noted that these strong data warnings are typically not definitive and require further examination, considering the client's particular conditions and acceptable risk levels. P2 mentioned that while some points were clear, especially for easily identifiable anomalies, others were not. Certain cases appeared like automated warnings without enough information to

understand the reasoning behind them. As a result, P2 stated that determining the rationale for each finding often takes hours. Overall, participants found that clear, data-based explanations, particularly those that specify the precise metric or threshold that triggered the alert, significantly improved the usability of the models. Nonetheless, AI's responses often remain too broad for many complex cases, requiring auditors to use their expertise and verify the original records before acting on any advice.

4.1.2 User Comprehension / Actionability (EX-UC)

Across all five participants, AI-generated explanations were generally viewed as valuable "first drafts" that guide the audit workflow, but none felt they were fully actionable without human intervention. P1 explained that for routine work such as reconciling account balances or checking journal-entry consistency, the system's explanations are "straightforward and usable," yet when it comes to more nuanced risk assessments (e.g., assessing the materiality of a variance), "the AI's rationale feels too high-level" and often omits the contextual details needed to decide on next steps. In P1's experience, this means taking the AI prompt as a starting point: "It points me in the right direction, but I always drill down manually" because the explanations alone do not map cleanly to specific audit procedures. P2 echoed this pattern by saying that AI outputs are "reasonable and reliable" for flagging apparent anomalies but "a little bit … too conservative," which results in numerous "false positives" that must be pruned. While the AI flags risk with a clear summary of why it flagged them (e.g., "unusual vendor payment patterns"), P2 still spends significant time deciding which flags warrant deeper investigation because "the explanation itself does not always clarify materiality or likelihood." P3 explained that even with clear data triggers, such as identifying cash anomalies, "the system tells me what to ask next, but I decide which lines of inquiry to pursue, based on client history and risk tolerance." Thus, while P3 found these explanations highly actionable, they still require interpretation through a domain-specific lens. P4 commented that AI-generated summaries of contract clauses or control-test results are "often quite accurate" but "lack the nuance of industry- or region-specific guidance." For instance, when the AI summarizes a lease-accounting query under IFRS, it may omit local tax consequences that P4 knows to be relevant. "It gets you 70–80 percent of the way there, but if you just took it at face value, you would miss those details—and that could change my audit approach." Therefore, P4 uses the AI output as a template: "I will copy it into my work papers, but then I edit to include the footnotes or policy references our firm requires." Lastly, P5 described relying on AI for initial drafting, particularly when summarizing accounting standards updates or generating lists of key control objectives, but "validates every line against official pronouncements." While explanations typically include "where they pulled the guidance" (e.g., citation to a FASB paragraph), the AI occasionally "forgets to add the reference" or misquotes a section, necessitating a manual double-check before sharing with senior reviewers. AI outputs commonly cover the "main points" needed for work papers but fall short of formal audit-reporting requirements without manual enhancement. P4 noted that while AI can cover "basic things" and reduce the drafting burden of manual tasks, P4 still adds notes and checks regional or industry-specific standards. P3 found AI adequate for documenting meetings and initial summaries, but it required "tweaking" when drafting final reports or findings to meet formal write-up standards.

4.1.3 Documentation Standards Alignment (EX-DS)

AI outputs commonly cover the "main points" needed for work papers but fall short of formal audit-reporting requirements without manual enhancement. P5 explained that although generated references align "to some extent" with standards, it is a must to verify citations to guard against outdated or nonexistent guidance. Otherwise, work papers risk citing invalid or superseded pronouncements. P4 also states that while AI can cover "basic things" and reduce the drafting burden of populating work paper templates (for example, inserting control objectives or summarizing standard-setting updates), P4 still adds notes and checks regional or industry-specific standards such as agricultural versus banking audit requirements to ensure completeness and compliance with current audit frameworks in each field or country.

Similarly, P3 found AI adequate for preliminary checks but requiring "tweaking" when assembling final work papers or narrative findings to satisfy formal write-up standards. Often needing to insert firm-mandated wording, detailed risk assessment rationale, or explicit linkage to assertion-level objectives to render the work papers "audit-ready." Together, these perspectives underscore that while AI can accelerate the initial population of work paper sections by identifying key headings, control points, or data triggers, auditors must still manually refine and validate the content to align with detailed documentation standards, ensuring that work papers not only include essential facts but also adhere to firm policies, regulatory requirements, and the explicit formatting conventions that underpin a compliant audit file.

4.1.4 Trust and Confidence (EX-TC)

Trust in AI outputs hinged on the presence of clear, logical explanations and the auditor's ability to verify them. P2 emphasized that without a "clear explanation" for why a flag or recommendation was generated, "I would not trust it," and as a result, always performs a human double-check before acting on any AI finding. Likewise, P5 relies on AI for initial references but "validates every line" against official pronouncements, recognizing that "AI can give you certain details that are not really in there," so "it is a must always to double-check the outputs of the model" and pay extreme attention to the explanations provided. At Rabobank, P3 explicitly distinguished between simple anomaly-detection models where trust is relatively high because the underlying rules are transparent. Moreover, large language models (e.g., ChatGPT or Deep Seek), whose outputs P3, are "more questionable" given their black-box nature. In P3's view, anomaly-detection tools that point directly to the specific data trigger (e.g., an unusual transaction pattern) earn "a level of confidence" by their clear data-driven logic.

In contrast, LLM-based summaries lacking an easily traceable chain of reasoning always require additional validation. P1 and P4 expressed similar opinions. They both noted that when explanations include concrete data points or direct references to policy text, they feel comfortable "leaning in" on the AI recommendation. However, neither ever relinquishes final authority to the system. P1 remarked that even when a GPT-style tool identifies the exact figure behind a variance, "I still check the raw ledger myself" to ensure nothing was misinterpreted. P4 agreed that a well-structured explanation "builds confidence" but "never leads me to skip my review" since P4 knows that an AI summary might omit critical nuance, especially around industry specific guidance.

4.2 Transparency

Transparency refers to the extent to which the inner workings, data sources, model architecture, and assumptions of an AI system are visible and accessible to users and stakeholders. It includes knowledge of how the model was trained, what data it uses, and what processes it follows internally. As described by Balasubramaniam et al. (2023) and Deloitte (2023), transparency is a broader system-level property that complements explainability by enabling traceability, regulatory review, and ethical oversight.

4.2.1 Model design (TR-MD)

Participants generally welcomed high-level overviews of AI architectures but expressed frustration at the lack of deeper technical insights that would enable them to understand how these tools reach their conclusions fully. P1 highlighted that GPT-style tools provide a basic level of transparency by allowing auditors to ask follow-up questions such as "Which data points influenced this output?" thereby offering a glimpse into the model's decision-making process. This feature gives P1 confidence that there is at least some traceability behind each recommendation. For instance, understanding that a particular journal entry inconsistency was flagged because of a mismatch in invoice amounts. Nonetheless, P1 contrasted this openness with the experience on proprietary audit platforms, which "do not let you do that." Without the ability to query underlying data sources or algorithms, P1 finds it "more difficult" to trust how these closed systems arrive at their risk flags or suggested audit procedures. P4, who has a programming background and, therefore, some familiarity with pattern-analysis techniques, also underscored the discomfort that arises from black-box components. Even though P4 "sees the value" in AI tools identifying outliers or summarizing contract clauses, P4 "does not see details like which statistical model or which training data was used." For P4, this gap means that while they might recognize that a machine-learning classifier has detected an unusual vendor payment pattern, they cannot be sure whether that classifier was trained on a representative dataset or which modeling technique it uses. P4 further worries that overly technical disclosures—like hyperparameters or feature selection—could overwhelm non-technical auditors, yet believes that without some understanding of model architecture, it is hard to assess relevance or generalizability to specific audit contexts.

Meanwhile, P3 described a more formal approach to managing model design transparency at the organizational level. Their firm has adopted a multi-stage vetting process: a committee reviews architecture and data flow, and compliance teams assess models for bias and data leakage before deployment. For P3, this is essential to ensure audit models do not compromise client confidentiality or regulatory standards. These experiences illustrate the balance auditors seek—between transparency that is detailed enough to validate outputs, but abstract enough not to burden everyday audit work.

4.2.2 Data Provenance (TR-DP)

Across all participants, traceability from raw inputs to AI outputs emerged as a persistent concern, reflecting the gap between auditors' expectations for complete data lineage and the practical limitations of current systems. P3 stressed that to "avoid any data leakage," the organization must know exactly which client or third-party sources feed into its models, particularly when opensource tools are used. In P3's experience, failing to map every data source into the model's training or inference pipeline risks exposing sensitive client information to unauthorized repositories. P5 lamented that AI outputs sometimes cite standards or regulations that have since been repealed, forcing P5 to perform manual back-checks against up-to-date regulatory databases. For example, P5 recounted an AI-generated summary that referenced a 2018 version of a banking regulation, which had been superseded in 2022; this "misleading citation" compelled P5 to cross-reference every AI-generated standard with the firm's internal update tracker, significantly eroding efficiency. According to P5, these lapses occur because AI models are often trained on static snapshots of policy documents. Without a mechanism to flag or refresh outdated content, auditors cannot reliably depend on the provenance of referenced guidelines.

By contrast, P1 observed that GPT-style interfaces at least enable a minimal form of provenance transparency. Users can query, "Which document did you use to arrive at this conclusion?" and receive the title or URL of the source text. While this feature does not provide a complete lineage or data-flow diagram, it gives auditors some assurance that the AI is not inventing rules "out of thin air" and allows them to pull the original document for verification. Nonetheless, P1 noted that proprietary audit platforms rarely offer this query capability, leaving auditors "in the dark" about which internal or external datasets inform a given risk flag.

Consequently, auditors across firms rely heavily on the quality of explanations—clarity about why a decision occurred rather than on a complete provenance trail. P2, P4, and P3 all echoed the sentiment that, in the absence of complete lineage, they must judge the trustworthiness of an AI output by how convincingly it articulates the logic behind a finding, even though "a well worded explanation may mask gaps in actual data sourcing." In practice, this means that auditors frequently revert to manual data tracing, pulling transaction logs, regulatory codices, or source ledgers. To confirm the AI's assertion, rather than depending on the model to document its data path.

In sum, while some level of provenance can be obtained through metadata queries in GPT-style tools or firm-mandated compliance reviews, auditors identify a clear gap between the transparency they desire (a whole, auditable trail from raw input through feature engineering to final recommendation) and the reality (opaque or static data snapshots with no real-time lineage verification). This disconnect intensifies the need for manual checks, undermines some of the efficiency gains AI promises, and highlights data provenance as a critical area for future improvement in audit-focused AI systems.

4.2.3 Bias & Fairness (TR-BD)

Participants highlighted both procedural safeguards and model level biases, underscoring the tension between formal governance frameworks and the subtle ways in which AI can skew audit outputs. P3 described a multi-layered governance framework at Rabobank in which every proposed gen-AI deployment must pass through a first-line review committee and a second-line compliance team. These groups apply comprehensive biasdetection checklists to examine training data for representativeness and ensure that no unintended demographic or transactional skew could taint the results before approving any model for live use. This process, P3 argued, is essential to "prevent any data leakage or inappropriate profiling," effectively embedding bias mitigation into the approval workflow rather than relying solely on post-deployment monitoring.

By contrast, auditors interacting with AI in day-to-day tasks reported encountering more nuanced biases within the models

themselves. P2 noted a pronounced conservative bias: the system tends to flag "every risk" as significant, compelling auditors to "tweak many things" to avoid over-flagging trivial issues. This thereby increases the audit team's workload and sometimes obscures genuine material concerns. Beyond risk-flagging, P2 also observed a confirmation bias when using large language models like ChatGPT: filtered to "be helpful," the AI often "immediately confirms your ideas" rather than acting as a critical sparring partner, leading P2 to suggest a "sparring-mate" mode that would proactively surface counterarguments instead of simply echoing preexisting beliefs.

P4 voiced skepticism that vendors' fairness claims are more than marketing. Although vendors assert that their models are tested on balanced datasets and subjected to bias checks, P4 "has not seen proof" of these assertions, no detailed fairness-audit reports or third-party verifications, leaving uncertainty about whether equitable outcomes are genuinely enforced or merely advertised. This lack of visible evidence, P4 argued, undermines confidence in the AI's ability to treat all clients or transaction types uniformly, especially in complex engagements like industry specific compliance audits.

By contrast, P1 reported never encountering identifiable biases in the firm's AI tools, attributing this to the "heavy oversight and regulatory testing" required before any rollout. Nevertheless, P1 acknowledged that neither P1 nor colleagues routinely scrutinize AI outputs for bias in everyday use, so latent issues could remain undetected unless formally audited. Similarly, P5 reported no clear bias when using AI for references and summary tasks. However, P5 stressed that validating AI outputs is crucial precisely because "you never know when a subtle data skew might creep in," reiterating that human review is the final safety net.

Across all firms, then, formal governance frameworks coexist alongside model-level biases conservatism in risk detection and confirmation effects in narrative summaries that auditors must actively manage through human oversight and validation. While institutional committees and compliance checklists help mitigate overt biases before deployment, the day-to-day interaction with AI continues to reveal a subtler skew that auditors must recognize and correct to ensure fair and reliable audit outcomes.

5. DISCUSSION

This study explores the extent to which AI systems used in auditing are transparent and explainable. The findings indicate that while AI integration in auditing has progressed, the current level of explainability and transparency is limited.

Based on the interviews explainability level was frequently indicated as intermediate. Although systems may flag anomalies or present results through interfaces, these outputs frequently lack the depth, audit-context relevance, and interpretive clarity necessary for auditors to fully understand or justify the system's conclusions. In many cases, explanations are either too technical or too generic, making them insufficient for stand-alone use in documentation or professional judgment. This confirms concerns raised in the literature, which emphasize that without actionable and domain-specific explanations, AI systems risk becoming black-box tools that hinder rather than support accountability in auditing.

Transparency, while somewhat more established, is limited by system complexity and confidentiality. Auditors do not have information about the model's inner workings, and data it was trained on, or how decisions are internally generated. This aligns with the literature 's view that the transparency of AI goes beyond output visibility. It includes access to model logic, data lineage, and operational assumptions. But in practice, access to such areas is not granted to the auditors, especially if third-party solutions are involved. Auditors know almost nothing about the inner workings of the models. However, based on the responses it was indicated that a better understanding of the model's inner workings or training data set disclosure would increase the trust and reliability of the model in the eyes of auditors.

To sum up, the extent to which AI systems in auditing are transparent and explainable is moderate. The systems serve a support and assistance role for auditors but should be closely overseen by the auditor. Right now, current existing systems are not able to provide the level of clarity, interpretability, or openness needed for high-stakes audit decision-making without humans in the loop. To improve the transparency and explainability of the model AI tools must be co-developed with practitioners and adhere to transparent design principles such as trying to be clear about inner algorithms, and relevant training data set disclosure. As well as embedding explainability that aligns with both regulatory expectations and the practical demands of the audit process.

5.1 Theoretical implications

This study advances theoretical discourse on explainable and transparent AI (XAI) by empirically validating how auditors perceive and interact with AI systems through six core dimensions: clarity of explanation, user comprehension, trust, model design transparency, data provenance, and bias detection. As outlined by Zhang et al. (2022), explainability must be tailored to user needs particularly in professional contexts where decisions carry regulatory weight. The interviews demonstrated that while auditors appreciated AI-generated summaries for routine tasks, they frequently found the explanations too generic or lacking the contextual specificity required for more complex judgments. This affirms Kroeger et al. (2022), who argue that explainability must be context-sensitive and readily understandable to end users. Moreover, the ability to comprehend AI outputs in plain, jargon-free language was repeatedly emphasized as crucial for enabling informed decision-making, in line with Shelf (2024), who stresses that comprehension is foundational for meaningful auditor interaction with AI tools.

Trust, another key dimension, emerged as dependent on the presence of clear and logically structured explanations that align with the auditor's expectations and mental models. This supports the findings of ACCA (2020), which showed that over half of audit professionals believe that explainable AI enhances professional skepticism. On the transparency side, interviewees highlighted persistent issues with understanding how models operate and what data sources they rely on. The inability to trace AI outputs back to source inputs, or to assess potential biases, undermines confidence in model integrity. These concerns directly reflect the work of Balasubramaniam et al. (2023), who underscore that data transparency—including the disclosure of training data origins and bias mitigation steps—is foundational to ethical AI deployment.

Additionally, findings indicate that AI systems used in audit lack technical depth. Auditors were typically not informed about the specific algorithms, model logic, or feature engineering processes behind AI generated outputs. This lack of visibility reduced Trust in the systems. Moreover, based on the interviews current AI systems are unable to trace AI outputs back to original data

sources. Interviewees cited cases where outdated or incorrect regulatory references were cited by AI. Which resulted in manual validation of AI generated output. This undermines one of the core expectations of transparency: being able to understand and audit the AI input-output chain. The results show that AI transparency in auditing is currently limited and inconsistent. While some features (e.g., quarriable explanations, basic model summaries) offer surface-level transparency, deeper insight into model logic, data provenance, and bias management remain restricted. This confirms what theory predicts: transparency requires more than visibility—it requires meaningful access to systems, data flows, and assumptions (Balasubramaniam et al., 2023; Deloitte, 2023). In practice, auditors often operate in a black box environment, relying on professional skepticism and manual controls to compensate for system-level opacity.

Overall, the study contributes to theory by grounding abstract XAI concepts in the real-world constraints and expectations of the auditing profession. It shows that for AI systems to be truly usable in high-accountability environments, explainability and transparency must be operationalized through interpretable outputs, source traceability, and alignment with professional judgment extending the theoretical literature into a domain specific, practice-based context. Therefore, this research has contributed to existing literature on explainability and transparency of Ai in the audit. The findings from this study can be used as input for further research.

5.2 Practical implications

The findings of this study offer several practical implications for auditing firms seeking to adopt or improve AI systems. Based on the interview responses and supported by the literature, it seems clear that the "clarity of explanation" of AI models must be prioritized. AI systems that are used for auditing should generate outputs that are not only accurate but are capable of clearly explaining the reasoning that led to a specific conclusion. As revealed in the interviews, auditors are more likely to act on Ai generated insights when they include identifiable triggers—for example, specific pattern deviations and transaction thresholds, rather than generic summaries.

Secondly, to improve user comprehension, organizations should ensure that explanations are communicated in plain, audit relevant language, avoiding technical jargon. This also aligns with Shelf (2024), who emphasizes the importance of language accessibility in fostering informed decision-making. Furthermore, firms should establish clear guidelines on how AI outputs can be effectively integrated into existing audit documentation procedures, including when and how human review should supplement AI-generated material.

Regarding trust, all auditors emphasized the need to verify AI outputs before taking any action, highlighting the importance of model outputs that reinforce but do not replace professional skepticism. One of the good suggestions for firms would be to start implementing training programs that help auditors understand AI decision logic. Thus, strengthening confidence and alignment with mental models. Additionally, developing domain specific, industry-based models for auditing purposes would enable increased transparency and better alignment with existing regulations, as well as improve visibility and data provenance. This includes disclosing the origin of training data and any known limitations or biases. This appeared to be one of the practices employed by an auditing company. However, most companies still use open LLM, which limits their ability since it is not allowed to attach sensitive information to those models. Moreover, bias detection and fairness checks should not remain isolated within development teams. However, they should be embedded in the audit workflow, enabling frontline auditors to flag anomalies and assess risk more effectively.

6. LIMITATIONS

While this study offers valuable insights into the role of explainability and transparency in AI auditing, several limitations must be acknowledged.

First of all, this research was based on interviews with five audit professionals, which came primarily from large international companies. While participants offered valuable expert knowledge. It is important to mention that a small sample limits the study's ability to capture a wider variety of experiences. Especially from mid, and small-sized firms. As such, the findings of this study represent practices and concerns of technologically advanced, resource-rich auditing environments and may not be representative of the broader industry.

Additionally, participants were drawn mainly from Big Four firms or similar, where the integration of AI is more advanced. Those organizations are more mature and richer in resources. This allows for better AI integration, traineeship programs for the employees, and more standardized procedures compared to mid and small-sized firms. As a result, the challenges and perspectives captured may underrepresent the struggles faced by firms at earlier stages of maturity.

Secondly, interviewees were professionals who had direct involvement with the responsibility of AI usage. This could have introduced some potential bias. Since the response may have described AI adoption in a better light. Meanwhile, downplaying risks, errors, or some implementation failures. Additionally, given that all interviewees are current workers at the active job position it could have had a direct link to a social desirability bias. Which may have influenced the participant to frame their responses in ways aligned with organizational norms rather than personal opinions. This is only a potential theoretical assumption that might have a place.

Last but not least, it is important to understand that AI is a rapidly growing field of technology that is evolving fast. As such, some findings may quickly become outdated as new tools with more advanced explainability, or transparency features might be introduced shortly. This temporal limitation means that conclusions drawn from current practices may not fully reflect the state of AI tools in the near future.

7.FUTURE RESEARCH RECOMMENDATIONS

Building on the limitations identified in this study, several promising avenues for future research emerge that could enhance understanding explainability and transparency of AI in the audit. Derived from the limitations, future studies should expand beyond large multinational audit firms and also include small and mid-sized adult firms. This will allow for a better representation of how resource constraints, organizational culture, structure, and regulatory requirements affect the ability to implement explainable AI. Comparative research across different firms could reveal potential information about how different constraints and resources shape transparency standards and expectations. Additionally, this research focuses solely on auditors as endusers.

Future work could explore how other stakeholders such as AI developers, risk officers, clients, and regulators influence the explainability and transparency of AI systems that are being used. This would allow for a better understanding of ecosystem-wide dynamics involved in AI adoption. Possibly revealing where alignment or friction occurs between technical design and audit needs.

Secondly, this study only captures a snapshot of perceptions at a single point in time. However, a longitudinal case study could offer deeper insights into how trust, transparency, and explainability evolve over time. Especially, as auditors gain experience, or as systems are being improved. Such research could potentially track changes in adoption levels and integration with audit standards over months or years.

Several interviewees highlighted potential risks involving algorithmic bias and unclear accountability, though these topics were not the main focus of this study. Future research could investigate for example how explainability tools can be redesigned in order to mitigate or communicate bias and fairness risks in AI-assisted audits, especially in light of ethical and legal responsibilities.

Given that this study is qualitative in nature, and the nontechnical background of the researcher. Future studies could aim to quantify the relative importance factors of explainability and transparency across various audit tasks. Survey-based or experimental design-based studies could assess how different types of explanations (e.g. visual, textual, generic, or case specific) affect auditor trust, and accuracy, how they influence explainability or transparency, and what are the effect on decision-making quality.

8. REFERENCES

A call for transparency and responsibility in Artificial Intelligence. (2023, September 14). *Deloitte*. https://www.deloitte.com/nl/en/services/consulting/perspectives/a-call-for-transparency-and-responsibility-inartificialintelligence.html

Association of Chartered Certified Accountants. (2020). *Explainable AI: Putting the user at the core*. https://www.accaglobal.com/content/dam/ACCA_Global/professionalinsights/emtech/Explainable%20AI.Narayanan%20Vaidyanatha n.pdf

Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkanen, K., & Kujala, S. (2023). Transparency and explainability of AI systems: From ethical guidelines to requirements. *Information and Software Technology, 159*, 107197. https://doi.org/10.1016/j.infsof.2023.107197

Bello, N. O. A., & Olufemi, N. K. (2024). Artificial intelligence in fraud prevention: Exploring techniques and applications challenges and opportunities. *Computer Science & IT Research Journal, 5(6)*, 1505–1520. https://doi.org/10.51594/csitrj.v5i6.1252

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: complex or simple? Research case examples. *Journal of Research in Nursing, 25(8)*, 652–661. https://doi.org/10.1177/1744987120927206

Datsenko, H., Kudyrko, O., Krupelnytska, I., Maister, L., Hladii, I., & Kopchykova, I. (2024). Innovative approaches to the use of artificial intelligence in accounting, control, and analytical processes to enhance enterprise competitiveness. *Salud Ciencia Y Tecnología - Serie De Conferencias, 3*. https://doi.org/10.56294/sctconf2024.665

Dittmar, L. (2024, November 8). What does transparency really mean in the context of AI governance? *OCEG*. https://www.oceg.org/what-does-transparency-really-meaninthe-context-of-ai-governance/

Fedyk, A., Hodson, J., Khimich, N., & Fedyk, T. (2022). Is artificial intelligence improving the audit process? *Review of Accounting Studies, 27(3)*, 938–985. https://doi.org/10.1007/s11142-022-09697-x Fotoh, L. E., & Lorentzon, J. I. (2021). The impact of digitalization on future audits. *Journal of Emerging Technologies in Accounting, 18(2)*, 77–97. https://doi.org/10.2308/jeta-2020-063

Hu, K., Chen, F., Hsu, M., & Tzeng, G. (2021). Construction of an AI-Driven Risk Management Framework for financial service firms using the MRDM approach. *International Journal of Information Technology & Decision Making, 20(03)*, 1037–1069. https://doi.org/10.1142/s0219622021500279 Jager, J.,

Putnick, D. L., & Bornstein, M. H. (2017). II. MORE

THAN JUST CONVENIENT: THE SCIENTIFIC MERITS OF HOMOGENEOUS CONVENIENCE SAMPLES. *Monographs of the Society for Research in Child Development, 82(2)*, 13–30. https://doi.org/10.1111/mono.12296

Kloosterman, R. (2021). *Artificial Intelligence and its influence on audit efficiency* [Thesis]. University of Amsterdam.

Kokina, J., & Davenport, T. H. (2017). The Emergence of Artificial Intelligence: How Automation is Changing Auditing. *Journal of Emerging Technologies in Accounting, 14(1)*, 115–122. https://doi.org/10.2308/jeta-51730

Kokina, J., Blanchette, S., Davenport, T. H., & Pachamanova, D. (2025). Challenges and opportunities for artificial intelligence in auditing: Evidence from the field. *International Journal of Accounting Information Systems, 56*, 100734. https://doi.org/10.1016/j.accinf.2025.100734

Kroeger, F., Slocombe, B., Inuwa-Dutse, I., Kagimu, B., Grawemeyer, B., & Bhatt, U. (2022). Social Explainability of AI: The Impact of Non-Technical Explanations on Trust. 150–156. Paper presented at *IJCAI 2022 Workshop on Explainable Artificial Intelligence (XAI)*, Vienna, Austria. https://drive.google.com/file/d/1TULeerUPQz2bIbKiyPMPtCm02G6lnr7-/view

Mitan, J. (2024). Enhancing Audit Quality through Artificial Intelligence: An External Auditing Perspective. *Accounting Undergraduate Honors Theses*. https://scholarworks.uark.edu/acctuht/58

Mulliqi, S. (2022). *Exploring the Challenges and Strategies of AI Adoption in Auditing: Insights from a Big Four Firm* [Journal article]. University of Twente.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation https://doi.org/10.1007/s10488-0130528-y

Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation, 27(2)*, 237–246.

W. Hannah (2024) What is AI transparency? A comprehensive guide *Zendesk*. https://www.zendesk.nl/blog/aitransparency/#

Yin, R. K. (2011). *Qualitative research from start to finish*. The Guilford Press.

Zhang, C., Cho, S., & Vasarhelyi, M. (2022). Explainable Artificial Intelligence (XAI) in auditing. *International Journal of Accounting Information Systems, 46*, 100572. https://doi.org/10.1016/j.accinf.2022.100572

Zhong, C., & Goel, S. (2024). Transparent AI in Auditing through Explainable AI. *Current Issues in Auditing, 18(2)*, A1– A14. https://doi.org/10.2308/ciia-2023-009

Patel, R., Khan, F., Silva, B., & Christ University, Lahore. (2023). Unleashing the Potential of Artificial Intelligence in Auditing: A Comprehensive Exploration of Its Multifaceted Impact. *Journal of Artificial Research, 4*. https://mpra.ub.unimuenchen.de/119616/

Risks of cognitive technologies. (n.d.). *ICAEW*. https://www.icaew.com/technical/technology/technologyandthe-profession/risks-and-assurance-ofemergingtechnologies/risks-of-cognitive-technologies

Seethamraju, R., & Hecimovic, A. (2023). Adoption of artificial intelligence in auditing: An exploratory study. *Australian Journal of Management, 48(4)*, 780–800.

Sharp, C. A. (2003). *Qualitative Research and Evaluation Methods* (3rd ed.). *Evaluation Journal of Australasia, 3(2)*, 60–61. https://doi.org/10.1177/1035719X0300300213 Shelf. (2024, November 14). *IT Survey on 2025 Outlook: The State of Enterprise GenAI and Unstructured Data – Shelf.* https://shelf.io/resource/it-survey-on-2025-outlook-the-stateofenterprise-genai-and-unstructured-data/

Appendix

Appendix A – Interview guide

The aim of this research is to determine the degree of explainability, and transparency of AI systems used by auditing firms. The objective of this interview is to understand whether certain characteristics foster trust, facilitate effective audit practices, and help to comply with relevant regulatory standards.

Every interview starts with an introduction. The research introduces themselves and the research. The interviewee introduces their function in the organization, level of experience with AI in auditing, and involvement in AI implementation. Before the introduction an explanation about the interview process is given.

Explainability

1. **Clarity of explanations:** How clear are the explanations provided by the AI system when it flags anomalies or risks?
2. **User comprehension:** Do you find these explanations actionable for your audit tasks? If not, what improvements would you suggest?
3. **Alignment with Documentation standards:** How do the explanations provided by the AI system align with audit documentation standards
4. **Trust and Confidence:** Does the AI system's ability to explain its decisions affect your trust in its outputs? Why or why not?

Transparency

5. **Model design:** Do you feel that the AI system is open about how it operates (e.g., its algorithms, data sources, and decision-making processes)?
6. **Information needs:** What kind of information about the AI system would help you feel more confident in using it during audits?
7. **Data provenance:** Have you encountered any challenges in understanding how the AI system processes data or arrives at conclusions?
8. **Bias & Fairness:** How does the AI system address potential biases in its predictions? Are these efforts sufficient?

General Perceptions

10. What role do you think explainability and transparency play in ensuring ethical use of AI systems in auditing?
11. How do these factors affect your ability to exercise professional skepticism during audits?

Appendix B – Interviewee table

Participant	Company	Role	Duration
P1	EY	Auditor	20:09
P2	Deloitte	Auditor	16:37
P3	Deloitte	Audit Manager	11:31
P4	EY	Auditor	14:13
P5	Rabobank	IT Auditor	22:07

Appendix C – Coding sheet

Possible criterion. Transparency: Data Provenance (TR-DP), Model Design (TR-MD), Bias Detection (TR-BD). Explainability: Clarity of Explanations (EX-CE), User Comprehension (EX-UC), Trust and Confidence (EX-TC), Documentation Standards (EX-DS).

Possible themes. Explainability or Transparency.

This coding sheet presents interview data categorized into themes and subthemes, following the Grad Coach qualitative analysis structure. Each quote is linked to a defined coding criterion under either Explainability or Transparency.

Theme	Subtheme/ Code	Quote	Participant	Assessment	Question Ref
Explainability	Trust and Confidence (EX-TC)	"Every time AI generates responses; I need to verify them again to ensure correctness and relevancy."	P1	Neutral	Q4
Explainability	User Comprehension (EX-UC)	"I use AI mostly for references; I might not be the best person to ask improvements."	P1	Neutral	Q2
Transparency	Data Provenance (TR-DP)	"Sometimes AI references are outdated; I must always validate the provided evidence."	P1	Negative	Q7
Explainability	Clarity of Explanations (EX-CE)	"AI clearly flags anomalies like unusual cash transactions, helping us take specific actions."	P2	Positive	Q1
Explainability	User Comprehension (EX-UC)	"Outputs help refine models and guide subsequent business inquiries effectively."	P2	Positive	Q2

Explainability	Documentation Standards (EX-DS)	"For anomaly detection, AI explanations are sufficient, but generative AI results require more scrutiny."	P2	Neutral	Q3
Explainability	Trust and Confidence (EX-TC)	"Confidence varies; anomaly detection is trusted more than generative AI outputs, which always require verification."	P2	Neutral	Q4
Transparency	Model Design (TR-MD)	"Transparency varies; generalized AI tools like Microsoft Copilot use broad, non-transparent data sets."	P2	Negative	Q5
Transparency	Model Design (TR-MD)	"Open-source AI models enhance transparency and trust; specific audit-trained models are preferable."	P2	Positive	Q6
Transparency	Bias Detection (TR-BD)	"Bias checks and extensive compliance oversight ensure AI models avoid biases, especially in sensitive areas."	P2	Positive	Q8
Explainability	Clarity of Explanations (EX-CE)	"System typically provides short, clear notes for anomalies but occasionally lacks sufficient details."	P3	Neutral	Q1
Explainability	User Comprehension (EX-UC)	"Adding summaries of key data points and references to similar past anomalies would improve usability."	P3	Positive	Q2
Explainability	Documentation Standards (EX-DS)	"AI systems broadly align with	P3	Neutral	Q3

		documentation standards but require human review due to varying industry-specific standards."			
Transparency	Model Design (TR-MD)	"Detailed understanding of AI's inner workings would increase my confidence significantly."	P3	Neutral	Q6
Transparency	Data Provenance (TR-DP)	"Occasionally difficult to understand data processing and need IT assistance or manual checks."	P3	Negative	Q7
Transparency	Bias Detection (TR-BD)	"Bias detection claims from vendors exist but explicit proofs are rarely seen."	P3	Negative	Q8
Explainability	Clarity of Explanations (EX-CE)	"AI clearly explains financial discrepancies in reports, explicitly stating the inconsistencies."	P4	Positive	Q1
Explainability	User Comprehension (EX-UC)	"Outputs are generally actionable but require precise questioning to ensure relevance and accuracy."	P4	Neutral	Q2
Explainability	Trust and Confidence (EX-TC)	"Trust in the AI outputs is conditional on continuous verification; anomalies flagged often require human oversight."	P4	Neutral	Q4
Transparency	Model Design (TR-MD)	"Model explanations regarding standards are clear, but inner algorithms remain opaque."	P4	Neutral	Q5
Transparency	Model Design (TR-MD)	"Better information about model	P4	Neutral	Q6

		training and validation processes would enhance confidence."			
Explainability	Clarity of Explanations (EX-CE)	"AI's reasoning and flagged risks are understandable but sometimes overly conservative, flagging minor issues."	P5	Neutral	Q1
Explainability	Trust and Confidence (EX-TC)	"Explanation clarity strongly influences trust; without clear explanations, AI outputs wouldn't be usable."	P5	Positive	Q4
Transparency	Model Design (TR-MD)	"Transparency about algorithms is limited; detailed processes remain mysterious even internally."	P5	Negative	Q5
Transparency	Bias Detection (TR-BD)	"Bias is evident in overly conservative risk assessments and confirmatory responses to user inputs."	P5	Negative	Q8