UNIVERSITY OF TURKU

ABSTRACT

☐ Bachelor's thesis
☒ Master's thesis
☐ Licentiate's thesis
☐ Doctoral dissertation

Subject	International Master in Management of IT	Date	06/06/2024
Author	Irene Manetti	Number of pages	77 pages + 9 appendices
Title	Enhancing Financial Audits through Deep Learning: Addressing Key Challenges and Improving Efficiency		
Supervisor(s)	Dr. Soreangsey Kiv		

The financial audit (FA) process, traditionally based on manual procedures and reliant on professional judgment, faces challenges in the era of digitalization, due to the requirement of analyzing large volumes of complex data. This thesis investigates how deep learning (DL) can address challenges in the FA process, particularly focusing on large data volumes, manual procedures, and the subjectivity of professional judgment. Using the Task-Technology Fit (TTF) theory as a guiding framework, the study explores DL's potential through a comprehensive research approach.

Through 13 EY expert interviews across various global locations, and a qualitative survey, the research identifies key challenges in current FA practices, and shows a fit with DL applications. DL shows promise in addressing these issues by automating tasks, managing data complexity and large data volumes, and providing auditors with data-driven recommendation.

Findings reveal that DL's capabilities in natural language processing (NLP), computer vision, anomaly detection, recommendation systems, and big data analytics can address the identified FA challenges. Additionally, DL models are suggested for alleviating each challenge.

This study not only validates existing DL applications, but also introduces up to date FA challenges. This thesis provides a solid foundation for future research and practical applications in the field of financial auditing. The implications of these findings suggest that adopting DL can lead to more efficient and accurate FA processes.

Key words	Deep Learning (DL), Financial Audit (FA), Task-Technology Fit (TTF)

UNIVERSITY OF TURKU
Turku School of Economics

Aix-Marseille Graduate School of Management
Aix-Marseille Université

TILBURG UNIVERSITY
School of Economics and Management

ENHANCING FINANCIAL AUDITS THROUGH DEEP LEARNING

ADDRESSING KEY CHALLENGES AND IMPROVING EFFICIENCY

Master's Thesis
In International Master in Management of IT

Author:
Irene Manetti

Supervisor:
Dr. Soreangsey Kiv

Tilburg, Turku, Aix-en-Provence

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

1.1.1 Company description

The following research has been conducted during an internship at Ernst & Young LLP (EY) within its Technology Risk department.

EY stands as a multinational professional services firm offering assurance, advisory, tax, and transaction advisory services worldwide and is part of the Big Four accounting firms (EY, 2022). According to the EY Value Realized 2023 Report, EY has been a preferred auditor for companies going public since 2012, showcasing its commitment to delivering quality audits and valuable insights. The company places a strong emphasis on technological innovation, evidenced by its substantial investment of US$10 billion in technology solutions in 2021 and ongoing efforts to integrate AI into its global technology offerings over the past decade. The company's commitment to AI solutions is showcased by several recently received awards, such as the AI Excellence Award in 2023 (issued by the Business Intelligence Group) and the Global AI Partner of the Year in 2022 (issued by Microsoft) (EY.com, 2024). A prime example of this commitment is the development of the Document Intelligence (DI) tool, employing computer vision, machine learning, and Deep Learning (DL) to streamline data extraction and translation (EY, 2021).

EY demonstrates a commitment to expanding the application of these solutions across various domains, with auditing being a notable example. Specifically, EY utilized the DI platform to aid auditors in reviewing lease contracts across more than 5000 audits (EY, 2024).

More specifically to the audit realm, EY's Digital Audit initiative highlights its dedication to addressing current issues in financial statement audits through technological advancements, marked by a substantial US$1 billion investment in 2022 in Assurance technology. At the core of this initiative, EY Canvas serves as a pivotal tool connecting global audit teams, enabling secure data exchange and real-time progress monitoring through a cloud-based platform. Additionally, EY Helix, a suite of analytics, contributes to overcoming the difficulties of data analysis and risk identification by providing advanced analytics tools. Finally, to find a solution for manual and time-intensive audit procedures, EY developed EY Smart Automation, showcasing the integration of AI and machine learning to automate audit processes, ensuring precision and compliance. Collectively, these

platforms illustrate EY's digital transformation efforts, specifically tailored to enhance auditing practices (EY, 2023).

By investing in emerging technologies and continuously improving audit quality, EY is well-positioned to address the challenges of FA and aligns perfectly with the objectives of this thesis, which explores the potential of DL in overcoming these challenges.

1.2 Problem indication

The rapid technological evolution observed globally has led to a transformative shift in the business landscape, particularly with the advent of artificial intelligence (AI) technology. According to McKinsey's report on digitization in response to the COVID-19 pandemic, companies across sectors and regions have sped up their digitization efforts, accelerating the adoption of digital interactions with customers, supply chains, and internal operations by several years (McKinsey, 2020).

As companies increasingly automate their operations using advanced computer systems, the accounting and audit profession faces new challenges brought by digitalization and automation (Werner, Wiese, and Maas, 2021). The complexities arising from automated transaction processing, heterogeneous source systems, intricate business processes, and the increasing volumes of data pose significant challenges for auditors (Werner, et al., 2021). Almufadda and Almezeini (2022) underscore the potential of AI technology in auditing to enhance the effectiveness, efficiency, and quality of audit work by reshaping the trade-off between speed, cost, and quality in human-performed tasks.

The imperative to enhance audit practices in response to evolving challenges highlights the need for innovative solutions in the audit profession. Traditional audit procedures, as highlighted by Werner et al. (2021), face inefficiencies and ineffectiveness in environments characterized by the high integration of information systems for transaction processing. The prevalence of information asymmetry in capital markets, as discussed by Jan (2021), further increases the challenges faced by auditors, undermining public confidence in the financial system and the capital market.

Inaccurate or incomplete audits can have far-reaching consequences, leading to financial losses and reducing stakeholder confidence in financial reporting (Jan, 2021). The complexity and scale of financial data pose formidable challenges for audit practitioners, including the detection of anomalies, identification of fraudulent activities, and assessment of audit risk (Werner et al., 2021).

Emerging technologies are believed to induce substantial changes in audit and assurance engagements, offering opportunities for auditors (Seidestein et al., 2024). Within this landscape, Deep Learning (DL), a subfield of AI, emerges as a promising solution for the audit profession (Sun & Vasarhelyi, 2017; Sun, 2019; Jan, 2021). DL algorithms hold the potential to mitigate the shortcomings of traditional approaches to financial statement fraud detection, particularly in the era of big data and AI (Jan, 2021). As auditors struggle with the complexities of modern financial systems, the integration of AI technologies, and specifically DL, offers a promising avenue for revolutionizing traditional audit methodologies. Studies by Föhr et al. (2024), Jan (2021), and Sun (2019) have highlighted the potential benefits of leveraging DL techniques in FA. Examples of application areas include the use of DL for fraud detection and anomaly detection in financial data (Jan, 2021 Föhr et al., 2024), capabilities of identifying information and support for judgment (Sun, 2019) and audit sampling process enhancement (Schreyer et al., 2020).

The existing literature presents limited empirical evidence regarding the application of DL techniques in FA. While theoretical discussions have explored the potential benefits of DL for enhancing audit processes, there is a gap in the literature highlighting the comprehensive understanding of its applications in the FA domain. This research aims to address this gap by conducting a comprehensive theoretical exploration of the fit of DL techniques in the FA domain to address its challenges.

1.3 Problem statement

The evolving landscape of FA necessitates a comprehensive understanding of organizational processes, risks, and controls to ensure accuracy and compliance with regulatory standards. Werner et al. (2021) delineate the audit process into four phases: (1) understanding the entity and its environment, (2) assessing material misstatement risks, (3) designing and executing responses to risks, and (4) forming opinions on financial statements.

DL represents a sophisticated machine learning technique that employs hierarchical artificial neural networks to abstract data features from raw data. Sun (2019) shows DL's potential in supporting audit practices by enabling the identification and extraction of insights from diverse data sources, including text, audio, images, and video. Sun illustrates the potential of DL in audit processes through its capabilities of text understanding, speech recognition, and visual recognition. Additionally, Ding (2022) explored the application of DL models in auditing, demonstrating the effectiveness of intelligent audit data

transformation models in automating data analysis and enhancing audit efficiency through the utilization of DL networks and automatic encoders.

Despite the potential benefits, companies have been slow to adopt DL techniques in audits leading to a disconnect between theoretical promise and practical implementation (Föhr et al., 2024). While specific examples demonstrate the potential benefits of DL, cohesive understanding of its capabilities and limitations in audit settings is missing. This research aims to bridge this gap by providing a comprehensive theoretical exploration of DL's opportunities in FA, showing a theoretical fit between specific FA challenges and DL, by showing the most appropriate capabilities and models that can address such limitations.

By focusing on theoretical analysis, this study aims to provide valuable insights into the potential implications of DL for the audit profession, laying the groundwork for future empirical research and practical implementations. By highlighting the unique capabilities of DL and exploring its potential applications in FA, this research aims to provide insights for audit practitioners. By exploring the theoretical possibilities and implications of integrating DL techniques, this research aims to contribute to the development of innovative approaches to address FA challenges.

1.4 Research question

This chapter introduces the research question guiding this study, as well as sub questions formulated for them to contribute to the main question. The main objective of this research is to investigate which inherent challenges in the FA process can be alleviated through the application of DL techniques. This exploration is guided by the Task-Technology Fit (TTF) theory, which argues that the effectiveness of technology in improving task performance depends on the alignment between task requirements and technology capabilities (Furneaux, 2011). To this end, the research is structured around a central question and three sub questions. Each sub question is designed to explore specific aspects of the FA process and DL applications, addressing the task characteristics, the technology characteristics, and their fit.

Main research question:

What challenges inherent to financial auditing could be alleviated through the application of deep learning techniques?

This overarching question seeks to identify the core challenges in the FA process that can be mitigated by applying DL techniques. By examining the FA process through the lens

of TTF theory, the research aims to determine the alignment between the challenges (tasks) and DL capabilities (technology). To comprehensively answer this question, three sub questions are formulated, each targeting specific components of the TTF framework.

SQ1. What are the current challenges encountered in a financial audit process?

This sub question focuses on understanding what the financial audit process entails and identifying the specific challenges within it. It will examine the nature of these tasks, the complexities involved, and the inefficiencies present in the current methods. By clearly defining the task characteristics and the specific challenges faced in FAs, SQ1 lays the groundwork for understanding which of these challenges can potentially be addressed by DL technologies.

SQ2. What are the main capabilities of deep learning?

This sub question aims to explore and delineate the specific capabilities of DL techniques that make them suitable for addressing the challenges identified in SQ1. It will examine the technical aspects and functionalities of various DL models. By elucidating the strengths and functionalities of DL technologies, SQ2 helps in understanding how these technological capabilities align with and can potentially address the task characteristics identified earlier.

SQ3. How can deep learning techniques be applied to address the challenges identified in the financial audit process?

This sub question bridges the gap between the task challenges and the DL capabilities by exploring practical applications and implementations of DL in the FA process. It seeks to determine the fit between the tasks and the technology. This involves mapping DL capabilities to specific FA challenges, demonstrating how particular DL models can be used to improve efficiency and accuracy in auditing tasks. It will provide concrete examples and case studies of DL applications in financial auditing.

Each sub question plays a critical role in systematically exploring the main research question through the TTF framework. SQ1 identifies the task characteristics and challenges in the FA process. SQ2 examines the technological capabilities of DL that can address these challenges. SQ3 integrates the insights from the first two to analyze the fit between the tasks and DL technology, providing practical applications and demonstrating the potential of DL to transform financial auditing. By addressing these sub questions, the research aims to build a comprehensive understanding of which challenges in financial auditing

can be alleviated through the application of DL techniques, ultimately answering the main research question.

1.5 Research design

The research design will involve identifying improvement areas in EY's audit processes, exploring the potential applications of DL techniques to address these issues.
The problem identified shows a gap in the literature providing a comprehensive exploration describing the possible applications of DL in FA. To address this gap, the a qualitative methodology will be employed, fitting with the exploratory nature of this research.
Data triangulation will be applied, by gathering data from, the existing literature, interviews with experts, and analysis of internal EY documents. Finally, a qualitative survey will be administered to specific experts, which serves the function of validating the results.
Three groups of stakeholders are identified: financial auditors, data scientists and those stakeholders who have knowledge in both domains. The first group is aimed at obtaining answers that elucidate the current financial audit challenges, considering their direct experience on the domain. The second group of data scientists has been identified to address technical perspectives and challenges in the application of DL. The group composed of experts with knowledge in both domains is selected to identify a general overview of the combination of these two disciplines.
Stakeholders will be chosen considering their availability, expertise, and rank. The objective is to include a range of ranks, from lower to higher, to gather both operational perspectives and higher, more comprehensive ones. To provide a robust guideline to gathering and analysing the data, the Task-Technology Fit theory, initially operationalized by Goodhue & Thompson (1995) will be applied as a lens through which exploring this thesis topic. The combination of three expert groups is aimed at covering the topic from different perspectives, namely from a task perspective (FA challenges), a technology perspective (DL characteristics), and a fit perspective (the combination between the two).
The theoretical background will be built through existing literature, which will then be combined with the results of interviews and the analysis of EY's documents, such as EY's audit methodology. By soliciting perspectives from various stakeholders, this thesis aims at achieving a comprehensive understanding of the integration of DL in audit practices.
The data obtained from interviews will consist of qualitative information, including opinions, insights, and expert knowledge regarding audit challenges and potential DL

applications. EY, as one of the Big Four accounting firms, provides a wide range of FA experts, who face FA-related inefficiencies daily. Their insights will add value to this research by providing a current view on FA problems to provide a fresh and direct view on them. Additionally, as mentioned in the company description, EY poses great importance towards innovation and new technologies, as showed by their recent investments in technology solutions (EY, 2021). Specifically, EY experts are working on applying DL in the FA process. This underscores the focus of EY towards technological innovations, meaning the company's employees are composed of a wide range of experts in AI in general, as well as in DL.

The data will be analyzed qualitatively to identify recurring themes, patterns, and areas of consensus or contention among the interviewees.

2 THEORETICAL BACKGROUND

This chapter provides the theoretical background addressing the topics of the research SQs. Section 2.1 presents the TTF theory, section 2.2 shows an overview of the FA process, while section 2.2 describe its challenges; sections 2.3 and 2.4 present DL and its capabilities; and section 2.5 discusses the possibilities of DL applications in FA. Finally, a conceptual framework and hypothesis formulation are explicated.

2.1 The Task-Technology Fit theory

To enhance the theoretical underpinnings of the study, the Task-Technology Fit (TTF) theory was selected as a guiding framework to draft the sub questions and to analyze the results. Rooted in the notion that Information Systems (ISs) are most effective when they align well with the tasks they are meant to support (Furneaux, 2011), the TTF theory provides a robust lens through which to explore the intersection of DL and FA. Initially operationalized by Goodhue and Thompson (1995) (Fig, 5), the theory has since been widely applied across diverse contexts to comprehend IS applications and their contextual nuances (Furneaux, 2011), making it highly pertinent to this research.

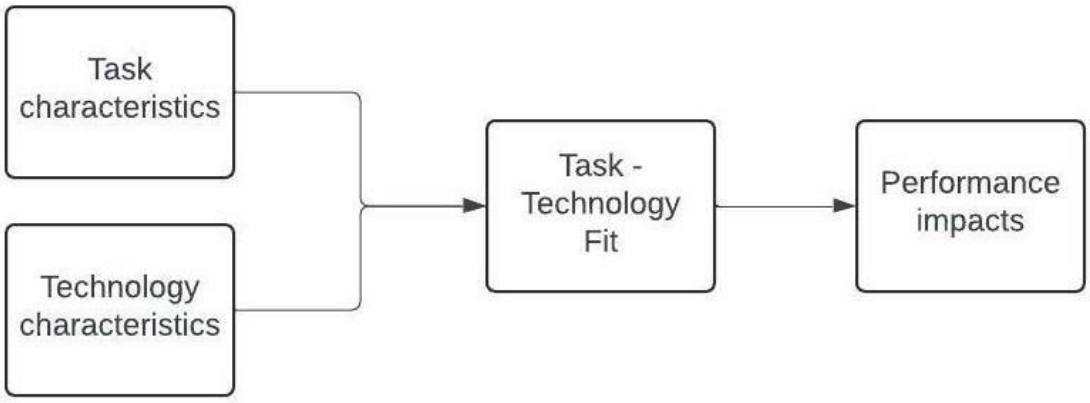

Figure 1. Task-Technology Fit model (Goodhue & Thompson, 1995)

The theory has been applied in various settings, including finance, healthcare, and education (Hattingh et al., 2020). For the present study, the adopted TTF definition is the one proposed by Lin and Huang (2008), framing TTF as "*perceptions that system capabilities match with the user's requirements*".

Various methodologies have been employed (Furneaux, 2011), including qualitative research. For instance, Nan et al., (2011) adopted TTF in their qualitative research using

interviews to explore the fit between task and technology in mobile applications. They performed two rounds of interviews, where the first was more exploratory and the second one was more in depth. Additionally, they involved stakeholders having different backgrounds, specifically stakeholders with a more technical background and those with a less technical one, which provides an additional match with this thesis.

Although direct applications of TTF to the intersection of DL and financial auditing are scarce, studies with similar applications have been found. TTF has been applied to research the auditor's acceptance of blockchain technology in the audit process (Li and Juma'h, 2022), a technology that was not widely used at the time in the context chosen.

Due to time constraints and extensive research coverage, the performance impacts section of TTF is not explored, which can be a starting point for future research.

The TTF theory served multiple purposes. It guided the research scope, by offering a perspective of fit that was used to choose and analyze the existing literature, interviews, and EY's internal documentation. The theory guided the formulation of the research questions, as clarified in chapter 1.3. Finally, TTF assisted the analysis of results, by determining how DL technologies can fit with the identified FA challenges.

2.2 The financial audit process

A financial statement audit, referred to as financial audit (FA) for brevity purposes, is a process performed to reach an independent opinion on how the financial statements of a company are presented (PwC, 2017). As defined by the International Auditing and Assurance Standards Board (IAASB), the goal of a financial statement audit is to "*enhance the degree of confidence of intended users in the financial statements*" (ISA 200).

The importance of audits is expressed by Knechel and Steven (2017), who underline how "*informed decisions should be based on information that is objective, relevant, reliable, and understandable*" - audits exist to provide reasonable assurance that this information is indeed valid.

Financial statements are widely used to support the decision-making process of various stakeholders, such as investors, banks, suppliers, and customers, to name only a few (PwC, 2017). Therefore, a high level of assurance must be provided on the fair representation of a company's financial performance through its financial statements (PwC, 2017).

Audits are guided by auditing standards that can be set by organizations both at the national and international levels (PwC, 2017). For this thesis, the standards issued by IAASB will be used as a source of reference. The reasons are twofold:

1. The IAASB issues ISAs, supported by the International Federation of Accountants (IFAC), which is globally recognized. At the time of writing, it comprises 135 jurisdictions and 180 members (IFAC, 2024) This allows the findings of this thesis to be more universally applicable, as opposed to being confined to the context of a single country.

2. The audits of financial statements performed by EY Netherlands, where this research is conducted, are based on the ISAs.

In their research on embedding process mining in the audit, Werner et al. (2021) divide the financial statement audit process in the four phases, following the IAASB standards: (1) obtaining an understanding of the entity, (2) identifying and assessing the risks of material misstatement, (3) designing and executing responses to the identified risks, and (4) concluding and communicating.

The first phase entails obtaining an understanding of the client, including its environment and the client's system of internal controls. This is performed through procedures such as inquiries, observation, inspection, and analytical procedures (ISA 315). Analytical procedures are performed to analyse the possible relationships between financial and non-financial information, to detect potential inconsistencies, odd transactions, with the goal to flag areas that require the auditor focus (ISA 315, revised 2019).

The second phase of the audit entails identifying and assessing the risk of material misstatement at the financial statement level as well as the assertion level for specific classes of transactions. This determines the audit procedures that will have to be applied to address this risk (ISA 315, revised 2019).

In the third phase, the audit entails designing and applying procedures to respond to the previously identified risks. Such procedures generally include test of controls and substantive procedures. The latter are aimed at identifying risks at the assertion level and include test of details – aimed at specific classes of transactions and disclosures - and substantive analytical procedures (ISA 330). Types of substantive procedures are substantive analytical procedures and test of details, which entail testing 100% of the population (Sekar, 2022). Test of controls have the goal of evaluating whether the system of the client's internal controls operate in a way to prevent, detect, and correct material misstatements (ISA 330).

The fourth phase entails forming an opinion on the financial statements depending on the audit evidence obtained and analysed until this point of the process. The auditors then need to draft a report including such opinion (ISA 700).

Finally, contrary to some literature findings, the IAASB standards define planning as "*not a discrete phase of an audit, but rather a continual and iterative process*", underlining how it starts at the beginning of the current audit - or even before - and ends with the completion of the engagement. Planning requires determining the strategy for the audit. (ISA 300). Figure 1 summarises this process.

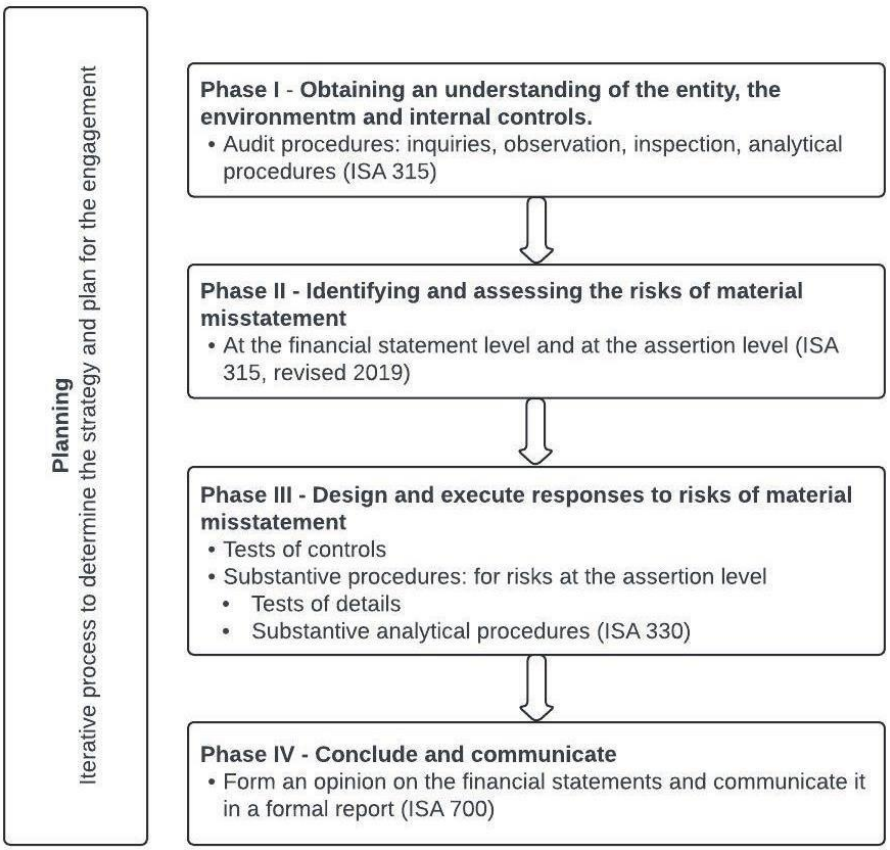

Figure 2 Audit process illustration

2.3 Financial audit challenges

Literature findings highlight that the FA process presents inefficiencies and challenges, deriving from the increasing volume of available data, the heterogeneity of the data, the number of manual procedures that traditionally characterise the audit process, and audit areas that require high-level professional judgment. These challenges have been emphasised by the era of digitalization, that has brought changes in many industries (Werner, 2021). Gartner defines in its glossary the term digitalization as the strategic use of digital technologies to transform business models, creating new revenue streams and

value-producing opportunities. This process involves prioritizing digital tools and technologies in operational tasks to enhance efficiency and effectiveness. Among the consequences of digitalization, the following can be mentioned: the massive data banks and variety of data created, the increasing diversity among source systems, and the evolving intricacy in business operations. This presents novel obstacles for auditors performing audits (Werner et al., 2021). For instance, the evaluation of the client's systems of internal controls and processes is performed manually, including inquiries, analysis of documents, as well as sample testing (Werner et al., 2021). There is a consensus that the efficiency of traditional audit procedures diminishes in environments characterized by extensive integration of information systems for transaction processing (Werner et al., 2021; Fotoh & Lorentzon, 2021).

Common themes in the literature have been identified. Consequently, they have been grouped according to the nature of the challenge, which led to the identification of three classes. The identified classes represent macro challenges deriving from traditional audit procedures, and they are (1) large data volume and big data analytics, (2) manual procedures, and (3) professional judgment. While categorizing the challenges for clarity, it is important to acknowledge the inherent interconnectedness among them within the FA domain. The complexity of audit tasks often means that challenges can arise from multiple sources simultaneously. For instance, a challenge attributed to manual procedures may be aggravated by the volume of data involved, or a judgment support issue may be further complicated by the extensive data requiring analysis. Therefore, while the classification provides a structured approach to addressing each challenge individually, it is important to recognize that these challenges often intersect and influence one another.

2.3.1 Large data volume and big data analytics

The introduction of big data analytics in FAs is recognized in the literature to bring quality benefits to audits (PwC, 2019). EY (2018) stated that effectively analysing Big Data has the potential to increase the quality of the risk assessment procedure. For instance, incorporating Big Data analytics into audits can empower auditors to better identify fraud indicators (Fotoh & Lorentzon, 2023; Tang & Krim, 2018), litigation risks (Sun & Vasarhelyi, 2018), business, and financial reporting risks (EY, 2018). However, processing big data presents challenges, not only in FA, as shown by the extensive literature focusing on finding solutions to big data integration (EY, 2018). Following the same definition that Najafabadi et al. (2015) adopted, big data is associated with four

characteristics that define it, namely volume, variety, velocity, and veracity. Volume refers to the immense amount of data generated, which requires scalable storage solutions and distributed processing strategies. Variety involves the diverse types of data, both structured and unstructured, demanding advanced preprocessing to create usable representations. Velocity denotes the rapid rate at which data is produced and the need for real-time processing to avoid data loss and ensure timely feedback. Veracity addresses the reliability of the data, emphasizing the difficulty in maintaining trustworthiness in the increasing complexity and number of data sources. Traditional data processing systems struggle with these aspects, requiring more sophisticated and adaptable approaches to handle the dynamic nature of Big Data effectively (Najafabadi et al., 2015). Considering only the proliferation of large databases due to digitalization, traditional audit methods become less effective, prompting a need to reconsider how audits are conducted (Dai and Vasarhelyi, 2016).

Specifically to FA, handling large data volumes presents several complexities that demand specialized approaches and tools. In numerous scenarios, reaching an informed decision through traditional audit methods becomes nearly unfeasible, as it ideally mandates analysing extensive data within a restricted timeframe (Almufadda et al., 2022).

2.3.2 Manual procedures

The challenge of manual procedures within FAs is multifaceted and encompasses various aspects of the audit process. Manual tasks are often characterized by their repetitive nature, consuming considerable time and effort, and contributing to inefficiency in audit execution (Sekar, 2022; Werner et al. 2021).

All the phases of an audit are affected by manual and routine work. As transaction volumes increase, so does the intricacy of conducting tests of detail (Sekar, 2022). Analysing millions of transactions requires specialized audit software, but handling such volumes presents challenges, with some tasks requiring hours for processing, particularly for complex operations like joining tables and summarizing transactions (Sekar, 2022). Sekar explains that related to the testing procedure there is another challenging procedure, as testing conventionally involves selecting a sample of data from a larger population for examination. Various statistical and non-statistical sampling methods can be employed to select a set of transactions to be checked by auditors. Traditionally, auditors test these samples manually. Several concerns arise: the procedure is time-consuming and error-prone, where only a percentage of data is tested (Sekar, 2022). Manually analysing a

sufficiently large sample of transactions becomes inefficient, and in some cases, impossible (Werner et al., 2021). Additionally, with a fixed random sample size, the likelihood of selecting a truly representative sample diminishes significantly. Particularly when dealing with millions or billions of transactions (Werner et al. 2021), this procedure only permits the examination of a fraction of the total transactions, potentially resulting in overlooked discrepancies and anomalies within the dataset. EY has been long advocating for a shift from sample-based to entire population testing of relevant data in an audit engagement, as expressed in a publication of 2018.

Finally, manual processes are utilized for extracting and researching unstructured data within the audit data market, often resulting in inaccuracies and low speed (Ding, 2022). Other examples of manual procedures include reviewing contracts, drafting reports, transcribing, and analysing interviews (Sun & Vasarhelyi, 2017)

2.3.3 Professional judgment

The IAASB (2021 Edition) requires auditors to apply professional judgment, which is defined as "the application of relevant training, knowledge, and experience, within the context provided by auditing, accounting, and ethical standards, in making informed decisions about the courses of action that are appropriate in the circumstances of the audit engagement." The ISAs request auditors to employ professional judgment in the entire planning phase and in the strategy execution of an audit. More specifically, it is requested for decision-making processes regarding the determination of materiality, the type and amount of audit procedures to be performed, assessing whether the audit evidence obtained is sufficient and appropriate, verifying the entity's application of the applicable financial reporting framework, and forming an opinion on the financial statements based on the audit evidence gathered, including the assessment of estimates made by the entity and included in the financial statements (ISA 200). In other words, professional judgment is a constant in the entire FA process. As Sun (2019) and Jan (2021) address, the challenge arises when tasks require high-level judgment, where auditors must make final decisions after considering vast amounts of evidence from multiple angles. For example, evaluating engagement risk requires integrating all gathered background information, making the process intricate and demanding thorough analysis (Sun, 2019). Fraud detection is also a particularly pressing topic, as financial statement fraud continues to occur sporadically (Jan 2021). The implementation of fraud examination practices relies heavily on auditor judgment, leading to inconsistent success rates. This is due to the interpretation and

execution of fraud detection standards, which may vary among auditors (Tang & Krim, 2018). Addressing this challenge requires the adoption of more effective procedures, (Tang & Krim, 2018; Jan, 2021). Professional judgment is also used to select the method for the sampling procedure, and Sekar (2022) underlines how years of extensive experience would make an auditor able to make such a decision. Finally, another area where professional judgment is required is expressing an opinion on the client's ability to continue as a going concern. Auditors are responsible to conclude on the correctness of the client's management going concern prediction (ISA 570). Through a survey that involved 175 investors and 198 business leaders, PwC (2019) underlined the importance of audit opinions on the entity's going concern, as investors are interested in the prospects of a business. Jan (2021) identified the problem of incorrect audit opinions, stressing the importance of having a sounder foundation on which to base auditors' opinion regarding the going concern of an entity.

To summarise, professional judgment is a constant and an integral part in FA. However, literature findings explain how this bring issues in areas requiring high-level of judgment, underlining inconsistencies in decision-making procedures.

Table 1 summarizes the findings of this section.

Challenge	Description	Key References
Large Data Volume and Big Data Analytics	Managing vast amounts of diverse and rapidly generated data poses significant challenges. Examples: advanced processing for structured and unstructured data, timely analysis or large and complex volumes of data, and performing tests of details.	PwC (2019); EY (2018); Najafabadi et al. (2015); Dai & Vasarhelyi (2016); Almufadda et al. (2022); Sekar (2022)
Manual Procedures	Traditional audit processes involve repetitive, time-consuming manual tasks that reduce efficiency. Examples: sampling methods, difficulty analyzing large data samples, inefficiencies in extracting and researching unstructured data, manual contract reviews, drafting reports, transcribing, and analyzing interviews.	Sekar (2022); Werner et al. (2021); Ding (2022); Sun & Vasarhelyi (2017)
Professional Judgment	Auditors must apply professional judgment throughout the audit process, which can lead to inconsistencies. Examples: decision-making on audit procedures, sufficiency of evidence, evaluating engagement risk, inconsistent fraud detection, selecting sampling methods, and forming opinions on going concern.	IAASB (2021); ISA 200; ISA 570; Sun (2019); Tang & Krim (2018); Jan (2021); Sekar (2022); PwC (2019)

Table 1 Summary of financial audit challenges**.**

2.4 Deep Learning

DL is a subset of Machine Learning (ML), which is in turn as subset of Artificial Intelligence (AI). Therefore, it is necessary to briefly introduce the concepts of AI and ML to provide the rationale behind the development of DL. Afterwards, the concept of DL is elucidated, then its capabilities and applications are discussed.

AI can be described in many ways. According to Fergus and Chalmers (2022), "AI can be loosely defined as incorporating human intelligence into machines". According to different speakers at IBM's course Applied AI professional certificate (2023), AI can be described in several ways, such as "a set of technologies that allows us to extract knowledge from data", to cite one. To address this issue, Berente et al. (2021) define it as being an idea rather than as specific phenomenon, stating that AI is "*The frontier of computational advancements that references human intelligence in addressing ever more complex decision-making problems*". This definition underlines AI's nature being an evolving concept.

ML is a subset of AI that enables algorithms[1] to learn independently, eliminating the need for humans to write each single line of code specifying elaborate rules (Fergus and Chalmers, 2022). Goodfellow et al. (2016) describe it as "*a technique that allows computer systems to improve with experience and data*". By providing large amounts of data to ML algorithms, patterns are detected, rules are automatically generated and subsequently applied to novel data, fostering an experiential learning process (Fergus and Chalmers, 2022). This iterative approach empowers algorithms to adapt and develop autonomously, extracting insightful information from data without explicit guidance. However, traditional ML presents a limitation called the selectivity-invariance problem, which makes it challenging for these methods to handle raw data effectively. This problem entails picking out the most important pieces of information from the data while ignoring less important ones. The selected information should be different enough from each other. This limitation motivated researchers to create a more advanced solution, hence DL (Chauhan and Singh, 2018).

---

[1] "A set of mathematical instructions or rules that, especially if given to a computer, will help to calculate an answer to a problem" (Cambridge Dictionary).

DL is an advanced approach to ML (Goodfellow et al., 2016). While ML is a general term that includes DL, the latter specifically refers to techniques that enable computers to learn from data through deeper and more complicated architectures (Fergus & Chalmers, 2022). Figure 1 shows the relation between AI, ML and DL.

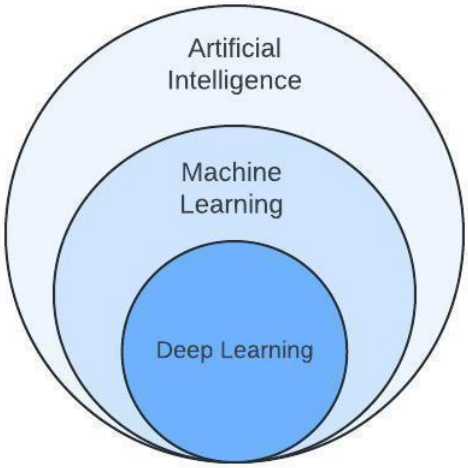

Figure 3. Relation between AI, ML and DL

DL is composed of a hierarchical structure, through which computers can grasp complicated concepts by assembling simpler ones (Goodfellow et al., 2016). These algorithms analyse data and create patterns; information moves through layers, with each layer using the output from the previous one as an input. The first layer is the input layer, the last one is the output layer, and the ones in between are referred to as hidden layers (Figure 2). When a network has three or more layers, it is called a deep network (Fergus & Chalmers, 2022). Visualizing this process reveals a deep graph with numerous layers, hence the term "deep learning" for this AI approach (Goodfellow et al., 2016). DL can extract complex features[2] from high-dimensional data[3] (Dargan et al., 2019) a common characteristic of Big Data (Schintler 2021). High-dimensional data brings challenges, including storage space, high computational time, and the risk of low output accuracy (IBM, 2024). DL

[2] A feature is a specific piece of information measured about each item in a collection of items, called a dataset. Each item, known as an example, has multiple features. For instance, in a dataset about plants, an example would be an individual plant, and features would include measurements like the length and width of its leaves and petals (Goodfellow et al., 2016).
[3] A dataset where the amount of examples, also known as observations, is smaller than the amount of features (Bobbit, 2021).

models are designed to address these challenges, due to their ability known as dimensionality reduction - the process used to reduce the number of features in a dataset while preserving its important structure or patterns.

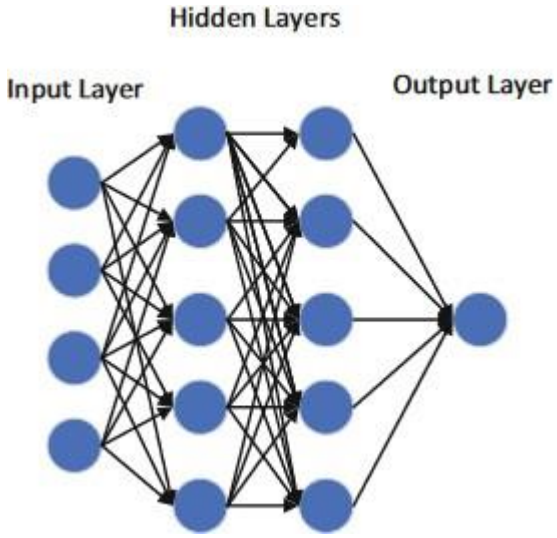

Figure 4. Example of DL architecture. Fergus & Chalmers (2022)

DL may seem to be a new concept because of its recent popularity. However, it was developed in the 1950s. It went through different names, where ‘Deep Learning' is the most recent one. As explained by Goodfellow et al. (2016), one of the names that used to describe DL was Artificial Neural Networks (ANN), which are algorithms that were developed to reproduce the way learning happens in biological brains. Drawing inspiration from how information is processed in living organisms, ANNs are composed of interconnected computational units known as artificial neurons. Similarly to the synapses found in the brain, the connections between these neurons transmit signals whose intensity can be increased or decreased by a weight that is continually modified during the learning phase (Janiesch, Zschech and Heinrich 2021). Today's DL represents an advancement from ANNs. The term ‘Deep Learning' is sometimes used interchangeably with Deep Neural Networks (DNN) (Janiesch, Zschech and Heinrich 2021).

The reasons for DL's current popularity are multiple. Among them is today's advancements in technology, which provides increased computational resources, leading to higher-performing computers that can process higher volumes of data. Moreover, the era of Big Data leads to a higher availability of large datasets that can be used to train DL models and therefore improve the model's accuracy (Goodfellow et al., 2016; Fergus and Chalmers, 2022).

There are several benefits of DL when compared to traditional ML models. For instance, DL performs better when dealing with large amounts of data, leading to algorithms that can construct more accurate models. Another benefit entails representation learning. Representations are features describing one item in a dataset. The performance of AI algorithms depends on these representations; however, it may be complex to understand which features have to be extracted. In response to this challenge, the extraction of such representation has been automated through DL algorithms, which perform the process known as representation learning (Goodfellow et al., 2016). DL eliminates the need for manual feature extraction, allowing the system to autonomously determine which features are most relevant for describing the data, accelerating the processes by removing the requirement for humans needing to manually define all the knowledge that the computer needs – differently from traditional ML approaches (Fig. 3) (Fergus and Chambers, 2022). Another area where DL excels is pattern recognition, which involves identifying regularities or patterns in data to make predictions or decisions (LeCun et al., 2015).

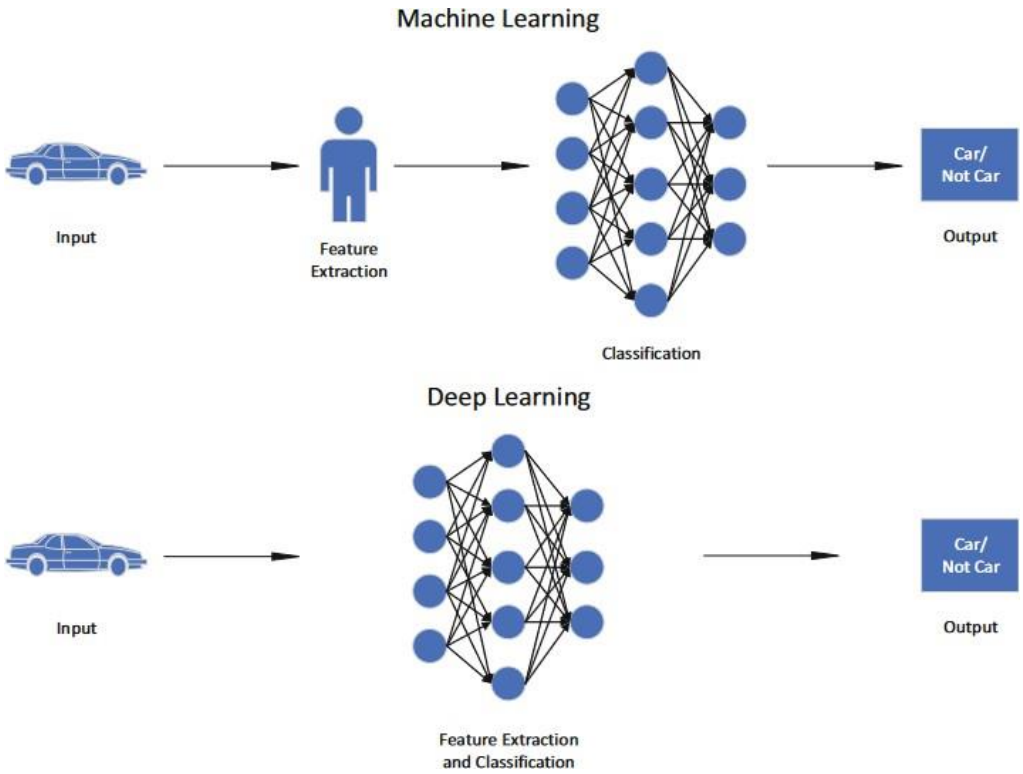

Figure 5. DL's automatic feature extraction (Fergus & Chalmers, 2022)

DL performs better that ML in analysing unstructured datasets (Shiri et al, 2023). Structured data comes in a strict format, where rows and columns are labelled. Semi-structured

data is provided with some structure, but without any predetermined format, such as emails and documents. Unstructured data does not have any predefined format, such as images, audio files, videos, social media messages (Fergus and Chalmers, 2022). DL has shown successful results in analysing structured, semi-structured and unstructured data types (Sun, 2019).

2.5 Deep Learning's main applications and models

Below, DL's main applications are presented, and the main models employed are briefly explained.

All the previously mentioned characteristics allowed DL to tackle limitations in several applications, such as Natural Language Processing (NLP) and computer vision.

DL played a significant role in advancing NLP, overcoming traditional ML algorithm limitations (Lauriola, Lavelli and Aiolli, 2021). NLP is a field within computer science and artificial intelligence that focuses on enabling computers to understand, process, and respond to human languages. Unlike conventional ML, which works mainly with numerical data, NLP deals with unstructured textual data, such as emails, reviews, and spoken commands. NLP transforms the vast, unstructured information contained in written and spoken language into structured data that computers can work with, making it possible for machines to access and utilize the knowledge humans have stored in text (Fergus and Chalmers, 2022). Some NLP capabilities that are enhanced with DL include: sentiment analysis – a form of text analysis utilizing unstructured data to classify the emotions expressed in text into categories such as positive, negative, and neutral - object recognition (LeCun et al. 2015; Fergus and Chalmers, 2022), speech recognition and language translation (LeCun et al. 2015). Other applications include retrieving information from flow text, generating natural language, and summarizing it (Deng and Liu, 2018). DL models make the summarization of multiple documents possible, generating summaries from a group of different documents (Ma et al., 2022).

DL models can also extract information from audio files (Fergus and Chalmers, 2022), convert spoken words into written text (LeCun et al., 2015), recognize different pronunciations (Dong et al., 2021), or identify the emotions from audio files (Abbaschian et al., 2021).

Computer vision is yet another successful application area for DL. It is the discipline of enabling computers to reconstruct the characteristics of data that comes in image or video formats. Examples are the ability to recognize and extract postal codes that are

handwritten, or automatically extract the number plate of vehicles from images (Szeliski, 2010). DL application to this area has been so successful that LeCun et al. (2015) state that it led to a revolution to the discipline. Examples of computer vision application are object detection - the process of finding and identifying specific types of objects in images and videos (Voulodimos et al., 2018), image classification, semantic segmentation, and object segmentation (Fergus and Chalmers, 2022). Image classification entails categorizing an entire image into a particular group or label, regardless of the number of objects it contains. Semantic segmentation labels every pixel in an image with a category, without differentiating between separate objects, while object segmentation detects individual objects in an image by marking their pixel areas. This has been proven especially helpful in medical imaging for tasks like measuring tissue volumes and finding tumours. One study by Kao and Wen (2020) is an example that computer vision applications have potential beyond its foundational applications. For instance, leveraging the capabilities of image recognition and detection, the study shows how a DL-based approach is successful in verifying signatures and detect forgery. Regardless of their use of one known sample of signatures, they proved the CNN-based model's accuracy in differentiating the legitimate signatures from the forged ones. Not only signatures, but entire images can be tampered, making it difficult even for human eyes to detect it (Camacho & Wang, 2021). The science that deals with identifying the legitimacy of images is known as image forensics. In their survey focusing on the DL methods used for image forensics, Camacho and Wang (2018) mention several studies that proved DL's efficacy in detecting falsification images. The authors consider falsifying an image as intentionally altering some part of an image to deceive viewers about the events depicted, typically by inserting or removing specific content to change its meaning.

Increased value can be obtained when DL applications are merged and used together. In the book Deep Learning and Natural Language Processing, He and Deng (2018) explain how the areas of NLP and computer vision can intersect with each other. For instance, DL can generate natural language from images, by providing a written description of an image.

Another challenging topic addressed by DL is anomaly detection (Fergus and Chalmers, 2022), which is the process of identifying unusual events or objects within a dataset (Goodfellow et al., 2016).

Literature stresses the success of DL applications to Big Data analysis, addressing the challenges of exploiting information contained in Big Data, characterized among other

characteristics by large volumes and heterogeneity (Dargan et al., 2019). More specific DL capabilities need to be mentioned, such as recommendation systems through which models can provide recommendation to users (Goodfellow et al., 2016). Successful DL applications in this field have been proven to be in recommending music to users, by extracting features from audio files (van den O ord et al., 2013). Finally, also thanks to NLP successful applications, DL can automatically generate reports from given data, such as images. This topic has been studied especially in the healthcare industry, to automate the generation of reports from medical images, as in the study conducted by Alfarghaly et al. (2021), who focused on the automatic generation of reports from radiology images. Table 2 summarizes the findings mentioned until this point. Subsequently, DL's main models are briefly described.

Capability	Description	Example applications	Key references
Natural Language Processing (NLP)	Enabling computers to understand, process, and respond to human languages, transforming unstructured text into structured data.	Sentiment analysis, language translation, summarization of documents, extracting formulas described in text, question answering, text summarization, image summarization	Lauriola et al., 2021; Fergus and Chalmers, 2022; LeCun et al., 2015; Deng and Liu, 2018
	Speech recognition: converting spoken words into written text and recognizing different pronunciations.	Transcription of spoken words, recognizing emotions from audio files	LeCun et al., 2015; Dong et al., 2021; Abbaschian et al., 2021
Computer Vision	Enabling computers to interpret and make decisions based on visual data such as images and videos.	Image classification, object detection, semantic segmentation, object segmentation, image forensics, signature verification	LeCun et al., 2015; Szeliski, 2010; Voulodimos et al., 2018; Camacho & Wang, 2021; Kao & Wen, 2020
Anomaly Detection	Identifying unusual events or objects within a dataset.	Credit card fraud detection	Fergus & Chalmers, 2022; Goodfellow et al., 2015
Recommendation Systems	Providing personalized recommendations to users	Music recommendation	van den O ¨ord et al., 2013
Big Data Analysis	Handling Large-Scale Data	Processing and extracting insights from large, heterogeneous datasets.	Dargan et al., 2019

Table 2. Summary of Deep Learning's capabilities and applications

2.5.1 Deep Learning's models

Several DL models have been developed through the years, tailored to address specific tasks and data types.

Convolutional neural networks (CNNs) are neural networks built with several layers and inspired by the biological visual cortex (Dargan et al., 2010). This DL architecture is

effective for dealing with data with spatial relationships, such as images (Goodfellow et al., 2016). CNNs are particularly suitable for computer vision applications (Janiesch, Zsech & Heinrich, 2022).

Deep Convolutional Neural Networks (DCNNs) are an advanced extension of CNNs characterized by their deep architecture comprising multiple convolutional layers (Fergus & Chalmers, 2022). They extend CNN's capability by stacking more layers, which allows for the extraction of increasingly abstract and complex features from the input data. One prominent example of a DCNN application is in facial recognition systems (Fergus & Chalmers, 2022).

Recurrent neural networks (RNNs) are mainly capable to analyse time-series data, as they have an internal memory, by considering the temporal sequence of the inputs (Shiri et al., 2023). Their architecture incorporates internal loops, facilitating the sequential learning of patterns and the retention of preceding events (Janiesch, 2021).

Long-term short-term memory (LSTM) networks were developed to overcome the limitations observed in traditional RNN architectures. (Fergus & Chalmers, 2022). In LSTM, a specialized unit known as a cell is employed, capable of retaining its state over an extended period and treating it as a function of its input. This feature enables the unit to effectively store and recall the most recent computed value (Dargan et al., 2019).

Autoencoders (AEs) structures typically include an encoding phase, where the input is condensed into a compact representation, and a decoding phase, where the network tries to recreate the initial input from the acquired features. As a result, the network is driven to keep important details in the condensed representation while removing unnecessary background noise (Goodfellow et al., 2016). They can be employed mainly for dimensionality reduction, both in sequential data spatial data (Fergus & Chalmers, 2022). AEs are DL models in the class of generative models, that aim to create samples that closely mimic the actual data distribution used to train the model (Shiri et al., 2023). The financial statement audit process and challenges. Table 2 provides a summary of the DL's most used models.

Large Language Models (LLMs) are advanced DL models capable of processing and generating text that resemble human-like coherence. They excel in various applications across different fields, including finance, medicine, and law, demonstrating flexibility and adaptability to different industry specific language and concepts. An example application is thematic analysis, where LLMs can be employed to generate explanations of legal terms in a legal text, answering legal questions, therefore improving efficiency and

quality in legal research. Another successful application area is in providing recommendations, such as in medicine, for providing evidence-based treatments, after an analysis of medical literature (Naveed et al., 2024).

Finally, Generative Adversarial Networks (GANs) consist of two competing networks: the generator and the discriminator. The generator creates data, often images, while the discriminator tries to distinguish between real and fake images. Example applications are in image generation (Fergus & Chalmers, 2022).

DL models	Description	Applications	Key references
Convolutional Neural Networks (CNNs)	CNNs are neural networks built with several layers, effective for dealing with data with spatial relationships, such as images.	Object detection, object recognition, object segmentation, image classification, image recognition, image processing, speech recognition, time-series prediction	Dargan et al., 2010; Goodfellow et al., 2016; Janiesch et al., 2022
Deep Convolutional Neural Networks (DCNNs)	DCNNs are an advanced extension of CNNs with multiple convolutional layers, enabling the extraction of complex features from input data.	Facial recognition, complex feature extraction	Fergus & Chalmers, 2022
Recurrent Neural Networks (RNNs)	RNNs analyze time-series data by considering the temporal sequence of inputs with internal memory.	NLP, representing a thought expressed in text	Shiri et al., 2023; Janiesch et al., 2022
Long Short-Term Memory (LSTM)	LSTM networks overcome RNN limitations by employing a cell capable of retaining its state over an extended period.	Machine translation, language modelling, sentiment analysis, speech recognition, multi-document summarization	Ma et al., 2022; Fergus & Chalmers, 2022; Dargan et al., 2019
Autoencoders (AEs)	AEs aim to create samples that closely mimic the actual data distribution used to train the model.	Fraud detection, dimensionality reduction, computer vision	Goodfellow et al., 2016; Shiri et al., 2023; Fergus & Chalmers, 2022
Large Language Models (LLMs)	LLMs are DL models that process and generate coherent text, adaptable across fields like finance, medicine, and law.	Thematic analysis, providing recommendations	Naveed et al., 2024
Generative Adversarial Networks (GANs)	GANs consist of a generator and a discriminator. The generator creates data, and the discriminator identifies real vs. fake data.	Image generation	Fergus & Chalmers, 2022

Table 3. Summary of Deep Learning (DL) main models, description, and applications

2.6 Exploring potential applications of deep learning in the financial audit process

The literature used for this section can be broadly categorized into less technical and more technical studies. Less technical studies serve the purpose of highlighting areas of an audit where DL can provide benefits, without delving into specific models or technological details. One example is Sun's (2019) illustrative framework of DL applications in the FA process, that sheds light on where, in a FA process, the DL's capabilities of text understanding, speech recognition, and computer vision can provide benefits. On the other hand, more technical studies focus on applying DL solutions to specific FA procedures, which serve the purpose to showcase the models used and their technical characteristics,. Finally, other studies leveraging DL to address similar challenges arising in FAs are considered, in situations where direct FA-domain research could not be found. The rest of the chapter is divided in five sections, one for each DL's class of application derived from chapter 2, namely: 2.5.1 NLP, 2.5.2 computer vision, 2.5.3 recommendation systems, 2.5.4 big data analytics, and 2.5.5 anomaly detection. At the end of each section, the classes of FA challenges addressed are briefly explicited.

2.6.1 NLP

The previous chapter highlighted the powerful capabilities of DL in NLP, including automated document analysis, sentiment classification, and speech recognition. These capabilities can be extended to FAs, where handling large volumes of textual data and inquiries efficiently is crucial. Sun (2019) underlines how DL's NLP capabilities can aid auditors to extract information and review documents and any other forms of written text. For instance, DL can automate the document analysis and review in a FA process. DL has already been employed for a similar purpose, as exemplified by the collaboration between KPMG and IBM Watson, an AI platform for business, employed for Research and Development (R&D) tax credit processes (KPMG, 2020). This DL-based solution automates the review and analysis of vast document volumes, streamlining qualitative documentation processing and facilitating comprehensive validation of R&D activities (KPMG, 2020). This solution could be extended to the FA realm.

DL enables the automatic review of multiple documents, to cross-check their numerical accuracy. Cao et al. (2018), built a DL-based system for automating the process of cross-checking the numerical accuracy in documents, where formulas are explained in flow text. Mathematical formulas explained in natural language is a characteristic present in

several documents, especially financial documents such as annual reports and disclosures. The system built allows the extraction of formulas described in natural language and automatically compares them to numbers, to check their accuracy, through an LSTM model. Sun (2019) also suggests the employment of sentiment analysis in FAs to analyse inquiries and management's documents, aiding the auditor to shed light on possible fraudulent areas. Yadav and Vishwakarma (2019) found, through a review of studies about sentiment analysis leveraging DL techniques, that different architectures excel in various aspects of sentiment analysis tasks. For document-level sentiment classification, CNN followed by LSTM demonstrates superior accuracy. LSTM is also reliable for multi-domain sentiment analysis, which involves analyzing sentiment across different topics. In multimodal sentiment analysis, CNN + LSTM, often complemented by fusion techniques, is preferred. Multimodal sentiment classification involves analyzing sentiment expressed through multiple types of data, such as text, images, audio, or a combination of these. This enables applying sentiment analysis to a series of documents varying in format and topic (Yadav and Vishwakarma (2019). As various formats of evidence are used in FAs (Sun, 2019), these types of DL applications show potential to address the burden of financial auditors to corroborate audit evidence by cross checking diverse documents, such as written and oral inquiries, and other documents to verify consistency and potentially identify mistakes or areas of financial fraud risk.

Studies show that manual procedures can also be aided by DL, through the automation of checking regulatory compliance within the financial statements. This is exemplified by Sifa et al. (2019), who developed the Automated List Inspection (ALI) tool. ALI utilizes various emerging technologies, such as ML and DL in the realm of NLP. Employing RNNs, it works by extracting relevant text passages from financial statements and matching them with specific legal regulations, effectively ensuring compliance and accuracy. The tool leverages NLP for document representation and language modeling, thus enabling it to understand the semantic context of the text.

Sun (2019) explains how the speech recognition function of DL can aid auditors to process oral inquires, phone calls, speeches, and presentations.

RNNs, powered by their sequential modelling capabilities and memory management, are instrumental in speech recognition tasks (Mehrish et al., 2023). By processing audio data over time, RNNs capture nuanced speech patterns, leveraging their ability to understand contextual information and long-term dependencies, and excel in transcribing audio input into text with high precision (Mehrish et al., 2023). Examples of RNNs used in speech

recognition domain are Google's search system, activated with vocal commands, as well as Amazon Echo's “Alexa”, working in a similar fashion (He et al., 2017).
To summarize, DL applications in NLP through models such a RNNs, LSTMs, and CNNs enable the automation of document analysis and sentiment classification, tackling challenges of large data volume and manual processes. LSTMs can be employed to cross-check mathematical accuracy by comparing various documents, and RNNs can be utilized to verify regulatory compliance in the financial statements. Oral inquiries can be automatically transcribed through RNNs; then, other DL capabilities such as document summarization, multi-document analysis, and sentiment analysis, can be applied to the transcripts for a faster and more comprehensive audit evidence analysis.

2.6.2 Computer vision

The earlier discussion explains how DL-powered computer vision excels in tasks such as image classification, object detection, and document interpretation. Sun (2019) explains how these capabilities can be crucial in FAs, for instance in performing substantive procedures, by automating the analysis of scanned documents to select relevant items. Studies demonstrate the potential of combining DL capabilities, such as NLP and computer vision, to create powerful DL solutions. For instance, Ding (2022) developed a DL-based model for internal audit, addressing challenges that can also be applied to FAs. Leveraging the automatic feature extraction and dimensionality reduction, the researcher developed an LSTM model that combines NLP and computer vision applications to enable auditors to detect relevant information from large volumes of data, to make meaningful and data-based predictions. The model is capable of processing and analysing data in different formats, including scanned documents transformed into pictures, accounting vouchers, minutes of meetings, or contracts. The LSTM-based solution performs data identification and classifications, by matching and comparing the different types of data (Ding, 2022).
Finally, NLP and computer vision applications can automate the generation of reports, as proved by Alfarghaly et al. (2021), who employed a combination of CNN and RNN models to automate the generation of reports from medical images.
To summarize, DL-powered computer vision aids in document interpretation, automating tasks and reducing reliance on manual procedures, addressing challenges in document analysis. Models employed can be based on LSTMs or CNNs. LSTMs can be employed to analyze a large amount of diverse data, both historical and industry data. A combination

of CNNs and RNNs can be employed to automate the generation of reports from given data, such as images. This addresses FA challenges such structured and unstructured data analysis, as well as manual procedures such as drafting reports.

2.6.3 Recommendation systems

As discussed earlier, DL's ability to process diverse datasets and generate tailored recommendations is exemplified in recommendation systems. These systems have the potential to support auditors in making informed judgments by providing data-driven insights and strategy suggestions. Sun (2019) states that DL solutions could be employed in FA and serve the function of judgment support, for instance by providing recommendations about the strategy for determining the nature and extent of substantive procedures, or regarding the sufficiency of evidence obtained. In the end, the auditor will consider whether following the recommendations or not. The literature proves DL's capability in providing recommendations. To address professional judgment-related challenges.

Da'u and Salim (2019) reviewed DL-based recommendation systems in e-commerce, social media, and movies, demonstrating the success of AEs in managing high-dimensional data and imputing missing information. Yashudas et al. (2024) developed a DL-based cardiovascular disease prediction and personalized health recommendation system using a type of RNN model. This system analyzes physiological data from various sensors to provide accurate diagnoses and classify cardiovascular diseases into distinct categories. By integrating historical and real-time health data, it delivers personalized treatment and dietary recommendations via a mobile application. Such advancements exemplify DL's capability to process diverse datasets and generate actionable insights, laying a strong foundation for its application in recommendation systems across various industries, including FAs.

By harnessing DL techniques such as AEs and RNNs methods, FAs can benefit from enhanced feature extraction, dimensionality reduction, and identification of underlying patterns in financial datasets, ensuring more informed decision-making and tailored recommendations in auditing practices. Hence, DL functions as a support for professional judgment, as stated by Sun (2019), who suggests this application to aid auditors in understanding the nature and amount of testing to perform as well as judging the sufficiency of audit evidence obtained.

2.6.4 Big data analytics

The previous chapter emphasized DL's strengths in handling big data, such as scalability, automatic feature extraction, and pattern recognition. These capabilities are vital for auditors dealing with large and varied datasets to identify risks and ensure comprehensive analysis.

Areas where big data analytics is recognized to have a value-added function in FAs are the identification of fraud indicators (Fotoh & Lorentzon, 2023; Tang & Krim, 2018), litigation risks (Sun & Vasarhelyi, 2018), business, and financial reporting risks (EY, 2018). Leveraging various DL capabilities, including automated representation learning and feature extraction, Najafabadi et al. (2015) provide a description of how DL can address various big data analytics issues. The issues identified in their study follow the four Vs with which big data is usually associated, namely volume, variety, veracity, and velocity. Najafabadi et al. (2015) provided a summary of studies involving DL's capabilities and applications to various domains. The results show that leveraging DL's hierarchical structure, automating feature extraction, pattern recognition capability, and dimensionality reduction, allow this technology to process complex and large volumes of data, diverse data formats, and being able to highlight the relevancy of data amidst large volumes.

The literature highlights other FA areas where big data analytics can provide a robust foundation for auditors to make more accurate and informed judgments, such as evaluating the going concern prediction of the entity (Jan, 2021). For instance, Jan (2021) constructued four hybrid going concern prediction models, including DL solutions, concluding that the RNN-based model was the most successful among the others. Data used is composed of samples about companies that received doubtful going concern opinions as well as those that received positive opinions.

To summarize, DL-based models can aid the analysis of big data, enhancing decision-making procedures in areas such as fraud, litigation, business, and financial reporting risks. Furtherly, RNNs can be employed to aid auditors evaluating management's going concern predictions. Professional judgment-related challenges can be alleviated, as auditors can base their decision on more relevant data.

2.6.5 Anomaly detection

Earlier, DL's abilities to detect anomalies within vast datasets and to identify irregular patterns have been discussed. These capabilities are critical for auditors to effectively identify financial discrepancies and potential fraud.

Detecting financial irregularities within vast datasets challenges traditional audit methods reliant on sampling, often failing to meet today's stringent accuracy requirements (Ding, 2022; Zhang et al., 2021). By analyzing 100% of datasets rather than employing sampling techniques, auditors can gain substantial benefits, by enabling auditors to spend their time on anomalies, and therefore high-risk areas (KPMG, 2021).

A recent study by Schultz & Tropmann-Frick (2020) validates the effectiveness of DL application consistent with this approach. Recognizing the limitations of sample-based methods in today's data-rich environment, the researchers utilized a real-world dataset encompassing the entire population of journal entries from three financial accounts of a single entity. They proved the efficacy of this application through an AE model, trained with a real audited dataset of more than 300-thousand-line items. The model has been evaluated qualitatively by experienced auditors, who identified the anomalies themselves and compared them to the ones found by the AE, resulting in a successful outcome (Schultz & Tropmann-Frick, 2020).

Anomaly detection can also address the challenge of detecting financial fraud, that is heavily reliant on professional judgment.

Jan (2022) demonstrates DL's efficacy in fraud detection through a study employing financial and non-financial data from 153 companies, using financial and non-financial variables. Comparing two detection models, namely LSTM and RNN, the results show that the former achieved a higher detection accuracy of 94.88%, indicating LSTM's superior performance in long-term memory retention and fraud detection accuracy. However, this study did not leverage the automatic feature extraction capability of DL, but the variables were manually selected by the researchers. To fully exploit DL's capabilities, automatic feature extraction should be performed, which would also help in handling large and high-dimensional datasets, where one difficulty arises from selecting the more appropriate features. Even if not explicitly directed to FAs, a study by Alghofaili et al. (2020) proves LSTM's high performance in detecting financial fraud, leveraging DL's automatic feature extraction to process credit card transaction data. The model addressed

the challenge of detecting financial fraud, particularly in the context of big data, where traditional methods struggle to capture complex patterns.

To summarize, through RNNs, LSTMs and AEs, anomaly detection techniques leveraging automatic feature extraction enhance audit accuracy by identifying irregularities and potential fraud, addressing challenges related to manual procedures, large data volume and professional judgment.

2.7 Conceptual framework

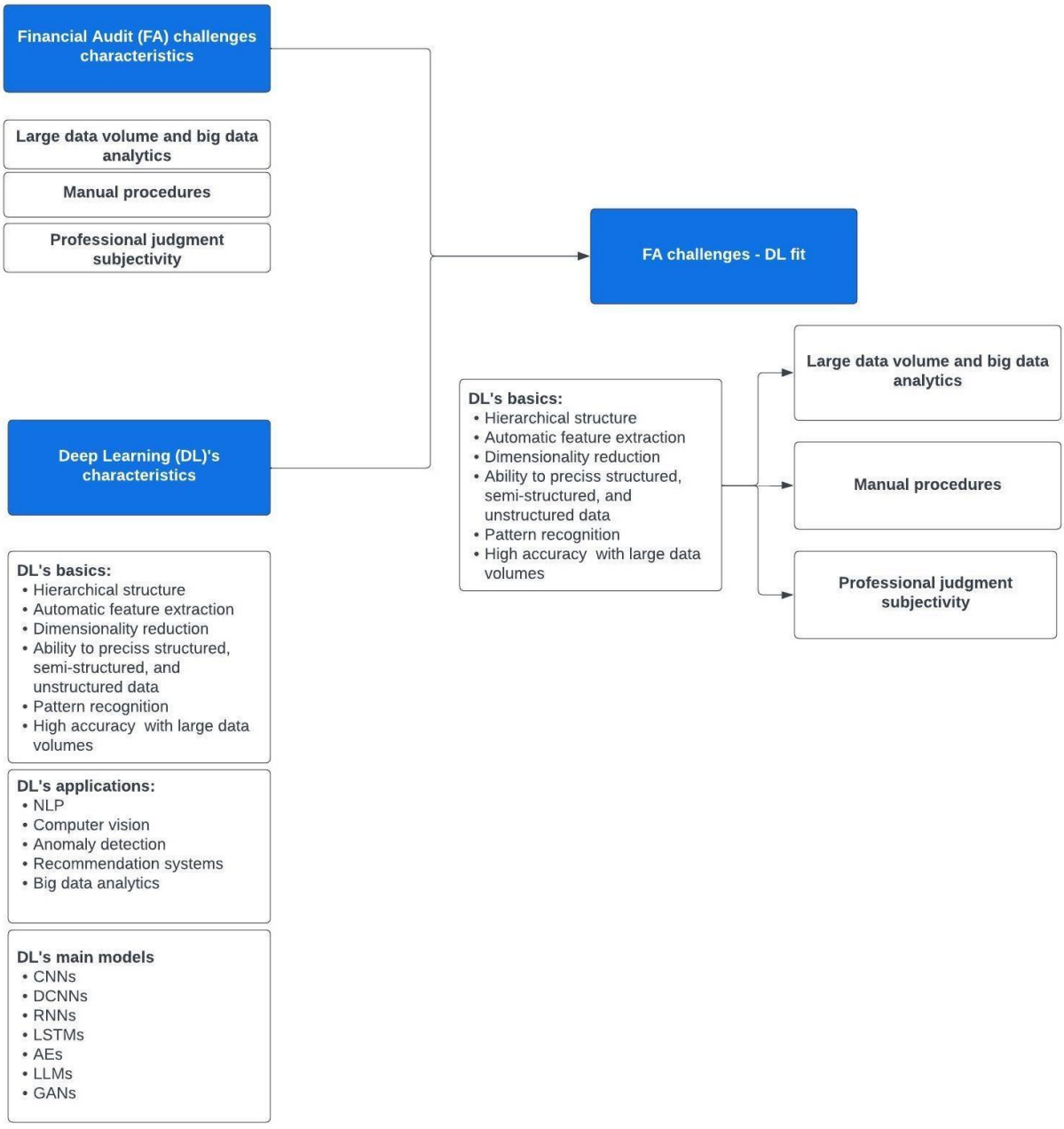

Figure 6. Conceptual framework summarizing literature findings.

Figure 7 illustrates the main findings, form a TTF perspectives, of the main literature research findings.

2.8 Hypothesis formulation

Based on the analysis of financial audit challenges, the capabilities of DL, and existing literature on their combinations, the following hypotheses were formulated. These hypotheses aim to explore the suitability of DL models in addressing key challenges in financial audits.

Hypothesis 1 (H1): DL models are well-suited to address the challenges of processing large volumes of structured, semi-structured, and unstructured data in financial audits. Their capabilities align with the needs of auditing large and diverse datasets.

Hypothesis 2 (H2): DL models are well-suited for automating manual procedures in financial audits. They can effectively streamline repetitive tasks, reducing the time and effort traditionally required.

Hypothesis 3 (H3): DL models are well-suited for supporting auditors' professional judgment in financial audits. They provide valuable insights and help identify potential discrepancies, aiding in more consistent and informed decision-making.

3 METHODOLOGY

This chapter explains the methodology employed in this thesis and it is structured as follows: 3.1 describes the data triangulation approach and illustrates an overview of the research; 3.2, 3.3, and 3.4 describe how each sub question is addressed from the perspective of the theory adopted. Section 3.5 explains the qualitative approach; 3.6 shows the interview procedure and analysis, and 3.7 briefly explains the document analysis. Finally, the survey procedure ispresented in section 3.8.

3.1 Data triangulation and chronological overview of the research

Data triangulation was employed to answer the research sub questions. In general, triangulation is the process of employing multiple methods or datasets to enhance the credibility and validity of findings. Triangulation provides greater confidence in the reliability of the conclusions drawn (Noble & Heale, 2019). One type of triangulation is data triangulation, that requires drawing conclusions from various data sources in research (Bans-Akutey & Tiimub, 2021).

Additionally, to clarify the general chronological order of events regarding the research, Figure 6 provides an illustrated overview.

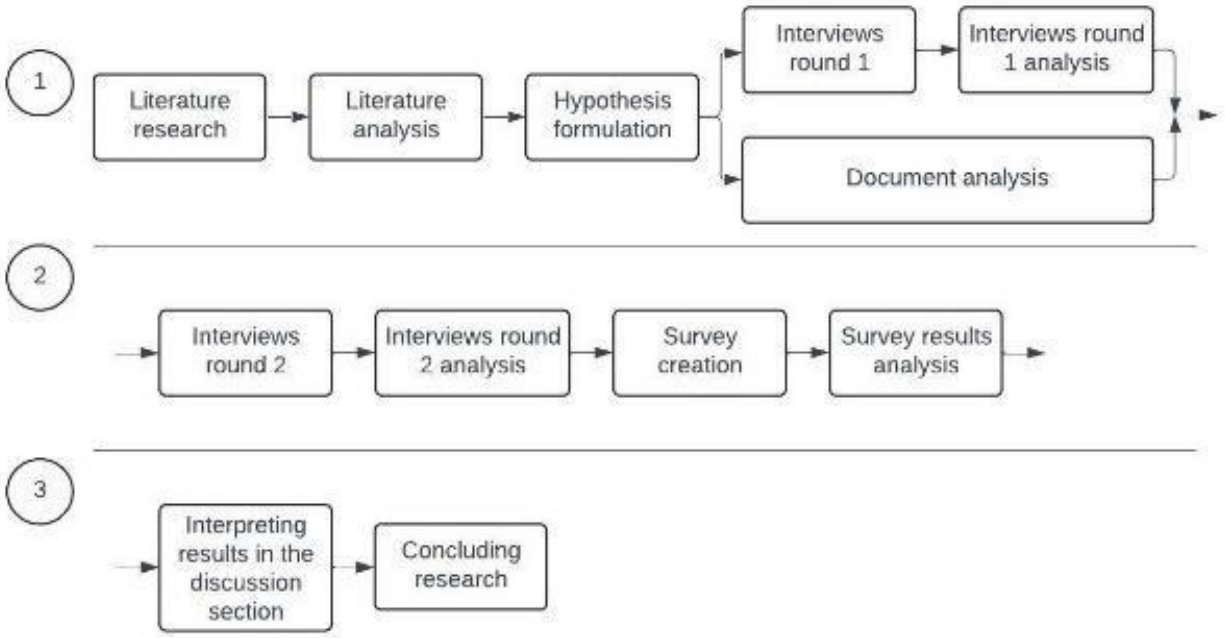

Figure 7. Chronological overview of the thesis steps

3.2 Task characteristics – SQ1

Tasks are defined as procedures performed to transform inputs into outputs (Goodhue & Thompson, 1995). In this context, the tasks are the specific procedures performed by auditors required by FAs. Selecting challenges as task characteristics is justified by the need to tailor technological solutions to the nuanced and specific obstacles faced by auditors. This targeted approach ensures that technology is designed to meet the exact demands of the FA process, directly addressing the areas that most impact auditor performance and task efficiency. By concentrating on challenges, it is possible to verify that a technology is highly relevant. A comprehensive literature review was conducted to gather knowledge on the FA process and to identify the current challenges and tasks in FAs. This included reviewing academic papers, books, and industry reports.

To provide an overall understanding of the FA challenges, the FA process was finalized. To refine the FA process, existing literature and EY's audit methodology were utilized. Due to the extensive existing literature and documentation and the time constraints during interviews, this topic was not explored further with experts, to allow more room during the interviews to the actual focus of this thesis – the FA challenges.

Interviews, existing literature, and EY's audit methodology were employed to explore and conclude on FA challenges. Interviews with financial auditors provided insights into the tasks they find challenging. These interviews helped to validate and expand upon the tasks identified in the literature, where they were usually described only in general terms. Interview questions can be found in appendix 4.

3.3 Technology characteristics – SQ2

Technologies are identified as instruments employed to facilitate the completion of tasks (Goodhue & Thompson, 1995). In this thesis context, the technology is represented by DL.

The capabilities and applications of DL technologies were mainly identified through a review of current research and books in the field, as the topic is extensively covered. Through interviews, data scientists provided technical perspectives on the capabilities and limitations of DL technologies. Their insights were crucial in understanding the feasibility of applying DL to financial auditing tasks. The research on EY's internal documentation did not provide significant new insights. Therefore, data triangulation was applied at a reduced extent for this SQ compared to SQ1 and SQ3, were more sources were employed.

However, the extensive and detailed existing literature, as well as the interviews, provided a large amount of data from which it was possible to base hypothesis and draw conclusions.
Interview questions can be found in appendix 4.

3.4 Task-Technology Fit – SQ3

TTF refers to assessing whether a technology supports an individual in carrying out their tasks; namely, it refers to alignment between task and technology (Goodhue & Thompson, 1995).
Data has been derived from existing literature, interviews, a survey, and EY's internal documents. A series of interviews were conducted with financial auditors, data scientists, and professionals with expertise in both domains. Interviews with the latter group of experts highlighted FA challenges and areas where DL is currently employed at EY, and where DL solutions are being piloted and researched. The other two groups of experts provided personal perspectives on where DL could be applied in a FA process, and underlining areas of fit and DL's feasibility to address them. To address the lack of expertise in one of the two domains, the thesis writer introduced the topic at the beginning of the interviews. A brief introduction on DL's basics, its capabilities, and examples of applications in FAs drawing from the literature, was given to financial auditors. This explanation lasted around 10 minutes. In the second round of interviews with financial auditors, specific DL's capabilities were introduced one at a time and questions were asked regarding that specific capability, to make sure not to overwhelm experts with information. However, due to time constraints, not all the DL capabilities could be covered during the interviews.
Data scientists were provided with a brief explanation of the FA challenges identified from the literature and the interviews. However, due to the long list of challenges, the interviews were insufficient to cover them all. To overcome this limitation, a qualitative survey including open questions was drafted. The survey is described in section 3.10 and it can be found in Appendix7. The thesis writer connected DL's capabilities and models to specific use cases in FAs, designed from literature, interviews, and document analysis. To validate these connections, the survey was sent out to four data science experts. Two responses were obtained.
Finally, internal EY's documents were analyzed. However, limited documentation was found on the topic. Interview questions can be found in appendix 4.

3.5 Qualitative Approach

Given the exploratory nature of this study and the desire to solicit diverse perspectives, a qualitative research methodology was deemed most appropriate. Myers and Avison (2002) describe in their book “Qualitative Research in Information Systems” how qualitative research has received attention in the information system area - the subject that is involved on the engineering, usage, and consequences of information technology in enterprise and business areas - by enabling the discovery of valuable results in the field.
As Myers and Avison (2002) explain, the general approach to pursue qualitative research is by collecting data through interviews, documents, and observations of participants, which is in line with this thesis methodology.
Given the dynamic nature of this thesis' domains, an exploratory approach ensures that the information gathered from interviews remains current and relevant.

3.6 Interview Procedure

A total of 13 interviews was conducted, details of which are presented in Table 8. The number has been chosen to reach data saturation, which happens when new primary data becomes repetitive; to reach this threshold, Saunders (2023) reports that with 9 to 17 interviews, data saturation is generally reached. To preserve confidentiality, the names of interviewees are withheld. Interview durations ranged from 45 to 60 minutes, all conducted remotely via Microsoft Teams. Consent was obtained from each interviewee prior to recording, with assurances provided regarding the utilization of results for research purposes. The interviews are numbered, and a letter has been assigned to them, to ease their citation in the following text. The number represents the chronological order in which the interviews have been performed, whereas the letter identifies the expert. Repeating letters indicate that the same expert has been interviewed twice.

Interview	Prossional profile	Interview type	Interview round	Area(s) of expertise	Geographical location
1A	Senior financial auditor	Semi-structured	1	FA	Netherlands
2B	Senior financial auditor	Semi-structured	1	FA	Netherlands
3C	Manager in the data team	Semi-structured	1	FA and data analysis	Netherlands
4D	Manager with data science experience	Semi-structured	1	Data science	Netherlands
5E	Senior manager with data science knowledge and AI applications in FAs experience	Unstructured	1	FA, data science, and digital solutions in FA	Germany
6F	Staff with data science knowledge and experience	Semi-structured	1	Data science	Australia
7G	Manager in financial audit	Semi-structured	2	FA	Netherlands
8A	Senior financial auditor	Semi-structured	2	FA	Netherlands
9H	Partner within EY Assurance Organization, in the capacity of financial or support of financial audit and performance of assurance-related engagements	Semi-structured	2	FA	Netherlands
10E	Senior manager with data science knowledge and AI applications in FAs experience	Semi-structured	2	FA, data science, and digital solutions in FA	Germany
11F	Staff with data science knowledge and experience	Semi-structured	2	Data science	Australia
12I	Senior manager with FA experience and now in the team building digital FA solutions	Semi-structured	2	FA and digital solutions in FA	UK
13L	Manager in data analytics team, with data science and FA knowledge	Unstructured	2	FA and data science	Netherlands

Table 4. Interview details

Experts sampling strategy:

1. Snowball method: initial interviewees and colleagues were asked to locate experts with specific knowledge required for the study (Fossey et al., 2016).
2. Internal platform: EY's internal platform aggregating employee profiles was used to identify experts. Keywords such as "deep learning," "AI," "deep neural networks," "digital audit," and "financial audit" were employed in platform searches. Leveraging EY's global reach, experts spanning two continents—namely, Australia, and Europe—were identified. Interviews with financial auditors and data analysts were confined to Europe due to variations in audit standards worldwide. Conversely, the geographical constraints were less stringent for data scientists, given the universal applicability of technological knowledge.

Interviewee profiles:

- Financial auditors and data analytics experts working with FA data: provided insights into the pressing challenges in FA.
- Data science experts: offered technical perspectives on DL applications and feasibility.
- Professionals with expertise in both domains: facilitated the synthesis of diverse insights, particularly regarding the implementation of AI and DL in financial auditing, providing real-life examples of DL applications in FAs and opinions on its challenges, benefits, and feasibility.

To ensure a multifaceted examination of the subject matter, interviewees spanning various hierarchical levels—ranging from staff to partners—were engaged.

Interviews were conducted using a combination of semi-structured and unstructured formats based on open-ended questions, to provide interviewees with the freedom of expressing their opinions. Semi-structured interviews are characterized by an exploratory nature, with a pre-defined guide that leaves room for the discovery of novel and uncovered aspects (Magaldi & Berler, 2020). They are particularly suited to capturing the evolving landscape of the two pivotal areas in this thesis: DL and FAs. DL research is experiencing rapid innovation, as noted by Fergus and Chalmers (2022), and FAs are undergoing significant challenges due to the pervasive effects of digitalization (ISACA, 2019). While remaining cautious by avoiding mixing FA challenges questions with opinions on DL applications, question sequencing remained flexible, and the result obtained was a collection of iterative discussions.

Two rounds of interviews were conducted. The first round was exploratory, focusing on gathering perspectives on FA challenges and DL applications feasibility. The second round was more in-depth, discussing specific FA challenges, DL applications in the context of the identified challenges, and more specific identification of audit areas where DL could provide benefits. Interviews have been automatically transcribed with the support of AssemblyAI, an online available AI tool. The transcripts have been meticulously reviewed, to verify for correctness of the outcomes.

The decision to not explicitly differentiate between round 1 and round 2 interviews in the results section is based on enhancing thematic consistency and integration, which allows for a fluid and coherent narrative, by focusing on overarching themes rather than fragmenting the discussion. This approach also facilitates the cumulative building of insights, presenting a comprehensive understanding of FA challenges and DL

applications without redundancy. In this way, the results emphasize the substantive findings, ensuring clarity and coherence in the presentation of the study's insights.

Transcripts have been manually analyzed to identify recurring themes, by applying a thematic analysis approach. The choice is dictated by the approach being flexible and a good method for creating a solid qualitative approach (Saunders, 2023), in line with the explorative nature of this thesis. The choice to manually analyze the interview data is justified to ensure familiarization with the data, which is encourages (Saunders, 2023). Labels have been created next to each passage that was deemed relevant; then themes have been created, making sure they are relevant with the SQs and main questions (Saunders, 2023). The transcripts, highlighted passages, identified labels, and themes are presented in Appendix 6.

FA challenges, DL applications in the FA process, and perspectives on DL applications derived from the interviews. Interview transcripts were also analyzed to discover patterns among the interviewees, which led to additional and valuable perspectives that go beyond the pre-selected themes and that were considered important to include.

3.7 EY's internal documentation

EY's internal documentation encompasses a comprehensive and extensive description of their Global Audit Methodology (EY GAM), which was meticulously analyzed to refine audit procedures and identify areas for potential DL applications, drawing from existing literature. Additionally, other EY documentation regarding DL solutions, although not extensive, served as illustrative examples.

3.8 Survey

To address SQ3, a qualitative survey was conducted after analyzing the interviews to validate the connections made by the thesis writer. A list of 20 FA challenges was created, each linked to relevant DL capabilities and possible DL models to address these challenges. These connections were established by the thesis writer based on insights gained from the literature review, document analysis, and interview analysis. To provide an example, Jan (2022) and Alghofaili et al., (2020) prove LSTM's success in detecting fraud. However, both Fergus and Chalmers (2022) and Goodfellow et al. (2016) deem AEs as particularly adept for anomaly detection, and Fergus and Chalmers suggest the application of anomaly detection in fraud detection applications. Therefore, the use case "fraud

detection" is associated in the survey to both LSTMs and AEs. To keep the survey clear and focused, specific DL capabilities were associated with each challenge, rather than using broad categories like "NLP" or "computer vision." This approach is like explaining the specific features of a car, such as "cruise control" and "backup camera," instead of just saying "automobile" to avoid overwhelming respondents and ensure clarity. This method made the survey more efficient and relevant by directly addressing the detailed applications needed in FAs. Broader categories are still associated in the results section to provide clarity and coherence in the thesis.

The qualitative survey was administered via email to four respondents using an Excel table. These respondents had already been interviewed, and their knowledge regarding the specific topic was well-established. The purpose of this survey was to validate connections rather than to collect primary data, which is why only four respondents were chosen. The minimum response expectation was 1 or 2, and ultimately, 2 responses were obtained from: Survey Respondent 1 (SR1), a senior manager possessing extensive knowledge and experience in AI and DL solutions in FAs, and Survey Respondent 2 (SR2), a manager expert in data analytics within FAs and with AI background. The survey included a list of 20 FA challenges, 20 DL models connections, and open-ended questions, detailed in Appendix 2. To validate the connections made, respondents were provided with Excel cells featuring a drop-down menu with four Likert scale options: Strongly Agree, Agree, Disagree, and Strongly Disagree. In cases where respondents chose Disagree or Strongly Disagree, they were asked to provide explanations. Two optional open-ended questions were included:

- "Do you have other propositions regarding DL capabilities and architectures for this challenge?"
- "Please explain your answer."

The survey required each respondent to evaluate 40 matches between 20 formulated FA challenges and corresponding DL models. Six of these matches were left blank for respondents to fill in due to the high uncertainty of the thesis writer.

Given the survey's length and the extensive time and specific knowledge required to complete it, a high response rate was not expected. To encourage participation, a follow-up Teams message was sent to ensure the email was received. For the same purposed, respondents were allowed to leave sections blank if they were unsure. Instructions for completing the survey were provided in the first sheet of the Excel file and are also detailed in Appendix 8.

The survey analysis is qualitative, supplemented by a table and pie charts to illustrate and break down the responses. Differing answers from the two respondents were cross-checked with literature findings and interview results. Open-ended responses provided valuable insights into the reasoning behind disagreements and suggestions, making it easier to evaluate and corroborate survey respondents' viewpoints.

4 RESULTS

This section presents the results and is structured as follows. Results are divided and presented according to the research SQ intended to answer. First, results regarding the FA process and challenges are described. Then, results regarding DL capabilities are addressed. Subsequently, survey results to validate connections to directly address SQ3 are addressed. Finally, a conceptual framework is illustrated.

4.1 The financial audit process – SQ1

For this research, the financial audit process of EY will be considered, as it adheres to universally recognized ISAs, ensuring its applicability across diverse contexts. Furthermore, EY's extensive experience and expertise in conducting audits offer a real-life example, enriching the depth and credibility of the thesis findings. The financial audit process at EY is divided into four phases: (1) initial planning, (2) risk identification and assessment, (3) designing and executing responses to risks and (4) concluding and communicating. Each of these phases includes several processes that need to be performed. The content of section 4.1 describing the financial statement audit process is retrieved from EY's Global Audit Methodology (EY GAM). When findings from EY GAM are directly derived from ISAs, the original sources will be cited after cross-checking, to ensure accurate representation.

4.1.1 Initial planning

The initial planning phase includes a series of processes and activities that are aimed at planning the overall audit. First, audit procedures to obtain sufficient and reliable audit evidence are designed. Audit evidence is data utilized by the auditor in formulating the conclusions upon which the auditor's opinion relies (ISA 500). These procedures include various techniques, such as inquiry and analytical procedures (assessments of financial data by analysing possible connections between both financial and non-financial information.), to name a few. In this phase, the service requirements and scope of the engagement are defined. Decision on the client and engagement acceptance is defined and the engagement agreement is finalised. According to the findings in this phase, the team, roles and responsibilities are established, and ethical and independence agreements are thoroughly determined.

4.1.2 Identify and assess risks

This stage, driven by a comprehensive understanding of the audited business and its internal control system, is aimed at detecting any potential risk of material misstatement within the financial statements.

A misstatement in financial reporting occurs when there is a difference between what is reported and what is required under financial standards, resulting from error or fraud (IAASB, 2021). Materiality can instead be described as the level of a misstatement that could have an impact on the economic decisions of financial statement users (ISA 320).

By identifying those accounts with a reasonable possibility of containing material misstatements, auditors can determine the areas that hold elevated risks. These accounts are referred to as ‘significant accounts'. Essential to this stage is understanding the entity's internal control and operations at a broader level (EY GAM). The procedure also involves discerning fraud risk factors, i.e. situations that present a motive or opportunity to commit fraud (IAASB, 2021).

Moreover, significant classes of transactions (SCOTs) – i.e. those classes of transactions that materially influence significant accounts -are determined to help identify risks affecting the significant accounts (EY GAM). Armed with this knowledge, an audit strategy is then designed that is responsive to the entity's risks of material misstatement. In essence, this phase aims to build a solid risk-oriented foundation for the subsequent stages of the audit (EY GAM).

4.1.3 Designing and executing responses to risks

The designing and executing responses to risks phase involves formulating strategies and procedures to address potential risks of material misstatement identified in the second stage of the audit (EY GAM). Several crucial steps are involved. External confirmation procedures are utilized to gather audit evidence regarding various financial statement assertions, such as bank deposits, investments, and related party transactions. Secondly, understanding controls enables the identification of risks of material misstatement, guiding the design of the audit strategy. Thirdly, responses to fraud risks are formulated to gather sufficient and relevant evidence, including evaluating controls addressing the risk of material misstatement due to fraud (EY GAM). Understanding the impact of IT on an entity's operations is vital for effective audit planning, as it influences financial information processing and reporting. Ensuring compliance with laws and regulations involves

obtaining evidence of adherence while assessing potential risks from litigation, claims, and assessments requires specific procedures. Additionally, substantive procedures are designed and executed to gather sufficient audit evidence tailored to identified risks, aiming to reduce audit risk, which is the risk of auditing materially misstate financial statements while issuing an inappropriate opinion (ISA 200), to an acceptable level and draw reasonable conclusions.

4.1.4 Conclude and communicate

According to EY GAM, in the final phase of a FA, ongoing discussions with the entity's management facilitate the confirmation of facts and allow management to respond to audit findings. Ultimately, a written auditor's report is delivered to the entity, expressing the auditor's opinion on the financial statements. The audit evidence is ultimately cross-checked, the risk of material misstatement is reassessed, and the control environment of the client is re-evaluated. During this phase, the remaining issues are addressed, and comprehensive documentation is finalized before issuing the report. This documentation serves to demonstrate compliance with auditing standards, provide support for conclusions regarding financial statement assertions, and ensure consistency between the underlying accounting records and the financial statements (EY GAM).

4.1.5 Additional findings

From the analysis of EY GAM, it is evident that the audit process cannot be strictly divided into sequential phases. The four phases represent a broad categorization and should not be viewed as a rigid chronological order. Many procedures are iterative and interdependent. For example, findings from phase II related to the risk of material misstatement can influence the scope and requirements of the audit service defined in phase I. Additionally, the performance of substantive procedures in phase III can impact the risk assessment procedures applied in phase II.

A simplified illustration of the FA process is provided in Appendix 1. This illustration includes arrows to visually represent the feedback loops both within and between the phases.

4.2 Financial audit challenges – SQ1

From the literature review, FA challenges were identified, which the writer of this thesis has categorized into three broad classes, namely large data volume and big data analytics, manual procedures, and subjectivity in professional judgment. The categories have subsequently been identified as recurring themes from the interview analysis.

All experts possessing knowledge in financial audits unanimously acknowledged the presence of several challenging tasks within the audit process, contributing to inefficiencies. Certain experts emphasized specific challenges over others and some opinions diverged on certain issues.

In general, the relationship among these three data sources is the following: literature explains issues in general or in extreme detail, lacking something in between. The document provides an overview and explanation of the whole audit procedure, from which relationships among audit tasks are clarified. The interviews served the purpose to obtain more tangible examples and prime opinions, which bring more currency to the topic analysed, that are used to corroborate literature findings.

4.2.1 Large data volume and big data analytics

EY GAM confirms that auditors need to process and analyze large volumes of structured, semi-structured, and unstructured data. When FA were asked their opinion on large data volumes being a challenge, they unanimously agreed. Interviewee 9H explained how a bank processing payment of its customers can translate into millions of transactions per day. Similar sentiments were echoed by other FA experts interviewed, including interviews 1A, 7G, 8A and 12I. However, interview results shed light on various perspectives, providing different examples that corroborate this statement and reveal additional dimensions to the problem.

EY GAM emphasizes the necessity of analyzing various data types, such as internal audit reports, business plans, external economic journals prior engagements data, board meetings minutes, business plans, control manuals, analysts reports, banks or rating agencies, industry sector data, competitors' financial performance, current market data, social and political factors, and interest rates. This extensive volume of data is crucial for identifying and assessing risks of material misstatements during initial planning, understanding business operations, going concern risks, and fraud detection. In other words, analysis of this data has pervasive effects on the whole audit engagement outcome.

EY GAM highlight that this type of data is fundamental for more specific procedures, such as tests of details, which are performed to ensure transactions' existence and that they are correctly disclosed, as well as evaluating accounting estimates performed by the client. Types of tests of details include inspection of documents, recalculation of amounts to verify their mathematical accuracy, and obtaining external confirmations, for which auditors need to construct expectations of recorded amounts, trends, and ratios, based on extensive data, including knowledge of the client, its industry, prior period information (EY GAM). Consistent with literature findings that highlight the challenge performing tests of details with large data volumes (Sekar, 2022). Interview results corroborate this problem, as FA and AI solution in FA expert 12I stated:

"*the challenge is learning how to apply it [data] effectively, have different approaches that are audit approaches that rely on the use of data and doing things like correlation analysis to identify anomalies or trends that are unexpected*" (Personal, communication, 2024)

This sentiment is echoed by a FA expert in interview 8A. Regarding the issue of lacking specialized software for analyzing large data volumes, results from some FA experts interview, such as 3C, 7G, 9H, 12I, indicate that this is no longer considered a significant challenge. They explain that data analytics tools, although non-AI based, are now implemented in audit procedures and can effectively process structured data. However, interview results with FA experts from 3C, 7G, and 12I highlight that this does not entirely solve the problem, as these tools are not well-suited for analyzing semi-structured and unstructured data.

In conclusion, the challenge revolves around effectively utilizing large data volumes to draft an audit plan responsive to risks of material misstatements. Additionally, such data is necessary for more specific audit procedures, such as tests of details and substantive procedures, which require extensive data analysis, including trend, ratio, and correlation analysis on structured, semi-structured, and unstructured data to corroborate evidence.

4.2.2 Manual procedures related challenges

The literature highlights that FA manual procedures are laborious and prone to errors, contributing to inefficiencies (Sekar, 2022; Werner et al., 2021). Respondents emphasized the repetitive nature of these tasks, highlighting the risk of oversight of critical evidence and data. Interviewee 1A stated:

"*We do too much manually, which could also be done by a computer. I think that we should use more techniques to make our work more doable, also to decrease the number of hours that we work*" (Personal, communication).

Tests of controls emerged as a challenging area due to the manual nature of inquiries, observations, inspections, and reperformance (EY GAM). Tests of controls are audit procedures performed to evaluate that the system of internal control to verify it can prevent, detect, or correct material misstatements. FA experts interviewed in 2B, 3C, and 8A, provided tangible examples, such as verifying authorized employees' signatures. Reconciliations - processes that require checking and verifying the consistency among two or more data sources - were also mentioned as inefficient, requiring verification of consistency among various data sources. Extensively performed during audits as tests of details and to verify the completeness of data populations, as well as during the financial statement close process (EY GAM), these procedures are a significant source of inefficiency and problems. As explained by data science expert, who has extensive experience in applying AI solutions to FAs, during interview 5E, the financial statement close process involves comparing opening balances to prior period financial statements, reconciling period-end amounts on the trial balance to the balance sheets, and ensuring consistency among financial statement disclosures and other information in documents containing audited financial statements, which was corroborated by EY GAM. FA experts interviewed in 1A, 9H, and 7G emphasized the extensive time needed for manual reconciliations, sometimes taking several weeks. FA expert 1A mentioned,

"*I just did a group engagement where we needed to reconcile a lot of documents, which were in different formats or in different tables from the same application. This procedure should be more efficient to perform with some kind of tool, instead of doing everything manually, because it takes several weeks to reconcile everything*" (Personal communication, 2024).

As mentioned in the previous section, some non-AI automated solutions exist but are not fully effective, expressed by FA experts during interviews 3C, 8A, 12I. For example, numbers can appear in different formats, like "six hundreds" instead of 600 (Personal communication). Current solutions can't handle these variations, requiring manual checks. This inefficiency is worsened by the increasing data volumes (Personal communication, 2024).

The sampling procedure was highlighted as problematic, especially with large datasets (Sekar, 2022; Werner et al., 2021).

Sampling is a procedure applied for tests of controls and tests of details (EY GAM). EY GAM highlights that the objective is obtaining a representative sample of the population. Sampling strategies include random sampling, systematic sampling—where the population is divided into intervals, and an item is selected randomly from each interval—and haphazard sampling, which lacks structure. However, EY GAM notes that assuming a homogeneous population for equal probability sampling can overlook riskier items. Population testing is sometimes used, such as for recalculations (EY GAM), employing automated techniques that experts suggest could be improved, as manual understanding of the recalculations is still needed, as explained by FA and data analytics expert in intervie 3C.

In general, interview results identified the sampling procedure as problematic and bringing risks of overlooking risky transactions, a theme recurring in FA expert interviews 1A, 2B, 3C, and 13L. Additionally, the risk of overlooking risky transactions is aggravated by the process of selecting a fixed sample size, as explained by FA expert 1A. The literature suggests population testing as an alternative, (KPMG, 2021; EY 2018) although opinions on its feasibility are divided. For instance, one partner raised concerns:

"*If you get millions of transactions for one day, and you're about to analyse all those transactions, if you find outliers to your presumed process or transaction scheme, then you would need to analyse all those outliers. So, there is a huge potential for a huge amount of additional work that someone needs to do*" (Personal communication, 2024), a perspective shared by another FA expert, during interview 3C.

The literature highlights tasks like reviewing contracts and documents, analysing information, drafting reports, and transcribing interviews as examples of manual procedures (Sun & Vasarhelyi, 2017). These tasks, although straightforward, occupy significant auditor time that could be used for higher-level tasks, a view shared by a senior financial auditor and a partner in assurance. Additionally, manual and routine procedures increase the risk of mistakes due to fatigue from looking at hundreds of invoices, expressed by several experts interviewed, including data scientist expert 5E. During interview 7G, a manager in FA highlighted that:

"*What is often a challenge is that you have a lot of manual procedures to do which are very factual in nature [...] that can be very time consuming, but it's not something that requires any form of judgment. So, the real challenge is getting the capacity there to perform these really simple and I would say, uninteresting procedures*" (Personal communication, 2024).

Inquiries, both written and oral, as well as interviews, are integral to audit procedures. They are used to assess risk of material misstatements, evaluate fraud risk, determine going concern, assess litigation and claims, ensure compliance with laws and regulations, and perform tests of controls (EY Atlas). A partner in assurance services noted that auditors need to transcribe oral inquiries or interviews, draft a process diagram, and have the client review it, highlighting the time-consuming nature of this task (Personal communication, 2024).

4.2.3 Subjectivity in professional judgment

Professional judgment is integral to the audit process, yet its subjectivity presents challenges, a view shared by a senior financial auditor and a partner in assurance, as well as the amount of information it should be based on. The literature addresses that lack of standardized guidance and the need for extensive data analysis to support decisions can lead to inconsistencies (Jan, 2021). While emphasizing the importance of professional judgment, interviewee 9H acknowledged the auditors' risk of bias in decision making, due to previous years' knowledge.

Areas involved are the evaluation of risk assessment procedures which require extensive data analysis and knowledge (Sun, 2019), fraud detection (Jan, 2021; Tang & Krim, 2018), decisions on audit procedures, such as the selection of sampling methods (Sekar, 2022) or forming an opinion on the going concern (PwC, 2017). EY GAM confirms the pervasive use of judgment in audit planning and substantive analytical procedures. However, it also highlights the difficulty in ensuring consistency across different audits. Determining the sufficiency of evidence is a delicate task, often subjective and contingent upon the individual auditor's discretion, leading to potential disparities in audit decisions (Interview 1A, 8A), consistent with literature findings of Jan (2021).

The risk of bias and the need for a more solid basis for decisions were highlighted as critical issues.

4.3 Deep Learning capabilities – SQ2

Results will be presented in these sections exclusively from interviews with experts with data science knowledge, therefore interviews considered are: 4D, 5E, 6F, 11F, and 13L.

The examination of internal documents provided limited results due to the unavailability of sufficiently technical materials. However, the extensive literature on the subject provided robust insights into the characteristics and capabilities of DL.
Interviews with data science experts underlined the ability of DL solutions to process structured, semi-structured, and unstructured data types, confirming their capability in identifying complex patterns within it. Interview 10E provided additional evidence regarding the numerous capacities of DL models, offering incredible value especially when DL's singular capabilities are stacked together. For instance, the expert expressed how applying a single DL capability, such as making documents machine-readable, would not provide much value. Instead, by adding NLP capabilities, such as identifying key relevant parameters, and presenting them in a table, the technology becomes more useful. The same expert stated that DL capabilities need to be stacked in the most meaningful way for different use cases, which is a sentiment derived from data science expert interviewed.

Interview results accentuate DL's competency in handling vast data volumes and diverse formats, its automated feature extraction capabilities, and its adeptness in pattern recognition from intricate data sets. Moreover, experts data science experts interviewed underscored DL's adaptability and versatility across varying use cases, highlighting its applicability to the multifaceted challenges encountered in FA.
Five principal macro-applications of DL were recurrently identified within the literature. Although the interview protocol did not explicitly investigate for opinions on these applications to conserve time, relevant insights were nonetheless obtained from expert responses regarding potential DL applications in the FA process. NLP capabilities emerged as a significant theme, with their relevance confirmed by interviews 5E, 6F, and 11F, alongside the utility of computer vision techniques. While big data analytics was not explicitly mentioned, the adeptness of DL in managing substantial data volumes was affirmed, particularly in by a data science interviewed in 4D, who revealed that while a minimum threshold of data is necessary for DL to function effectively, there is no upper limit to the volume of data DL can process. The proficiency of DL in analyzing both semi-structured and unstructured data was unanimously recognized by all interviewed data science experts.
The potential of DL to generate actionable recommendations was highlighted and deemed particularly useful within the FA domain. DL's capacity to autonomously process unstructured data types, such as news articles or social media posts, conduct sentiment

analysis, and synthesize an overview of the findings to discern trends was illustrated. Should such trends exhibit significant fluctuations, DL has the capability to pinpoint the underlying event or cause, assuming the relevant information is present within the input data. Anomaly detection was also affirmed as a DL capability, as evidenced by the discourse in an interview with a data science expert.

These applications encompass NLP, computer vision, anomaly detection, recommendation systems, and big data analytics. Interview findings corroborated these observations. Despite its less prominent depiction in literature, DL's automation capabilities emerged as a focal point in interviews, with data science experts 4D and 5E accentuating its potential to streamline processes within FA, including specific use cases and reconciliation procedures.

Prevalent DL models highlighted in literature include CNNs, DCNNs, RNNs, LSTMs, AEs, LLMs, and GANs. While LLMs were less emphasized in literature concerning FA applications, interviews shed light on their significance in DL applications within FA contexts. GANs were not mentioned in interviews. However, it is important to note that due to time constraints, this question was asked of only two respondents, and one of them revealed insecurity in extensive knowledge about models other than RNNs, CNNs, or LLMs. Comparative discussions with traditional ML underscored the distinct advantages of DL models, particularly their deep architecture facilitating rapid learning and high performance in FA contexts, as corroborated by data science expert in interview 4D. Interview results from 6F and 10E mentioned these models when asked about the main DL classes of models used to address FA challenges.

4.4 Deep Learning applications in financial audits – SQ3

Below, SQ3 results are presented and divided in two sections. The first explains the survey results providing a brief analysis. The second section connects survey results with literature, document, and interview findings, ensuring a thorough analysis.

4.4.1 Survey results

Survey respondents referred to the FA challenges as “use cases”, therefore the two terms will be used interchangeably in the following text. Table 5 provides an illustration of survey results.

	Survey respondent 1 (SR1)	Survey respondent 2 (SR2)	Total of actual answers	Expected answers
N. of **strongly agree** answers	2	4	**6**	/
N. of **agree** answers	32	20	**52**	/
N. of **disagree** answers	0	5	**5**	/
N. of **stronlgy disagree** answers	0	0	**0**	/
Total of Likert answers	34	29	**63**	68
N. of filled **blank cells**	6	0	**6**	12
N. of answers to **open questions**	11	10	**21**	/

Table 5. Survey results

A pattern emerged in the open-ended answers of SR1. Out of 11 answers, 5 pointed out possible inconsistencies with FA challenges classification and aggregation, as well as with the aggregation of DL capabilities. Examples of SR1's feedback include comments like “Strong overlap with use case 3” and “Strong overlap with use cases 3 and 6. Consider homogenizing DL capabilities.” SR1 highlighted that word prediction includes text generation, indicating a need to bring DL capabilities to the same level of specificity. Two of SR1's open-ended answers emphasized the necessity of making scanned documents machine-readable through Optical Character Recognition (OCR). SR1 and Interviewee 10E described this as a computer vision capability that allows making scanned documents, including handwritten text, machine-readable. This capability was also mentioned in Interview 7G. Although this was not specifically found in the considered literature regarding DL, it was corroborated by internal document analysis. As a matter of fact, EY already employs in its operations, mainly outside of FAs, but now slowly incorporating this process as well, a ML and DL powered tool that allows automation and support for document management purposes, including data extraction and feature identification. Employing automatic feature extraction, classification, and document generation, this tool exploits ML and DL, with DL mainly employed for OCR and language models (EY Discover, 2021).

Finally, some of SR1's answers provided specific insights. For instance, SR1 noted that anomaly detection does not directly help select representative samples, but rather checks the sample's representativeness after selection. Additionally, SR1 clarified that web scraping algorithms – algorithms employed to extract data from the web (Interview 6F) - do not necessarily require AI or DL, a point consistent with Interview 6F's statement that DL can be applied to organize the data after web scraping algorithms are employed to extract data from the web.

SR2's open-ended answers clarified use cases and added inputs regarding DL capabilities. For instance, SR2 highlighted that minutes of meetings might also be recorded. SR2 also suggested that for the use case of filling standardized forms, DL needs to scan and extract patterns from pre-filled documents. Additionally, Regarding the use case of filling standardized forms, SR2 added the DL need to scan and extract patterns from pre-filled documents, which is an additional explanation and step that should indeed be added in the explanation SR2, like SR1, pointed out that report generation includes data not only from images but also other types of documents, aligning with EY GAM's findings. Regarding the use case in planning using historical and external data, including market reports and news articles, SR2 suggested adding the DL capability of sentiment analysis to detect fake news. This addresses the concern of Interviewee 1A about the reliability of audit evidence, consistent with EY GAM's explanations.

Regarding processing oral inquiries, SR2 suggested exploring other DL models, such as BERT and XLM-R, which are part of the transformer model class, other models such as the ones discussed. These models were encountered in the literature, but not extensively in DL applications directed specifically at FA problems. Therefore, they were not included in the thesis discussion, but are noted as potential areas for future research. Both respondents expressed concerns about anomaly detection's effectiveness in selecting representative samples, highlighting the importance of non-outliers in fraud detection. SR2 also recommended CNNs and RNNs for fraud detection, depending on the data and specific problem, such as fraud in signatures. However, since detecting fake signatures was addressed in a separate use case, CNNs will not be added to this use case.

Towards another direction, SR2 also suggested that for the use case of determining the sufficiency of evidence based on historical data of past similar engagements, simpler ML models may be applied.

In summary, the survey results showed positive outcomes, with many connections made by the thesis writer being confirmed. The pie charts in Figure 8 illustrate the distribution

of responses. The pie charts are created through Python code, built with the support of ChatGPT-4.

Apart from providing validation to the connections between FA challenges – or use cases – and DL capabilities and models, the main inputs from the survey helped restructure the division of FA challenges and DL capabilities to bring them to a similar level of specificity. Additional useful suggestions regarding DL capabilities were obtained and corroborated by further document analysis and interviews. The survey validated the approach of matching DL capabilities and models to FA challenges, providing a solid foundation for further research and practical applications.

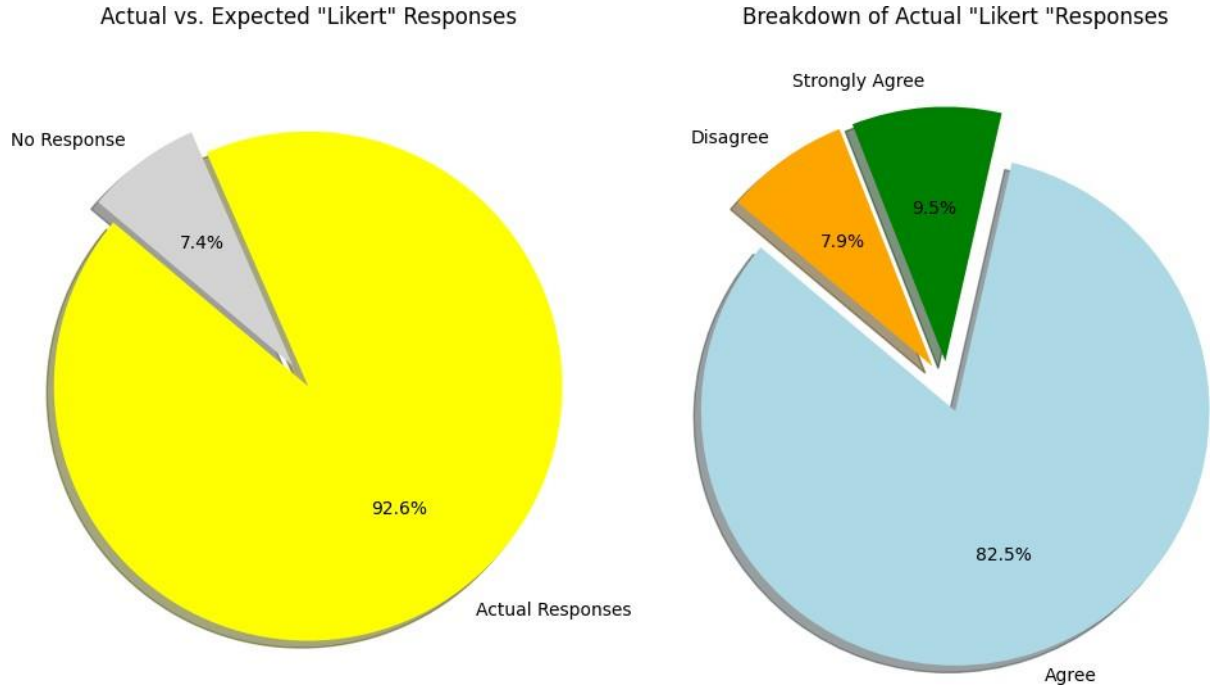

Figure 8. Breakdown of survey responses

4.4.2 DL applications to FA challenges

This chapter delves into the intersection of DL and FA, exploring how DL solutions can address longstanding and newly identified challenges within the field. Drawing upon a comprehensive research approach encompassing literature review, EY's audit methodology document analysis, and expert interviews, the connections between specific FA challenges and DL solutions have been made by the thesis writer and later validated by experts through a survey.

The challenges delineated have undergone a categorization procedure facilitated thorough document analysis and survey inputs. This examination has allowed for the recognition

of patterns and relationships among various FA challenges, enabling the grouping of related challenges into coherent categories.

4.4.3 Control Testing and Validation

Understanding controls involves applying audit procedure to gain knowledge of how the client's internal control system operates (EY GAM). Interviews and document analysis have highlighted these procedures as being challenged by manual, inefficient processes. DL can automate the verification of authorized personnel's signatures on documents. To address this, SR1 and SR2 validated DL capabilities in text categorization, image recognition, and image classification. SR1 suggests including the capability of making scanned documents machine readable. This translates into applications in NLP and computer vision. Suggested models are RNNs, CNNs, and DCNNs, validated by both respondents.
FA experts noted the issue of counterfeit signatures during interviews 2B and 3C. Document analysis shows that forensic involvement is necessary in such cases. Drawing from the literature (Camacho & Wang, 2021), this procedure can be automated using DL applications in NLP, computer vision, and anomaly detection. Extending to FA is possible and validated by both survey respondents. Leveraging the capabilities in writer identification, image analysis, and pattern recognition to verify the authenticity of signatures and detect image tampering, validated by both survey respondents. Also in this case, SR1 suggests including the capability of making scanned documents machine readable. Suggested models for this application include RNNs, CNNs, and AEs, validate by both respondents.
Manually processing oral inquiries is another identified inefficiency in FAs (Sun & Vasarhelyi, 2017). Results from interview 8A and 9H emphasize the potential benefits of tools to assist auditors in focusing more on the interview itself rather than on subsequent questions. Additionally, expert 9H recognized the benefit of applying an automating tool to process oral inquiries. DL-powered NLP can streamline this process by automatically translating speech to text, tagging parts of speech, performing sentiment analysis, and text classification (Fergus & Chalmers, 2022), validated by SR1 and SR2. Through sentiment analysis, the tone of inquiries can be detected and cross-analysed, helping auditors streamlining the process of comparing audit evidence. Suggested models for this application are RNNs. SR2 suggests the application of BERT and XLM-R for multi-language processing. Given that oral inquiries are a constant in FA procedures, this application can benefit various parts of an audit, including tests of control.

4.4.4 Tests of Details

Tests of details are another type of challenging manual procedure exacerbated by large data volumes. Interviewee 5E illustrated this challenge with the example of cross-checking invoices to verify accurate accounting. DL can automate this process, enabling a larger sample of items to be analyzed timely, thereby addressing the issue of overlooked data with fixed sample sizes. DL models based on CNNs, DCNNs, RNNs, and LSTMs are suggested for alleviating manual procedures while processing large data volumes, leveraging NLP and computer applications. Validate by SR1 and SR2, DL capabilities that can be employed are text categorization, image recognition, image classification, pattern recognition, and OCR, with the latter being an input by suggested by SR1. This reduces the risk of overlooked transactions as highlighted by literature (Sekar, 2022; Werner et al., 2021) and interviews (Interview 1A, 2B, 13L).

Tests of details are usually performed on samples (EY GAM). As addressed in the literature and interviews, selecting representative samples is an issue. While all data science experts interviewed agree on DL's capability in addressing this problem, inconsistencies derived from survey results. For instance, both respondents commented on the role of anomaly detection in this use case. Consistent with SR2 stating that in case of fraud detection non-outliers are equally important to be checked, SR1 suggests the use of anomaly detection not for selecting samples, but rather for checking whether the sample is truly representative. In any case, SR1 agrees on the use of AEs for this application, consistent with literature findings. As a results, it can be inferred that AEs can select representative samples in large data volumes, but the role of anomaly detection lies in checking how representative the sample is compared to the data population.

4.4.5 Document and Data Reconciliation

Reconciliation procedures are challenging due to the lack of appropriate automated tools – as mentioned in interviews 3C and 7G - and the increasing volume of data. Common procedures in FA, such as control testing and financial statement close processes (EY GAM), require the performance of reconciliations. Interview results validate DL's ability to automate the process of reconciliations, according to the opinion of expert 13L, and specifically in the realm of the financial close process, according to expert 5E. DL, leveraging applications in NLP, computer vision, and anomaly detection, can automate reconciliations, flagging inconsistencies for auditors to focus on. Validated by SR1 and SR2,

capabilities such as multi-document analysis, document classification, and automated mathematical checks among numerical data and formulas explained in flow text. Suggested models include CNNs and RNNs.

Interviewee 6F shared a differing perspective, suggesting that DL solutions might be overengineered in this context. However, survey results support DL's efficacy, especially for addressing different data formats and eliminating manual procedures still required by automated but non-AI-based tools.

4.4.6 Risk Identification, Assessment, and Audit Planning Procedures

Risk identification and assessment procedures are performed in the second phase of an audit, involving the identification of risks of material misstatements. This process requires analyzing immense volumes of diverse data types, necessitating focused analysis and cross-checking. Manual procedures in this phase are sources of inefficiency and, as derived from interview 5E, a source of risk, as auditors cannot thoroughly analyze all relevant available data in a timely manner.

Audit planning is an iterative and continual process (ISA 300; EY GAM). Planning entails defining an audit strategy that is responsive to the risks of material misstatement identified and requires planning the audit activities to be performed throughout the whole audit, which includes determining the sufficiency of evidence to be obtained (EY GAM). Results confirm that DL can assist auditors in the phase of identifying risks of material misstatement and planning in various ways.

Interviewee 5E explained that a DL-based recommendation system can support the auditors' judgment while planning an engagement, consistent with Sun's (2019) findings. This technology can analyze past engagements, gathering data from the audit platform where documentation of all engagements is stored. DL can detect past engagements that are similar in terms of certain features, such as the size of engagement, effort required, geographic and industry specific paraments, and it can detect the planning of those engagements, explained in interview 5. This technology can provide recommendations to auditors for procedures to consider, enhancing audit planning. Leveraging DL applications in NLP and recommendation systems, for pattern recognition and multi-document summarization, models such as CNNs and AEs, filled in the survey by SR1, can be leveraged for these applications, validated by both survey respondents.

DL applications promise significant value in the process of gathering relevant knowledge about an entity, particularly in the phase of understanding the business, which, as

extensively described in previous sections, requires a substantial amount of data analysis. Data science expert interviews 4D and 5E concur on DL's capability in such applications. Expert 5E highlights that not only is gathering all relevant information challenging, but additional difficulties arise from staying current with new information emerging during the engagement. Interviewee 5E provided an example to illustrate this issue. For instance, if a client operates in the automotive industry in Europe, a shortage of rubber due to an event in America could impact the client's business during the current period. This information is crucial for auditors as it helps in forming expectations, which serve as a basis for their judgment and conclusions, however it bears a high risk of being undetected. DL solutions can efficiently gather, highlight, and flag the relevance of such data through NLP, recommendation systems, and big data analytics applications. By recognizing complex patterns and utilizing capabilities such as information retrieval, text summarization, and text classification, DL models can effectively perform these tasks. Additionally, SR2 suggested that sentiment analysis can detect fake news. Suggested models for these applications include AEs, addressing the need for dimensionality reduction, and LLMs.

Interviewees 5E and 12I also noted that DL can compare industry, competitors, and client's previous year analytics, in addition to past engagements. DL can extract basic analytics – such as EBITDA - from the financial reports of the previous year—since the current year's report is still being audited and thus not yet finalized—and compare these analytics with similar companies or competitors, highlighting relevant data and potential outliers. This comparison provides crucial information for auditors to focus on, streamlining the process and allowing auditors to concentrate on more valuable procedures. As explained in interview 5E, and corroborated by literature results, if an analytic deviates significantly from industry values, auditors may consider this a risky area warranting further investigation. Leveraging NLP and anomaly detection capabilities, DL capabilities in semantic matching, time series analysis, and numerical cross-checking, can be employed (validated by SR1; no response by SR2). Models such as CNNs, LSTMs, AEs, and LLMs are recommended for this application.

Another FA challenge, identified through interviews and corroborated by EY GAM's analysis, pertains to areas involving estimates, which inherently involve a significant amount of judgment and uncertainty. As interviewee 12I explained, when estimates are involved, uncertainty increases. Auditors play a critical role in evaluating accounting estimates to ensure these estimates accurately reflect the financial reporting framework (EY Atlas). Extending the previous capabilities of DL to create sound expectation ranges

based on industry, market, and entity-specific knowledge, CNNs, RNNs, LSTMs, and AEs can assist auditors by leveraging DL's capabilities of automatic feature extraction, pattern recognition, and anomaly detection, thus enhancing the evaluation process, which are results validated by SR1. However, diverging feedback was obtained, as SR2 suggests GANs as more appropriate for this task.

Evaluating the going concern predictions of the audited client is another high-risk area identified in the literature (Jan, 2021). This is an example of procedure where estimates are involved. Through DL applications in NLP, models can construct going concern prediciton models based on RNNs (Jan, 2021), leveraging time-series analysis and time-series forecasting. As SR2 suggested, LSTMs can also be appropriate for this task.

Manually scanning meeting minutes is another challenge that DL-powered NLP applications can address. As validated by SR1 and SR2, using capabilities such as text categorization, text classification, sentiment analysis, and multi-document summarization through LLMs, RNNs, and LSTMs, can address this challenge.

Identified in the literature as another pressing challenge in FAs due to its complexity, fraud risk identification is an iterative process that emerges from procedures performed in all risk assessment activities (EY GAM). EY GAM itself underscores the challenge, as distinguishing errors from fraud involves discerning intent, with fraudsters attempting to conceal their actions (EY GAM). This difficulty is echoed in interviews, such as respondent 1A, who noted the challenge of detecting fraud. Data scientist expert interview 4D identified fraud detection as an application area for DL during the interview. Literature has highlighted LSTMs as suitable DL models for fraud detection in large datasets of financial and non-financial data (Jan, 2022), enabling automatic feature extraction (Alghofaili et al., 2020). Anomaly detection, a critical step in fraud detection, can also be effectively addressed using AEs. According to SR2, CNNs and RNNs are valid solutions for fraud detection, depending on the type of document involved, with CNNs being particularly suitable for detecting fraud in signatures. However, this use case is addressed separately and already discussed. Therefore, RNNs, LSTMs, and AEs are suggested models to enhance fraud detection in large volumes of financial and non-financial data, thus alleviating the burden of professional judgment.

4.4.7 Regulatory Compliance and Reporting

FA is a highly regulated process, involving compliance with ISAs, reporting frameworks such as the International Financial Reporting Standards (IFRS), and audit methodologies

(EY GAM). Expert insights, derived from interviews 3C and 5E, reveal the laborious nature of tasks such as completing standardized forms and adhering to prescribed audit protocols, as stipulated by ISAs. Manual completion of standardized forms is characterized by inefficiencies inherent in repetitive, time-consuming tasks. For instance, filling standardized word forms is a manual process that can be automated using DL. NLP applications, leveraging capabilities in word prediction, and pattern detection from already-filled forms, with the latter suggested by SR2, can streamline this process. Suggested models are AEs, LLMs, and GANs, validated by both survey respondents. Still resulting from interviews, the verification of financial statements against regulatory frameworks demands meticulous scrutiny and consumes considerable time. Rule-based checklists are used to verify financial statement compliance with reporting frameworks. As explained in data science and FA expert in interview 5E, DL can automate these checklists, particularly in the financial statement close process. NLP and computer vision applications can leverage capabilities such as information retrieval, text categorization, text classification, multi-document summarization, and OCR. LLMs, RNNs, and LSTMs are suggested models for this application, while SR2 did not provide opinions on this use case.

Finally, the last phase of an audit requires drafting reports based on all the audit evidence documented in the audit platform (EY GAM). This is a yet a highly manual procedure that requires processing large amounts of data. Scholars propose DL solutions to automate the drafting of reports (Alfarghaly et al., 2021), and interviewee 5E confirms that DL can automate this process by automatically generating a draft of the report, being able to extend it to the FA domain. Then, the auditor will need to evaluate it and finalize it. Both survey respondents agree that generation of reports does not require processing data from images only. Therefore, because of an additional review of the literature, new capabilities have been connected to this use case. As proved by the literature, image processing, word prediction, information retrieval, text categorization, leveraging applications in NLP and computer vision, are all DL capabilities that can address this FA challenges. In this case, they can be leveraged to automate this report generation. Suggested models are CNNs, and LSTMs, models validated by both survey respondents.

4.4.8 Conclusion

DL offers significant potential to address various challenges in FA, characterized by large data volumes, manual processes, and over-reliance on professional judgment. By

leveraging capabilities in NLP, computer vision, anomaly detection, big data analytics, and recommendation systems, DL can streamline control testing, substantive procedures, document reconciliation, risk assessment, and regulatory compliance. Suggested models such as CNNs, DCNNs, RNNS, LSTMs, AEs, LLMs and GANs provide robust solutions to enhance the efficiency and accuracy of FA processes, ultimately contributing to more effective and reliable audits.

4.5 Conceptual framework of results

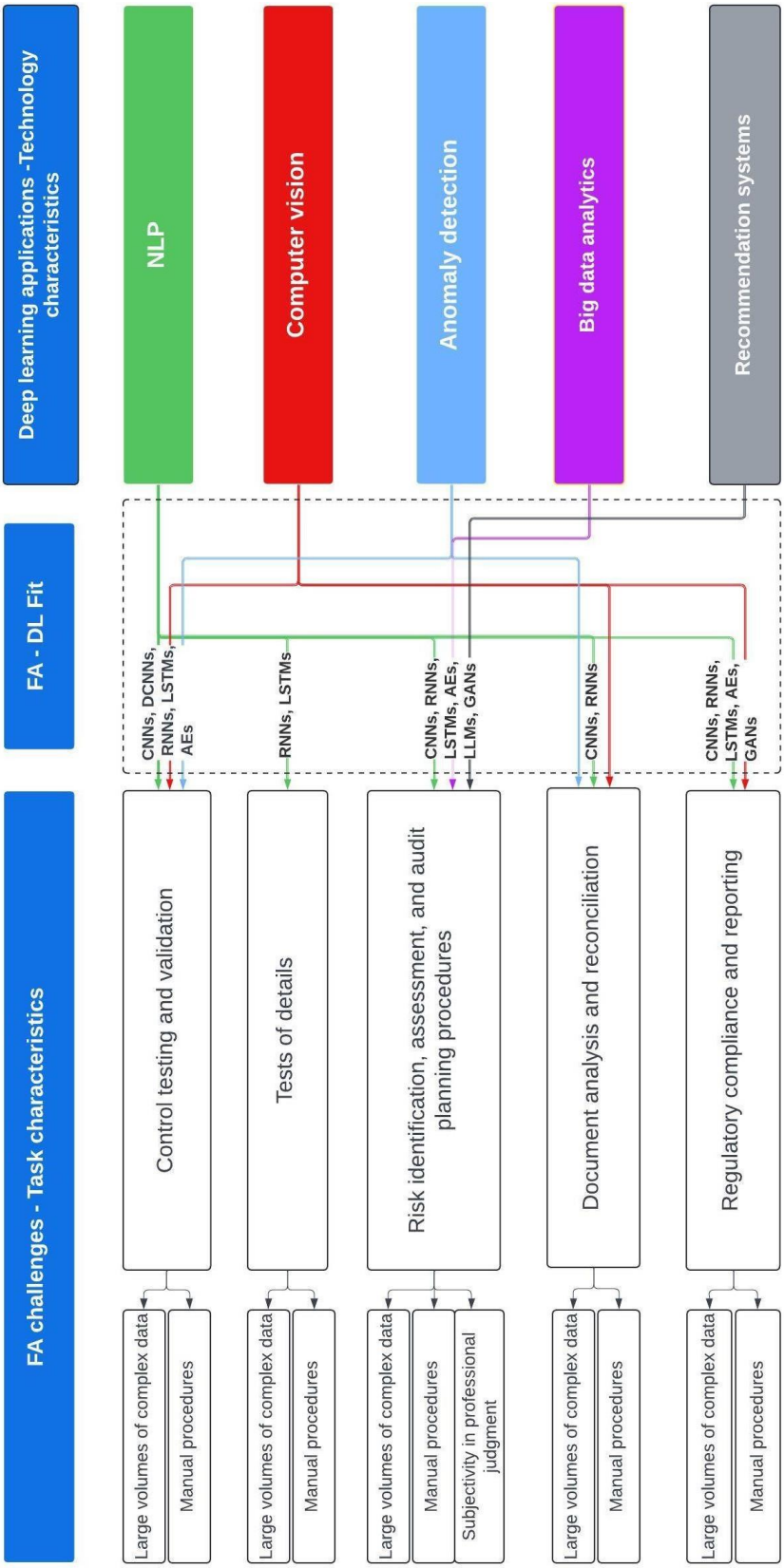

Figure 9. Results conceptual framework.

Figure 9 represents the conceptual framework of results, providing an overview. Due to further details discovered and validated through the research, presenting all details in an illustration would have been complex, hindering the comprehension of connections for the reader. Figure 9 has been designed. to remain coherent with the TTF model's illustrative representation and to create a visual continuation compared to the conceptual framework presented at the end of the literature background. More detailed connections, including specific FA challenges, are presented in Table 6. A larger image of the table can be found in appendix 8.

Challenge category	Example	Challenge type	DL's application	DL model(s)
Control testing and validation	Verification of signatures made by authorized personnel	Manual procedure, large data volume	NLP, computer vision	CNNs, DCNNs, RNNs
	Veryfing legitimacy of signatures and documents	Manual procedure	NLP, computer vision, anomaly detection	CNNs, RNNs, AEs
	Oral inquiries	Manual procedure	NLP	RNNs, LSTMs
Tests of details	Tests of details	Manual procedure, large data volume	NLP	CNNs, DCNNs, RNNs, LSTMS
	Verifying representativeness of samples	Large data volume	Anomaly detection	AEs
Document analysis and data reconciliation	Reconciliation procedures	Manual procedure, large data volume	NLP, computer vision, anomaly detection	CNNs, RNNs
Risk identification, assessment, and audit planning procedures	Audit planning and audit procedures selection	Manual procedure, large data volume, professional judgment	NLP, recommendation systems, big data analytics	CNNs, AEs
	Understanding the business	Manual procedure, large data volume, professional judgment	NLP, recommendation systems, big data analytics	AEs, LLMs
	Evaluation of competitors and client's historical analytics	Manual procedure, large data volume, professional judgment	NLP, anomaly detection, recommendation system	CNNs, LSTMs, AEs, LLMs, GANs
	Evaluation of going concern prediction	Manual procedure, large data volume, professional judgment	NLP	RNNs, LSTMs
	Scanning minutes of meeting	Manual procedure	NLP	RNNs, LSTMs, LLMs
	Fraud risk identification	Large data volume	Anomaly detection	CNNs, RNNs, LSTMs, AEs
	Evaluation of accounting estimates	Large data volume	NLP, anomaly detection	CNNs, RNNs, LSTMs, AEs
Regulatory compliance and reporting	Filling standardized forms	Manual procedure	NLP	AEs, LLMs, GANs
	Verification of compliance through checklists	Manual procedure	NLP, computer vision	RNNs, LSTMs, LLMs
	Drafting reports	Manual procedure, large data volume	NLP, computer vision	CNNs, LSTMs

Table 6. Results conceptual framework, with detailed use cases

5 DISCUSSION

This chapter discusses the results found in chapter 4, briefly summarizing them as well as comparing the literature findings. The chapter is divided following the same structure in the whole thesis, therefore section 5.1 addresses results gathered for SQ1, the task characteristics; section 5.2 discusses results and literature about answering SQ2, technology characteristics; 5.3 discusses the fit pars, aimed at answering SQ3. Finally, additional findings, implications, and research limitations are presented.

5.1 What are the main challenges in a financial audit process? – SQ1

This section first presents the discussion regarding the FA process. Then, the discussion regarding the FA challenges is addressed.

5.1.1 The financial audit process

The audit process is divided into four macro phases, although it should not be viewed as strictly chronological due to the iterative and interdependent nature of the procedures. In the initial planning phase, the audit scope and service requirements are defined, and ethical and independence agreements are determined. The risk identification and assessment phase involves detecting potential material misstatements and understanding the entity's internal control system. The design and execution of responses to risks involve strategies to address identified risks. Finally, the concluding phase focuses on communicating findings, reassessing risks, and ensuring comprehensive documentation.

The results highlight the flexibility in defining FA phases, as different literature sources present varying divisions despite adhering to the same ISAs (Werner et al., 2021). This indicates that the interpretability of standards can differ, allowing professional judgment while still complying with regulations. However, the content of the phases remains consistent. Notably, there is an agreement that the planning phase is an iterative process. However, EY GAM analysis reveals that most audit procedures are interdependent and iterative, further emphasizing the complexity of the process and the importance of auditors thoroughly analyzing and cross-checking all audit evidence. The literature often underrepresents a detailed description of the FA process. The results from the document analysis provided a deeper understanding, which was crucial for conducting interviews with financial auditors and data scientists, offering a solid grasp of FA challenges to

present to them. Understanding these detailed procedures was essential for locating challenging processes within the FA framework and comprehending their nature and impacts. While these results do not have a direct implication on the hypothesis or SQs, they are vital for ensuring a comprehensive understanding of the topic, ultimately aiding in answering SQ1, SQ3, and the main research question.

5.1.2 What are the current challenges encountered in a financial audit process?

The results of the current challenges in the FA process reveal a consensus on the main classes of issues, as identified in existing literature. However, different levels of detail and varying perspectives emerged.

The research reveals that the handling of large volumes of structured data, such as journal entry testing, has seen some resolution through non-AI automated techniques. This development addresses literature concerns regarding the absence of specialized software for processing such data. Nevertheless, these solutions are not comprehensive, as they fail to effectively manage semi-structured and unstructured data. The interviews highlight the need for better data utilization, such as comparative and prior year data analyses. While automation for analyzing unstructured documents like news articles is emerging, other data types, including social media content, remain untapped in practice.

Manual procedures persist as a significant challenge, with interviewed auditors emphasizing the monotony of repetitive tasks and the underutilization of their skills. Specific examples derived from interviews, such as verifying authorized personnel signed documents and filling standardized forms, underscore the need for automation to alleviate the burden on personnel.

Population testing via automated methods is a topic of debate, as interview results provided diverging perspectives. While population testing can identify outliers in large transaction datasets and address the risk of overlooking risky transactions, this may inadvertently increase workload with potentially negligible benefits, as many exceptions might have already been addressed by management. This dilemma was not covered in the literature and requires further consideration.

Professional judgment is integral and a necessary component of the FA process, as confirmed by the interviews and EY GAM. However, the application of judgment is filled with complexity, particularly when evaluating vast amounts of data to determine evidence sufficiency. This leads to varied auditor behaviors and the potential for bias, which is a concern that both literature and practice acknowledge.

The research findings contribute to a more nuanced understanding of the challenges in FA, aligning with the literature while also presenting new perspectives and specific examples. For instance, while Sekar (2022) underscores the need for specialized software for data processing, interviewees note the existence of suitable IT solutions, despite limitations in handling diverse data formats and the necessity for manual intervention in certain cases.

The research enriches the literature by detailing specific challenges, such as the reconciliation procedure, which was not emphasized previously. This discovery highlights the importance of targeted applications of DL, to address well-defined problems within the FA process. Given the high computational costs and engineering complexity of DL models, as found through interviews 4D and 6F, a clear understanding of the challenges is crucial for directing DL applications more effectively.

Finally, these results have a direct implication on the task characteristics component of the research, adopting the TTF theory as a lens. TTF theory posits that ISs are most effective when they align well with the tasks they are meant to support (Furneaux, 2011). In the context of FA, the task characteristics can be understood as the specific and challenges inherent in the audit process. The intent of exploring the task component is translated in SQ1: "*What are the current challenges in the financial audit process?*".
In summary, the research corroborates the literature on the main classes of problems in FA, but also introduces additional challenges and practical insights. This contributes to a more comprehensive picture of the FA landscape, providing a foundation for future improvements and the application of advanced technologies like DL.

5.2 What are the main capabilities of deep learning? – SQ2

The key findings from the interviews with data science experts provide a refined understanding of the main DL capabilities in the context of FA. These capabilities include adaptability to various data types and the ability to stack multiple DL functions for enhanced performance. Specifically, the ability to process structured, semi-structured, and unstructured data, while discerning patterns from it, proves to DL's versatility in managing the diverse data tasks inherent to the FA process. Moreover, the confirmation of DL applications in NLP, computer vision, anomaly detection, large volumes of complex data, and providing recommendation demonstrates its capacity to automate complex analytical tasks, which are characteristic of the FA domain.

These results provide a basis for continuing research by hypothesizing a fit between task and characteristics, which was performed and validated through a survey. The interviews served to corroborate literature findings and offered more specific insights. In general, the interview results confirmed DL's characteristics and capabilities, providing motivation for hypothesizing a fit in the subsequent survey.

The literature provided a strong foundation on this topic, with DL being well-documented by highly reliable sources. Books were the primary sources for the literature review (Fergus & Chalmers, 2022; Goodfellow et al., 2016), enhancing credibility. The interviews did not reveal differences, but rather confirmed and elaborated on DL's applicability in the FA context. Additional results highlighted the potential for stacking DL capabilities to derive greater value from its application.

In the context of TTF theory, these results have direct implications for the technology-characteristics component, which is addressed in SQ2: "*What are the main deep learning capabilities?*". The identified DL capabilities align well with the intricate, data-intensive tasks characteristic of FA, suggesting a strong task-technology fit and good foundation for testing the fit, although on a theoretical basis, through the survey. The ability of DL to handle large volumes of data and complex data types corresponds to the task requirements of comprehensiveness and accuracy in financial data analysis. Additionally, the stacking of DL capabilities, such as combining NLP with computer vision, enhances the technology's fit by enabling more sophisticated, multi-faceted analytical tasks that are often required in FA. This suggests that DL technologies are well-suited to meet the task characteristics of FA, potentially leading to improved efficiency and effectiveness in financial decision-making processes.

5.3 How can deep learning techniques be applied to address the challenges identified in the financial audit process? – SQ3

This study explores the intersection of DL and FA, revealing how DL solutions can address longstanding and newly identified challenges within the field. Through a comprehensive research approach, including literature review, EY's audit methodology document analysis, and expert interviews, the study establishes and validates connections between specific FA challenges and DL solutions.

The primary challenges identified include manual procedures, large data volumes, and support for professional judgment. Manual procedures, cited as the most burdensome by FA experts, are prevalent in FA use cases. DL applications, particularly in NLP, are

frequently mentioned, appearing in 14 out of 16 use cases. Other DL applications like computer vision and anomaly detection significantly impact the audit process by addressing high-impact procedures. The prevalent DL models identified are CNNs, RRNs, LSTMs, followed by AEs, LLMs, and GANs.

The results have a direct impact on SQ3, which asks: "*How can deep learning techniques be applied to address the challenges identified in the financial audit process?*", connecting DL applications and models to FA challenges.

The findings address the fit component in TTF theory. Results highlight that different DL applications and models are suitable for different FA tasks, emphasizing the importance of aligning DL capabilities with specific FA challenges. For instance, while anomaly detection is not ideal for selecting samples, it is effective for verifying the representativeness of selected samples.

The research significantly extends existing literature by providing more detailed and validated use cases of DL in FA. While previous studies, such as Sun (2019), generally described DL-powered judgment support in FA, this study corroborates and expands on these findings. The validation by experts in both DL and FA fields adds robustness to the results, presenting new connections and detailed applications not previously documented, including tasks like reconciliation and more precise use case division in risk assessment procedure. For instance, literature highlighted the use of DL for image forensics and signature verification (Camacho & Wang, 2021). This research confirms these applications in the FA context, adding details and validating them, such as confirming the use of NLP, computer vision, and anomaly for verifying the legitimacy of signatures and documents, which are critical for control testing and validation. Additionally, the inclusion of OCR capabilities for making scanned documents machine-readable, suggested by SR1, was not widely covered in the literature. This research emphasizes OCR's importance in enhancing DL applications in FA, particularly for automating the processing of physical documents. Finally, reconciliation procedures, often manual and data-intensive, can be addressed with DL applications in NLP, computer vision, and anomaly detection, as validated by experts. This extends the literature by providing concrete models and applications to automate these newly identified challenging tasks.

The results discussed in this section provide the final information to comprehensively address the three hypotheses formulated. By finding current FA challenges in SQ1 and uncovering DL capabilities in SQ2, SQ3 functions as a bridge to the findings of the first

two questions, providing a more technical nature to the research, as well as the details and validations required to answer to the main research question.

The findings confirm **H1**, demonstrating that all DL applications mentioned in this thesis significantly streamline the procedures of handling large and complex datasets. The use of all the reported DL models are adept at processing and analyzing vast and complex data volumes.

H2 is validated by identifying DL applications that automate tasks like analyzing news, checking regulatory reporting framework compliance through checklists, and completing standardized forms, which traditionally require substantial manual effort. All DL applications and models discussed in the thesis facilitate the automation of these procedures. This is consistent with literature findings, such as those by Sun & Vasarhelyi (2017), which discuss the potential for automation in FAs

H3 is validated through detailed use cases where DL supports professional judgment, such as planning engagements by comparing past data and evaluating accounting estimates. Again, all DL application areas and models discussed in the thesis can address providing judgment-support to auditors. These findings extend the work of previous studies (Sun,2019), providing concrete examples of DL's application in professional judgment support.

Attention needs to be drawn on the fact that each use case within the classes of challengess addressed by the hypotheses is optimally managed with specific DL applications and models, ensuring that the unique strengths of various approaches are leveraged effectively.

5.4 Business relevance

The research presented in this thesis has significant business implications, addressing practical and current issues in the FA domain. By exploring the application of DL solutions, this study theoretically proves DL's capability in providing valuable judgment support to auditors, as well as assisting them in analyzing large and complex data volumes and automating routine procedures. This enables them to focus on high-value procedures in FAs and ensuring a strong data-driven decision-making process. The findings suggest that by leveraging the unique strengths of various DL approaches, the FA process can be made more efficient. This is not only beneficial for auditors, but also for stakeholders that rely on accurate and reliable financial statements for decision-making. Companies like EY are already investing in DL for FA, indicating a trend towards innovation and the

adoption of new technologies in audit practices. Finally, by specifically addressing recognized FA challenges, targeted applications of DL can prevent the wasting computational resources on unnecessary tasks.

5.5 Scientific relevance

Scientifically, this thesis contributes to the body of knowledge by providing a comprehensive overview of DL applications in the FA process, a topic that has been relatively unexplored with fragmented literature. Few studies offer a comprehensive analysis of DL's benefits and applications within FA. While some existing works do address particular use cases in FA, there is a noticeable absence of research exploring the application of DL across a range of FA scenarios. This research aims to fill this gap by providing a thorough and inclusive overview of the FA process, defining various use cases, and addressing them through DL. In doing so, it also establishes connections to more technical aspects, such as specific models and capabilities that are suggested for these applications. This study lays the groundwork for future empirical research and practical implementations, by highlighting the unique capabilities of DL and exploring its potential applications in FA. Finally, by presenting the most current and pressing challenges in FA, the study not only suggests DL solutions, but also opens the door for researchers to consider and apply alternative technologies to tackle these issues.

5.6 Additional findings

5.6.1 AI in the loop

AI and DL in financial audits represent a shift towards an "AI in the loop" approach, as interviewee 5E stated, rather than the traditional "human in the loop" model. This paradigm, which is consistent with literature such as Sun (2019) and Jan (2022) even if not clearly stated, suggests that AI, including DL, are designed to support - not replace - human auditors. According to expert 5E, AI solutions are not built to perform the entire audit independently, but are integrated into the process to handle data-intensive and repetitive tasks. Moreover, DL plays a crucial role in judgment support, as it does not merely automate high-volume routine tasks but also assists in complex decision-making by providing data-driven insights and recommendations. This allows human auditors to focus on areas requiring high levels of critical thinking. Expert 12I highlighted that the traditional audit model, which relies on a hierarchical pyramid structure where manual

work is extensively reviewed at various levels, is fundamentally altered by automation. DL can prepare and suggest actions, but ultimately, human auditors review and make final decisions, ensuring the data-driven application of professional judgment. While automation changes the traditional hierarchical audit model, the essential role of human auditors remains intact. This synergy ensures a more efficient, accurate, and insightful audit process, demonstrating that AI are indispensable tools in modernizing audits, not as replacements, but as powerful allies in achieving greater audit quality.

5.6.2 Data Scarcity

DL models are known for their high accuracy, especially when trained with large datasets. The recent surge in digitalization has contributed significantly to the popularity of DL, as the vast amounts of data required for training these models are now available (Fergus & Chalmers, 2022). However, in the highly regulated field of financial auditing, accessing this data remains a substantial challenge. Interviews with experts knowledgeable in both the FA process and DL (Interviews 5E, 10E, 12I, 13L) raised the concern of data scarcity, revealing that it is the most critical issue when applying DL or any AI solution to audit procedures. The problem is twofold:

- Client hesitation: this poses a problem not only for using DL with a single client but also for training the models effectively across multiple clients.
- Legal constraints: the use of data is tightly regulated by law.

The traditional nature of audit procedures amplifies these challenges. According to interviewee 5E, the current cooperation between auditors and clients typically involves the client sharing the minimum amount of information necessary. Experts like 12L noted that clients are often reluctant to provide auditors full access to their data. As an example, this hesitancy limits the feasibility of selecting truly representative samples, as the entire data population cannot be considered. Clients are aware that audits can still be performed with limited data using traditional methods. For instance, audits can be completed by testing samples of just 25 items, so clients may not see the necessity of allowing auditors to test 10,000 items. Even when the audited party is willing to share data, training DL models remains problematic as it requires millions of data entries collected from various sources, making it even more challenging to convince multiple clients to participate.

Moreover, as expressed in interviews 12I and 13L, data usage is constrained by legal policies. The audit engagement letter, established at the start of an engagement, clearly

defines the specific purposes for which client data can be used. Although firms like EY are progressing towards updating these terms and conditions, the process is still ongoing, and clients have a say in the negotiations. However, as expert E pointed out, with the increasing integration of AI in business operations, clients are becoming more aware of the benefits of such innovations and the requirements to make them work. This awareness is leading to better education and upskilling of clients, which may reduce their skepticism over time.

As interviewee 5E explained, clients stand to benefit significantly from the shift to AI-embedded audits. These more efficient audit procedures would drastically reduce the time needed to complete an audit. The movement towards making audit a service and conducting audits several times a year instead of one could offer substantial benefits. The audit report could state that the client has been audited using AI methods, with extensive data, and multiple times a year, thereby enhancing trust in the market and providing a competitive edge.

In conclusion, while DL offers significant potential to revolutionize the FA process by improving accuracy and efficiency, its implementation is hindered by data scarcity due to client hesitation and legal constraints. Overcoming these challenges requires a collaborative effort to educate clients about the benefits of AI and to update legal frameworks to facilitate data sharing. If these obstacles can be addressed, the adoption of DL in financial auditing could lead to more reliable and efficient audit processes, ultimately benefiting both auditors and clients.

5.6.3 Black box and trust

In exploring the issue of interpretability and trust surrounding complex AI approaches, particularly DL models, it becomes evident that the concept of the "black box" nature poses a significant challenge. This term encapsulates the difficulty in understanding the internal workings of these models, hindering their application in critical areas such as auditing (Kokina & Davenport, 2017). While differing opinions exist, with some arguing that conceptual understanding is sufficient for adoption (Sun, 2019), others emphasize the ongoing need for transparency and alternative approaches to ensure trustworthiness (von Eschenbach, 2021).

Interview findings corroborate the recognition of DL models as black boxes, with varying perspectives on their explainability. Explainability refers to the ability to articulate the reasoning behind a decision, suggestion, or forecast made by an AI system. To cultivate

this attribute, one must grasp the inner workings of the AI model (Grennan et al., 2022), hence the close connection between the concepts of explainability and black box. This suggests that the level of transparency and trustworthiness can be influenced by the scope of application, with narrower applications lending themselves to clearer explanations of model decisions.

Moreover, interviewees also highlight the importance of transparency in providing insights into the accuracy metrics of DL models (Interview 6F). Tools like confusion matrices are instrumental in facilitating a deeper understanding of model performance, biases, strengths, and weaknesses, thereby enabling refinement and risk assessment.

While DL models may not be infallible, they can achieve high levels of accuracy (Interview 4D, 6F), often surpassing 90%, as evidenced by literature findings. Despite occasional errors, the argument stands that human fallibility also exists, suggesting a parallel in the acceptance of errors in both AI and human decision-making processes. (Interview 4D, 5E).

In conclusion, while literature and interviews present varied perspectives on the interpretability of DL models, a common thread emerges: optimism tempered by recognition of present challenges. Data scientists express confidence in the potential for future advancements to address these concerns, while recognising possibility of explainability with the use of narrow-scope models.

5.7 Research limitations

There are limitations to this research that need to be addressed. First, the methodology involved conducting interviews and administering a qualitative survey exclusively with EY employees, which could restrict the generalizability of the findings. However, the diverse geographic locations of the interviewees somewhat mitigate this limitation, enhancing the potential generalizability of the results. Second, the audit methodology analysis focused on EY's audit methodology (EY GAM), which, despite being based on internationally recognized standards (ISAs), may not be uniformly applied in all contexts. Furthermore, the qualitative nature of this research, which involved analyzing interviews, could have inadvertently introduced bias. Additionally, the extensive range of use cases and details analyzed, combined with time constraints during the semi-structured interviews, meant that not all intended questions were asked of every interviewee. This limitation highlights the challenge of ensuring comprehensive coverage of all relevant topics within the limited time available for each interview. The semi-structured nature of the

interviews, while allowing for free discourse and the exploration of unexpected insights, sometimes led to deviations from the planned questions, further contributing to this issue. This could result in a partial view of some aspects of the research topic, potentially overlooking certain insights that might have emerged if all questions had been uniformly addressed to all participants. Finally, some use cases presented in the results were validated by one respondent only, reducing the robustness of such validations.

6 CONCLUSIONS

This research aims to find solutions for addressing challenges encountered in FA. DL stands as a promising technology to alleviate several identified challenges within this process. The problem is approached through the lens of the TTF theory, which guided the formulation of research questions, the drafting of the methodology, and the analysis of results. Three SQs have been formulated, each addressing one component of the TTF framework, namely task characteristics, technology characteristics, and the fit between the two.

Through data triangulation, qualitative data was gathered and analyzed via interviews, audit methodology analysis, and a survey. The intersection of FA challenges and DL solutions has been theoretically investigated, providing an answer to the main research question. This chapter synthesizes the findings, answering the three SQs and leading to providing an answer to the main research question.

The primary challenges identified in the FA process include handling large volumes of structured, semi-structured, and unstructured data, which traditional non-AI tools manage inadequately. Another persistent challenge is identified by the presence of several manual procedures, like verifying signatures, performing reconciliations, and transcribing interviews, that are labor-intensive and prone to errors. Additionally, the subjective nature of professional judgment, especially in evaluating extensive data to determine evidence sufficiency, is still an issue, aggravated to the requirement of basing decision on large data volumes, which can lead to potential inconsistencies and biases. This defines the task characteristics, phrased as FA challenges, answering to SQ1: "***What are the current challenges in the financial audit process****?*".

DL, an advanced subset of traditional ML, emerged as a promising technology to address inefficiencies and issues in FAs, during the preliminary literature review of the topic. To evaluate a fit between this technology and the FA realm, technology characteristics had to be researched, hence SQ2: "***What are the main capabilities of deep learning?***". The answer can be summarized as follows. Through its hierarchical structure, DL can identify patterns in complex data and automate feature extraction. Its characteristics enable DL's success in several areas, including Natural Language Processing (NLP), computer vision, big data analytics, anomaly detection, and providing recommendations, making it a powerful and versatile technology. As a solution, DL can process structured,

semi-structured, and unstructured data types, such as documents, images, and audio data, more effectively and with higher accuracy than traditional ML. DL offers numerous automation possibilities, including streamlining the analysis and cross-checking of large volumes of documents and performing sentiment analysis. These capabilities can be combined to provide high-value solutions. Additionally, DL is a vast domain, and it encompasses various approaches, therefore an analysis of the most employed models was performed. The most commonly used models are CNNs, DCNNs, RNNs, LSTMs, AEs, LLMs, and GANs, all being more appropriate for certain applications than others. Through these capabilities, DL offers a more thorough and faster analysis of large data volumes, deriving insights that would be challenging for humans to achieve within the required time frame, as in the case of FAs. These results already provided a sound basis for a fit between this technology and the tasks identified. However, researching existing DL applications in FAs, or DL applications to problems similar to the ones found in FAs, provides a more robust result for this thesis, including the survey which served the purpose to validate such fit. For this reason, SQ3 was formulated as "***How can deep learning techniques be applied to address the challenges identified in the financial audit process?***". DL techniques can address issues connected to the three classes of FA challenges. More specifically, it can alleviate control testing and validation procedures, substantive procedures and tests of details, document analysis, and data reconciliation. Additionally, it can streamline pervasive FA procedures, such as identifying risks of material misstatements, planning an audit responsive to risks, and automating factual procedures required to assess regulatory compliance and reporting. DL also supports auditors' judgment by providing data-driven recommendations regarding the selection of audit procedures or highlighting critical information about the client for understanding the business. For the latter use case, models such as AEs and LLMs are particularly useful for these tasks. A detailed list of applications can be found in Table 6 at the end of section 4.4, The table is a comprehensive list of classes of challenges, challenging areas, specific examples, DL capabilities leveraged, and suggested DL models.

After a thorough and comprehensive research, the main research question, stated as "***What financial audit challenges can be alleviated with deep learning applications***?", is finally answered. Based on the findings from this study, it is evident that DL can significantly alleviate several challenges inherent to the FA process. These challenges are primarily categorized into three broad types: large data volume and big data analytics,

manual procedures, and subjectivity in professional judgment. These types of challenges, individually or combined, emerge in several FA procedures, including control testing and validation; substantive procedures and tests of details; document analysis and data reconciliation; risk identification, assessment, and audit planning; and regulatory compliance and reporting. Each of these procedures is exemplified with more specific use cases, that each require a different set of DL capabilities and models to be addressed with. The research provides the theoretical proof to affirm that DL's capabilities in NLP, computer vision, anomaly detection, recommendation systems, and big data analytics can address these issues. Finally, by interpreting the results, it can be inferred that adopting DL solutions can significantly enhance the accuracy and efficiency of FA processes.

6.1 Future research

This thesis provides a solid starting point for future research. The application of TTF theory has been partial, as explained in the methodology; due to time constraints, it was not feasible to test the performance benefits of each DL application. Future research could address this gap by empirically evaluating the benefits of these applications in the FA process, for instance through experiments or case studies. Given the breadth of use cases identified, it is impractical to test all of them within a single study. Additionally, testing DL applications on one use case at a time allows for the possibly of building different models and test them to validate the most promising ones with more detail, as performed by several of the studies considered for the literature background. For instance, a CNN model and an RNN model could be built to automate the reconciliation process, by testing their accuracy and draw a conclusion on the most appropriate one. Prioritizing the most pressing challenges, such as reconciliation procedures, is advisable. These procedures not only represent significant challenges, but are also underexplored in existing literature, making them prime candidates for initial empirical testing.

Finally, research could address the data scarcity challenge, presented in the additional findings. As the fit of DL applications in the FA process is proven, addressing the challenge limiting such application would streamline the adoption of DL solutions. Specifically, future research could be addressed to examine the current legal constraints and work to propose updates to regulatory frameworks, to better accommodate AI applications in auditing.

REFERENCES

Alfarghaly, O., Khaled, R., Elkorany, A., Helal, M., & Fahmy, A. (2021). Automated radiology report generation using conditioned transformers. *Informatics in Medicine Unlocked*, *24*, 100557.

Almufadda, G., & Almezeini, N. (2021b). Artificial Intelligence Applications in the Auditing Profession: A Literature Review. *Journal Of Emerging Technologies in Accounting*, *19*(2), 29–42. https://doi.org/10.2308/jeta-2020-083

Appelbaum, D., Kogan, A., & Vasarhelyi, M. A. (2017). Big Data and Analytics in the Modern Audit Engagement: Research Needs. *Auditing-a Journal Of Practice & Theory*, *36*(4), 1–27. https://doi.org/10.2308/ajpt-51684

Bans-Akutey, A., & Tiimub, B. M. (2021). Triangulation in research. *Academia Letters*, *2*, 1-6.

Balios, D., Kotsilaras, P., Eriotis, N., & Vasiliou, D. (2020). Big data, data analytics and external auditing. *Journal of Modern Accounting and Auditing*, *16*(5), 211-219.

Berente, N., Gu, B., Jan, R. and Radhika, S. (2021). MIS Quarterly Vol. 45 No. 3 pp. 1433-1450 / September 2021 DOI: 10.25300/MISQ/2021/16274

Bobbit, Z. (2021). What is high dimensional data? *Statology.com* https://www.statology.org/high-dimensional-data

Cao, Y., Li, H., Luo, P., & Yao, J. (2018, April). Towards automatic numerical cross-checking: Extracting formulas from text. In Proceedings of the 2018 World Wide Web Conference (pp. 1795-1804).

Chauhan, N. K., & Singh, K. (2018, September). A review on conventional machine learning vs deep learning. In *2018 International conference on computing, power and communication technologies (GUCON)* (pp. 347-352). IEEE.

Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2019). A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. Archives Of Computational Methods in Engineering, 27(4), 1071–1092. https://doi.org/10.1007/s11831-019-09344-w

Da'u, A., & Salim, N. (2020). Recommendation system based on deep learning methods: a systematic review and new directions. *Artificial Intelligence Review*, *53*(4), 2709-2748.

Deng, L., & Liu, Y. (2018). Deep learning in natural language processing. Springer.

Dignum, V. (2019). Responsible Artificial Intelligence: How to Develop and Use AI Responsibly. *Genetic Programming And Evolvable Machines*, 22(1), 137–139. https://doi.org/10.1007/s10710-020-09394-1

Ding, R. (2021). Enterprise Intelligent Audit Model by Using Deep Learning Approach. *Computational Economics*, *59*(4), 1335–1354. https://doi.org/10.1007/s10614-021-10192-9

Dzuranin, A. C., & Mălăescu, I. (2015). The Current State and Future Direction of IT Audit: Challenges and Opportunities. *Journal Of Information Systems*, *30*(1), 7–20. https://doi.org/10.2308/isys-51315

EY, (2023) EY's commitment to ethical and responsible AI principles | EY - Global

EY GAM

EY (2015). How big data analytics are transforming the audit. ey.com.

Fergus, P., & Chalmers, C. (2022). Applied Deep Learning: Tools, Techniques, and Implementation. *Springer Nature*.

Fotoh, L. E., & Lorentzon, J. I. (2021). The Impact of Digitalization on Future Audits: Journal of Emerging Technologies in Accounting. *Journal of Emerging Technologies in Accounting*, 18(2), 77–97. https://doi.org/10.2308/JETA-2020-063

Fotoh, L. E., & Lorentzon, J. I. (2023). Audit Digitalization and Its Consequences on the Audit Expectation Gap: A Critical Perspective. *Accounting Horizons*, 37(1), 43–69. https://doi.org/10.2308/HORIZONS-2021-027

Furneaux, B. (2012). Task-Technology Fit Theory: A Survey and Synopsis of the Literature. In Y. K. Dwivedi, M. R. Wade, & S. L. Schneberger (Eds.), Information Systems Theory (Vol. 28, pp. 87–106). Springer New York. https://doi.org/10.1007/978-1-4419-6108-2_5

Gartner. IT Glossary. Information Technology (IT) Glossary - Essential Information Technology (IT) Terms & Definitions | Gartner

Goodfellow, I., Y. Bengio, and A. Couryille. (2016). Deep Learning. Available at: http://www.deeplearningbook.orgKokina, 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

Goodhue, D. L., & Thompson, R. L. (1995). Task-Technology Fit and Individual Performance. *MIS Quarterly*, *19*(2), 213–236. https://doi.org/10.2307/249689

Grennan, L., Kremer, A., Singla, A., & Zipparo, P. (2022). Why businesses need explainable AI—and how to deliver it. *QuantumBlack AI by McKinsey*. Retrieved from https://www.mckinsey.com/capabilities/quantumblack/our-insights/why-businesses-need-explainable-ai-and-how-to-deliver-it

Hattingh, M., Matthee, M., Smuts, H., Pappas, I., Dwivedi, Y. K., & Mäntymäki, M. (Eds.). (2020). *Responsible Design, Implementation and Use of Information and Communication Technology: 19th IFIP WG 6.11 Conference on e-Business, e-Services, and e-Society, I3E 2020, Skukuza, South Africa, April 6–8, 2020, Proceedings, Part I* (Vol. 12066). Springer International Publishing. https://doi.org/10.1007/978-3-030-44999-5

He, Y., Prabhavalkar, R., Rao, K., Li, W., Bakhtin, A., & McGraw, I. (2017). Streaming small-footprint keyword spotting using sequence-to-sequence models. In *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)* (pp. 474-481). IEEE.

IBM, (2024). What is Dimensionality Reduction? | IBM

IAASB (2021). Handbook of International Quality Control, Auditing, Review, Other Assurance, and Related Services Pronouncements. Volume I. www.iaasb.org. https://www.iaasb.org/publications/2021-handbook-international-quality-control-auditing-review-other-assurance-and-related-services

ISACA (2019). Trends, challenges and strategies for effective audit in a rapidly changing landscape. Trends, Challenges and Strategies for Effective Audit in a Rapidly Changing Landscape (isaca.org)

Jan, C.-L. (2021). Detection of Financial Statement Fraud Using Deep Learning for Sustainable Development of Capital Markets under Information Asymmetry. Sustainability, 13(17), 9879. https://doi.org/10.3390/su13179879

Jan, C. (2021a). Using Deep Learning Algorithms for CPAs' Going Concern Prediction. Information, 12(2), 73. https://doi.org/10.3390/info12020073

Jan, B., Farman, H., Khan, M., Imran, M., Islam, I. U., Ahmad, A., Ali, S., & Jeon, G. (2019). Deep learning in big data Analytics: A comparative study. *Computers & Electrical Engineering*, 75, 275–287. https://doi.org/10.1016/j.compeleceng.2017.12.009

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Knechel, W. R., & Salterio, S. E. (2016). Auditing: Assurance and Risk. Taylor & Francis.

KPMG (2021). KPMG Audit Technology Evolution content series: Machine Learning. the-rise-of-the-machines-machine-learning-and-the-audit.pdf (kpmg.com)

Lauriola, I., Lavelli, A., & Aiolli, F. (2022). An introduction to deep learning in natural language processing: Models, techniques, and tools. *Neurocomputing*, 470, 443-456.

LeCun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Li, Y., & Juma'h, A. H. (2022). The Effect of Technological and Task Considerations on Auditors' Acceptance of Blockchain Technology: Journal of Information Systems. *Journal of Information Systems*, *36*(3), 129–151. https://doi.org/10.2308/ISYS-2020-022

Lin, T.-C., & Huang, C.-C. (2008). Understanding knowledge management system usage antecedents: An integration of social cognitive theory and task technology fit. *Information Management*, 45(6), 410–417

Ma, C., Zhang, W. E., Guo, M., Wang, H., & Sheng, Q. Z. (2022). Multi-document summarization via deep learning techniques: A survey. ACM Computing Surveys, 55(5), 1-37.

Magaldi, D., & Berler, M. (2020). Semi-structured Interviews. In V. Zeigler-Hill & T. K. Shackelford (Eds.), Encyclopedia of Personality and Individual Differences (pp. 4825–4830). *Springer International Publishing*. https://doi.org/10.1007/978-3-319-24612-3_857

Mathew, A., Amudha, P., & Sivakumari, S. (2021). Deep learning techniques: an overview. Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020, 599-608.

McKinsey (2020), How COVID-19 has pushed companies over the technology tipping point—and transformed business forever.

Mehrish, A., Majumder, N., Bharadwaj, R., Mihalcea, R., & Poria, S. (2023). A review of deep learning techniques for speech processing. *Information Fusion*, 101869.

Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. A. (2020b). The Ethical Implications of Using Artificial Intelligence in Auditing. *Journal Of Business Ethics*, *167*(2), 209–234. https://doi.org/10.1007/s10551-019-04407-1

Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of big data*, *2*, 1-21.

Nan, Z., Guo, X., Wang, F., Chen, G., & Wei, Q. (2011). Task-Technology Fit in Mobile Work: Exploring the Links between Task Attributes and Technology Characteristics. In Proceedings—2011 10th International Conference on Mobile Business, ICMB 2011 (p. 274). https://doi.org/10.1109/ICMB.2011.47

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., ... & Mian, A. (2023). A comprehensive overview of large language models. *arXiv preprint arXiv:2307.06435*.

Noble, H., & Heale, R. (2019). Triangulation in research, with examples. *Evidence-based nursing*, *22*(3), 67-68.

PwC (2017). Understanding a financial statement audit. www.pwc.com. Understanding a financial statement audit (pwc.com)

Saunders, C. H., Sierpe, A., von Plessen, C., Kennedy, A. M., Leviton, L. C., Bernstein, S. L., ... & Leyenaar, J. K. (2023). Practical thematic analysis: a guide for multidisciplinary health services research teams engaging in qualitative analysis. *bmj*, 381.

Seidenstein, T., Marten, K., Donaldson, G., Föhr, T. L., Reichelt, V., & Jakoby, L. B. (2024). Innovation in Audit and Assurance: A Global Study of Disruptive Technologies. Journal Of Emerging Technologies in Accounting, 1–18. https://doi.org/10.2308/jeta-2022-02

Sekar, M. (2022). Machine Learning for Auditors: Automating Fraud Investigations Through Artificial Intelligence. Apress. https://doi.org/10.1007/978-1-4842-8051-5

Sifa, R., Ladi, A., Pielka, M., Ramamurthy, R., Hillebrand, L., Kirsch, B., ... & Loitz, R. (2019, September). Towards automated auditing with machine learning. In *Proceedings of the ACM Symposium on Document Engineering 2019* (pp. 1-4).

Sharifani, K., & Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. World Information Technology and Engineering Journal, 10(07), 3897-3904.

Shiri, F. M., Perumal, T., Mustapha, N., & Mohamed, R. (2023). A comprehensive overview and comparative analysis on deep learning models: CNN, RNN, LSTM, GRU. *arXiv preprint arXiv:2305.17473*.

Sun, T. (2019). Applying Deep Learning to Audit Procedures: An Illustrative Framework. *Accounting Horizons*, *33*(3), 89–109. https://doi.org/10.2308/acch-52455

Sun, T., & Vasarhelyi, M. A. (2018). Embracing Textual Data Analytics in Auditing with Deep Learning. *The International Journal Of Digital Accounting Research*, 49–67. https://doi.org/10.4192/1577-8517-v18_3

Sun, T., & Vasarhelyi, M. A. (2017). Deep Learning and the Future of Auditing: How an Evolving Technology Could Transform Analysis and Improve Judgment. *CPA Journal*, *87*(6).

Szeliski, R. (2022). Computer vision: algorithms and applications. Springer Nature.

Tang, J., & Karim, K. E. (2018). Financial fraud detection and big data analytics – implications on auditors' use of fraud brainstorming session. *Managerial Auditing Journal*, *34*(3), 324–337. https://doi.org/10.1108/MAJ-01-2018-1767

van den O ord, A., Dieleman, S., and Schrauwen, B. (2013). Deep content-based music recommendation. In NIPS'2013 . 427

Von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, *34*(4), 1607-1622.

Werner, M., Wiese, M., & Maas, A. (2021). Embedding process mining into financial statement audits. *International Journal Of Accounting Information Systems*, *41*, 100514. https://doi.org/10.1016/j.accinf.2021.100514

Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, *53*(6), 4335-4385.

Yashudas, A., Gupta, D., Prashant, G. C., Dua, A., AlQahtani, D., & Reddy, A. S. K. (2024). DEEP-CARDIO: Recommendation System for Cardiovascular Disease Prediction using IOT Network. *IEEE Sensors Journal*.

APPENDICES

Appendix 1. The financial audit process at EY

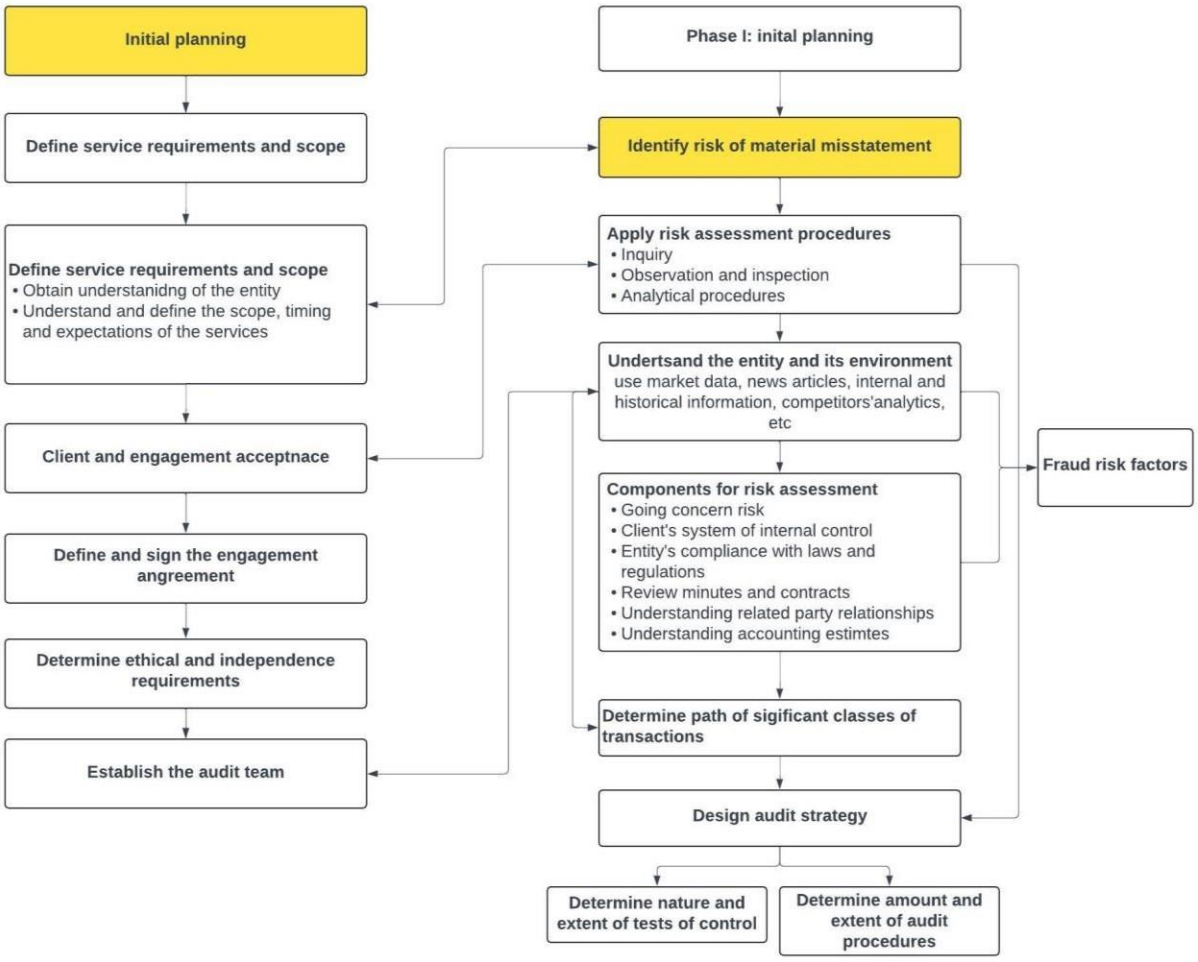

Figure 10 Phases I and II of the FA process at EY (EY GAM)

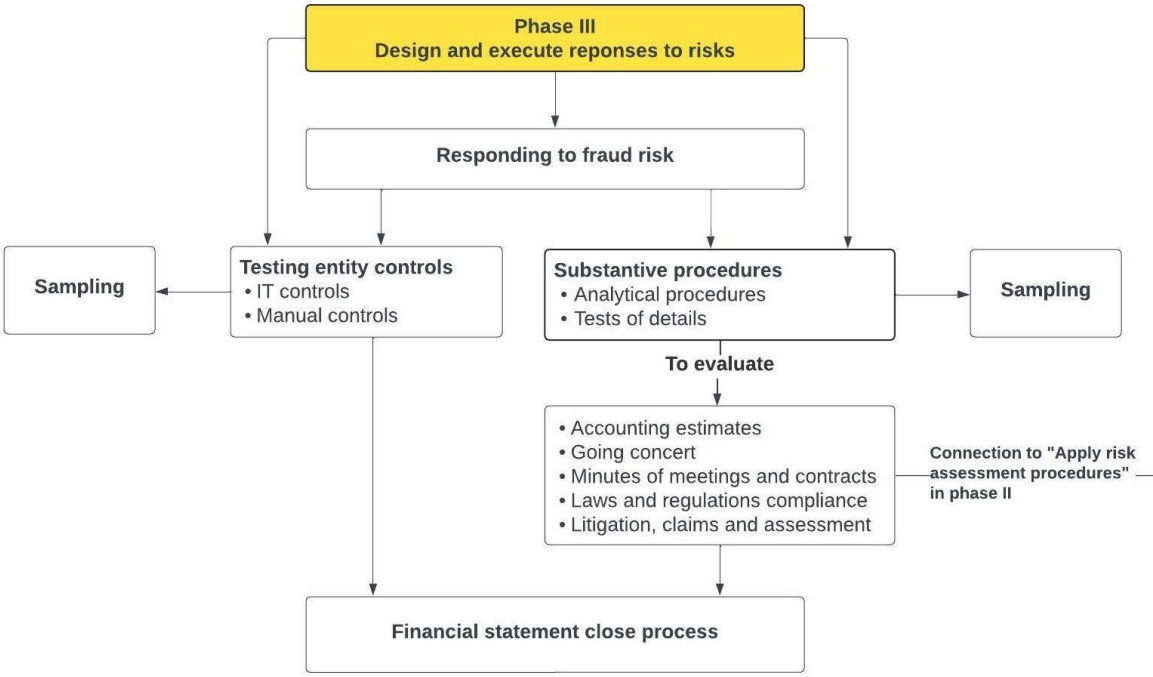

Figure 11 Phase 3 of the FA process at EY (EY GAM)

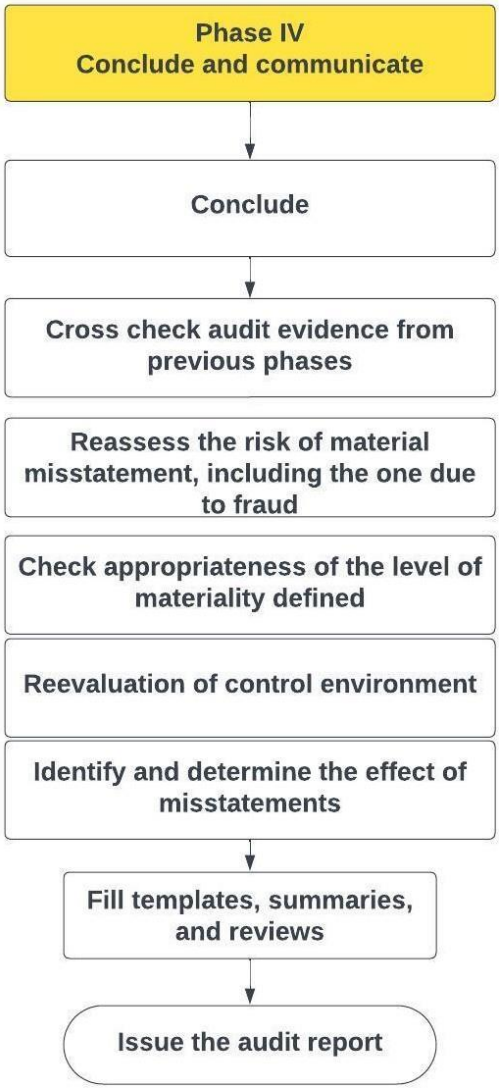

Figure 12 Phase IV of the FA process at EY (EY GAM)

Appendix 2 – Result conceptual framework with use cases

Challenge category	Example	Challenge type	DL's application	DL model(s)
Control testing and validation	Verification of signatures made by authorized personnel	Manual procedure, large data volume	NLP, computer vision	CNNs, DCNNs, RNNs
	Verifying legitimacy of signatures and documents	Manual procedure	NLP, computer vision, anomaly detection	CNNs, RNNs, AEs
	Oral inquiries	Manual procedure	NLP	RNNs, LSTMs
Tests of details	Tests of details	Manual procedure, large data volume	NLP	CNNs, DCNNs, RNNs, LSTMS
	Verifying representativeness of samples	Large data volume	Anomaly detection	AEs
Document analysis and data reconciliation	Reconciliation procedures	Manual procedure, large data volume	NLP, computer vision, anomaly detection	CNNs, RNNs
Risk identification, assessment, and audit planning procedures	Audit planning and audit procedures selection	Manual procedure, large data volume, professional judgment	NLP, recommendation systems, big data analytics	CNNs, AEs
	Understanding the business	Manual procedure, large data volume, professional judgment	NLP, recommendation systems, big data analytics	AEs, LLMs
	Evaluation of competitors and client's historical analytics	Manual procedure, large data volume, professional judgment	NLP, anomaly detection, recommendation system	CNNs, LSTMs, AEs, LLMs, GANs
	Evaluation of going concern prediction	Manual procedure, large data volume, professional judgment	NLP	RNNs, LSTMs
	Scanning minutes of meeting	Manual procedure	NLP	RNNs, LSTMs, LLMs
	Fraud risk identification	Large data volume	Anomaly detection	CNNs, RNNs, LSTMs, AEs
	Evaluation of accounting estimates	Large data volume	NLP, anomaly detection	CNNs, RNNs, LSTMs, AEs
Regulatory compliance and reporting	Filling standardized forms	Manual procedure	NLP	AEs, LLMs, GANs
	Verification of compliance through checklists	Manual procedure	NLP, computer vision	RNNs, LSTMs, LLMs
	Drafting reports	Manual procedure, large data volume	NLP, computer vision	CNNs, LSTMs

Table 7. Results conceptual framework with use cases

Appendix 3 – AI use

In line with this thesis results, AI has been used as a supporting tool. As the results of this research show, it is crucial to always review AI outputs. Additionally, results show AI's fit for providing recommendations, even if in the realm of FA. Therefore, any AI use has been reviewed and considered as a suggestion, meaning that no text has been copied and pasted from AI tools. The use of AI is clarified below.

During the literature review, some highly technical studies have been considered. Due to the difficulty of terms and complexity topic, that also went beyond my previous knowledge, ChatGPT has been used to explain difficult sections. Additionally, sometimes it was not clear in the existing literature when traditional ML or DL were used, as the terms are usually used interchangeably, and the architecture of the models studied are not always disclosed. Questions asked were: “Can you explain this research section in simple terms?”, and “Can you clarify whether this part is describing a DL approach or a traditional ML one?”. This approach helped me with the selection of the most appropriate studies for my literature review.

When encountering difficulties in finding specific literature, Copilot has been used to locate studies related to what I was looking for. However, this approach has been used limitedly, as the results did not provide positive outcomes. Indeed, the studies linked were either ones that I had already found, were too old, or were not meeting the characteristics of deeming them a reliable source.

Two pie charts are included in the results, in section 4.3. GPT-4 has been used to create those charts. Providing the conversational AI with the characteristics needed, the goal I wanted to reach, and the specific numbers to represent, I asked it to create a Python code to address the layout. Modifications to the code regarding the colors of the charts as well as its layout have been subsequently made. In a similar, but less pervasive manner, I used GPT-4 to obtain opinions regarding how to represent the conceptual framework in the results, still in section 4.3, as it has been challenging to find a way to comprehensively and effectively visually present them. Various ideas have been provided, and in the end they only served the purpose to inspire me, by using more colors, to make my own illustration.

An AI tool, named AssemblyAI, has been used to automatically transcribe the interviews. Care has been taken in order to review all the transcripts to check for correctness of the outcome, correct mistakes, as well as modify the layout.

Finally, ChatGPT has been used in all the chapters of the thesis to improve my writing. As English is not my primary language, sometimes I needed to fill some vocabulary gaps. Questions such as “What are synonyms of *word*?” were asked. Otherwise, once I wrote a part of a text, I would feed it to the AI and ask questions such as “Check the spelling, English grammar, and write it in academic style”. Then, I would copy the result in a separate document and correct my original text based on the outcome of ChatGPT.

Appendix 4

Interview questions in round 1

Round	1
Aim of interviews	Exploratory
Knowledge of experts	Financial audits
Interview	1A, 2B, 3C
Questions	1. Based on your experience, what are the issues in the financial statement audit process?
	2. What specific difficulties do you encounter with large volumes of data?
	3. Can you describe some of the key challenges related to large volumes of data, especially when it comes in different formats?
	4. In what form is audit evidence typically received or obtained (e.g., text, emails, pictures, videos, social media content)?
	5. What is your opinion on sample testing and the potential to switch to population testing?
	6. Can you list the major manual procedures that are very time-consuming and inefficient?
	7. Do you think there are inconsistencies in how different auditors assess the sufficiency of evidence?
	8. Have you encountered any difficulties in detecting financial statement fraud using traditional methods?
	9. What type of methods do you use to detect financial statement fraud?
	10. Based on my introduction on deep learning and your previous knowledge, what are your thoughts on integrating deep learning technologies into the audit process?
	11. Where do you see potential applications of deep learning to alleviate challenges related to large data volumes or manual procedures?
Explanation	These semi-structured questions ensure that key topics are covered while allowing flexibility to explore individual insights and experiences. These interviews have the aim to explore the FA challenges from FA experts, while gathering technical linformation on the challenges, such as the type of data involved, to provide such information to data science experts.

Round	1
Aim of interviews	Exploratory
Knowledge of experts	Data science and AI
Interviews	4D, 6F
Questions	1. How much do you know about financial audit, and have you worked with deep learning?
	4. Financial audit evidence and data comes in unstructured forms such as images, graphs, and different formats. Do you think DL could be used to address this, and can DL help with data extraction from unstructured data as well as filtering and review?
	5. Financial auditing entails sampling. What is your opinion on automating the testing process of the transactions in the sample, and do you think manual work could be alleviated through the application of deep learning?
	6. In order to apply AI in financial audit, there needs to be assurance that the AI performs consistently, without bias, and is explainable. Is there a DL algorithm or model that could address this?
	7. What are the technicalities and specifics that need to be known in order to understand whether deep learning can be applied in certain use cases in financial audit?
	8. Regarding the possibility to extract more data from online sources, do you think deep learning could help the extraction and processing of such digital data, as well as assessing the reliability of the sources of this data?
	9. Reconciliation procedures require manual work and are time-consuming. Would you think there's room for DL solutions here?
	10. What are the most commonly used DL models?
Explanation	These questions aim to explore the interviewee's familiarity with financial audit and deep learning, assess the feasibility and challenges of implementing deep learning in financial audits, and understand the technical requirements and potential benefits of automating various audit processes using deep learning techniques.

Interview questions in round 2

Round	2
Aim of interviews	Deeper understanding of FA challenging procedures
Knowledge of experts	Financial audits
Interview	7G, 8H, 9I
Questions	1. Do you agree that the available and potential large data volume is one of the challenges in the financial audit process?
	8. Would you agree that the amount of routine and manual procedures is a major challenge in the financial audit process as of now? Can you give examples?
	9. Can you briefly explain how the reconciliation process works? What types of formats do you encounter, and is this process manual?
	10. Could you explain how the process of checking if the IFRS standards are actually met works?
	12. Who ultimately decides the amount of documentation to use? Are there any rules that can help with this decision?
	15. Do you see possible applications of deep learning for text understanding in financial audits?
	16. Would you see possible applications of visual recognition connected to deep learning in the financial audit process?
	18. Would you see an area of application in financial audits for judgment support using deep learning?
	19. Can you give an example of where there is available data, but auditors are unable to go through all of it?
	21. Can you describe a manual and repetitive procedure you encounter in audits? Is the way you do it right now entirely manual?
	22. What can you say about fraud detection, for example, trying to understand whether some fraud has been perpetrated?
	26. How is the process of understanding and analyzing huge data sets currently addressed?
	28. If population testing is applied and a big number of outliers are discovered, would choosing samples from those outliers increase audit quality?
Explanation	The aim of these interviews is to gather more specific knowledge regarding the single FA challenges that were mentioned by FA experts in round 1, in order to clarify the challenges to data scientists, subsequently. The questions build on previous inquiries by delving deeper into specific technicalities, practical examples, and expert opinions. Also, they were formulated to corroborate the specific examples of FA challenges derived from previous interviews. Finally, they were designed to provide data scientists the required information for them to provide opinions on DL applications to FA process.

Round	2
Aim of interviews	Deeper understanding
Knowledge of experts	Data science
Interview	11F
Questions	1. What do you think about using deep learning to help with the reconciliation process?
	4. Checking the reliability of data, how could deep learning be applied here?
	5. The amount of regulations in the financial audit realm often transforms into checklists. Do you think deep learning could help with checking compliance with these regulations?
	6. In the first phase of an audit, the auditor needs to collect as much knowledge of the client as possible. Do you think deep learning could help with automatic extraction and relevance finding of this data?
	7. There are many procedures that are heavily based on the so-called professional judgment of auditors. What do you think of deep learning's role in judgment support?
	8. What are some or one challenge that you could see in applying deep learning in financial audit, apart from the black box thing and the fact that sometimes it seems an over-engineered solution?
	9. You mentioned the need for trust in the machine's decisions. Is that different from the machine being a black box?
Explanation	Providinng more details to data science experts gained during the previous interviews, to gater a sounder opinion regarding the fit of application of DL inthe FA process.

Round	2
Aim of interviews	Deeper understanding
Knowledge of experts	FA and AI/DL, or FA and knowledge of AI solutions for Fas
Interview	10E, 12I, 13L
Questions	Would you agree that large data volume is a major challenge in financial audit? Why or why not?
	2. Would you agree that the number of manual procedures is a major challenge in the financial audit process? Can you list the major manual procedures that are very time-consuming and inefficient?
	3. Do you think that many decisions are left to the auditor's professional judgment and that these could lead to inconsistencies? Would you see AI or deep learning being able to give some recommendations or some basis on which to base a professional judgment?
	4. Can you think of any AI, or specifically, deep learning applications right now in the audit procedure?
	5.. How would you see deep learning being applied to reconciliations to be helpful? Do you think it could help with the reconciliation process without considering the black box, explainability, and current regulation constraints?
	6. Among the challenges in the financial audit process is the sampling procedure that seems to present high risks of overlooked transactions. Would you agree that this is a challenge as of now? Would you see room for deep learning applications to draw more representative samples out of the entire population?
	7. What is your opinion of including other external sources, like market data, in the financial audit process?
	8. Would you see an application of deep learning in order to quickly analyze the recordings of interviews?
	9. What are the most important challenges in applying deep learning in financial audit?
	10. What are the main deep learning architectures used for the majority of financial audit applications?
Explanation	These questions aim to understand how DL/AI can address issues such as handling large data volumes, automating manual procedures, and supporting professional judgment, while identifying current applications, benefits, and obstacles in implementing these technologies.

Appendix 5. Survey

The instructions are presented first, then the actual layout and content of the survey is illustrated subsequently.

Instructions

Thesis topic: researching how deep learning (DL) can address the challenges arising in financial audit (FA) processes.

Purpose: to gather your opinion on the connections I have made between DL capabilities and FA challenges. Your feedback will help validate these connections.

You will find the survey in sheet 2. Below are two questions for you to fill, the instructions of the survey, the deadline, and some remarks.

Q1: What is your position?_____

Q2: What are your experiences (area of knolwedge/expertise)?_____

Instructions

1. Only fill the green columns.

2. Columns E and G contain drop-down menus with Likert scale options: strongly agree, agree, disagree, or strongly disagree.

3. Column D lists DL capabilities that may address the FA challenges in column C. Rate your agreement on this connection.

4. Column F lists DL architectures linked to the capabilities in column D. Rate your agreement on these architectures' appropriateness.

5. If you disagree or strongly disagree, please explain your reasoning.

6. Optional: Provide additional suggestions or explanations if you have any input beyond the provided connections.

7. Yellow cells are for suggesting DL capabilities and architectures for challenges where I was unsure to provide an answer. You can fill these with your input. In that case, you will not need to answer to the Likert scale and the other optional questions.

8. Leave blank if you cannot answer.

9. There is no need to be very specific; only general capabilities and the main classes of DL are sufficient for my research purposes.

Remarks: there is no required form, length, or any other requirements for answering the open questions.

By returning this survey, you consent to your responses being used for my thesis. Your name will remain confidential; only your position and experiences will be disclosed. Thank you, and good luck with the survey! Feel free to contact me with any question.

ID	Challenge type	Challenges
1	Manual procedures + large data volume	Control testing - checking the documents are signed from who is actually authorized to do so
2	Manual procedures + large data volume	Test of details - cross-checking invoices and making sure they are correclty written down in the bookkeeping
3	Manual procedures + large data volume	Reconciliation of different formats documents (checking mathematical accuracy among the documents; checking the presence of information in both document sources)
4	Manual procedures + large data volume	Investigation of signatures, forgeries
5	Manual procedures	Scanning minutes of the board and cross-checking them to understand consistency or differences
6	Manual procedures	Analyzing documents to cross-check information among each other (Structured, semi-structured and unstructured, different layouts)
8	Manual procedures	Inquiries (oral and written)
9	Manual procedures	Financial statement close procedure (verifying consistency of data and mathematical accuracy in the whole document, among numerical and flow text parts)
10	Manual procedures	Automation of checklist-based regulation and framework compliance of financial statements (comparing the checklist requirements with the financial statements content)
11	Manual procedures	Filling standardized word forms
12	Manual procedures	Drafting reports based on data from documentation in the audit platform
13	Large data volume	Identify anomalies and trends on large data sets
14	Large data volume	Extraction and analysis of large volumes of unstructured data (news articles, social media messages, emails, market reports)
15	Large data volume	Comparing analytics (from internal data and competitors)
16	Sampling	Selection of representative samples from large datasets
17	Professional judgment	Determining the sufficiency of evidence based on historical data and similar engagements
18	Professional judgment	Planning: determining audit procedures connected to the identified risks based on historical data and external industry, market and news data
19	Professional judgment	Construct going concern prediction models
20	Professional judgment	Accounting estimates: create expectation range, also using information from the market, competitors, past data from the company, etc
21	Professional judgment	Fraud detection

Figure 13. Survey excerpt, part I

DL capabilities that can address the challenges in column C	Likert scale. Please select one of the opions from the drop down menu of each cell in this column
Text categorization, image recognition, image classification	
Text categorization, image recognition, image classification, extracting formulas from flow text, pattern recognition	
Anomaly detection, pattern recognition, athomated calculations, cross-verifications, document classification, report generation and vizualization	
Writer identification, image analysis, feature extraction, anomaly detection, pattern recognition	
Information retrieval, text categorization, text classification, text summarization, sentiment analysis, topic extraction, mullti-document summarization	
Information retrieval, image classification, image to text, text categorization, text classification, text summarization, data integration, feature extraction, data representation and vizualization, multi-document summarization	
Voice recognition, acoustic to text, parts of speech tagging, sentiment analysis , text classification	
Image to text, formula extraction from flow text, consistency checking, numerical cross-checking	
Word prediction, NLP, Text generation	
Report generation from images	
Anomaly detection, trend detection	
Data extraction (webscraping algorithms), information retrieval, feature extraction, dimensionality reduction, text summarization, text classification	
Pattern recognition, anomaly detection, named entity recognition, semantic matching, time series analysis, vizualization, numerical cross-checking	
Pattern recognition, anomaly detection	
Recommendation system, pattern recognition, multi-document summarization	
Recommendation system, pattern recognition, multi-document summarization	
Time-series analysis, time-series forecasting, feature extraction	
Feature extraction, pattern recognition, anomaly detection, time-series forecasting	
Anomaly detection	

Figure 14. Survey excerpt, part II

ctures	Likert scale. Please select one of the opions from the drop down menu of each cell in this column
CNNs, RNNs, DCNNs	
CNNs, RNNs, DCNNs, LSTMs	
CNNs, RNNs	
CNNs, RNNs, Autoencoders	
LLMs, RNNs, LSTMs	
DNNs, LSTMs	
CNNs, LSTMs	
LSTMs, Autoencoders, RNNs, CNNs	
AEs, LLMs	
Autoencoders	
DNNS, RNNs	

Figure 15. Survey excerpt, part III

Appendix 6. Interview themes and edited transcripts

6.2 Interview themes

Theme	Interviews	Description
Large Data Volume - FA Challenge	1A, 2B, 3C, 8A, 9H	Challenges related to handling large volumes of data in financial audits, including issues with data reliability, different data presentations, and difficulties in effectively using the data. Examples include millions of transactions per day, issues with processing structured and semi-structured data, and the need to enhance audit procedures to better exploit large data volumes.
Large Data Volume - DL Solutions	4D, 5E, 10E, 12I	The ability of deep learning to manage and analyze large volumes of structured, semi-structured, and unstructured data effectively. This includes DL's ability to process vast data volumes, dimensionality reduction capabilities, and handling diverse data formats.
Manual Procedures - FA Challenge	1A, 2B, 3C, 7G, 8A, 9H	Manual and time-consuming procedures in financial audits, such as reconciliations, checking validity of signatures, and performing sampling. Issues include inefficiencies, the risk of errors, and the need for automation.
Automation of Manual Procedures - DL	4D, 5E, 6F, 10E, 11F, 12I, 13L	How deep learning can automate various manual procedures in financial audits, enhancing efficiency and accuracy. Examples include automating reconciliations, verifying signatures, and improving the sample selection process.
Subjectivity in Professional Judgment - FA	1A, 2B, 7G, 8A, 9H	Challenges associated with the subjective nature of professional judgment in financial audits, leading to inconsistencies and potential biases. This includes issues in evaluating risk, detecting fraud, and ensuring consistent application of audit procedures.
Judgment Support Role of DL	4D, 5E, 6F, 10E, 12I, 13L	The potential of deep learning to support professional judgment in financial audits by providing data-driven recommendations and insights. Examples include assisting in risk assessment, fraud detection, and evaluating accounting estimates.
DL Adoption Challenges	4D, 5E, 6F, 10E, 11F, 12I, 13L	Challenges related to adopting deep learning in financial audits, such as the need for high computational resources, data scarcity, regulatory compliance, and client hesitation to share data.
Use Cases of DL	4D, 5E, 6F, 10E, 11F, 12I	Specific applications of deep learning in financial audits, including anomaly detection, risk assessment, regulatory compliance verification, and generating audit reports.
The Role of AI and auditors	4D, 5E, 6F, 10E, 12I, 13L	The complementary roles of AI and human auditors, emphasizing that AI can enhance but not replace human judgment. AI can perform tedious and repetitive tasks, allowing auditors to focus on higher-level decision-making.
DL's Technical Characteristics	4D, 5E, 6F, 10E	Technical aspects of deep learning relevant to financial audits, such as explainability, dimensionality reduction, and the capabilities of different DL models (e.g., CNNs, RNNs, LSTMs, GANs).

6.2 Interview 1A

Senior financial auditor **Round 1** **Semi-structured interview**	**Labels/interpretation**
Q: Based on your experience, what are the issues in the financial statement audit process? We do too much manually, which can also be done be either by a computer, I think that we should use more techniques to make our work more doable, to also decrease the number of hours that we work. For example, I make use of the [tool], I use it for the financial statement, because with it you can just upload the financial statements in an excel file and then [tool] you can just extract all the numbers in the report and it calculates the total. By one click, you have validated the mathematical accuracy of the whole financial statement. So that's what I use, but I don't know if it's in line with deep learning.	Manual and time-consuming procedures challenge. Non-AI tools are available to ease the auditor's life.
Q: Does this tool require to check the results, or do you trust the results? From experience, I need to check the result, I cannot fully rely on the application. From a deep learning perspective, I think you should fully rely on the model. As auditors, we have to verify whether a tool or a model result in the result we have. That's the challenge.	Auditors always need to review results.
Q: Even if you use this tool, do you think this activity is still time consuming? The tool itself is not time consuming, as it requires just a click of the mouse, then the totals are reconciled. But again, if you have to check because you cannot fully rely on the tool, it also depends on the settings of the financial statements, sometimes it doesn't pick the numbers correctly, so that will take time.	Non-AI tool employed presents issues dealing with different settings and document layout, plus inaccuracies in numbers representation
Q: Can you think of any other challenge in financial audits? I just did a group engagement where we needed to reconcile a lot of documents, which were in different formats or in different tables from the same application. This procedure should be more efficient to do with some kind of a tool, instead of doing everything manually, becuse it takes several weeks to reconcile everything. I have done it for 2 or 3 years and we have to do it twice a year. Maybe something that can be done by a model could help.	Manual and time consuming procedures: reconciliation. Automation is considered necessary.
Q: Can you describe some of the key challenges related to large volumes of data, especially when it comes in different formats? I'm currently working on pension funds and we have a lot of data from participants within the pension funds that we have to verify. There's also something where I use IT members who processes big data and transpose it and analyse it. It's big data from the client perspective.	
Q: In what form may the audit evidence be received or obtained? For example, text, emails, pictures video, social media content, etc. From emails and just standard of word, pdf, excel files that we receive. I don't know if you're familiar with [platforn] that we're using. There's not really like images or social media, that's not really the evidence that we use from an auditor perspective.	Data types involved in a FA: structured and semi-structured data

Q: Why do you think social media messages or other types of data are not use in the first phase of an audit, where evidence is collected in order to understand the business? That's true, we just search on the internet, if there's something in the news that is relevant. That is actually somehing that can be enhanced, media or social media search, if there is something which mentions the client you have.	The use of external data, such as news articles, social media messages, is considered beneficial and its use to be inhanced.
Q: What type of data do you currently use in the first phase? Just search on the internet if I can find something on the client, but I also include the website of the client.	
Q: Why do you think other types of data are not used? Is it because it cannot be processed, is it because the data you have is sufficient or are there other reasons that you can think of? On the internet you can find basically everything, it's also a matter of what the relevance of the data is, and also what is the source, where the data comes from. The data used need to come from a reliable source.	Making sure the data gathered is relevant. Interpretation: the interviewee, by stating that on the internet everything can be found, seem to address the large data available, increasing the difficulty of veryfying the relevancy of such data.
Q: When it is time to determine the sufficiency of the evidence received, what factors do you consider and what challenges do you encounter during this process, if any? Basically we have to check whether or not we have performed sufficient procedures. It is difficult to answer what the challenges are - when the evidence is enough then it is enough, based on judgement. In our profession, everything is professional judgement. It also depends on the person, sometimes someone says if you have the confirmation, then it is sufficient, but someone else could say the confirmation alone is not enough, you have to include the email, etc. And the contact you have with the client... So, it depends person per person.	Subjectivity in the excercise of professional judgment.
Q: In financial audits, there is the sample testing practice. However, literature is pushing towards population testing. What is your opinion on this? Our amount of samples is 25 or 60, and that is ifxed. If the population is 500 or 1 million, the maximum sample size remains 60. If the population is 1 million, I don't know if the sampling technique is the right method to use. Maybe there could be a method to allow the test the whole population.	Sampling strategy: risk-bearing procedure.
Q: Why do you think the sample size is fixed, regardless of the popualtions size? It's just our methodology.	

Q: Is selecting samples a manual procedure? We have tools that you have to include certain criteria from your population and it provides you the sample size that you need to use.	
Q: Have you encountered any difficulties in detecting financial statemet fraud using traditional methods? And what type of methods do you use to detect financial statement fraud? No. We use the journal entry testing that we have. It's really difficult to detect fraud. It was a one time thing for me, there was a case of fraud in one client, but it was already known at the client. It's not one of our main focus areas, we have to perform procedures, but it's not the purpose of our work. It's also relly difficult to detect, because someone who's perpetrating fraud becomes difficult to encounter. I'm in the insurance sector, so they are quite good at detecting fraud themselves, they have their own frameworks and control environments. Maybe it's the type of client, because insurance sector is highly regulated.	Detecting fraud is challenging. However, it is not the main purpose of the auditor's work.
Q: Based on my introduction on deep learning and on your understanding, what are your thoughts on integrating deep learning technologies into the audit process? I think that would be really helpful becasue there's so much information, so much data, you can retrieve evidence from picture or social media, there's a lot of information that we do not do much with, but we should. I'm not really familiar with the subject (deep learning), but it's really interesting. I think there should be more ways to do an audit.	The need to enhance audit procedures is highlighted, due to large data volumes that can be exploited better.

6.3 Interview 2B

Senior financial auditor **Round 1** **Semi-structured**	**Labels/interpretation**
Q: Based on your experience, what are the challenges in financial statement audit process? It's very client and sector specific. One of the challenges that we encounter is regulation for the financial statement. For example, IFRS regulation in the insurance sector. That's one difficulty we ecounter with the implementation of new regulations for specific sectors.	FA challenge: checking financial statement regulation compliance
Q: What does this difficulty entail? Complying with the regulation, but also the financial statements, can lead to a different set of table, set of numbers, when applying other regulations. For us it's difficult to see what is the old regulation, what is the new one and what is the correct one, because I think when an insurance company or a bank implements a new set of regulations from the governmenl, accountants and the client itself also needs need to learn what the regulation exactly is.	Different layouts and presentation of financial statements can derive depending on the regulatory framework applied.
Q: How do you currently address this issue? Mostly, what we do is having conversation with the client and discuss what they think, what we think, and together we come with a solution.	

Q: Would you say this is is a time consuming task Yes, it is time consuming because you have to address the risks. The communication between the auditor and the client is time consuming because you have to get to right person and have to get a view of the regulation, so coming to a conclusion is time consuming.	Inefficient and manual procedure: checking regulatory compliance in financial statement.
Q: Can you describe some of the key challenges that you face, particularly in terms of dealing with large volumes of data and unstructured data? First of all, it's time consuming. If we have big data, our computer cannot always handle the large data set. That's one of the issues, but not always the case. Getting a straight overview of what you actually see is the greatest challenge when you have big sets of data.	Large data volume challenge: machine difficulty in processing large volumes of data; getting overview of the data.
Q: Do you mean the visualization of data is a challenge? Yes, for example if you have and Excel with 6 thousand rows and you want to translate that in how the client transported that in the financial overview, it's time consuming and you don't alsway know if you have the right data to get to that overview.	
Q: How do you currently address this problem, especially if you ecountered it yourself? I'm encountering it right now with a client. We ask the client about how we can do it the best way. So they send us instructions about how they have done it. So we can see if a table, for exmaple, was used to make a set of tables.	
Q: Can you think of other types of time consuming tasks in your profession? The extraction and the filtering of data is the most time consuming part of the job. What's also time consuming is reviewing some tasks. If someone prepares some task, for example with large data sets, I have to review the work of someone else to see if the work done is complete, accurate, if all the data has been used and if the correct data has been used.	Time consuming procedures: extraction and filtering of data; reviewing tasks.
Q: Can you make a practical example of how you deal with it? For example, when I review the investments part of an entity and someone else has prepared the reconciliation between the investment administration and the finanical statement, it's mostly a large data set of investments with certain codes, with all kind of invetsments. So I have to see if the source of investment is mapped as it should be, if all the data is in the table it should be and if the reconciliation can be made with the financial statement. So the mapping, for the accuracy costs time.	
Q: What are the most data intensive parts of a financial statement audit process? It's a difficult question because all the financial statements are based on one set of data, so the most intensive part is when we have the data of the financial administration, that data has to be mapped into the right account numbers and I think that part is the most data intensive part of the financial audit. Mostly, we outsource it to our IT colleagues, for the [tool]. It's a [tool] which has all the data, all the bookings from the company, made it physical, and balance sheet.	Time consuming procedure. Interpreted as a reconicliation process.
Q: When you otsource to the IT colleagues, do you trust the results they give or do you have to review them? We trust the results, but we also make a reconciliation with the financial	Reviewing is intrinsic to FA. Also when using non-AI tools,

statement of the client, to see if the mapping is right, if we're missing something, if there's a difference between certain numbers. So we also make a recount to see if we have the same numbers as the ones the client used.	reviewing tasks is necessary.
Q: Is this done manually as well? Yes. On the [tool] we make a sort of copy of the sheet and then we manually reconcile the numbers.	
Q: How long can this reconcilitiation process generally take? The reconciliation is not very time consuming because you have two files, one is our own file from EY and the other one is the file of the client, and we just look if the numbers are the same. I think this would be no more than one hour work, except when there is a difference.	Reconciliation is not time-consuming.
Q: What do you do when the are differences? When there are differences, we go to our colleagues from the data team to see if the mapping is correct. Sometimes there's a difference between two lines, so you see there's one wrong mapping which means there's plus in the first line and a minus in the second line. So it's the same difference which is mapped wrong. So we go to the data team which changes it and we hope the reconicliation can be made.	Non-AI automation tools presents accuracy issues.
Q: What is your opinion on sample testing an what do you think about the possibility of switching to population testing? I think with sample testing there's a good chance that you catch get the mistake when there actually is one. So, when you have 100 invoices and you check 10, for example, you cannot see the other 90. But if you have a population testing, then there's certainty that there's no exception in your population.	Population testing enhances the risk detection and overview of data.
Q: Do you perform any kind of population testing right now? For some of our examples, we have some data sets which include non financial data which relates to one person, for example in the pension sector, like the gender, the birth date. In a pension fund it's important that the data of one person is correct in the system- birth date, gender, what kind of pension they have, are they married or not, is someone divorced. This kind of data is imported in a data tool of us and then we have the whole population that we can see differences about, for every single person, in our data set. When we test the whole population and then we only investigate some differences we did not expect, but if we have invoices like financial data, then it's mostly a sample that we investigate.	For financial data, sample testing is the norm, rather then population testing.
Q: 10. In what form may the audit evidence be received or obtained (text, emails, pictures, videos, social media content)? Mostly, the evidence we receive is pdf file, but we also have an email or a scan of an invoice and sometimes also an excel file.	Data formats: structured and semi-structured.
Q: Do you receive any video or social media content, for example? No. I've never had videos or social media content.	
Q: How about the data used in the first phase of an audit, when you need to understand the entity? I've never used	

social media content, but we do use newspapers or magazines, but mostly online.	
Q: When determining the sufficiency of evidence during an audit, what factors do you consider and what challenges do you ecnounter in this process? Everything is coming digital and we can receive an invoice or evidence in pdf. It should be signed by some director or manager or someone who's authorized to sign the paper. Sometimes the same signature is in multiple files. We have to investigate if the signature is copied from another file, and the risk that the signature is copied from someone else. Sometimes when I open a file, I can move the signature and I could copy it myself, for example.	Manual procedure: checking validity and potential forgery of signatures.
Q: Do you need to outsource this investigation procedure to another team? No, I mention it to a manager and we contact the client, so the person who should have signed it should approved that it is his/her signature.	
Q: Have you encountered any difficulties in detecting fraud in financial statements? I have never encountered it before. We have journal entry testing and all the bookings made by the client in the [non-AI tool]. We have an option to check some bookings for fraud, but I think it could be done better. So, maybe with deep learning you should be able to check the whole bookings made and the journal entries and red flag some issues that you encountered.	No issues in detecting fraud. Non-AI tools in place do not solve the issues completely.
Q: Based on my introduction on deep learning and on you previous knowledge on it, what are your thoughts on intergrating deep learning in the financial audit process? I am excited to see what options are there and where we can implement it, because I think in time financial audit could be done with machine learning, or deep learning or automation tools. I think when we have such applications, the audit can be done much more efficiently and less time consuming, so the audit quallity will go up, because you see more of the data and more of the risks which you encounter.	Including automating tools in FA procedures has the potential to increase the audit efficiency and quality, by reviewing more data.

6.4 Interview 3C

Manager in data analytics for FA **Round 1** **Semi-structured**	**Labels/interpretation**
Q: Based on your experience, what are the main challenges for financial statement auditors? I would say the amount of work they need to perform, because in financial statement audits there's a lot of regulations to follow. In EY there's the global methodology, EY GAM, wih says how the audit should be performed at EY. There are a lot of forms that need to be filled, words form: what's the scope of the audit, what do we know about the lcient, do we understand the business and also if they need to performm to procedures. The standardized forms are something that usually is delegated to GDS, the offshoree team	FA challenges: amount of work to perform, due to regulations. Examples: filling standardized forms, as well as processing Excel files.

in India, or newjoiners in the team, first or second year, which is quite time consuming sometimes. Manual procedures also in excel, that would be way more automated if everyone would apply it.	
Q: Can you make an example of manually done and time consuming works that you mentioned? We have a tool to automate based on the data, but not everyone is using it, but it still requires some manual procedures. Financial statements are built based on the data of the client, so the trial balance data, so the client basically adds which is the mapping in the financial statement, but what auditing do, they get the trial balance data also and they create working papers of that account in the fin stat that are relevant in the audit, but some of them are manually created. That would take couple of days, but it could be done in one day. Another example is testing. Testing selection, in case it's a quite straightforward documentation that they need to check, for example comparing pdf documents with numbers in excel, this is still done manually. Probably this could be done by AI. So they have to check the results. This is the case in IT audit as well. So it's more about checking and reviewing the reuslts that is timeconsuming. If parts of this that are automated anf taken over by AI this would be better.	Non-AI tools are employed for automation, but they do not solve the problem completely. Manual adjustments are still required. Sample selection for tests of details is done manually, but has potential for being automated. Time consuming procedures are reviewing tasks.
Q: What is your opinion on sample testing and the potential to switch to population testing? How it qorks in fin aud, based on the methodology of EY, you have a population and first you need to apply a threshold. Base on the threshold, all the items in the population have to be tested. All the items below that threshold, depending on the setting, such as the risk, the sample size need to be modified. Sometimes this happens on a random base. In data analytics, we can already visualize outliers, so sample items can be chosen better. Probably with AI this can be easily automated. That would definitely help to make a better sample selection, that would be very useful. Maybe with keywords they can search it up with text, they can find text, if they want to focus on journal entires made by the CFO for example of management, then they can just ask the question and then some outputs can be given, without the need to do everything manually.	With non-AI base data analytics it is easier to visualize outliers, but AI can potentially make better sample selection, for instance by searching by keyworkds in the data, or asking questions to conversational AI and receiving insight.
Q: Do you think there would be benefits regarding the quality of audits if population testing is applied, instead of sample testing? That could be, but it depends on the data. If it's a recalculation, we can do it already with data analytics on the full populations. But we need to figure out ourselves how the recalculation works, which fields to use based on interaction with the client, but maybe if there's some model that can do it itself, then that would help. In the case of external evidence needed from the client, that would be more difficult because you still rely on external evidence.	Population testing: already applied in case of recalculations through non-AI automate techniques. However, manual adjustments need to be employed.
Q: So do you think the clients would not allow the auditor to collect all the evidence? What if the collectoin of evidence is automated? If EY says we need all the evidence, like hard	Population testing risks to increase the workload of auditors

copies of invoices, then at some point they need to give it to us, because to finish the audit. But if the client has everything digitally, such as in the cloud, then why not, we can do everything. The risk is that if you look at the whole population, there might always be exceptions and those excpetions might have been followed by the client, internally. Maybe they are emailing, callign each other and agreeding on that. The amount of work to figure out all the differences might not be feasible from the auditor side.	without actually inreasing detection of risky transactions.
Q: What do you mean exaclty with 'exceptions'? For example, we might expect the client to receive a payment, so if the system says we should expect the payment next week for exaple, but the payment didn't get received, but the client is aware of it becaue they spoke to the customer via email or through the phone, then we need to trace back all the communication, internal documentation, this becomes lots of work.	
Q: In which format can the evidence be received generally form the client? It depends on the client. Some clients are able to provide data direcly from their system, also depending on the scope of their work. If the value of the account can be validated via external sources, let's say client has a back account, so we call the bank of the client and request the bank statement fro confirmaiton and this is usually provided in pdf, nowadays also in excel, bu we need to make sure it's received from external party, to validate it and rely on it, to make sure it's not something the client made up. Usually it's excle, pdf, csv (so text file). The outlooks are different as well. For example if you receive invoices from the client';s clients, each invoice may look different.	Structured and semi-structured data is involved.
Q: How do you deal with it right now? It's just a manual process. Some invoices are more straightforward, some other are more difficult. There's a tool in EY called [name], a copmany in collaboration with EY that developed this tool. In excel, if you select some amount or section in the invoices, you can upload the invoices in this tool and then you can select the seciton and it will automatically recognize if the invoice has the same layout for example, it can get all the amounts from the same invoice. I doubt if it's working. Maybe now it works better. It doesn't automatically collect the data. There are still manual requirements.	Non-AI tool for automatically select relevant sections in documents exists, but does not work effectively. I does not automatically collect the data, so manual procedures are still involved.
Q: Financial audit is a broad topic. According to your experience, do you think there is an area of financial audit, where it could be interesting to focus from a deep learning solution lens? You can focus on the financial statement itself. Let's say some sort of financial statement reviewer that basically says, is this fin statement made based according to the reporting standards that apply to this specific fin stat or client? It can tie all the amounts with the notes of the financial statement. So basically reviewing the financial statement like an auditor would do. Or, the manual procedures, mandatory forms and all the	Financial statements tie out procedure, where consistency is checked, as well as regulatory compliance, is an interesting use case for deep learning application. Same with automating the filling of standardized forms.

procedure, they could be automated with deep learning, that would help also.	
Q: What are your thoughts on integrating deep learning technologies into the audit process? I would say the digital assistant, we already have with EYQ (EY's chatGPT). I would say fraud detection is an interesting area for application. If you receive evidence, invoices or signatures, you can see if it's been modified to some extent, with paint for example – which I saw in the past. To what extent has the management been involved in the data, or for example do we see any people who should have been involved in the process of the data for example. We don't necessarily record meetings to my knowledge, but maybe in the future, I don't know to what extent.	Fraud detection could benefit from deep learning applications, for instance by checking invoices or signatures are legitimate and not tampered.

6.5 Interview 4D

Manager data science knowledge **Round 1** **Semi-structured**	**Labels/interpretation**
Interviewee introducion You cannot generally term DL better than classic ML algorithms. For exaple, if it's a classificaiton task, when I say classification it can be anything – boy or girl, fraud or non fraud, all these kind of things – generally, they work much better than DL itself. Second thing is, in ML you will have a lot of explainability. If you just care about the end results, than DL will be much better, because it can connect the dots in many comninations than ML. But if you want to explain to your business: this is the result I'm getting and this is why I'm getting it, DL is kind of a black box, but ML is much more explainable. Now people have also started working on the explainability of DL models, but it is in a very naive phase. I don't want to move you're direction, but when you explain it to someone, you know you have these points to keep in mind. So, these are the limitations, but since in financial audit you have a lot of unstructured data and DL algorithms have much better chance of understanding that data. For example, NLP can be done with traditional ML models, but if you consider LLM, they're much more deeper in the learning, so they work much better and much faster. Also financial audit is a domain, it's not a use case, so this is something you can consider in your pros and cons. DL is very focusing on specific cases, but maybe you can find specific areas in financial auditing where you can use DL to help. DL is deep and it requires a lot of efforts and computation which costs. People never talk about this, but this is also something to consider.	ML is more explaiable, but DL is more accurate. DL explainability is being addressed by research. DL is suitable for unstructured data, and FA procedures deal with it. DL is costly to implement, it requires computational efforts.
Q: You mentioned financial audit is a domain and not a use case. What does this entail? Inside financial audit you might have different use cases. I'm not from financial audit, but from my understanding, it could be for example that if a firm's balance sheeu is working properly or not, this could be a use case. Maybe you don't need a DL algorithm. If you have a use case where you have to find out that a fraud is carried out, when some numbers don't make sense, that's a proper classic use case for DL, because there's a lot of data. There can be multiple use cases. You cannot use one model to solve the whole domain.	DL is case specific.

Q: Do you think that having several use cases could be a limitation for the actual DL implementation? A lot of people think from a direction that: I have AI, where do I use it? This should not be the approach. The approach should be: this is where I am, what are the pinpoints, how do we sove them. Rather than forcing that to be solved by AI, we should look for if AI is needed or not and what value can bring in this, because there is a cost, effort, which sometimes is so trivial, that if you just put an if else statement, the results will be much better then a DL, because it can never be 100% accurate. But if there's a problem which can be solved with an if else condition, it is deterministic, you can just get the results. I'm simplifying it, but there can be cases. It's important to think from where the pinpoints are and understand whether we actually need AI and if it can bring benefits.	Thinking process should start from identifying the challenges and finding a solution, rather then forcing AI to solve issues.
Q: When you mentioned that DL can be used in a case where there's a large volume of data, how large are we talking about? There's no limitation. High volume data is not a problem. We're also doing an audit, fraud detection for one of the utilities in Portugal. We started dealing with 10 thousand customers for 10 years of data. We are talking about billion of rows. But when we were doing the prediction, it was just 400 customers that we had to use in the first phase, so we were working on azure cloud system, just to give an example. A very basic version of machine was able to handle it. Now we have moved to 8 thousand customers that we have to capture and that's not possible with the current scnario. Our model and training remains the same, but to get the results we had to move from a very basic machine toa very heavy machine, so that it cas manage the data. It doesn't itself impact oon the DL model, but getting results out of it is where the volume of data matters.	There is no upper limitation to the volume of data that DL can handle.
Q: I made a list of financial audit challenges. Financial audit evidence and data comes in unstructured form: images, graphs, different formats. Do you think DL could be used to address this? Structured is something that you can directly read. Unstructured could be a recording, so the machine can understand the reconrding. So it has to understand natural language. Then it can also check the expression, so if we're for example both comfortable in this call. Text, like a pdf, is unstructure. Structured data is tabular. This is not a challenge, this is an input. Challenges can be: anomaly - unexpected spikes, discrepancies, non regular financial statements. AI can connect many more dot than a normal software and a human can. Humans doing anomaly detection in financial audit, if you have good data, with DL you can have 90% accuracy, which is an increase. Another example is risk assessment. What risks are in the financial audits: considering the market, maybe it can tell the client is going throuogh a lot of risk because of this factor. If you can predict something. What's going to be the market trend and all this stuff. Another use case could be compliance: most of the companies, ever EY, use rule based engines. If this is checked, then it is compliant. There are multiple things that are not straightforward. Then you can check if in the future you can get non compliant.	AI can connect many more dots than a normal software or human can. Add in results of SQ2. Use cases identified for DL application: anomaly detection; risk assessment procedures considering the market dara, predicting market trends; veryfing compliance through checklists.
Q: Financial auditing entails sampling, to collect a sample of transactions on which to perform test of details or substantive tests. However, regardless of the population size, the maximum sample size does not change, as performing these tests is time consuming and	DL can be employed to automate the testing process. DL can be

manual. Reserach proposes population testing, which requires automatizing to a certain degree these tests. Otherwise, having a better sample selection, rather than random sampling, would also be an improvement. Would DL be able to address this issue? We can have a process setter that can consume more data than the one included in the fixed sample size. So if the population increases, also the sample size can be increased. This entails automating the testing process of the transactions in the sample. This can be done with deep learning. So if you a data population of 600 and test a sample of 60, you don't see 90% of data. Isn't this a risk? Also, if the data population is 6 million they still check 60, this is like 99.9% unseen data. So in the current process there is trust in unseen data, but no trust in technology?. I'm pro human in the loop, but if you use technology as a complement, it can do magic. If you try to replace humans, there arer a lot of risks. There are can be a lot of sampling techniques based on DL that can be used	employed to select representative samples. Human is uspposed to be in the loop. Technology as a complement can do magic.
Q: How about abouting autmating the testing process? Then there will be multiple use cases in this section. You can apply DL and combine everything to have a whole system of audit. Maybe somewhere you need to have human in the loop, but it will become more accurate and more data oriented.	Automation of tests of details entails various use cases. DL can lead to higher accuracy.

6.6 Interview 5E

Senior manager experienced in leveraging AI solutions in FA process **Unstructured** **Round 1**	**Labels/interpretation**
Introduction I'm directly from the scientific community data science cognitive science and I've been kinda pushed into the AI topic within EY for the last 7 years within Germany. [...]	
Q: Based on your introduction, I would like to have a dig deeper in your focus area. What have you been working on and what are you currenlty working on regarding AI applications in audit? There are multiple areas where you can work with AI in audit. One is the planning phase, one is the execution part and one is the reporting part. The planning phase, we've been working on recently on recommendation engines. At the beginning of each audit, you plan the audit based on the information that you have about the client and the industry. There you can use the AI in different ways to process way more context information and give recommendations. The AI takes the information in our audit platform about this specific engagement, find prior years engagements that are very similar in terms of size, effort, other geographic or industry specific paramenters and then compares your planning to the average planning of all of the other similar engagements found. Then it can give you recommendations such as: you have not identifie cash as a significant account, wherease the other engagement team in one engagement that is 95% similar did, so think about	AI in the audit: planning phase, execution part, and reporting part. Planning is about providing recommendations based on similar engagements, past anlaytics, market and competitors' analytics. Exectution phase is about automatin manual and tedious procedures: checking the financial statement content consistency; automating tests of details; automating verification of

it, just give it another thought. The human in the loop is always important, so it's always just recommendations. Also regarding understanding the business, the industry, what has happened in the meantime when you plan your engagement, collect way more information, highlight the relevance. If you're in an automotive industry, maybe there's a shortage in the rubber based on something that happened in Brazil. You might not have found it when you did your research. The AI can flag it, it could affect the business of the audit client this year. Also some basic analytics, from financial reporting that goes out every year. It takes the financial reporting of your client form last year and then pierce similar commpanies or competitor or whatever and automatically does a busnch of analysis and highlights the relevant bits. Up until the analysis it's mostly just automation, but highlighting the relevance, finding the outliers – look, your EBITDA is off the charts compared to your competitors, maybe that's a field you need to focus on during the audit. Execution phase is automating very tedious work, manual work. Financial statement tie out: it can take the financial statement, usually the current year financial statement that is in draft version and it double checks its content on consistency, so if you're talking about 200 million revenue in chapter 1 and then you're talking about subsets of those revenues in different business units in subsequent chapters, do the subsequent chapters mentioned actually add up to the 200 million? In the table, in the document, the math and the tables get automatically checked to see if it's correctly reported or if there are mathematical errors, compares the currently in draft financial statments to the prior year statements, because in the financial statement you always reference the prior year. So in the financial statement from the prior year, the current information aligns with it and you can double check the difference this information.Also,and that's a lot of AI, is in the flow text. Sometimes you don't write the number in a table or have a numebr in the table but you also say in wirtten words ‘six hundred million'and the other figure says ‘6000000'in the text, then it double checks if the revenue table there's an entry of 6hundred million even tho here there's a number and there is words. AI makes it possible to compare. It double checks everything and highlights to the auditor just where it ofund inconsistencies so it saves the auditor a lot of time, because usually you the auditor needs to chek manually. Similarly, there's a lot of checklists, you need to go from the documentaiton to the checklists and check – is this ocndition from the IFRS? Yes, no... – the AI does it for you and the auditor only needs to double check if the AI did a good job. This saves a bunch of time. These are just automation steps or, if you have to a test of detail, say like I need to draw a sample of 100 invoices of paper and cross chekc if they are correclty written down in the book keeping, with AI you can do thousands or ten thousand of that, because you can just let it run through. Instead of working on a sample of a 100 documents and create a	compliance with checklists. Final phase: AI can automate drafting reports. Auditor always needs to review the result. These applications improve quality an efficiency.

judgement based on that, you can create a sample of 10k documents which improves the quality. In the end, when you did all of your steps, tests, procedures, you need to fill out and right a lot of forms and reports. This autofilling base on the data that you have the audit platform where you work and document it. You ask it – ok this is all the tasks that I did for this engagement: draft the first report. AI will do it based on the content. Human in the loop always checks the quality, but definitely saves the time.	
[Confidential information]	
Q: Is there any type of DL application that you've been working on or in which you could see potential? All of the applictions are DL. Everything has DL in it. GenAI uses a neural network that has a couple hundred of hidden layers, so it's DL.	The applications are all based on DL.
Q: How about trust in these tool? The audit business has to change to a certain degree. AI is never 100% perfect, there's 95% accuracy. You make a 100 samples, then 5 out of them will be processed the wrong way, which is something you need to account for. It also means that as an auditor you need to learn how to interpret results of AI solutions. So, if the AI solutions for instance tells you Í've tested 100 invoices of test of details and those 5 are suspicious', I as an auditor could say 'I know the performance of this tool is very good, 95%, so I can trust it for the 95%'then I need to double check those 5. I canonot just stop there and say: ok 5% of your invoicing is worng, what did you do wrong? Because you need to double check if it's an error from the lcient or from the AI. That's the methodology change that you need to work on, work on the itnerpretation of AI solutions. Also, a thing of exposure experience, we need to upscale for auditors. They need to use those tools on a regular basis to collect experience with the tools themselves and then it will trickle into regulations, internation standards of auditing etc, because we will rely more and more on those things. Risk based model: you have a very easy formula. Inherent risk x control risk x detection risk. Inherent risk: how likely is some error to happen in the process, how error prone is the process. Control risk: how likely that th einternal control of the client will fail to catch an error. Detection risk is how likely is it that our own audit method will also fail to catch it. Multiplying these 3 probability values, the result should be below 5%. So you don't also expect 100% results from the auditors. That's also why in Germany, theprofessional practice department we're working on integrating the accuracy of an AI model into this risk model. So if you say my AI model is part of the detection risk calculation and I can add the accuracy calculation from my AI model as part of the detection risk factor and if the AI model performs poorly, then I have to perform other substantive procedures to balance it out, just like how the audit risk based model would work. That's a development process, it'll take a little longer, but the business will get there.	The audit business has to change and auditors need to learn how to intepret AI results. AI is never 100% accurate, but it can be 95% accurate.

Q: Is this focus area more centered on improving the audit quality rather than the efficiency? The thing is that we're definitely improving the quality because we cna process a larger amount of data than a human could ever do. You also could get tired looking at hundreds of invoices, eyes get tired, but that doens't happen to the machine. It also saves the auditor time. So the idea is that we don't want to automate the audit as it is framed now. We don't want to reduce the workload because then you coome into the unfortunate position that the client also pays the auditor less, because we have a service based model. If we suddenly take half the time for the service and we're charging hours, then the client could ask why are you chargin the same amount? The monetization would change, which is a big challeng. This monetization has been around for a hundred years. One way is to increase the quality. We still charge you the same amount than last year, but then we use AI solutions to get higher quality service, because we check more, we dig deeper, instead of our manager looking through a 100 entires, it just focuses on the 2-3 suspicious and spends more time digging deeper into those examples, so the quality increases.	The mentioned AI solutions can improve audit quality and efficiency.
Q: How can the black box and explainability issue be addressed? A lot of tools are easily explainable because they have a very narrow scope. The easiest example is the automation of test of details. Here is the invoice, here is an x line and we just tell it to compare. That's it. It's still a black box, but it's much easiere to adjust. We're not asking the AI that here is all the data we have from the client, give the result., because we have the human in the loop. It's not really the human in the loop, but rather the AI in the loop. The majority of the work is still done by humans. The levels done by AI are very limited in scope and easy to encapsulate and to explain. Another point is the risk model integration. We're not just the AI blindly, we have the whole operations around the AI that need to fit and we're upscaling our professionals, who need to understand what's going on, you cannot just drive the car, you need to understand that it needs gasoline, oil, what happens when your car breaks down, what could be the reason. We have smarter people and very tight documentation and human in the loop. With genAI it can get full circle because you can ask genAI to explain why it did something the way it did. It's not trustworthy yet, but at some point, just like you ask a human why did you decide left and not right and they will give you an answewr without looking into the human's brain and poke it, at a certain point we'll be able to ask the AI in the same way	Easy explainability for narrow-scoped AI models. AI in the loop view.
Q: In the first phase of the audit, the planning phase, do you think with AI retireving more data could be feasible? For sure. That's part of the data volume of the AI. You can drop instruction data, news articles, you can drop in documents like reports etc and the AI can kind of mash it all otgether and give a nice overview and do that for a much larger volume of documents, reports, news artciles, that you as a human would	AI can process large volumes of structured, semi-structured, and unstructured data, meshing it and providing an overview. AI implementation

never be able to read in a lifetime. Regarding the sampling, something that is challenging for the auditors is data scarcity. Currently, the cooperation with the client is based on sharing as little information as possible with the autiro, only as much as absolutely needed. So, there's also a change in eduction of our clients or a model where we an do it in the traditional way, where we test 25 with random sample testing, we only have humans, no AI, then you get a work quality, or you let us put an API that pulls all the documents and then in the same time frame for the same money you get amuch higher quality product. At the moment, when I have a new AI product and I want to test it, I need to go to engagement teams who need to talk to the client who needs to consent to this. There's still a lot of hesitation around it. Since all of those client are now starting to use AI more and more themselves, this blockage is softening. They understand the domain a bit better, so they understand the auditors want to use it themselves, because if we use high tech solutions in our financial reporting, we want the auditor to be at the same level as us.	challenge: data scarcity due to the traditional audit model.
Q: Thinking from the client perspective, such as a client who wants to hide something, or is just hesitant, if the audit price is the same and the audit quality could be higher or lower depending on the AI use, why would a client choose the higher quality audit where they need to share much more documents? From the change of business perspective, thinking more on the longer term, if the client agrees to open up their data walls and share everything, then we can say 'hey by the way at the end of the year you can say to all of the shareholders, to the stockmarket, everyone, EY has tested us 4 times in this year with their automated processes'. The competitor could say 'we got tested once, the old school way'. What kind of image does that give to the market? What kind of level of trust in your business does that give to the market? There is an intrinsic motivation to get more automated because in the end you can show off your better quality audit to everyone who is interested in the audit. If you come to a situation where people want to hide someting, that basically directly feed into your audit planning and in your expectations about the client. It turns into a source of information about the client.	Client's benefit in choosing to be audited with AI solutions
Q: Then, is it shown in the audit report that a client has been tested more? Yes, this can be added to the audit report and specify that highly automated modern AI approac - that gives the possibility to run the entire audit process 3 times in the year because so much time has been saved- has been used. It's more an audit as a service. If we get far enough with the automation, we can do it every quarte, we can do it every month, or whenever the client want to push on the button on the audit portal and ask to be audited now, where 2 weeks later after the human in the loop is done you get the results. Whenever the company faces any kind of scrutiny of questions, they can just quickly get audited, in 2 weeks you get your results. That's the future that	

we're envisioning, but you need to get everyon emoving into this direction and it will take some time.	

6.7 Interview 6F

Staff – data science expert **Semi-structured** **Round 1**	**Labels/interpretation**
Q: In order to apply AI in financial audit, there needs to be assurance that the AI performs consistently, without bias and that it is explainable. Is a DL algorithm or model that could address this? It is a porblem for machine learning everywhere. In terms of the black box, it's true, we don't exactly understand how the model came up with the answer form the inside, but we still have a good understanding of what happens on a conceptual level. I don't know if you've seen those images which show you how the model is creating an image layer by layer. It starts off by making dots, the next layers will be making lines between dots, the next layer will be making a shape of the lines and it goes on. In terms of bias and data, the best way you can give yourself assurace, in my opinion, would be by using confidence levels and accuracy. If you use these metrics and reach a high enough number, then it should be fine, because humans are likely to be bound to make the same level of mistakes. All we can do is just use the confidence levels and the error levels, the ratings. On top of that, just relaly understanding the data set that you're using to train the model. At the end of day, almost every model will be a little biased to the data that you give it. All you want to do it to make a model that doesn't overfit on the data, but that's something that you take into account when creating the model anyway. So I genually feel that AI has a place here. If you use these confidence intervals to represent the AI to the management, it should provide that assurance that you're looking for. But it all depends on whether it's been trained well enough. It's all about quality at that point.	Black box around AI can be circumvented with solutions, allowing its application to areas where assurance that the model performs consistently is needed. Understandining the data the model is trained with is also crucial.
Q: Can you expand on the use of confidence levels and accuracy? Sure. When you're training a model,you have validation and accuracy levels at each stage of the training. You get accuracies at each stge of the training. At by the end you have 2 types of accuracies that are useful – the accuracy on the testing data, which is a fundamental stage. If that number is high enough, then you have assurance the system works. Then there's validation accuracy, which is useful for people during the training phase. You're not actually supposed to reach a 100% with these numebrs. If you reach 100% too fast, it probably means the model has a problem of overfitting and is overly biased towards the data. The point of the testing stage, the last stage in training, is to make sure the model doesn't get biased. Training properly entails avoiding such situations, which is why data from any type should be used. You shouldn't use data from the same dataset. Other forms of data or output that you could	Explanation of accuracy levels and confusion matrix, that are metrics and solutions to ensure enhanced understanding in the working of AI models.

use is the confusion matrix. A confusion matrix is just a matrix where each row is every class where you're trying to identify your data. In every column you have the same thing. What happens is that the Y axis is the predict label, whereas the X axis will be the true label. It's useful to see if every time you predicetd class A, it was actually class A. So, high numbers will form in the diagonal, which will mean the model is really good and able to predict class A when you're given class A. But you can also use it to figure out where your model might get confused. It gives you insight into whether there is any bias that is significant. You don't even have to say "the whole model is biased", but you can say "the model is biased because it keeps confusing these classes", for example. The relevance of the confusion matrix is here is to provide more assurance, that you're not only aware the model can be biased, but you know exaclty where it is actually biased. With that you can do a risk assessment analysis. Depending on where the bias is, the risk could be higher or lower, when using the model.If the model confuses certain classes, it could happen that that area does not lead to a high risk situation. For example, consider management matrix. Maybe it is acceptbale to work a little bit overtime. Maybe it's not ok to work too many hours overtime, but your model is assessing your employees to . With the confusion matrix realises that the model often makes mistakes regarding small or large amounts of overtime. The manager sees the situation, and says there's a high risk we lose a lot of money because many employees work overtime. So the main source of information is accuracies. When you want to go deeper, there are many artifacts that can be used, such as the confusion matrix, that gives more insight into where the model can go wrong. I personally believe that it's a lot safer if you know how the model can go wrong, rather than saying it doesn't go wrong. A confusion matrix is not AI, it's not even a model, it's just a table. When you're training the model, you can just write a quick script on the side, to create a confusion matrix.	
Q: You mention the overfitting problem. In case deep learning is applied in financial audit and the data used to train it is data from the same client from past engagement, would this lead to this issue? That would be a scenario to watch out, yes. If you're certain the results of the audits in the past were all up to par and great, then you don't really care if it's biase, becase it would be biased to something that's correct. But that's never really the case. It's always smarter to get more diversification in the data set you use for training. Another thing to avoid the overfitting problem is to reduce the number of epox in the training, but this is more technical.	Type of data to use for training the model in FA scenarios.
Q: Do you think manual work could be alleviated through the application of deep learning? Yes, I believe it should be able to do it. It's a bout engineering to the best solution. Using AI isnt always the right answer, sometimes a too grand solution. It depends on how specialized the tasks are. If they are general	Manual procedures can geenrally be addressed by DL. However, caution is suggested to verify whether simpler

tasks that have hundreds of different ways to go about it and humans considers the right one that does it, then it's a bit difficult for machine learning to do it. If it's a specific task, then a specific DL can certainly do it, probably even better than humans can. But you should consider if simpler solutions could be applied. When it comes to the volume of data, AI can 100% be a solution. Visualization of data, ML can be used. Using more data in the first phase of the audit: if you can write it in a plan, then you can ceryainly teach the ML algorithm to learn that.	solutions could also do the job. In case the task is specific, then specific DL models can approach it even better than humans.
Q: One example of manual task is data extraction from unstructured data, as well as filtering and review. Can DL help here? You can use deep learning here, specifically LLM, which can be used effectively to provide structure to data received. If the text, for example, is in an image, then you wan to incorporate some type of character recognition software along that. That stuff has come a long way now.	DL, especially LLM, can provide structure to the data received.
Q: Regarding the possibility to extract more data from online sources, do you think deep learning could help the extraction and processing of such digital data, as well as assessing the reliability of the sources of this data? One way to improve the sampling process. If you get stuck with the same sample size, we can pick the best 60 rather than random sampling. There are ML algorithms that can be used to find the data points that are most different to each other. That allows you to get a taste of all the different types of extremes you see in the data. For example, it's called principle component analysis (PCA). When you apply a PCA to your data set, it tells you what the significant trends are in the data. When it comes to your sample size, you can the 60 most significant trends in the data and pick 60 data points, to get a diverse sample size as possible. PCA is party used in the field of DL to actually reduced your data, funnily enough. It tells for examples, these 2 folders are very differernt to each other so it's interest to keep them, for these other folderes are similar, so let's remove one of it. So we have less data to train the model. Sometimes you have too much data and not enough time.	ML can select representative samples.
Q: Once the PCA is used and we can reach a more representative sample, is there room for DL? In terms of using DL for test of details on transactions, the more significant role DL can play would be to provide the auditor with tools to address specific things in all the steps need to be taken. For example, if the auditor needs to trace a transaction back to verify it's free of mistakes, a DL model can be made. It has to be fed with million or thousands of transactions with the correct stages being traced and then you can give the model examples of cases where the procedure was not traced properly, for whatever reason. The DL model can certainly trace the transactions back properly. In order to identify fraud, a separate DL model is needed. I don't think the DL model can take over the job, but if it focuses on different aspects of the audit, it can definitely speed the work of the auditor. Depending on the sensitivity of the task, it may not be feasibile. For example, assessing whether a transaction is	DL needs to be specific for use case. AI cannot replace the auditor, but it can flag areas where the human should focus more.

fraudulent can't be just left to the computer. What can be done is to use a DL model that flags transactions with high probability of fraud cases. In this scenario, the auditor can focus on these flagged transactions. An example from one of my engagements. I was working at a university where they had a manual payroll method. People would receive the timesheets from the employees and logged the hours into the computer in order for them to get paid. The issue was that just one mistake led to delayed payments, it was a very sensitive task. Due to that, the responsibles had a lot of pressure on them and they would keep quitting. It was such a manually intensive job and if one little mistake leads to someone else not getting paid. This turned into high turnover rate. So, automating very manual tasks, you can reduce turnover rate as well, not just the work.	
Q: Reconciliation procedures require manual work and is time consuming. Would you think there's room for DL solutions here? I think DL would be an overengineered solution. I think using just some scripts or simple algorithms would be enough. Probably DL is more than you need here.	DL is overengineered for reconciliation procedures. However, in hindsight, not much information was given on this procedure. The word reconciliation is too general, and for someone not familiar with FA procedures, it could provide very little information.
Q: Could you tell me what specifics need to be known in order to understand whether DL can be actually applied? What is the nature of the task, for example when it comes to trying to understand the data, instructions of what the data analytics people do to understand it to verify the quantity of manual work. Another example would be about fraud, understanding how someone figures out that one is a fraudulent transaction. Once the steps are known, then we can see whether ML can elevate those or alleviate the work.	Required information to proceed with the second round of interviews.
Q: What are the most commonly used DL models? When it comes to visual data, images, CNNs (great for spacial data) dominate that area. When it comes to audio related data, time series data (sequential data), then transformers are the general go to. These are the 2 main ones. When it comes to understanding text and trying to bring a meaning to it or structure it, then you are probably using LLMs. I like to think these are the 3 cornestones of models that DL is using these days. RNN is a type of CNN. It's about the timeseries data. An RNN is good at analysing sequential or timeseries data. There's some good performance in video and audio data as well, of course. If you want to go deeper into pure audio, then transformers are going to be better. It's about using the right model for the right problem. CNNs, RNNs, LLMs and transformers are all types of models. Then you actually have models. In CNNs you have BGG, restnets. Within each of them, depending on the type of	DL models cornerstones: CNNs, RNNs, LLMs, transformers.

problem, you find the better or worse solution. I don't think it's possible to rank the models in an absolute way. Once we have a specific problem, then it's possible to rank them.	

6.8 Interview 7G

Manager financial auditor **Semi-structured** **Round 2**	**Labels/interpretation**
Q: I classified the main challenges that I found in the literature four areas. The first one is large data volume, the second one is the manual procedures. The third one is the sampling process, and the fourth one is professional judgment. So, the first question would be, do you agree that the available and potential large data volume is one of the challenges in the FA process? Yeah, I would say yes. I think for two reasons. One, there is more and more emphasis on the reliability of data, and the larger the data sets, the harder it is to determine whether the datasets are reliable. And two, is transforming datasets we receive to usable data sets for the audit. I think a good example is for one of the bank audit them. We try to determine the existence of the mortgage loans by reconciling the mortgage administration to incoming cash flows. But these are billions of transactions, so it's always a challenge to them, first, being able to work with the data because it's so much data, and then second, transform it to a usable format. So, yes, I really agree that that's a challenge.	Large data volumes present challenges in FAs for: verifying data reliability and transforming received data into usable formats.
Q: And what do you mean with transforming it into usable format? How do you address that process as of now? So for the example I just mentioned, with the mortgages, we get two different types of data sets. You have a data set with all the mortgages in the books of the bank, and you have a data set with cash flows which are in different forms, formats. So now, and I'm not into all of the details because I don't perform procedures myself, I only review them on a higher level. But what we now do is that we determine which of the data points in all those sets are relevant for our procedures. [confidential data] We use a tool to link the relevant data fields to each other. But what you then quickly notice, for instance, is that let's say you have a mortgage number in one of the data sets, and it's in a bit of a different format than the other data set. So then you have to transform one of the data fields to reconcile to the other data set. And I can imagine that if you have some sort of, don't know if I'm using the right terms now, but deep learning of machine learning or whatever. And you can tell that so that one of the data fields need to be in a specific format, that it automatically transforms all the data for you to reconcile to the other data.	Non-AI tool employed does not completely solve the issue. For instance, data points can be presented in different formats, which is something that has to be adjusted manually.
Q: And what type of formats can you have? So what you can have is just all numbers next to each other, but sometimes you also receive it with numbers with dots in between. So that's simple to remove the dots. But then it can also be that on certain points within the string of numbers, there needs to be dashes because there can be	Different data presentation is a challenge.

subsets, and it's not always on the same position within these strings where the dash needs to be.	
Q:And how do you address this problem right now? Is it a manual procedure? Yeah, partly. So we can usually reconcile most of the data, and then we just get this set of remaining data which we can reconcile, and then it's just manual labor, more or less.	Manual labor.
Q: And then you also mentioned that you need to check the reliability of the data you receive. Can you expand on that? Yes. So there is information produced by the client, and there's a couple of dimensions. So do we have a complete data set? Is the data intended data captured? Were the correct parameters used? Is the application processing the data reliable at all? Which is, of course, also a big question. And with the bigger data sets, the queries to get them are becoming larger and more complex. And to review those queries can be quite complex. And a lot of times I can't even read the queries myself, read them on the basic level. But to be able to read the queries for paradox data sets, you need an it auditor who can read SQL or other types of codes to be able to conclude on reliability of data.	
Q: I've heard from other interviews that one difficulty still connected to the large data is trying to get an overview or comprehensive, like a straight overview or comprehensive understanding of a data set. Do you agree with that? No, not really, because, well, it depends as always, because usually what we do in the audit process is that we split up the audit in two parts. So one is understanding processes, and then based on the understanding of the process, we perform the other procedures. So understanding also how a data set is constructed should be part of your procedures, of your understanding. So I can imagine, yes, that there can be difficulties in that first part of the audit where you confirm your understanding, but once you have that, it shouldn't be difficult in the audit itself because you already have the knowledge of how the data.	Getting an overview of the data is not a major challenge.
Q: At first I mentioned the available and potential large data volume. With available, I mean the data that you receive from the client and the data that you are currently analyzing. With "potential" large data volume, I'm referring to the big data available. The literature states that including big data in every phase of the audit, for example, the first phase when you need to understand the entity, could increase the quality because you have the potential to include several types of data, like emails or news articles or social media message. Would you see a benefit in doing this? Yes, there is already some benefit. So as part of the planning of the audit, we already receive usually quite large data sets with these old transactions of their ledgers, financial ledgers. So it already gives some information on what they are doing and how certain transactions are recognized and processed. But I can really imagine that for completeness of your procedures, external information can be very valuable. So as an example, one thing we always need to consider is whether there is any form of litigation with the client, or if they are non compliant with laws and regulations. What we usually do is we ask a client, is there any form of litigation	Including external comprehensive data provides benefits for understanding the business, for example to uncover litigation claims or regulatory compliance.

or non compliance, and read their internal reports. But if you can also include large amounts of news articles which may name your client that they are non compliant or they, in a litigation case, you get a lot more information, external information. So you also remove the risk of being steered, so to say, by the company, to a certain answer. You get external information to also design your own procedures. So I can really see the benefit there.	
Q: I've read that the validity of the evidence needs to be checked. And specifically, sometimes it's not clear whether the signature on a document is actually original. Would you say that this could present challenges in the audit process? Well, I think we can get back to what we just discussed. For larger data, yes, it can be a challenge, but for something simple, as we received a contract which is signed by a CEO, and then the question, is that signature valid? I don't really see a challenge there because you can always inquire whether they really signed the contract or receive an internal document with all the signatures of all the employees and then match it. Of course, you can imagine if you have a lot of documents that have been signed and you have one authentic document with all signatures of employees, you can match those together automatically. That you can, then that will be helpful.	Verifying signatures is challenging only with large data volumes.
Q: Would you agree that the amount of routine and manual procedures is a major challenge in the financial aid process as of now? Yes, it is a challenge, but mainly from a resource perspective. So I think that's also one of the challenges you mentioned and which we will come to later. But I think the two chances that I saw were, on the one hand, capacity for manual procedures, and on the other hand, professional judgment. And what is often a challenge is that you have a lot of manual procedures to do which are very factual in nature. For instance, we select a few invoices, and we need to reconcile invoices to what the company recognizes in expenses. That can be very time consuming, but it's not something that requires any form of judgment. So the real challenge is getting the capacity there to perform these really simple and I would say, uninteresting procedures. And if you can also link it to, I would say, deep learning or machine learning, I think that's one of the parts where it can really help by already performing such mundane procedures for you, because it's some very specific data fields that need to be matched. For instance, the amount of the invoice to what's being recognized in the general ledger, a date on the invoice to what's being recognized, etcetera. So, yes, it's a challenge, but I think it's also a challenge that can be overcome.	Routine procedures that do not require professional judgment are factual in nature and time consuming, introducing challenges.
Q: You mentioned the reconciliation. This is something that has been mentioned several times. Can you briefly explain how the reconciliation process works? Yeah. So it can vary very much. So it can be that we have to reconcile something in an Excel file to a PDF for invoice, or sometimes we receive the invoices in a JPEG format or PNG format. So you have to reconcile that. This can be excel file to excel file. So it can be all kinds of reconciliations.	Reconciliation details.
Q: Something else that came up regarding the manual procedures realm is that there's a lot of regulations in financial audits	There are more factual checklists,

and many checklists that need to be filled, for example, to check if the IFRS standards are actually met or not. But this is a bit unclear to me, so could you explain that to me? Yes. It also depends on the type of checklist. So maybe start with the most mundane version of checklists is we have a checklist to determine whether the financial statements are compliant with the laws and regulations. And these are also quite factual checks that we do. So as an example, let's say a company has buildings on their balance sheet and it's required to disclose what is the value at the start of the year. Have there been any additions to the building which increase the value? Has there been any depreciation? And what's the ending value per end of the year of the building? And basically the only thing we then do is, okay, the checklist says it needs to have all these items in the financial statements. Do we see those items in the financial statements? That's just a check in the box. Okay, so that's very factual and simple. But another form of checklist we have is for the board report. And part of the board report is giving a description of what happened during the year. So what drove the financial results? Now one of the requirements in that board program or the checklist is "has it been described what drove the result during the year?". But of course, determining whether or not the description is correct and gives a fair view of what really happened within the company is a matter of interpretation and judgment. So that can be a lot harder to go through that checklist.	such as for determining financial statements compliance, and more judgment-based checklists, such as the ones regarding the board report content.
Q: What type of data and information do you use to support the professional judgment for this checklist that you mentioned last? So it can be that we read internal reports. So, for instance, that the board receives a monthly report with all the results. We read those reports so we know what happened during the year. We can have had interviews with the board or with the financial department or their internal audit department, but also sometimes external information, for instance, something from the news which we read. Okay, so it can from. Yeah, it can be a mix of reading documents, inquiries, etcetera.	
Q: You mentioned that it can vary the type of evidence or documentation that is used. Who ultimately decides the amount of documentation to use? Are there any rules that can help with this decision? There's no, as far as I'm aware of, no formal list that states, for instance, what you need to have, it's all judgment. So, as an example, let's say in that board report, it is being described that a new loan was granted to the company from a bank, which would then want to have a supporting documentation is the loan agreement itself. But if there's a description of, yeah, we had an income of this, expenses of this, etcetera, then it can be just a mix of monthly reports that are being drafted within the company, but also in the future hat. So it's also more or less judgment, proficient judgment, to determine whether or not the information is sufficient.	Example of application of professional judgment.
Q: Now, moving on from the challenges to the actual deep learning applications, let's start with one introductory question. The literature believes that including recordings could increase the amount of evidence used. Recordings of interviews, for example,	Recording oral inquiries leads to benefits and challenges. The benefit

conference calls, phone calls as well. So there is consensus that including these recordings could give more evidence and, of course, increase the audit quality. Do you think this should be done in the future? Do you see some benefits connected to this practice? I can imagine that if you record that, it's easier to determine what to do with the information, because often what happens, you have a long conversation and then you forget part of you interpret it in a certain way or you remember it differently. So, yes, I can imagine that it helps. On the other hand, a challenge that I see is that, let's say I have a conversation with a finance director upfront. I would ask him to record this conversation. And I can imagine if someone knows that they're being recorded, that they could give you less information, that they're less open. So that could decrease then the information that you get. So I think it's a double edged sword.	entail making sure no information is forgotten, while respondents may be less open in case they know they are recorded.
Q: One of the main deep learning capabilities is speech recognition. And speech recognition with deep learning spans across multiple functions. For example, it can be a simple speech to text, it can also be about feature extractions, highlighting content, keywords, and then comparing with other interviews that you had. Sentiment analysis is a function that provides rating regarding the conversation or the text as well. Positive emotion, negative or neutral. In a paper I read there was this example. It could happen that a CFO answers to an analyst question in a way that during the meeting seems positive. But then if you go to analyze directly, like specifically the words that have been used, you can realize that the specific words are more negative. So this could give a bit more insight. What do you think about this? What is the exact purpose of such analysis? Is it then to determine whether someone is being truthful?	
Q: One benefit could you can have is more information for judgment support. You can have another opinion to understand if a conversation is deceptive, you can flag that conversation or what has been talked about as a high risk area, then you know that you should focus on that, because there is a chance that something could have happened, fraud or something that should be hidden for some reason. Yes, I do see a benefit. I wonder how that would work operationally. Because you can imagine that if you want to analyze, for instance, the CEO, that you need to have a large data set with recordings to feed the model, how they speak usually, and also on the relationship itself with the client, I can imagine it can be a bit hard. So for instance, a CEO said something in a shareholder meeting, and then from that model it's determined, well, maybe it's not being fully truthfully or we see a risk based on what he said. If you then need to go back to the client and say, well, we think you're not, let's say, speaking the truth, or with an analysis of your voice, and we think we see a high risk. I don't know how that conversation would go with a client. So I think theoretically, yes, I do see the benefit, but I don't know how it would work in practice.	The benefit of DL's speech recognition function is identified and there is agreement. However, skepticism arises for corroborating evidence derived from sentiment analysis.

Q: Good point. One question, is there a need to go back to that person and explicitly disclose that you found your conversation strange, so you need to dig deeper, or can you just dig deeper without disclosing why? The second, of course, you can dig deeper without saying why. But often when you focus on an area in which you haven't focused yet in the past, the client usually asks why. And something I also just thought of is what I also can imagine where it would be helpful is in respect of your independence towards the client. We can only be part of the audit for seven years due to independence regulations. And one of the reasons is that if you are involved for a longer term with the client, get acquainted with the client. Maybe you like someone within the company. They're just friendly people, so your judgment can become clouded. And I can imagine that if a system looks on the background, if someone is being truthfully, or there's some potential for deception, that it would help in your own judgment making, because maybe you are too acquainted with the client.	Speech recognition functions can alleviate potential auditors bias.
Q: Text understanding means that deep learning is able to automatically review and extract information from text data. it can also classify the documents and again, perform sentiment analysis. One example that I read was analyzing one document from the previous year and one document from the current year from the management and, through sentiment analysis, you can just understand how conservative the management is. So instead of just searching for deceptiveness or just fraud or hidden, it can just give more. But apart from sentiment analysis, yes, there's the capability to classifying documents, identifying outliers or odd contracts that should be better analyzed. Do you see some possible applications of deep learning considering this capability? Yes, and I think also going back to what we discussed earlier, for instance with text recognition and all the manual reconciliations, text based, I would say yeah, there's a possibility. And maybe on the other hand, which you may mentioned earlier, the sentiment analysis on text, that can also be helpful, but then only in judgmental areas. So if you have a judgmental area, we always ask management to draft the decision paper. And again, imagine if you have some sort of sentiment analysis and you compare last year to current year, maybe you see some sort of trend that due to wording that's being used, that they are being more conservative than last year, even though if you just look at it plainly, you wouldn't see that they're being more conservative. So I would say that there's a split, there's possibilities for split in between the subjective procedures and the more factual manual work. So for instance, the invoice reconciliations.	DL's text understanding application benefits are recognized by the auditor interviewed, for example in reconciliation procedures. It can be used for judgment support.
Q: Another deep learning capability is visual recognition, which is the capability of automatically extract information from images, also videos. Focusing on images, you can automatically extract insights. And if with a reconciliation procedure, for example, you have an excel table and you have a screenshot, this can be analyzed automatically. So would you see possible applications of visual recognition connected to deep learning in the	Not many images are involved in audits.

financial audit process? In my audit, not very much because we don't use a lot of images. So I gave an example earlier that sometimes we receive, instead of a PDF invoice, we receive a JPEG invoice. But the amount of times that that happens is quite rarely. And the financial statements themselves usually also don't contain a lot of images. In that sense they are quite boring. So I would say no, at least for my clients. No, we don't use a lot of images and I can't recall ever having used any video in audit.	
Q: Okay, just to clarify, images can also be in PDF form. For example, if it's a screenshot of a table with all rows of data and the screenshot is taken and then transforming PDF that is still consider an image. Okay. Yeah, that can be helpful. And we already used that. So what sometimes happens is then receive a PDF, but then you can't select the text because it's basically just an image. And then we do the OCR, so it converts the image to text. So to a certain extent we already use it.	Screenshots of document and invoices are present, so an automatic image processing tool can be helpful and alleviate manual procedures.
Q: Okay. And once you convert that, what is the next step? Well, usually I convert it because it's a large document in which I am looking for something very specific. So let's say a loan agreement. And I'm looking for the notional amount or the interest percentage. So I just convert it to text so I can do a control-f to search. And then in that format I also retain it in the audit file. So others can also perform a search if they want to.	
Q: All right. Yeah, that's the first step and I understand why it can help a lot, but then it still remains the manual procedure behind, such as after you have converted the file. Are there still manual procedures to be applied? Yeah, that's correct. Yeah. So it's still then a manual search for the data I'm looking for.	Non-AI based tool requiring manual procedures.
Q: I mentioned the judgment supports benefit that deep learning could bring. It cannot remove the human in the loop. By removing routine and time consuming tasks, the auditor then can focus on highly value areas. But with deep learning, there could be some judgment support areas, for example, in fraud detection, or in determining the sufficiency of evidence, as we were saying before. For example, if you compare the current audit and you have trained the deep learning model with historical data of past audits, the deep learning model can tell you in the past with this similar task. Would you see an area of application or areas of application in the financial audits realm? Yes, but then I think mostly by collecting maybe external information to support certain judgments. So an example I can think of is for banks in certain models, they use so called forward looking information. So expectations towards the future, for instance, how the economy is going to perform. Okay, so the bank itself makes an expectation of how the economy is going to do, let's say, next five or ten years, and then we have to review those expectations. And I can imagine that if you have some sort of model that can also pull expectations from other parties. So let's say you're auditing *bank name*, but you can also pull expectations from *other bank name* from the web or maybe some sort of rating agencies, and then you can then benchmark what is our client expecting and what are the rest of the	DL's benefit in judgment support in FA procedures: checking and using information from the market on which to base audit expectations, used for evaluating the client's own expectations.

market participants expecting? And if then that's being presented very effectually to you. So client done does this, the rest of the market says this. To what extent is our expectation of the client valid? Or is there maybe some sort of bias in what the client is doing?	

6.9 Interview 8A

Senior financial auditor **Semi-structured** **Round 2**	**Labels/interpretation**
Q: Starting from the first challenge, do you agree that the available and potential large data volume is one of the challenges in the financial audit process and could you explain why or why not? Yeah, I think it is because we receive a lot of data and sometimes well, we do not use all the data, simply because the fact that we do not have the time to verify all the data, to check all the data, and maybe we leave some data out which could be relevant for audit procedures.	Large data volume is a challenge, not all data available is exploited.
Q: Can you make an example of where you see that there is available data but you, as auditors, are unable to go through all of it? Well, yeah, let me check. It's difficult to come up with like an example. we are we are performing data analytics on the for, for pension funds. On the provision. Well, we receive like a lot of data. But we do like basic checks. It can be like enhanced. We do not use like, all the data, it's just several variables we are using.	Data usage can be enhanced.
Q; And what are some variables that you think could actually be analyzed or used, but there's just no time to do it? Well, maybe that it's going to be used like to, to provide like more insights into the into the data. Maybe like because we do like the procedures every year maybe we now do like a year to a year comparison, but if, well, we retrieve the data every year, so you can make like a comparison with like well, five years in time. So maybe you can see like more like developments or changes maybe it's maybe to make a more in depth like analysis. Instead of looking only to prior year.	Comparing current data to previous year and in general providing more insights is something currently lacking in the FA process.
Q: Another challenge that came out last time and also in other interviews and literature was the amount of manual procedures. Can you give me one example that you can think of, of manual and repetitive procedure? Yeah, that was something I mentioned last time as well for the reconciliation that we perform. The last two or three years I was part of a group audit team and we have to do the reconciliation for all the reporting packs, which we received during half year and during year end. So it's two times a year we had to do the same procedures and it's all the data is coming from one source. But the format it's provided us in changes. So they select several parameters and they get a different type of report but they still use the same data.	Reconciliation is a challenging manual procedure.
Q: Okay. Can you guide me through this process step by step without going too much into detail? Well, we receive a reporting pack which we send out to all the several component teams that we have. And they need to audit their numbers in their financial	

statements. Once they have ordered the numbers, they report the numbers back to us. So we receive an Excel based file, but they report it obviously in reporting packs, which is like a PDF, so that format we need to reconcile to our overall data file, which contains all the numbers for all our overall financial statements.	
Q: So if I understood correctly, you have two data files in this type of reconciliation or three? we have just one because it's like a group audit. So all the several component teams, they need to report to us for the numbers that are their, like, responsibility, which they are required to audit. So they send through *audit platform* their audited numbers which is in a PDF format and we have to reconcile those numbers to our overall Excel. So that's the whole reporting package, which is input for the financial statements, the group financial statements.	
Q: And is the way you do it right now all manual? Yeah, it's manual. We try to use like the *tool name*, which I explained last time. But that's not all, that's not always that convenient to use. Because the data, it's just such a large file. And they have several reporting packs of; it is an Excel with tons of sheets, worksheets. So yeah, it's not I'm not very happy to work with those documents.	Manual procedures.
Q: I've read and heard that that many decisions are left to the professional judgment of experts. You made the example of the sufficiency of evidence. Some auditor would be okay with a certain amount of audit evidence, but some other auditors would like more specific evidence, like emails you mentioned. Would you agree that this is a challenge? Yes, it is. Because we have one methodology, but then still we have several different types of people who have different types of procedures in a similar way, given the methodology we use. But the procedure are not carried out evenly amongst different teams. So it also depends on like the manager, or the senior manager or the partner even what they think should be required or at least documented. It, indeed, should all be a similar, but sometimes people will say, that's fine for me. It satisfies what I expected. According to the procedures that need to be performed. But sometimes, yeah, there's like, it differentiates. It depends on the people that you're working with.	Professional judgment procedures are not carried out evenly among auditors, as the audit methodology can be interpreted differently.
Q: All right. And can you think of any other examples in which this professional judgment is used? No, not at the moment.	
Q: What can you say about fraud detection, for example? Like trying to understand whether some fraud has been perpetrated. Is that also left to professional judgment? Yeah, that's also something for professional judgment. But it in the end, it's not the purpose of our work to detect fraud, of course, what I mentioned last time as well. In the end you hope that the client when they communicate it with at least with us.	
Q: Now we can move on to the specific deep learning questions. Deep learning has many capabilities and applications. It's a vast technology, so I collected a few of the applications that are most connected or could be most connected to the financial audit. The first one is the text understanding capability. Text understanding refers to the ability of extracting	

information, create patterns automatically from text. Review the text, find topics, keywords, all done automatically. But also classify text and classify documents. For example, having a group of documents and divide it into classes or flag something as an outlier. But for example, if you want to use use it, is it, can it be used for like minutes or something from a board?	
Q: Can you make an example of what you mean? Well, minutes as from for an audit, we need to like scan through the minutes of meetings from a board or for pension funds we have to do it as well. And then we just scan the minutes on things that are interesting for us to know. Maybe some legal cases, maybe some fraud. That can be well discussed in there as well. So maybe that's something to scan those documents. It's maybe easier to find some kind of pattern, or maybe the tone of the meeting.	Use case for DL text understanding: scanning minutes of meetings
Q: Also, you can perform sentiment analysis. Of course deep learning would give an idea. It can take topics and keywords or flag something as having a negative tone. And then the auditor will have to focus more on that. For example, if you perform sentiment analysis and you see that something has a negative tone, then you go to check it and then you can find out that that could be a risky areaWwould you see possible applications in financial audits in general? Yeah. I think, I think that's the example I mentioned should be like something there where it can be used. Because yesterday I talked to a colleague and he was trying to use like the *EY tool* to see if it was possible to make a summary of the minutes of meetings. Because it takes a lot of work if you have to read through all the minutes of the meetings, they have a meeting every month and the minutes are 20 pages. Well, it takes like a day to scan through all the, all the documents. My colleage tried to because he had to fill in some key data. But then he gets a reply that it was not able to process his request. So it was not working yet. he tried to do it with like the minutes of the meeting from from the board, but it was like a reply from that it had not been sufficient capabilities to make a summary of the content.	Scanning minutes of meetings is time consuming.
Q: Another important deep learning capability is visual recognition. This gives the deep learning model the ability to automatically analyze images and videos, but let's leave videos aside for the financial purpose. And this gives the ability to extract insights from the images. Examples are scanning documents or another one is checking for signature approvals that a certain control has been performed, verifying that the signature is original and it's not counterfeit. Can you think of any area in financial audit where this could be useful? Yeah, I think what you mentioned already for control testing I've done a lot of control testing which we have to go through checklist to verify where the documents have been signed by the people who are authorized to sign. So I think that's something that could be very helpful.	Use case detection: control testing, verifying the documents have been signed by who is authorized.
Q: Is this control done manually, checking the signatures? **Ye**s.	

Q: Another important deep learning capability is speech recognition or voice recognition. And this gives the ability, for example, to transform automatically speech to text as well as to perform sentiment analysis. For example, to flag some conversation as deceptive and, if it is flagged as deceptive, that could be an interesting area to check, because maybe some or something that shouldn't be going on is actually going on. So would you see any benefit or possible application in financial audit? Yeah, it could be a benefit if you have meetings with the CFO or something. But then again, I'm wondering how that would work. Because if you have one conversation and the tone is like very happy or something, I don't know. And you tell about the results whether are really poor. Or you've detected fraud or something and you tell it with like a really happy tone or something, I don't know. Maybe if you interview like a CFO or something, you have to do like several interviews to maybe get a Pick up some kind of tone or something that you can say, he's lying or it's not telling the truth or he's hiding some information.	The benefit of speech recognition is confirmed, but skepticism is shown regarding the application of sentiment analysis. However, the skepticism is probably to be addresses to the knowledge gap regarding the workings of speech recognition. At the time, due to interview-time-constraints, I could not spend too long explaining each capability, indeed showing a limitation to this approach, leading to the final choice of integrating a qualitative survey.
Q: It has to be trained. Yes. The model has to be trained with a lot of data. If it's possible and if it can be trained, well, it can be a valuable addition, if you have like an interview, you have to pay attention to what the other one is saying. And maybe you're preparing your questions or you have like at least what you want to ask, you focus on the questions and maybe not so much on the answers. So maybe it's good to have like an addition to analyze the meeting.	Speech automatic processing benefit is recognized, as long as there are no doubts on how the model has been trained.
Q:. And the last one of the important deep learning capabilities is the ability to provide recommendations. This is done through the deep learning ability to extract features from any type of data, documents images, voice recordings. Deep learning can provide recommendations. In order to give you an example, if you have data from the past engagements with similar companies, similar clients, or similar types of engagements, then deep learning could analyze all the types of controls and all the types of evidence that have been collected. And at that point you can have a comparison and see that last year you collected also this evidence or you analyzed this type of documentation. So, this is the function of judgment support, to give something more that you can base your decisions on; professional judgment would still be applied. So, considering this capability, would you see deep learning applications a possibility in order to address the professional judgment issue that we mentioned before? Yeah, it could also challenge, for example,	Risk of previous year's engagement information being outdated.

pension funds. It doesn't matter what kind of client you're working on. The less experienced colleagues, they just look at what we have done prior year and then they're just like roll forward it for like this year. But then again we need to ask ourselves are these procedures only applicable to like last year, or is it also applicable for this year? Is it sufficient enough or do we need to do more? It's also sometimes a bit of a risk to well only look at what we did like last year because maybe for this year it's not not correct to perform the, those procedures anymore or it's not even applicable.	
Q: What if you can compare not only last year but previous engagements in general or similar engagements? Yeah, that would be helpful to then we'll see what kind of what kind of evidence you need to request from like a client, for example. So maybe it can be helpful if you do like an initial audit if you have like for every type of client you have like well, sort of example from all the data that you are required to ask from your client. Well, to make our profession maybe a little bit more interesting to work in because we are quite used to stick to the old habit. So it's really interesting to know something about what can be done. And hopefully it will be done eventually. but the regulations and all the reporting standards that we have to comply with, that makes it a little bit difficult, I guess, because they have to adjust the standards then as well.	DL can be accepted as judgment support, by collecting and analyzing data form similar engagements. Audit profession is too traditional, there is hope it will be made more interesting. However, regulations present an important constraint nowadays. This is to be interpreted as FA procedures being filled with several repetitive manual procedures, linking it to what interviewee 7G mentioned, by referencing “uninteresting procedures”.

6.10 Interview 9H

Partner, experience in financial audit and financial support in assurance engagements **Semi-structured** **Round 2**	**Labels/interpretation**
Q: Do you agree that the available and potential large data volume is one of the challenges in the financial audit process? I think it is currently. We see huge data volumes coming by. Obviously, some types of transactions are less common than others. But if you, for example, take a a bank that processes payments throughout the day of its customers. Yeah, that goes into the millions of transactions per day resulting in maybe even a higher amount of money. Transactions through all kinds of other systems to register all those transactions. So it's not the actual payment that I do at *supermarket*, but it has an effect in multiple systems. Ending up in the general ledger, but also the transaction on my	Millions of transactions per day can be received. Testing entire populations provides more knowledge, but there is a huge risk to increase the workload of auditors.

bank account. So that's another system. So then it's a transaction to the mobile app. So these are all individual, separate data transactions resulting from this one action that I did. So it's often one transaction doesn't remain one transaction throughout a organization's IT infrastructure. Maybe the technical challenges are less because I think nowadays IT infrastructure components can handle huge amounts of transactions. Obviously there will be some limitations at some point and we need to figure out some new technology to address this. But I don't think currently the technology is a bottleneck. I do think there are challenges in the way on how to analyze the data. If you get like millions of transactions for one day, and you're about to analyze all of those transactions if you find outliers to your presumed process or transaction scheme, then you would need to analyze all of those outliers. So, there is a huge potential for a huge amount of additional work that someone needs to do. And that someone is then the audit team to understand, in order to form expectations, find outliers from these expectations and check whether they are false positives potentially and why. There can be all kinds of. Of situations where the huge amount of data is almost a negative into comparison to selecting whatever 25 transactions for a period of time to analyze if everything went okay. So on the one hand, it brings a lot of interesting knowledge but you need to be able to handle and process. So not technically process, but more in a material way, process the data and to understand what is actually happening.	
Q: And how is the process of understand and analyze this huge data set currently addressed? So as far as I know, it is addressed that currently our methodology does not allow to use, for example, full populations. What I understand is one of the main reasons is because of the expectation that there can be a huge number of outliers that one then needs to analyze, which is potentially significantly more time than just validating.	
Q: Would you say that picking just few items and ignoring all the rest is a risk? Yeah. So in doing our work, we're not providing a hundred percent assurance, right? We, we will never do a hundred percent. And the whole thinking of select a number is based on all kinds of assumptions, calculations that we are using, well, as long basically as I can remember. And I would assume that a lot of people have given this a lot of thought. Because it's not only us in doing so, right, it's all the other practices and basically our method, the methodology that allows it international standards that focus on it like that. So the exact details on why a certain number I just don't know. But with doing a sample testing. we are able to provide reasonable assurance that something is okay. So, yes, there is always that risk, and that's why we call it reasonable assurance and not complete assurance.	Auditors are not meant to provide reasonable assurance. Sample testing brings risk, but it align with the audit objective of providing reasonable assurance that a financial statement is free of material misstatement. Interpretation: from this answer, the role of samples is justified. One possible interpretation would be that testing more samples is not an

	auditor's main pressing concern.
Q: For example, if population testing is applied and a big number of outliers actually is discovered, what if the sample of items is chosen from those outliers? Would you say that the audit quality quality could increase? So you say, let's assume you have a million transactions and out of those million, you see a hundred outliers, And you're, you're saying if you then select 25 from those hundreds, but then you will have 25 outliers, right? So basically, you're then confirming that your entire control is ineffective. So it doesn't help. If because if you expect that, let's say for all invoices above a thousand euros, two people need to sign off electronically in a system, right? Let's say that you want to test if that happened. You're going into the system, you see that there are a million popular million transactions for invoices above 1,000 euros. And then you pull the data and you see only 900,000 have a sign off by two people and the other 100, 000 don't. So you can't just unsee that. I say, technically you could, but that would not be ethical. And really against our internal policies of doing a proper job. So, if you see something being wrong you cannot unsee that. Okay. So you need to, you need to then investigate.	What auditors uncover must be investigated, as not doing so would go against ethical requirements.
Q: And still regarding the large data volume, but now focusing the potential one, for example, in the first phase of an audit when you need to plan the audit or decide whether you're accepting or not the engagement, I've read that it's important to learn about the client the environment and get as much information as you can. Nowadays there's a lot of information on the internet, social media, news articles, but I've read that this data extraction and analysis is performed manually. So, it has some limitations because of course a human can just extract a limited amount of information. Would you see this as a missed opportunity or as a challenge as of now? So the data we extract to be used in a audit is usually not data from social media. The data we extract in large volumes comes from the client systems, so their internal processing systems, the data that we use. For example on social media, the internet, et cetera, et cetera, that is used in a different process, which is called the engagement or client acceptance process. If we have a new client that says, Hey, I would like you to do our check on our audit on our financial statements. Then we see what kind of client this is. Is this a client that that does well? And not financially well, but that is sound, that is not dealing in arms or not dealing in, whatever in in all kinds of complex structures. So we're basically assessing the risk of this client. It could also be that the owner of the client is a known criminal, right? Or a politically exposed person. And in those cases, We will do further due diligence on those clients and or persons to evaluate the risk if we were going to service that client. And it can be that we're, that we're saying, well, we're not going to provide this client any services. For example, if Vladimir Putin would call EY and say, hey, I would like to use one of your services. Yeah, I'm very positive that the alarm bells will go off in each and every office	

around the world. And we will happily decline to provide any services. Obviously not everyone is as known as, as him. So we need to do very, very large due diligence exercises, but just so that data is used in a totally different process, right? The data that we get from the client is the actual transaction data, and we perform all kinds of procedures to validate that we get all the data. And that we get the correct data, right? We do procedures to verify the completeness and accuracy.	
Q: Okay. How is this performed right now to address the correctness and completeness of the data received? So we do all kinds of procedures. But basically, it very simply, it comes down to understanding how the client manages its systems and what is the process to extract the data. And is it, for example, that a person from the IT team or so goes into the database, says some types of script to the database, and then gets something back. and then maybe there are like hash totals on it or like a row count or whatever. And we use those kinds of elements to validate, Hey, on screen or on the data that we got, we got 1,000,001 rows of transactions. Then we go into the system and see, hey, do a calculation there automatically. Obviously not by hand. That would not be possible. But through other means to validate that we get the complete list and then with hash totals, we can validate the correctness of it. That is one way nowadays where, especially for the larger IT systems, we have yeah, let's call them connectors that have been certified from our side from our side to be used that are extracting. The complete set of data and also accurately. So like API's. Obviously there are, there is one side to this. So if you imagine that you are for example, processing invoices, there's someone on the finance team that basically process the invoices and then at some point in time it's in the system and it can be processed incorrectly. So maybe one should have been a two. But we are extracting that two. Which still means that our extraction process is correct and complete because, yeah, we see that too and we copy that too, basically. But then the initiation and processing of the transaction itself is not appropriate. But that is, that is something that that we cover off with other procedures. All right. So with them, when we look at the at the actual process from initiation to reporting, those are different procedures, right? And the data can validate it. For example, again, if you have like the invoice data then you can see, well, maybe whatever you pull data from the invoice system and you pull data from the payments system. Right. And then potentially you could say, Hey, I've got this invoice number here and on my payment system, there's always this invoice number, right? Yes. So then you get max on invoice number, which should be identical of a unique, sorry. And then pros and then validate. Hey, I've got 10 euros here, 11 euros here. That is strange, right? So you can also validate transactions. the processing of transactions using data from different systems.	Reconciliation explanation.
Q: Okay. Would you say this is part of the so called reconciliation process? Audit as of know Yeah. Or reconciliation could also	

be that they're expecting and testing controls in the the actual process.	
Q: It's an example of one of the processes that present challenges or complexities. From the literature and from previous interviews I found that one of the issues is that many decisions are left to the so called professional judgment of auditors. Of course, this is part of the auditing procedure and it will always be there. However, sometimes this can lead to some inconsistencies between the audit quality of one audit and another audit if they are led by different people who are following different standards. For example identifying the sufficiency of evidence, I've heard that someone can say that the evidence itself is sufficient. Someone else wants also the email confirmations from who sent the evidence. And another area where I read and heard that this could have been an issue is also the fraud detection, which is very much left to the professional judgment of experts. Would you agree that this could be a challenge or a procedure whose quality could be enhanced? Yes, it is. Yeah. It could be. And to cover that potential quality issue, we have our default, like, someone prepares it, someone reviews it. But still there is a level of, indeed, professional skepticism and professional knowledge and, and judgment in there. Sometimes led by bias or knowledge that you had from prior year. And a third person might not have that knowledge from prior year. And then, yeah, have more difficulty in understanding the evaluation. Why the audit team thought that the for example, not adding more evidence is sufficient. So, there are a couple of things that are super important, is documenting everything that you have. And don't assume anything and then having that validated by at least one more senior person. Processes are in place, which are called the engagement quality review which you can request. And in some cases they're mandatory to have. And one of the objectives of the [] is to review those areas in an audit that have a high or higher level of professional judgment in them.	Professional judgment is addressed with review procedures and engagement quality reviews, however a level of bias or knowledge imbalance could introduce risks and inconsistencies.
Q: [Brief explanation of DL capabilities in test understanding] Would you think that there is space for deep learning application in financial audit considering this capability in order to address some challenges? I think it can definitely be a very interesting input into a financial audit in many ways. And I think what, at least what I know is that they are using some of these examples that you're providing. That they're developing, So yes, this is the very short answer. The challenge always is how is the model doing the exact same thing today as it is going to do tomorrow or the week after, right? So if you feed it 20 reports from management, from different organizations or from different, even geographies, right? Will the language model, then learn from that. And, and let's say the one is even more negative than the other. So. Will the bar of what is negative change of that language model, right? Well, regarding the sentiment analysis, what is considered negative in one analysis is also considered negative in another analysis. That's a legitimate question. It depends on which data is used to train the	DL's text capabilities, some already in use at EY, can provide benefits in FA procedures. However, challenges regard understanding whether the model behaving consistently.

model. If the same data is used. then the bar will not change. The problem, however, now could be different. Like can this amount of data actually be collected from different engagements from different countries? That could be the issue, but if there's a lot of data used to train the model and the data is diverse, then whether the model is applied to an engagement or another engagement, the bar theoretically shouldn't change, but the training data needs to be so specific, I guess, and not that I'm an expert on this, not at all, but, so that the bias that the model can have is huge, and I can imagine that, for example a management statement from a organization in Japan is totally different than from a organization in Germany, right? And the model needs to understand that a Japanese language is different or the sentiment is different than German, for example. But yeah, I think you're absolutely right. Without being an expert on AI and language models.	
Q: Moving to another capability. Another important deep learning capability is speech recognition. Now I think that not many recordings are actually used in financial audit as of now. I know that we are now starting to use voice recording to provide review notes or questions, but that is more like instead of me typing into my senior manager in *audit platform* saying, "Hey, please have a look at my review comments in this Word document." I will just hit a record button and say, Hey, dear senior manager, please have a look at the review comments for this document. So, I don't think we're using speech recognition in anything else. And again, I'm not aware of it.	How speech recognition capabilities of DL are being introduced now.
Q: Considering the capability of deep learning to, for example, transform speech to text or make a review again, perform sentiment analysis without considering all the bias that we just mentioned or finding keywords mainly just deep learning can make use of the recording. So increasing the evidence in an automatic way. So without increasing so much the workload of auditors as of now, would you see a possibility of application in order to improve the audit quality? Thinking of a scenario where we could apply that and that could basically be when we are conversations with our clients to obtain a understanding of their processes, then a person or group of people will talk to our work teams to basically telling the process from cradle to grave for some processes that can be very lengthy. And currently we are using like whatever, a staff member and then a senior staff. And then they all try to make notes as much as possible. And then sometimes they make a flow chart diagram. Then they provide that information back to the client saying, Hey, did we understand this correctly? You could potentially think of running a voice recognition module thing to make a transcript of what the interviewee was saying. And having that automatically transformed into a process diagram or so. Yes, definitely. That processes all the all that has been said in a certain interview including like the the questions or follow up questions from the, from the interviewer. to make a diagram. I think that technically that should be possible. So that is an application of this that I could see happening.	Use case detected: DL-based speech recognition in tests of controls, during oral inquiries.

Q: [Brief introduction on computer vision]Would you see an opportunity in general to apply this? Yeah, very much.	
Q: Can you think of an example? Yeah, well, you gave a perfect example, right? We already have tooling where structured data elements are validated, against an expectation. So we trained the model to say, okay, this this element over here on the top left corner is whatever the date and this element on the lower right side is the name of the approver. And then we can just feed it like whatever PDFs tables and then do the analysis, right? Should be possible. I think that should be possible and it can be very beneficial, because it's boring that a, a person who has been to university. is super intelligent that she or he looks at a screen with a signature and then validates that with a list of signatures that she or he got from the client. Right. Yeah. It's just like my 10-year-old daughter can do that as well. You don't have to go to university to do so. If we can use technology to do that appropriately again, the challenge will be how to make sure that the model does that appropriately. Hey, this is the signature, and it needs to be, right. Always a signature from the head of accounting or whatever. I don't think that is difficult, but I think the challenge is more on the methodology side so that we can allow us ourselves to rely on the model.I think that's the, the bigger challenge. Rather than creating the technology to do so.	Benefit recognized in application of DL-based automation, to solve the challenge of manual and uninteresting procedures.
Q: [Recommendation systems introduction].Would you see a benefit or a way to apply deep learning thinking about this functionality? Yeah, I think so. That would be and I see potential usage of that in many phases, so many steps of our entire audit process from client acceptance all the way through like reporting, right? Because we have so many smart people around the world that are doing similar or exactly the same things. So why not learn from them? And those people can be like in Houston the Americas and the other can be in Milan and the other one can be in in Amsterdam. So you're not sitting next to each other. So, the deep learning can then say, Hey, you are on a, whatever, car manufacturing engagement, and all of your peer engagements over the last, whatever, three years had this revenue recognition as a fraud risk, for example. Suggest that you do so as well, or. Full stop basically just giving you the information and then you could say, well, do you want me to include this in your campus file, including all the required procedures or the suggested procedures that were performed by all the other peer organizations? that the audit team can select. Oh yeah, good idea. And then I need this one and I need this one and I need this one and I need this one to add to my procedure list and then I'm done. And we were even doing, thinking about this a few months back for our, our own little IT audit thing. And the example there that we were trying to use. is that people type in the name of the application that the client is using. And then AI could see, Hey, are you, do you mean SAP version one, two, three, or do you mean SAP version two, six, eight, and then the team can select, Oh, well, two, six, eight. And then the model can say, Oh, Hey, if you, if this is the application that these are	Recommendation systems can potentially used in several phases of an audit, from client acceptance through reporting. “We have so many smart people around the world that are doing similar or exactly the same things. So why not learn from them?”.

like the presumed risks. These are the pre presumed or these are then the controls that can be in place. These are definitely in place because they're like standard controls for this application and this is the way on how to validate those controls. Alright, so you can basically have the model create the entire flow of activities based on only the application name.	

6.11 Interview 10E

Senior manager experienced in leveraging AI solutions in FA process **Semi-structured** **Round 2**	**Labels/interpretations**
Q: What are the main deep learning capabilities in the financial audit domain? I think in financial, in financial audit, it's mainly natural language processing. You do need to some extent computer vision in the sense of optical character recognition, OCR. So whenever you have content that first needs to be made machine readable to be able to apply natural language processing. Then you need some computer vision. But the main focus is on natural language processing and generation.	In FA context, NLP and computer vision are the most employed.
Q: And what do you think about speech recognition capabilities of deep learning? Are they applied as of now, or if they are not, do you see a potential in speech recognition capabilities in financial audit? What do you mean by speech recognition? Like audio signals, like we are talking now and then transferring? **Q: Yes.** Yes, in the sense, if you have recordings of conversations and they are relevant to the financial audit, then obviously speech recognition would be nice to again documentor transfer it into written text. So for instance, a feature of copilot in teams is that you have this automatic system to do meeting notes and memos generated by AI. A feature that's not yet available for auditors, obviously. But if, let's say, clients use this functionality significantly, then it's obviously relevant for the audit. Or if they're not using it yet, but they have extensive video tapings, audio tapings of meetings and conversations, and it's not on paper yet, then I can see speech to text because that's basically speech to text. Another application area is just to improve user experience for the auditors. So again, if the auditor wants to interact with a automation system, computer system, and they want to use speech to do so. So for instance, for *EY tool*, we're currently thinking about implementing also like speech to text functionality. So you can just push your record button on your phone and ask *EY tool* a question without having to type it down. That's just like an additional efficiency game. But beyond that, not so much because that's mainly the scope.	Speech recognition function can be a good addition to FAs if it is relevant to the audit, meaning if recording are available. Additionally, it can be implemented for the purpose of improving the auditor experience, by enabling them to ask a question using their voice and having it automatically transcribed.
Q: All right. And connected to both speech recognition and all the other capabilities of deep learning, as well as NLP applications. What do you think about applying sentiment analysis in financial audit? I think it's a good tool to give the	Sentiment analysis can be used to evaluate relevant news articles,

auditor a better understanding or impression about the vast amount of information. So I'm thinking about, let's say [audit platform] news application that we have, which basically does the same. So it goes into our trusted news feeds, evaluates the individual news articles that are online or wherever about relevancy to your specific engagement, and also gives a sentiment analysis in the sense of is the majority positive negative? But that's just for more application areas in the sense of understand the business. So just provide more insights to the auditor. I don't see yet proper application area beyond sentiment analysis for the audit. So I wouldn't trust it beyond this specific use case.	useful for understanding the business.
Q: Okay, but this specific use case, is it already in use? Yes, we do have that in *audit platform* is basically evaluating more scoring articles about relevancy. It's also scoring them about whether they're positive or negative in nature, like binary or neutral. We have some projects that are in development that I've heard about from other ares who want to go into social media. So Twitter x, Instagram, all those social media accounts to do basically the same. So count the number of tweets regarding your client and then do some general business intelligence analytics. So frequency of posts, sentiment of posts, changes in sentiment and frequency most frequently terminology. So like those word clouds, like what kind of terms have been used	Social media messages can also be enhanced.
Q: We talked about some deep learning capabilities, or we could call also applications such as computer vision, NLP and so on. Can you think of other capabilities, for example, dimensionality reduction or feature extraction? Can you mention other capabilities that could actually be used or are already trying to be used in financial audit? The general application area where deep learning could also be applied to is definitely in scope. So if you're going like with the [tool name], you want to have as much information as possible about, let's say a retailer and then to see patterns on what impacts, for instance, revenue. And is it like geographic factors, is it store size factors, is it inventory that you're selling? Is there an interaction between those factors and how does it affect, let's say, the revenue of individual stores? Are there any outliers then across stores? Is the strategy maybe not working in, let's say, Canada when it works in the US, because the stores in Canada are generally smaller with more staff or whatever. And to go into basically this multi factor analysis, if you don't have those tools, then basically it's very difficult for a human to take into account more than two or three variables at most. But with the time series regression analyzer, you can go into much more complex patterns and to find outliers. And then where the deep learning comes in, which is not yet in application with us, but in the discussion is to broaden the data sources for this kind of analysis, because if you go into the purely statistical realm, as we have now, then it's structured	DL's main application area is in scope. The goal is to exploit data to extract as much important information as possible. With analytics and DL tools, more variables can be taken into account and compared, to understand and uncover relationships, causalities. Without these tools, it would be very difficult for humans to uncover such patterns. DL plays a role when such analysis is performed on data that can be structured, semi-structured, or unstructured.

data number crunching. But if you want to keep add to this analysis, let's say, to a time series analysis or something, information that is more unstructured in nature. So we're going about tweets about news articles, about contracts or something. Then you go, you end up in the deep learning domain again, because you're mixing the structured information that you get, let's say, from an ERP system with, let's say, unstructured information that you collect from a different data source, like the open Internet or some trusted information provider.	
Q: And regarding the dimensionality reduction, I found that in a book and they were describing that there's this capability of deep learning to reduce the dimensionality of data. And that recommendation systems were the ones that benefited the most from dimensionality reduction. So my question is, is this recommendation given through dimensionality reduction or through any other type of deep learning capability or application? I think that currently in financial audit, we're still at the stage of expanding information sources to just give way more information to the auditor than was previously feasibly accessible to them. Because again, an auditor cannot spend hours and hours sifting through data or through the Internet to find the relevant information. However, the dimensionality reductions then is basically a logical, immediate consequence. And that meaning, again, if you're thinking about the *EY tool*, there is a lot of dimensions or factors that we're using for the search. So it's not just look for the name of our client, but also industry, sub industry, geography, products, maybe go along products to find information that are relevant. So that's all dimensions. But then we're reducing all of this into basically one factor where we say, how relevant is this? Like a relevancy score? Because we're basically saying, okay, we're opening up the sources, but it's way too big for an auditor to understand why the whole scope. So we do need dimensionality reduction to some KPI's, to some meta score that is more palpable and more like the auditor can handle much better. And hat's like the relevancy score of zero to 100. They don't really need to know. Like, is it, has it the relevancy score of 95 because the company was mentioned or a product was mentioned or it's from the same industry sector or something like that, because all of that plays a role. That's the dimensionality reduction. Like all of it contributes to the score. But the auditor themselves, they don't really need to look into each and every one of those anymore. So yes, whenever we create new meta scores, we do dimensionality reduction, but that's mostly, again in information sourcing. So it's again about collecting information about the clients, collecting information about the industry sector, about the peers, about the competitors, about understanding the world the client is basically living in. And if we go into more internal projects like dimensionality reduction in terms of product ranges, that is usually already done by the client, you	Dimensionality reduction is the ability to detect various dimensions and highlight the relevant one. For instance, DL searching for similar engagement will use not only the name of the client, but various dimensions, such as industry, sub industry, products, geography, etc. Then, it will present the similar engagements with a percentage score, such as stating: this engagement has 95% relevancy, therefore showing one dimension only. This is because the rest of dimensions do not add value to the end users. This can be interpreted by stating that DL provides value to recommendation systems as it is capable to analyze many dimensions, but still present the most relevant ones, ensuring easy use.

don't need to do it by yourself because the client themselves, they also somehow need to keep an oversight over their own products, for instance. And if they create 2000 different products, then they will aggregate them into product categories, product clusters, et cetera. And we usually rely on that from the client side. So it's more about external information that is not already pre processed in a sense.	
Q: Okay, but if I understand correctly from your explanation, basically dimensionality reduction is used next to some other deep learning capability, for example, next to natural language processing. Do you generally use deep learning capabilities in combination with each other? Yes, yes, because I mean, the individual capabilities are not so much relevant. Like the impact is not so large. So for instance, if you go into computer vision, if you go only into the base capability, then it can make a scanned document machine readable. That means when you could, in a scanned document where you were not able to mark the text and copy it to somewhere else, now you can, that's computer vision, OCR. And that's a very limited scope. But if you combine it with natural language processing and tell it, okay, now make this document readable and then in the next step identify the key parameters that are relevant for me and give them to me in a table, then it's a little bit more useful. And so you definitely stack deep learning capabilities in the most meaningful way for different use cases and each individually to themselves. Like natural language processing is a very broad term. So sentiment analysis would be part of it. But again you would say like, okay, first transform the data source, then do the sentiment analysis. But then the next step is also like, just because you gave me for each individual tweet that it's positive or negative, it's not super useful. Until I have an overview, until I have a dashboard, until I see, oh, there's a trend, there is a change, and maybe you should connect. Like, let's say positive reviews went up and up and up, and then suddenly they dropped the real value. That is like, can you identify the single event that made it drop? And then if you can connect it to a certain newsarticle, maybe they went public, IPO or something like that. If you can connect different data sources, and that's where you need to combine different deep learning capabilities, that's where the real value is. But yeah, the individual small capabilities are all the foundation. Without those, you cannot build on top.	Without individual capabilities, you cannot build on top, but they represent the foundations. Individual capabilities alone are not of much value. Stacking DL's capabilities together increases the usefulness of the model.
Q: The DL in FA framework that I found builds on three capabilities only which are text understanding, speech recognition and computer vision. And from these three capabilities, applying that to audit analytics, but augmented with deep learning, they find that deep learning basically has two functions, which are judgment support and information extraction. Would you agree with these? If yes or no, why? And would you say that there are more functions	DL is able to generate information. A detected use case is drafting reports at the end of the audit, based on all the documentation.

that could be added when we say that deep learning can be applied to financial audit? Yes. So at the current stage or up, let's say up until half a year ago for sure, because judgment support is basically information extraction, because you support the judgment by providing more information in a digestible manner. So it's kind of connected and it's definitely a big, big deal. I mean, I told you, like the news article thing, comparing even the time series analyzer, it's always just making more information digestible and supporting the professional judgment. But with generative AI, we kind of broaden the scope now in the sense that we can all, we can move into output production as well. In the sense of whatever I have done in my audit, I need to document in a form, in a predefined form based on my documentation, I can create a generative AI tool that already makes a draft. So I'm not in front of, front of an empty form. I'm not starting with a k ten form that is completely empty. But I already have key points out of data pre filled. Again, it's professional judgment, it's only supporting because you cannot trust the tool to document perfectly. But it's again a comfort feature and a time saving feature. And it also kind of adds to a standardization process because no two auditors would fill out the same form the same way. But if the draft that is auto generated from your data is to somewhat standardized through the AI, then you get a certain level of quality assurance, because a certain standard is already, like, pre given. It's very rare that a user would say, oh, I don't like this proposition. Deletes it all and starts from scratch. So you can definitely go more into, like, not just into decision support and information extraction, but also in the direction of information generation. But again, again, in a support function. There's potential, and it's definitely gonna be used. So the framework is fine. I would just expand it that we're now also going into information generation.	
Q: If you think of the challenges of applying deep learning in financial audit as of now, what would you mention as the most important challenges? I mean, the biggest challenge still now is data availability for clients. So if we have an awesome new tool that makes things easier, but we require extensive data from the client that we did not need before. Yeah, we need to get the client provide this data and to approve this data exchange. But the biggest challenge for AI and deep learning in the financial sector is you need a lot of data to train these models. You need to get this data from somewhere. You need the approval from the data sources, probably your clients. They need to be comfortable with i, they need to approve that you're allowed to use the data, they need to provide that data in the first place. So my best example is still the german *name* tool for reconciliation between documents and structured data. It's like in the current business. The audit team says, I need to do a test of details. I do [tool name use] which tells me you need 20samples. Then they go to the client and	DL implementation challenges derive from data scarcity and from clients' hesitation.

say, I need 20 samples. Give me randomly selected entries from the ERP system. And then they manually look at it. Now if we want to apply *name tool* and select we want to test 10,000, then the client needs to be capable and willing to provide you with those10,000. And then they are asking you, why should we do that? Because we're open up, we're opening up basically our engine hood. We're showing you way more that we need to than we needed to show before. What's the benefit for us? And this, this thing is even worse if you're at the stage of training a model, because then you don't need the 10,000, then you need a hundred thousand, a million, and not just, and then you're, because the data is limited, you cannot go to one client and say, give me a million of your invoices, but you need to go to a hundred clients and say, each of you give me 10,000, and then you need to convince 100 clients to give you 10,000. Wherein the audits of the last 30, 50 years there was just an auditor coming in and saying, if we please 20. And this is a process that is limiting deep learning significantly. It's just data scarcity and how to get this data from somewhere. It is. But I mean, the thing is that there's also like change in mentality also with clients because they're doing it within their own companies, they want to use AI internally. And that kind of does the education or the upskilling on the client side because they're like, oh, we had to collect hundreds of thousands of, of documents to be able to enable this internal AI application. It kind of makes sense that if we want UI as our auditor to use AI and be more modern, we need to provide them with such data access as well. Because you cannot ask for one thing but not provide the fuel. So to say this learning is happening just, just by the fact that they are trying to use AI internally by themselves as well. So it's a changing process. It will take a few years more, but it will open up the gates for audit, hopefully. Currently still a big challenge.	
Q: Okay. And you cannot use data from past engagements, that's not enough, right? You can, but if you go into a little bit more complex areas, then this data might be outdated and you're training the system on out. So if you have like this recommendation system that looks into risks and significant accounts that you've identified, if you use data from ten years before, there might have been an update to the accounting standards. There might have been an update to financial accounting standards in those countries in the meantime, and then these recommendations would not be up todate anymore. Then it would recommend to our auditors like consider cash to be a significant account because exceptional. And the auditor thinks like, yeah, but theaccounting standard changes from two years ago don't match with that anymore. You know, like it's not up to date. And because the framework, the limitations, theregulations are changing, we need to stay up to date. If you have a little bit more complex applications than computer vision, making a document readable.	Past client data may be outdated, therefore it is not sufficient to use historical data only.

Q: Could you say which are the main deep learning architectures used for majority of the financial audit applications? Let's make it simplerI just mentioned in the literature review and theoretical framework, the main ones, such as the classes, CNN, RNN's, autoencoders, just the mainones, not without going into detail. You will have to go with mainly recursive ones like RNN's or LSTMs or even more, more sophisticated versions of that. Because you have, if you're working with text, if you're working with natural language, it basically processes word by word, but it's not, it should not forget what the first word of the paragraph was. So you need systems like, like a CNN is always just like piping it through and you don't need to remember what was before. If you classify one picture, is it the wolf or is it the dog? It's irrelevant how you classified five pictures. But if you have a text that is a page long, what was set at the beginning of the page is still relevant to what is set at the end of the page. So with natural language processing, if you go into architectures that kind of have a memory function like an RNN LSTM, and more sophisticated versions of that, that's the main focus that you, you should use. And if you're not dependent so much on what was before, then you can use the other architectures. And then like autoencoders, adverse networks like GANs or transformers, those things go into the generation of stuff that's like the new thing. Like if you want to auto fill a template, then you need a transformer.	RNNs, LSTMs, CNNs, GANs, transformers are the main DL models used or with potential for FA applications.

6.12 Interview 11F

Staff – data science expert **Semi-structured** **Round 2**	**Labels/interpretation**
Q: One of the first challenges that came up several times is the reconciliation process. [reconciliation explanation] Do you think DL could automate this task? You mentioned a few things, right? So the one that caught my eye was how in different reports you'll have the numbers written down different. They have a dash somewhere, they have a comma or something. In these areas, you certainly can use machine learning. You can actually use deep learning as an image classifier, which helps you read that data. So it'll come up with the right answer as to what's actually written there in a more standardized fashion. And then you can use general algorithms like they're using upgraded Excel to then work with that data. You can at least use it to kind of simplify or extract the data, and you can make that happen faster with a machine than you can with a human being. So certainly deep learning has some relevancy in extraction and understandings. Document understanding and data extraction. Then you mentioned a couple other things. It was like the whole reconciliation process, they use different files together. Like, I'm trying to think about what would be relevant for machine learning there. Where deep learning could help you is	Reconciliation can be addressed with DL, such as reading numbers presented in different formats, as well as processing invoices coming in screenshots.

like, let's say you used document understanding, machine learning to understand what it says, and you have all the numbers. You can certainly train a machine learning classifier to kind of figure out whether the way, you know, the money distributes in the reconciliation statements seems fraudulent or not. It can definitely help you, give you something called maybe a confidence rating of some sort, but it would be challenging to train a classifier to do that. Certainly within the realms of possibility. You kind of keep feeding the algorithm lots and lots of reconciliation statements where there is no fraud and you tell the algorithm there's no fraudulent here, recommendation statements which are fraudulent or which do need to be flagged, and you can assign it a class for that. So you have class one being completely fine, and class two can be likelihood of fraud is high, and you can certainly feed all this information to it. They can give you a confidence rating. I wouldn't say that it will do the whole job, but it'll at the very least flag for whatever human operator is there to look into this. Yeah.	
Q: The next challenge is checking the reliability of data. [brief explanation] Would DL be able to address this process? Yeah, yeah. So look, in the same vein as in the previous issue, you can definitely use machine learning and even deep learning to kind of like scan and extract the data, even if it's not structured. Maybe it's written with hand or whatever, but at the same time, so you would basically just use a classifier, some type of character recognition classifier. So it would certainly be in deep learning. The only issue with it is that you can't always trust it to be able to figure it out every single time. So the best case scenario here is that you train a classifier to be able to read all of this data and put it in a structured form and make a report out of it. But you also ask the classifier to give you a confidence rating. And what we can is then that that entity's employees can only look at the data points that were deemed as not a very high confidence rating from the deep learning classifier, and that they can specifically look at those because the classifier won't always enter everything for you correctly. But at the very least, you can ask it to do what it knows for sure is correct and then flag for you the other. So it can kind of reduce load by a good amount if there are more cases than not when it's straightforward.	DL can scan and extract data, also handwritten, and classify it. The auditor will need to check the ones flagged as outliers or with low confidence ratings.
Q: The next audit challenge is presented by the amount of regulations that are in the financial audit realm. There are regulations and this often transforms into checklists [brief explanation and checking understanding is done] Would DL be able to address this procedure? Yeah. Okay. So look, if the work is super repetitive, then, then it has a high potential to be automated. Now I'm gonna have to understand what the activities are in this, like, high volume of work to know whether deep learning or machine learning is relevant. But at the very least, you can certainly, it's a good candidate for automation. Okay. And so, and the challenge is like, what's the main challenge here in this context, is it the fact that there's just too many cases. Like the issue here isn't, is that, is there something specific that's like, I can write down the rules for	The interviewer states the potential of DL of automating this task. However, the knowledge gap with regarding the specific FA procedures, make it difficult to collect a definite answer. Automation is again stressed as being possibly addressed

and will happen over and over again, or is there something specific to look for? Then you can certainly pawn the job away to machine learning. And the fact that you said that there's a very high volume of this work means that there's clearly a very large data set that can be used to train it. So, like that's, that means that there might be a machine learning solution here. But if the task is so general that you have to first look at the task and then do your research looking at the flow text or the numbers, decide which numbers you need to look at and then go to another document. Like it's. Yeah, it's. There's too many. What's the word I'm looking for? It requires it to think more than it can. But like, what I will say is that there's certainly a very high potential for automation in general, and automation doesn't always have to have artificial intelligence. You know what I mean? It could just be some code on a computer, it could just be that simple. And you could probably simplify a lot of the little jobs in the process, but it would be hard to automate the whole thing, even with machine learning.	with simples tools that do not necessarily involve AI.
Q: But wouldn't the simple code with automation without AI be a challenge when we have different types of data that are everything but structured? Yeah, absolutely. That would be a little bit challenging. And in that area, just like in the previous two scenarios we just went through, you can certainly use machine learning to extract data from files or from documents and structure it for you. You can certainly train it to do that. It's not even necessarily like deep learning. It goes more towards generative AI too, but it can do that. It can automate these simpler tasks that take time, but you can't automate the whole process. But you can automate using machine learning, all the little things that you need to do. You just have to decide when they need to be done and click a button and make them do it. The machine won't know which one to do when. But certainly, you can simplify the processes that, the steps that you have to go through to finish the job. So, you can like documenting, I think, or validation of data, etcetera, etcetera. Yeah, okay. But it's not like a new, it's nothing new from what I said earlier, though, if you know what I mean. The other two scenarios.	The interviewer confirms that processing non-structured data would be challenging for simples non-AI based solutions.
Q: All right. In the first phase of an audit, the auditor needs to collects much knowledge of the client as possible in order to understand the company that they are going to be auditing/ [brief explanation of the procedure and the requirements, especially the volume and data types required] What do you think about this procedure to be addressed with DL? Yeah, my first question is that this is pretty complex because how do you teach a machine to go on Internet and think like a human being and figure out what to get relevance? Like, you could certainly create some type of web scraping tool that scrapes for data on the Internet relevant to the firm that they're auditing. And it can probably quickly download a lot of the information that it thinks is useful. But we couldn't analyze the data by itself, and you can't even necessarily trust that we'll have downloaded all the relevant information. But at the very least you could use machine learning to	Automation of data extraction is not a good use case for DL.

start it off, start the process off, do the web scraping, download all the relevant information, extract it however necessary so that it's for the operator. But at some point a human being will have to look at all the data and decide whether it needs to go back on the Internet and look for more information by himself, manually or if. I think that's the main one here. Yeah, so you can do some web space extraction, but you can't be sure it'll always get all the right data.	
Q: There's many procedures that are heavily based on the so called professional judgment of auditors. [brief introduction] So what do you think of deep learning's role of judgment support? Yeah, yeah. I think you summed it perfectly in that he can't do the job of the auditor, but machine learning can certainly set it up. And you would do this in a similar fashion to the first solution that we talked about, but at the very least you're going to need. So say there is seven different types of documents that are relevant to a firm when they operating them. And say, not every firm has to have all seven parameters that are relevant to begin with, but you have to set up the machine learning algorithm so that it has like limited number of types of things that it's going to look at. So let's say you have seven, and what you can do is you can use massive amounts of all the audits that you've already completed. You can feed the specific document in the machine learning algorithm and mark the document as fraudulent. It has to be document type a. So you'll have seven or eight different types of classifiers. Then it wouldn't just be one deep learning algorithm. Each one would be in charge of giving you an opinion on whether it's, whether it has something fraudulent in it or not. I mean, certainly it could be that it's a classifier that cross checks between more than one document, but it has to be cross checking between the same types of documents every time. And what the machine learning can do is it can just use previous opinions that were used, previous professional judgment, and then try to copy that professional judgment, and maybe tell the operator or the auditor that, hey, I flagged this case over here. It has a high likelihood of fraud in it, have a look at it. But. So, yeah, look, that's useful. And it's, yeah, it's probably possible too. And if I had to quickly give you a disadvantage and then an advantage, one disadvantage would be that it would be using professional judgment that was already used by human beings before to begin with. So you don't, you don't get to bypass the bias of professional judgment. The bias has now entered the machine by all the cases that were completed in the past. Now, the advantage that addresses this disadvantage is that your data set is going to have cases from more than one auditor. And when you have cases from lots of different types of professional judgment, then what you tend to get is an averaged out like response to all the judgment. And it kind of does somewhat bypass it in the sense that it's no longer the bias of one person. It's now relevant to the bias of the whole cohort. And the bigger the cohort, the less the bias, if you know what I mean. Yeah, because my bias is less relevant if there's 99 other zip ons who are also being asked question in	

different scenarios or situations or etcetera. So, yeah, it can set it up, but you just have to remember that when the machine learning algorithm does this, it's performing a very specific task. So you don't just give it all the documents and say, hey, let me know if you have. If you think something's fraudulent, you got to give it like, oh, hey, I have document type a and b here, put it in to algorithm a, and then you have algorithm and you put in documents type c and d in it, etc. Etcetera. And you have to pick which algorithm you want to use, depending on which type of documents that you receive, which you can, I guess, also automate to some extent. But, yeah, the point is, is that it's never gonna be a holistic approach. Machine learning is always gonna be a very specific response.	
Q: All right, but when you say machine learning, do you refer to traditional machine learning, or are you also including deep learning? Yeah, I'm including deep learning.	The interviewer refers to DL even when mentioning ML, unless clearly differentiated.
Q: Generally, what are some or one challenge that you could see in applying deep learning in financial audit, apart from the black box that we already discussed last time, and the fact that, probably, as I think you were suggesting, sometimes it seems an over engineered solution? I think the last thing you said is definitely something I was going to touch on, which is a lot of things don't need machine learning. They can just be, you know, simple algorithms, just lines of code, math, logic, and things like that. And a lot of the time, these things will aid you in doing the small things in a much cheaper fashion, I guess. Moving on from there, what's another challenge that someone might have with deep learning? Like, the most obvious challenges are always, do you have enough of a data set, do you have enough of a labeled or annotated data set to teach your algorithm to begin with? Yeah, I guess another one could be. I'm just thinking whether it requires a lot of computation power to use it, you don't really need it. You only need the power when you're training. It's. That's not that big of a deal.	It is important to consider when DL is really needed, because sometimes simpler solutions could do the job. Considering if there is enough data available to train the model is a typical challenge for any AI model.
Q: But just to clarify, when you say we need trust that the machine, like, performs every time in the same way for every client, for example, is that a different thing from a machine being a black box? A little bit. Because even if it's a black box, you can still trust it is giving you good enough results, if you know what I mean. So the black box is certainly the biggest hurdle that we'll have to resolve when it comes to getting people to trust it. But I think that can very easily be circumvented by just telling them how accurate it is, what the accuracy numbers are. Compare those accuracy numbers to when humans do the job, and in that sense, you'll be able to prove to people that this is better. But I also believe that a great way to keep the general public's trust when using deep learning tools is to probably reassure them that there will always be a human at the top that makes the last final call. You don't have to make the call of every single piece of data, otherwise,	Black box can be circumvented with accuracy metrics. It can be a black box, but there can be trust in the model anyways. To provide trust, the interviewer suggests to underline the human role of reviewer of the AI model's output.

you're not solving anything. But, like, when a machine learning algorithm flags to you that these 50 cases might be fraudulent, then a human should look into it before you, you know, accuse them of fraud.	

6.13 Interview 12I

Senior manager – 10 years FA experience and experience in AI solutions applied to FA **Semi-structured** **Round 2**	**Labes/interpretations**
Q: Would you agree that large data volume is a major challenge in financial audit? Why or why not? Yeah, I mean, I guess it's a challenge. I can approach it from two directions, I guess. There is the challenge for the team. We're pushing a lot more for data to be gathered and large data sets to be gathered for each team that's doing client service. So that starts with the general ledger data. And many teams are now able to capture that there have been challenges in the past, in particular, when you get to the largest clients, even being able to extract and transform and consume that volume of data that some of our clients has been a challenge. I think we have some technology that allows us to do that a little bit more effectively, and there's some structures that the firm is setting up to help the individual engagement teams to do that. So specialist teams that have data scientists and all that on there to help effectively get access to and use the data. But it's been a challenge in that regard, I think, and still remains not perfectly solved for. And that's only the first level of data. So general Ledger, the subledgers, for example, have much, much greater volumes or can have much greater volumes of information. So I think the other side of the challenge, not just getting the data, but in learning how to apply it effectively, have different approaches that are audit approaches that rely on the use of data and doing things like correlation analysis to identify anomalies or trends that are unexpected. I think that's sort of the direction that we're all trying to head in, because now that we do finally have made maybe an access or a way to get access to those data sets, there's a realization that looking at the full set of transactions that a company is recording is a much better basis upon which to base our opinion than doing sort of tested details on a sample basis. So, yeah, I think there's challenges. There have been challenges in getting the data. There's also challenges in how do we effectively use the data. And part of that's because, in my opinion, you look at the history of the profession, and it's generally almost pretty consistently been, you know, this sample based test approach. You have an account, you have a set of transactions, and you test some of them to the underlying supporting detail, whether that be invoices or contracts or whatever it is. And now we're sort of faced with this data. Can we do something more comprehensively? Are there ways to use this that sort of change the fundamental approach to an audit? And that's not only something we need to figure out, but it's something that our	The challenge arises from actually getting the data. There are IT solutions partially solving the problem of processing large data volumes. The challenge is using the data more effectively. Finally, regulation compliance is still to be addressed.

regulators need to be comfortable with. And so I think that's sort of the part of the journey that we're on is trying to carve those new approaches and hopefully get to a better result. But we're sort of figuring it out right now, is my sense.	
Q: Would you agree that the amount of manual procedures is now a challenging financial audit process? And again, why or why not? Yeah, the process is highly manual. It's become slightly less manual. We have more tools available to us to automate, for example, some of the procedures for sure. But again, I think that's still part of that journey path that we're on is trying to get to that place where the manual, especially the really repetitive manual work, is taken care of by technology more than people. And people are sort of put more in a review type capacity from the beginning. That's because the model of audit teams from a resourcing and staffing perspective has been, I think, for a long time sort of this pyramid type model. You have a lot of people doing a lot of manual work and then that gets reviewed by the next level up, which then gets reviewed and reviewed several times along that path. But as we continue to automate, I think that it changes the structure a little bit or the model of the team and sort of puts those automation pieces. And I think this is where AI comes in as well in terms of first cut drafting of things might be able to enter that sphere past just like kind of very basic automations that it puts our people more in that reviewer seat as opposed to that preparer seat. But yes, very manual. I mean, it just depends on what tools you have available to you. But theoretically, you know, most of the work is manually completed. I would say at this point still. Okay, large majority.	Manual procedures are being partially addressed for, but they are still challenging. The auditors are still reviewers, but instead of the humans preparing the work, it would be prepared by machines.
Q: You mentioned something that is already being automated right now. Can you make some examples if you can think of one? [confidential information]	
Q: I've read and heard from other interviews that, for example, assessing the sufficiency of evidence is a result of professional judgment. And sometimes someone could say that a certain type of audit evidence is sufficient, whereas someone else could say that more is needed. Do you think that many decisions left to the auditor's professional judgment and that these could lead to inconsistencies? And also, would you identify this as a challenge, and why or why not? Yeah, I think that's an interesting question. I think, at least the way I think of it now, and I'm guessing here, but I'm going to guess the way that our regulators think about it is that it's important for people still to be in the shoes of reviewers and applying their professional judgment to things. It's, you know, I think there's not. I don't think we're to the point yet where people – regulators and the firm -are willing to totally give up the reins. Because it's going to be incredibly contextual to that particular client, maybe to your past knowledge of that client. Like, do you think there's maybe more of a risk based off of what you've seen in the past? And there's just not, you know, to my knowledge. The ability for even deep learning to really apply that at this stage. So I think generally speaking, even as we move d	Auditors need to be in the shoes of reviewers and still apply professional judgment, according to the interviewee and the regulations.

down this path for at least a while, it's going to be pretty important that humans are still applying professional judgment to these things as opposed to kind of giving it over to AI to make the conclusion. So when we think about our products, it's always framed in this sort of, "the AI might be the preparer or might be suggesting things to you, but there's always someone there that reviews because it's just not to the level where we can confidently say it's going to work better than people". And I think maybe there is a point at which that comes into play. But, you know, I'm not sure when that is, I guess. It'll probably It's just going to depend a lot. The challenge is, can you actually measure how accurate a deep learning based suggestion or application is as compared to real people to demonstrate that it's better? That might be a challenge I think.	
Q: The idea would be, according to what I read, to not completely use AI to remove the human in the loop, but for example, to give some recommendations or some basis on which to base a professional judgment. Would you see AI or deep learning to be able to do that? Yes, absolutely. I think in terms of putting forward recommendations or doing an initial draft, that's sort of the place that at least I see it coming into play and it could be extremely powerful. I think the other side of that is kind of interesting. And I don't know what you've seen, but you almost have to kind of combat the human tendency to trust in the machine, trust in the outputs. everyone is or I guess, in position in a way that really makes people critically think about it and not just read it. That's just something that we as like society get better at recognizing as everyone becomes more comfortable with these technologies that you can't just entirely rely on it. But absolutely the main opportunity is for it to be in that recommendation or like initial pass kind of element, bringing surfacing information that maybe you wouldn't have otherwise thought about to your attention and enabling the ability to go back and I guess validate that.	DL is suitable for providing recommendations in FA processes.
Q: Can you think of any AI, or specifically, if you can, deep learning applications right now in the audit procedure? Yeah. There are a couple. So one is you've got this kind of a module within *audit platform*. So effectively, *audit platform* is where all the documentation is stored for the client. Everything exists within this workspace. You sign off on things within there. There is everything that our methodology would say based off of the profile your engagement needs to. So there's a lot of different screens and features and documentation elements within *audit platform*, but also associated with that, there's a new module. It's just called *tool name*. So its is to help with the risk identification process. So early on in the audit, you go through and you say, "here's my understanding of the nature of the business, the nature of the industry, the nature of their context. What are the risks to the financial statements based off of that information and based off of past experience?" And you define specific risks that you address. You always do a certain amount of procedures, but the risks are where you would maybe do incremental, additional things. And so some	Current DL applications are already in use, proving their feasibility in the FA process. These are the recommendation systems that uses DL to compare similar engagements. From this answer, another use case is identified: evaluating accounting estimates, which include uncertainty due to their nature. Also, DL is applied to help for the

common areas are where there's a lot of judgment. So maybe you have an estimate involved, and so they're estimated future event or some provision. And there tends to be uncertainty associated with that since it's more forward looking or whatever that is. So this tool looks at the profile of your engagement, looks at the similar profiles of the other engagements with similar profiles and tags or flags to you where maybe you have a different subset of risks just to be aware of. Hey, you're operating in XYZ environment. Typically clients that are operating in that environment have a risk in this area. Obviously you need to manage. It's not like other client information is being shared, but it's more just like a flag to say maybe reconsider or think about what risks might exist in this area. And that's, as I understand it, powered by some sort of AI. I'm not sure exactly how. So that one that exists, there's another product that has been piloted, not yet released, that's around *tool name*. There's tie out as a process at the end of the, typically near the end of the engagement where the financial statements. Ultimately our opinion is on all of the numbers. There's a couple different percentages. But *tool name* specifically is looking at the numbers within the financials to see if those tie into what we've done all of our work on. So is there actually a linkage between what's going on to the market and what we've done all of our stuff on? That's done through the tie out process to say each and every number in the financial statements, does it tie, does it agree with the numbers that we have in our engagement and the trial balance that we have audited? So there's a whole sort of match up numbers. You also do things like just the prior year numbers. Do they actually agree to what the prior financial statement said? Sometimes, surprisingly, you get things that magically change that you need to flag. There's also an element of, and this is where the AI comes into play in particular in this product, is, are the numbers consistent with one another? So you might have on the financial statements, but, you know, many, many pages. I think for IFRS there's a lot of disclosures. It could be 100 plus pages. And then on top of that, you sort of have to do this consistency check as well, outside of the financial statements. So you say if revenue over this segment is noted as 100 million here, but it's also mentioned in three other spots, do those numbers all agree to one another? And so there's a machine learning module, machine learning model that was trained based off of like snippets of annotations from different financial statements that say, I guess it sort of orients around the figures and says, we think, you know, this 100 million here should be the same. It can't base off the number. But, based off the description and maybe the bias or the movement that that sentence is describing. So is it change, is it $100 million change in revenue over the period? We think that this number should be the same as what's said, you know, page 53 and page 75. But it's not. Or maybe it is. And so it groups those things together and says, you know, here are, here's what we think that should be internally consistent. And hey, it is based off of the automation, AI or whatever, or it's not. And so maybe it needs to be fixed. So that's one application	reconciliation process at the end of the audit, where the content of the financial statements are checked to verify consistency.

that is, again, it's been piloted but not yet released. And yeah, there are more coming as well.	
Q: If someone still needs to review, what would you say is the major benefit of these AI deep learning applications? It's a lot faster. So even if, and you know, I guess you might be able to make the argument that it might end up being more accurate too. But it's certainly a lot faster to have a suggestion sort of initially identified and then documented. So that's the other thing. It sort of automatically puts in all these things that we need to add into the file to document, you know. Yes, this agrees here and there and there. So even if you're more in a position of review, it's a lot, it's faster. And much like if you think about just the experience of our people, it's a lot more pleasant to, to review that as opposed to be the one doing it.	AI-DL solutions make the audit faster, more accurate, and makes the audit job more pleasant.
Q: I read one deep learning application framework. It's an illustrative framework and it's based on the so called data warehouse. [brief explanation of the paper's findings] Would you say that this *audit platform* could be defined as a data warehouse similarly to the way that I just described? So *audit platform* as it currently exists holds a lot of data. It's not structured data, not all of it. Some of it is. A lot of it still exists in excel files, word files, PDF's. There's also data like actual structured data. When you think about what you've just said around, can we access sort of this historical data and information about the client and generate some initial starting points that sort of improve, I guess, are built upon or approve, leveraging the past experiences. The thing that gets tricky for us in particular is kind of data usage, like acceptable, like policies for acceptable data usage. So what we think the one of the issues around doing something like that, well, I think that that could be really interesting. The challenges are, do we have the permission from the client to actually use their data in that way? I think right now the answer is no. And that's all determined based off of like the terms and conditions of our engagement letter with different clients. Cause even if you think about applications beyond that, like can you, can you use like client data even if you're thinking about the structured data to train sort of anomaly detection models that can be applied across clients, but we don't currently have permission to use data in that way. I think the data issues are what come up and a bit from that, because even for *audit platform*, it's capturing structured data and there is a clear mapping, for example which accounts have risks associated with them across engagements, but no more details than that are gathered. Right. It's just which financial statement account are there a higher inherent risk associated with? You know, and so there's this really fine line that we have to walk, and I'm not an expert on it, but in terms of, like, how do we use our clients data? Is it an acceptable use of that data? That's kind of a big open question mark. And I know the firm has tried to put out new, like, new terms and conditions. I think it's called the client information clause, into our engagement letter templates. That would allow more use of that. But clients, you know, are also on the negotiating	Client hesitation to share the data, as well as policies, limit DL's applications.

table when it comes to what they agree to in the engagement letters. And I'm pretty confident many of them would say no and have said no to that use.	

6.14 Interview 13L

Manager in data analytics dealing with FA data + data science background **Unstructured** **Round 2**	**Labels/interpretation**
Q: How would you see deep learning to be applied in to the reconciliation process to be helpful? Yeah, I think that's quite a hard question. If we're going to apply determining ourselves, then you have to be line with the regulations, rules and regulations. So then you have to be sure that there is, yeah, urgency on your algorithm and that your algorithm is correct and that you also can explain it when the AFM, for example, controls your audit. So, yeah, I found this actually quite hard question to answer.	Regulatory compliance makes this question hard to answer.
Q: Future research would address ensuring regulatory compliance, while it is beyond the scope of mine at the moment. So leaving that aside, would you think that deep learning could actually help with the reconciliation process without considering the fact of a black box, explain ability, etcetera? Yes and no. Yes. In terms of where, for example, other team is now checking. Okay, what I see in the finished statement, do I see that also? The bank account, do I see that also over here? I think that can be automated with some deep learning where it automatically reads invoice, for example, can reconcile it to bank account and reconcile it back to the financial statements. But yeah, there are still worse the parts where obtaining the data. I don't see that happening with deep learning because all from the client side where we should get the right data.	Reconciliation can be automated, but obtaining the data from clients would be difficult.
Q: Among the challenges that I found in the financial audit process is the sampling procedure that seems to present high risks of overlooked transactions. Would you agree that that is a challenge as of now? I think that always has been a challenge. The sampling part. So that's why for a lot of clients, at least from the point I started till now, we strive for a digital audit. And that means that, for example, in the general ledger, you will obtain all journals and the complete trial balance, profit and loss, and then you will reconcile it, so you will actually have all data. And I think that will be more and more the case also for some custom things. Yeah. The part where they now draw samples on. It's, I think the part where they did not have the time yet to further make it digital or. Yeah, maybe some availability of data from the client side.	The traditional sampling process is problematic.
Q: And considering that the whole set of data is available, would you see room for deep learning applications in order to draw more representative samples out of the entire population, for example, in order to analyze those transactions that seem to be outliers? Yeah. I think that we definitely think the only thing where there is issue or not an issue which might cause difficulties	Considering only the characteristics of tasks and the solutions DL can bring, there is a fit.

is that when we make the letter for clients, we write which data we want to use and for what purpose we want to use it. So that also means that if we want to build some kind of deep learning application, we are not allowed to use that data set to train on, for example, for another client.	But the problem are regulations.
Q: What if you use data from previous clients though, in order to train the model? Can that be done? No, you're actually not allowed. At least that's for my understanding is because we are not allowed, because we know the data, we write to the client, an engagement letter, what data we want to use, for which purpose. You're not allowed to use it outside of that purpose. So if you're creating a deep learning model. So that's also a bit of the black box in this whole thing. And I'm not sure how they maybe go around these kind of rules when developing more AI models, but I know that's something which can hold it back a little bit.	The client data cannot be used outside of the specific purpose expressed in the audit engagement letter.
Q: And the reasons why. Yes, when you say that the client has to be aware how the data will be used. Can you modify the terms on the engagement letter? Yeah, maybe if you check it in the. So maybe if you write it at the beginning of the year in your engagement letter, then, yeah, you might be able to do that. Yeah. This is is a bit of a gray area in what data to use.	
Q: Okay. Another challenge that I found mentioned a lot literature and interviews was the great amount of manual procedures that have to be performed by auditors. Can you, for example, list the major manual procedures that are very time consuming and inefficient? Yeah. Between several documents. So the example, what I mentioned before, you have your invoice and then literally tick that on the bank account, you see the same invoice and you see the same amount. Yeah. That's quite time consuming. Also, for example, if you have 200 samples with a client and there is one account and it has, I don't know, ten transactions, and you have to manually obtain for each transaction a screenshot and put that in. Those things will take a lot of time.	
Q: And how about all the standardized word forms that need to be filled at the beginning of an engagement? Or also all those forms that need to be filled at the end of an engagement when the audit report needs to be published based on all the audit documentation? I've heard that that's also manual and repetitive and time consuming process. Would you agree with that? Yeah, I guess so. I don't know. I fortunately don't have to do that. So I'm not really aware of that process of all those forms. But I know there are a lot of forms to fill in. So probably there can be some time efficiency when using an algorithm who can already prefill it. And in the end, I think also in terms of errors, if you are able to build this deep learning model, which is able to already pre fill it, or take a document and you only have to verify still that it's correct, I think it's less, yeah, it's less prone to errors because when you have to do it manually, you will probably also make some errors, you will miss some things or stuff like that.	With algorithms as well as DL solutions, manual procedures can be automated, reducing the risk of errors.
Q: Regarding the large volume of data, can that happen that there is a large volume of data received many transactions, and	

in financial audit there is due to time constraints, difficulty to actually draw conclusions from it? I don't know if they encountered such things because we are, for example, using [tool name], and then we are already able to, I think, loads up to 100 million rows. Otherwise we have different kind of applications where you can read all those rows. So I'm not really sure.	
Q: More on a general question, as of now, where would you see deep learning possible application in order to alleviate some challenge regarding to the large volume of data or regarding automation or the sampling. Now, I would say what I said before, the tick and tie part, because that's often done on quite large volumes or when they follow an account with screenshots, etcetera, just to verify that all the things are the same, that will make their life much easier, because then they only have to verify that it's indeed correct what the algorithm says and then they're done instead of having to manually do that. And also your quality will go up if you would have.	Use case: reconciliation process.
Q: Would you see a benefit in incorporating big data in the financial audit process, for example, in extracting knowledge in the first phase when the auditing company wants to get an understanding of the clients? Yes, but that also depends on what you can and cannot do, because then, for example, if you indeed can use multiple clients data to train your algorithm on your, because you find a way to do that, then you can also identify in advance already. Within this group of clients or with the specific characteristics. So if you are, for example, you find that in a bank of a specific size, there are always the same kind of risks, etcetera. So they have to, or they found out that during the years they had to audit specific kind of things a bit more in depth or etcetera, then if you noted in advance, you can determine your strategy on that as well. So I think can be elaborated already in the understanding the business, the scope of strategy phase.	
Yeah. And what is your opinion of including other external sources, like market data? Yeah. Then you can indeed, for example, what they are building now. Newscraper. I would say that is just some kind of tool which already summarizes, indeed, if there is a stakeholder behind the company, etcetera, stuff like that, and on the news and. Yeah, just some information on the company itself already, which indeed goes faster than doing it yourself.	
Q: Definitely. And is that also deep learning based? Could that be deep learning based or not? The information scraper that you mentioned, like the kind of. Oh, yeah. Also we did learning. Okay, perfect.	
Q: I found that many procedures are bound to the so called professional judgment in financial audit and for example, fraud detection procedures, as well as assessing the sufficiency of evidence. What is your opinion on the fact that some important processes of an audit are bound to the professional judgment of audit experts? I think in the end that will always be professional judgments. So I think within this area of work, we are, there will be limitations on using AI and deep learning, because what you can	The role of AI in audits is exclusively for judgment support. Humans will have to review the AI output.

see, for example, as an outlier at one client might be completely normal for another client, or the whole infrastructure of your client is completely different than other clients, and you still need your professional judgment and your knowledge to verify, okay, we have to do more, or this is all we have to do, or we have to do something different. So even when you have, for example, deep learning model, which already pre determines your strategy. I think you still need professional judgment to verify that the strategy is indeed as expected or if you miss something. But yeah, I think it will mainly cover the part where it will save you time so you can focus on the things or the interesting things that can be the conclusion. Is this something we have to do? Is this something we want to, etcetera.	
Q: So, more for judgment support and giving recommendations, would you say you're, I don't know, you're lacking a little bit, I was saying. So you would suggest that you can see AI for judgment support, like in order to support a little bit differential auditors? Yeah, I would say indeed. But in the end, yeah, you will need professional judgment because if, for example, your algorithm is not giving a proper outcome and no one will check it, then you're doing your own audit strategy based on this.	