Degree Project in Information and Communication Technology and Industrial Management

Second cycle, 30 credits

# Adoption of Generative Artificial Intelligence in Professional Services: A case Study of an Established Firm

**ROBERT MINASYAN**

TRITA – ITM-EX 2025:237
Stockholm, Sweden 2025

www.kth.se

# Adoption of Generative Artificial Intelligence in Professional Services: A case Study of an Established Firm

by

Robert Minasyan

Master of Science Thesis TRITA-ITM-EX 2025:237
KTH Industrial Engineering and Management
Industrial Economics and Management
SE-100 44 STOCKHOLM

# Adoption av Generativ Artificiell Intelligens i Professionella Tjänster: en Fallstudie av ett Etablerat Bolag

av

Robert Minasyan

Examensarbete TRITA-ITM-EX 2025:237
KTH Industriell teknik och management
Industriell ekonomi och organisation
SE-100 44  STOCKHOLM

**Master of Science Thesis TRITA-ITM-EX 2025:237**

Industrial economics
and management

# Adoption of Generative Artificial Intelligence in Professional Services: A case Study of an Established Firm

Robert Minasyan



| Approved | Examiner | Supervisor |
|---|---|---|
| 2025-06-05 | Cali Nuur | Emrah Karakaya |
| | Commissioner | Contact person |




## Abstract

This study explores the adoption of Generative Artificial Intelligence within an established Professional Service Firm, focusing on its implementation in tax operations through a single case study. The findings reveal that while Generative Artificial Intelligence automates some menial tasks, its primary function is to augment the capabilities of tax professionals across various use cases. Despite a gradual transformation and anticipated disruptions yet to fully materialize, improvements in efficiency and quality of deliverables are reported. As professionals navigate challenges such as hallucinations, potential data breaches and difficulties allocating time, two distinct strategies emerge: implementation-methods and learning-methods. Implementation-methods involve procedures for integrating AI into workflows, characterized by validating and revising AI inputs and outputs. Learning-methods instead describe how professionals develop expertise in using AI through experimentation and collaboration, highlighting the significance of social support in the adoption process. Ultimately, while the integration of AI enhances client deliverables and offers valuable learning opportunities for the firm, concerns arise regarding the acquisition of domain knowledge by junior staff, potentially impacting their future effectiveness in leadership roles.



**Key-words**

Professional Services, Professional Services Firms, Artificial Intelligence, Generative Artificial Intelligence, Knowledge, Knowledge creation, Automation, Augmentation

**Examensarbete TRITA-ITM-EX 2025:237**

# Adoption av Generativ Artificiell Intelligens i Professionella Tjänster: en Fallstudie av ett Etablerat Bolag

Industriell ekonomi
och organisation

Robert Minasyan

| Godkänt | Examinator | Handledare |
|---|---|---|
| 2025-06-05 | Cali Nuur | Emrah Karakaya |
| | Uppdragsgivare | Kontaktperson |

**Sammanfattning**

Denna studie utforskar adoptionen av Generativ Artificiell Intelligens inom ett etablerat professionellt tjänsteföretag, med fokus på dess implementering i skatteverksamheter genom en fallstudie. Resultaten visar att även om Generativ Artificiell Intelligens automatiserar vissa rutinuppgifter, så är dess främsta funktion att komplementera arbetet hos skattejuristerna inom olika arbetsområden. Trots en gradvis transformation och förväntade förändringar som ännu inte har fullständigt materialiserats, rapporteras förbättringar i effektivitet och kvalitet på leveranserna. När juristerna navigerar svårigheter relaterat till hallucinationer, potentiellt dataintrång och svårigheter att avsätta tid, framträder två distinkta strategier: implementationsmetoder och lärandemetoder. Implementationsmetoder involverar procedurer för att integrera AI i arbetsflöden och kännetecknas av validering och revidering av AI-inmatningar och -utmatningar. Lärandemetoder handlar istället om hur juristerna utvecklar expertis i användningen av AI genom experimentering och samarbete, vilket framför vikten av socialt stöd i adoptionsprocessen. Trots att integrationen av AI förbättrar kundleveranser och erbjuder värdefulla lärandemöjligheter för företaget, uppstår det även oro kring lärandet av kunskap inom skatteområdet för juniora skattejurister, vilket kan påverka deras framtida effektivitet i ledarskapsroller.

**Nyckelord**

Professionella tjänster, Professionella tjänsteföretag, Artificiell Intelligens, Generativ Artificiell Intelligens, Kunskap. Kunskapsproduktion, Automation, Augmentering

**Acknowledgement**

Firstly, I would like to extend my deepest appreciation to my supervisor, Emrah Karakaya, for his invaluable guidance, support, and encouragement throughout this journey. This work would not have been possible without your insight and assistance.

I would also like to thank my examiner, Cali Nuur, for organizing valuable feedback sessions that not only contributed significantly to the improvement of this work but also enhanced my academic writing skills.

Lastly, I am deeply grateful to the firm that welcomed me to conduct this project. In particular, I wish to thank the Tax Technology team for their warm reception, their support throughout the project, and for making me feel like a valued part of the team. Your openness and collaboration provided meaningful opportunities for me to learn and grow.

# Table of Contents

# 1 Introduction

Innovation has long served as the cornerstone of societal development. From early agricultural tools to modern digital infrastructure, transformative ideas and technologies have continuously reshaped how people live, work, and connect. These advances not only solve immediate problems but also expand what is possible, laying the groundwork for new industries, institutions, and cultural shifts.

Among the various types of innovation, technological innovation holds particular significance in driving economic progress. Technological breakthroughs enable productivity, open new markets, and fundamentally redefine existing business models. One of the most powerful illustrations of this is the emergence of the digital economy (Carlsson, 2004), which has revolutionized commerce, communication, and labor. Digital technologies have enabled companies to scale rapidly, automate complex processes, and unlock new sources of value, thereby becoming a critical force behind modern economic growth.

While the rapid advancement of technological innovations within many fields is significant, the adoption of it is equally crucial. Innovations require adoption to realize their potential, yet the process of integrating new technologies is rarely straightforward. As Everett Rogers (1962, p.7), the pioneer of the diffusion of innovation theory, observed: *"Many technologists think that advantageous innovations will sell themselves[...] Unfortunately, this is very seldom the case. Most innovations, in fact, diffuse at a surprisingly slow rate."*

At the same time, *artificial intelligence (AI)* comes in with a lot of promises of transforming societies and economies. However, AI's economic and organizational impact has so far been limited, and on an economic level, the adoption of AI in industries and organizations globally has not led to any major increases in measured productivity. In fact, since the late 1990s, measured productivity growth has slowed significantly, despite rapid technological advancements (Brynjolfsson et al., 2017)

Alan Turing laid the foundation for AI in his seminal work "Intelligent Machinery" (1948), where he theorized about using the human brain as a model for designing intelligent machines. In this work, he proposed the concept of electrical neural networks inspired by human cognition as a means of processing information, emphasizing the necessity of "educating" machines (Turing, 1948). These ideas could be translated into the training of artificial neural networks, a fundamental principle underlying modern AI technologies. Further, Turing suggested the need *"to try and see what can be done with a 'brain' which is more or less without a body,"* arguing for the game of chess as a suitable domain for exploration of intelligent machines (Turing, 1948, p.13). This concept was perhaps realized 49 years later, in 1997, when IBM's Deep Blue chess computer defeated reigning world champion Garry Kasparov (Copeland, 2025).

Turing is also renowned for introducing the Turing Test, a method for evaluating whether a computer exhibits human-like intelligence. In this test, a human poses questions to both a computer and another human without knowing which is which. If the questioner is unable to distinguish between the responses of the machine and the human, the computer is considered to have passed the test, thereby demonstrating a level of intelligence comparable to that of humans (Copeland, 2025). Recent advances in *generative AI (GenAI)*, exemplified by rapidly spreading tools like ChatGPT, highlight this phenomenon. ChatGPT 4.5 achieved a significant feat by being as of yet the only AI technology to pass the Turing Test, with 73% of participants mistakenly identifying it as human after a five-minute conversation (Jones & Bergen, 2025). These developments in generative artificial technologies may signal a paradigm shift in intelligent computing, as the boundaries between human and machine capabilities become increasingly blurred.

Indeed, GenAI stands out with its potential to revolutionize sectors by generating content comparable to human output. According to Goldman Sachs (2023), GenAI is expected to boost the global gross domestic product (GDP) by 7% within the next decade. This transformative potential positions GenAI as a key driver of innovation, playing a crucial role in economic growth through the introduction of new processes and products.

However, expectations that AI, and specifically GenAI, will drive substantial economic productivity gains in the near future are increasingly being questioned. Acemoglu (2025) argues that while AI may yield cost savings and productivity improvements at the task level, its overall macroeconomic impact is likely to be modest. He estimates that AI could increase productivity and GDP by no more than 0.55% and 0.90%, respectively, over the next decade, suggesting that these predictions may even be overestimated.

In the context of knowledge-intensive work, AI is predicted to transform the very nature of it. Knowledge work is characterized by processing information to generate output that adds specific value, something that is increasingly being performed by GenAI. Hence, due to the generative technologies' ability to produce intellectual capital similar to that of humans, the work of knowledge workers is predicted to transform (Alavi & Westerman, 2023; Kronblad, 2020).

Given that a key characteristic of professional services firms (PSFs) is knowledge-intensity (Von Nordenflycht, 2010), this industry presents an industry prone to be disrupted by AI. For instance, Dell'Acqua et al. (2023) highlight a case in which management consultants, an essential segment in professional service providers, achieved significant efficiency and quality gains through the use of GenAI tools. AI has also been subject to transforming the very characteristics of these professional service delivering firms, through extended digital service and product offerings (Kronblad, 2020) and hence serve as a compelling case to study of how GenAI transforms knowledge-intensive businesses.

Established PSFs, one such segment being the so-called Big Four accounting firms, exercise a firm grip on the accounting market, making them possess vast financial resources. Thereby, established PSFs are leading the charge in AI adoption, investing billions of dollars in these technologies and transforming traditional workflows (The Finance Story, 2024). However, workplace GenAI adoption in Sweden has been found to lag behind the global average (PwC, 2024), which is surprising given Sweden's position as being one of the most digitally competitive countries (IMD, 2024). This makes it necessary to study GenAI adoption within PSFs in Sweden specifically. Furthermore, the shift in PSFs toward AI generated capital necessitates an understanding of how GenAI affects the work of knowledge workers and transforms professional processes within PSFs.

Growing literature on information systems has addressed AI adoption in organizational contexts for decades. The following body of literature shows that GenAI is distinct in its design and behavior, due to its ability to generate seemingly novel knowledge, as well as posing new limitations such as hallucinations and difficulties in detecting bias (Feuerriegel et al., 2024; Zhou et al., 2024; Janiesch et al., 2021). Furthermore, knowledge creation with AI can utilize vast amounts of data, redefining the domain-specific knowledge (Agarwal & Dhar, 2014; Henriksen & Bechman’s, 2020), necessitating human and AI collaboration in form of mutual learning mechanisms (Raisch & Krakowski, 2021; van den Broek et al., 2021; Grønsund and Aanestad, 2020).

Nevertheless, GenAI adoption is deemed to differ in its organizational adoption through its bottom-up approach as opposed to earlier AI-systems showcasing top-down introduction (Retkowsky et al., 2024; Waardenburg et al., 2022; Strich et al., 2021). Furthermore, knowledge-workers’ use of the technology differs from previous AI adoption, posing importance in studying GenAI adoption specifically (Retkowsky et al., 2024). However, the adoption of GenAI in knowledge-intensive businesses through a qualitative case-study approach remains understudied (Nesemeier, 2024; Blazques et al., 2024; Yang et al., 2024).

Given GenAI's distinct characteristics, behavior, and its relevance to knowledge-intensive businesses such as PSFs, this thesis contributes to the literature on how AI is adopted in an established PSF, and how it transforms knowledge-intensive work through a qualitative single case study within tax services.

The research is guided by the following questions:

- *How are generative artificial intelligence technologies adopted in a large professional services firm?*

- *How does generative artificial intelligence adoption transform the work of professional services professionals?*

- *What are the potential opportunities and concerns associated with the adoption of generative artificial intelligence in professional services?*

To explore these dynamics, a single case study of an established multinational PSF is conducted, focusing on the factors influencing GenAI adoption and its impact on the work of PSF professionals in its Swedish entity. The firm studied is one of the Big Four accounting firms, and is heavily investing in the adoption of GenAI, hence serving as an interesting firm to study. Furthermore, the thesis specifically focuses on its tax operations, as this business unit is leading the transformation, but also employs most GenAI technologies in its processes. By addressing these questions, the study aims to provide insights into the adoption and impact of GenAI in the professional services industry, contributing to a deeper understanding of this emerging technology.

The thesis is outlined as follows:

- Chapter 2 – Literature Review: Provides an overview of GenAI, and the role of AI in knowledge-intensive business.
- Chapter 3 – Method: Describes the empirical setting and the case study approach used.
- Chapter 4 – Empirical Findings: Presents the key findings from the study.
- Chapter 5 – Discussion: Interprets the findings in relation to the existing literature.
- Chapter 6 – Conclusions: Summarized the study's conclusions, offers managerial recommendations and suggests areas for future research.

# 2 Literature Review

This literature review begins by providing the necessary technical overview on AI, specifically focusing on GenAI, as its technical architecture is a critical factor in distinguishing it from earlier AI applications, capabilities, and limitations. Next, the review explores the role of AI in knowledge-intensive businesses, which is the context in which GenAI is expected to transform professional services. Lastly, given the technology's significant transformative potential, the ethical dimension of AI is explored.

## 2.1 Overview of Generative AI

This study adopts a capability-focused definition, as it is qualitative and explores the socio-technical aspects of AI adoption in professional services. The definition by Copeland (2025) describes AI as *"the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings,"* serving as a useful framework for exploration. However, understanding GenAI is also necessary due to its unique generative abilities, enabling the creation of seemingly new knowledge from past data.

Feuerriegel et al. (2024, p.111) conceptualize GenAI within socio-technical systems and provide a contextually fitting definition, describing GenAI as *"computational techniques that generate seemingly new, meaningful content, such as text, images, or audio, from training data."* A *generative AI system* instead encompasses the complete technical infrastructure, including the model, data processing, and user interface. These systems are often multimodal, meaning a single model can handle inputs from different modes—such as speech, text, and images—and generate different outputs. For example, ChatGPT processes text and generates text, while DALL-E processes visual inputs to produce images (Feuerriegel et al., 2024).

### 2.1.1 Architecture of Generative AI

Understanding the architecture of generative AI systems requires establishing a foundational understanding of machine learning and artificial neural networks. *Machine Learning (ML)* refers to a system's capacity to learn from data to solve analytical problems via model building (Janiesch et al., 2021). The three types of ML are supervised, unsupervised, and reinforcement learning. *Supervised learning* involves training a model with labeled datasets, *unsupervised learning* allows the model to detect patterns without predefined labels, and *reinforcement learning* involves the model exploring allowed actions within environmental constraints to maximize its utility (Janiesch et al., 2021).

*Artificial neural networks (ANN)* are a subset of ML and consist of interconnected mathematical processes similar to information processing in human brains. Similarly to brain synapses, the connections between the artificial neurons process signals that can be strengthened or weakened, which are continuously adjusted throughout training. An activation function sets a threshold that, if unmet, prevents signal processing. ANNs include an input layer, an output layer, and hidden layers for learning non-linear mappings. *Deep neural networks (DNN)* instead have multiple hidden layers to perform complex operations, which independently discover relationships in raw data, a process called *deep learning*. This complexity can render the system a *"black-box*", as decision-making becomes challenging to explain (Janiesch et al., 2021).

Popular generative AI systems include conversational chatbots such as ChatGPT and Gemini, both of which can be classified under the term *large language model (LLM)*. An LLM is based on a large-scale deep neural network using a transformer for self-attention, enhancing performance in language modeling compared to earlier methods (Vaswani et al., 2017). As shown in Figure 1, the neural network is then pre-trained through unsupervised learning using extensive datasets (Feuerriegel et al., 2024). Lastly, the LLM is fine-tuned for specific tasks by human oversight, using supervised learning and reinforcement learning mechanisms (Feuerriegel et al., 2024). The reinforcement learning mechanism, also referred to as *Reinforcement Learning from Human Feedback (RLHF)*, is especially important as it employs human feedback loops to improve the AI generated output. For instance, ChatGPT uses RLHF by generating example outputs to certain input prompts, and lets the user rank their quality, thereby training the model to maximize its output scores (Feuerriegel et al., 2024; Ziegler et al., 2019).

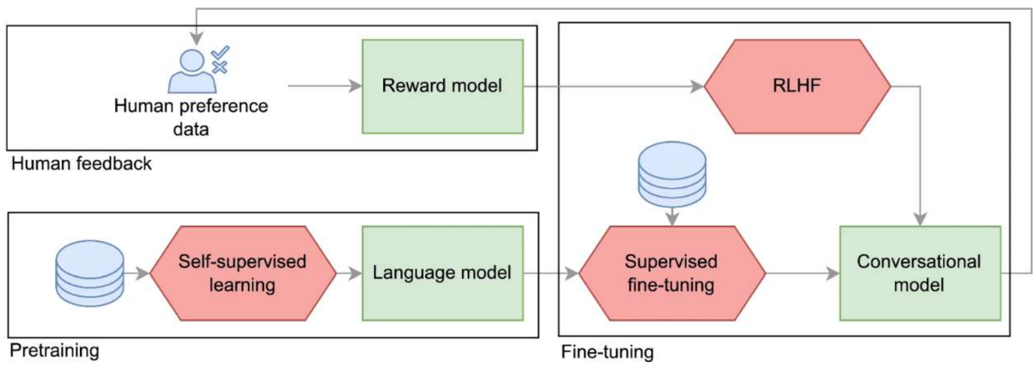


**Figure 1: Training process of conversational GenAI models, adopted from (Feuerriegel et al., 2024)**

### 2.1.2 Technical Limitations of Generative AI

While there are numerous technical limitations in generative AI systems, such as copyright infringement and excessive energy use (Feuerriegel et al., 2024), this section focuses on two relevant technical limitations to this study, namely, bias and hallucinations.

#### *2.1.2.1 Bias*

Ferrara (2024, p.2) defines *bias* in AI as *"a systematic error in decision-making processes that results in unfair outcomes"*, occurring due to various reasons such as data collection, algorithm architecture and human interpretation. In ML, biases generally learn to replicate biases in the training data, making the resulting model unjust in its output (Ferrara, 2024). While bias is not exclusive to GenAI and is evident in other AI technologies, it remains a concern in the generative technology as well. Zhou et al. (2024) revealed biases in the image generation tools Midjourney, Stable Diffusion, and DALL-E, which displayed biases against women and dark-skinned individuals, while subtle prejudices in emotions and appearance were also found. However, the generative nature of GenAI complicates bias detection and quantification compared to traditional AI technologies like ML, which focus on classification and modelling. This complication is due to GenAI generally being absent of a sole "true" output (Zhou et al., 2024).

#### *2.1.2.2 Hallucinations*

*Hallucinations* is a seemingly novel concept within the context of AI, introduced with the rise of GenAI, as this is the first technology to consistently report this behavior. Maleki et al. (2024) reviewed 333 definitions across 14 databases, illustrating significant variability in how hallucinations are defined. Hallucinations are mainly evident in LLMs, which Feuerriegel et al. (2024, p.117) defines as *"mistakes in the generated text that are semantically or syntactically plausible but are actually nonsensical or incorrect"*. These arise because generative models make inferences based on probabilistic algorithms, generating outputs that seem correct but may not be (Feuerriegel et al., 2024).

While the definition provided by Feuerriegel et al. (2024) is broad and encompassing enough for this study's scope, a practical example of hallucination within the legal context may help clarify the concept. Numerous instances of AI systems hallucinating exist, but one particular case stands out, especially relevant to professionals in fields like tax who operate in the legal realm and rely on accurate information.

In this case, Roberto Mata, an airline passenger, sought compensation for an injury caused by a metal cart during a flight with the airline Avianca. In response to Avianca's motion to dismiss the case, Mata's lawyers submitted a 10-page brief justifying their position by citing prior legal cases and documents. However, upon review, neither the defense nor the judge could verify these citations. This discrepancy arose because the lawyers had used ChatGPT to draft their brief, and the AI had fabricated all the sources, including fictional cases such as Martinez v. Delta Air Lines, Zicherman v. Korean Airlines, and Varghese v. China Southern Airlines (Weiser, 2023; Emmerich, 2023).

One of the GenAI technologies examined in this study, Harvey AI, utilizes *Retrieval Augmented Generation (RAG)*, a common approach to counteract hallucinations (Es et al., 2023). RAG can significantly reduce hallucinations in conversational models (Shuster et al., 2021), by grounding the model in a textual database as its reference point for output generation (Es et al., 2023).

## 2.2 AI in Knowledge-Intensive Business

The industrial revolution saw vast replacement of human physical labor by machines, referred to as the machine age (or at least the second part of the industrial revolution is). Brynjolfsson & McAfee (2014) suggest this should be considered the *first machine age*, as we are now entering the *second machine* age where machines replace human intellectual labor. An industry heavily reliant on human intellectual labor is knowledge-intensive businesses, which focus primarily on providing knowledge-driven resources to enhance the business processes of other organizations (Muller & Doloreux, 2009). PSFs exemplify such organizations due to their high levels of knowledge-intensive professional service provision. Considering the knowledge-intensive nature of PSFs, and that they provide value through expertise, this very industry is prone to be affected by the proposed second machine age.

Therefore, the role of AI in professional services is examined, with a focus on how it transforms PSFs in light of current advancements in GenAI. As GenAI enhances the relevance of AI-generated knowledge, the following section addresses AI knowledge creation and its industrial context. Given the limited peer-reviewed literature on GenAI due to its novelty, this literature review also incorporates studies on other AI technologies. These are deemed applicable due to the qualitative and human-centric nature of innovation diffusion, which shares commonalities across various AI advancements adopted in organizations. This discussion further necessitates an understanding of automation and augmentation, two key concepts central to the introduction of AI into processes. Finally, the factors that influence the successful adoption of GenAI in PSFs are presented.

### 2.2.1 AI in Professional Services

PSFs are characterized by being knowledge-intensive, low capital intensive and by having a professional workforce (Von Nordenflycht, 2010). Considering that law firms are an essential segment of PSFs, the work of Kronblad (2020) who discusses digitalization and AI in Swedish law firms becomes relevant. Kronblad's (2020) findings suggest that the increased digitalization of law firms has transformed their inherent PSF characteristics, showcasing lower levels of knowledge-intensity and professionalization. The services and products that legal firms provide their clients have increasingly become a result of AI and technological tools, as opposed to solely being produced by the professionals' intellectual capital. The shift from human intelligence to artificial intelligence as the basis for knowledge creation is also found to challenge the business model based on the firms, as the relevancy of billable hours in revenue generation is argued to be decreased (Kronblad, 2020).

Dell'Acqua et al. (2023) provide insights from a study on management consultants at Boston Consulting Group (BCG), highlighting the impact of GenAI on productivity and quality in PSFs. It was found that consultants using GenAI improved the quality of their work by 40% compared to non-users, while completing the tasks 25% quicker. The study also revealed that while GenAI significantly enhances performance in tasks requiring creative product innovation, its effectiveness diminishes for tasks involving business problem-solving. Interestingly, the findings suggest that technology is often mistrusted in applications where it contributes significantly, while trusted in applications that it is incapable of performing. By relating this transformative potential to Kronblad's (2020) suggestion regarding declining relevance of billable hours, Cook et al. (2024) anticipate that GenAI adoption will further reduce capital intensity, leading to a transformation of knowledge-intensive firms' business models.

In a related exploration, Kronblad et al. (2024a) present preliminary findings that specifically examine the transformation brought about by GenAI in various professional services, studying three specific PSFs in law, architecture, and audit. While their findings suggest that GenAI has been seen as an opportunity to be integrated into their work as opposed to being perceived as a threat to their profession, the actual effect of the technology remains to be realized, as the anticipated business model transformation proposed by Cook et al. (2024) has not yet taken place.

### 2.2.2 AI for Knowledge Creation

As PSFs increasingly deliver AI-enabled services (Kronblad, 2020), understanding AI-based knowledge generation becomes essential. AI-created knowledge involves deriving insights and understanding from data using AI techniques, potentially differing significantly from traditional forms of knowledge production by functioning independently of domain experts (van den Broek et al., 2021). Shollo et al. (2022) examine this process in a study of 56 ML projects across 29 firms, highlighting that ML algorithms generate knowledge by modeling data and identifying trends to support strategic, data-driven decisions.

As the ability of creating knowledge is shown to be an essential aspect of AI systems, the design of these is highly dependent on machines being able to derive a "truth" or right conclusion. Due to the ability of AI systems to utilize vast amounts of data, they could be used to discover knowledge that is "truer" (Agarwal & Dhar, 2014). Henriksen & Bechman's (2020) analysis of a Scandinavian AI company employing ML in healthcare, finds that the AI application does not only turn the discovery of truth into a process, but also conceptually redefines the truth within the domain.

However, abandoning AI generated knowledge for one's own opinion is evident within the organizational context, as AI-generated knowledge can lack considering secondary organizational objectives. Kim et al. (2024) studied through a case study, an Inspectional Services Department of a major metropolitan city in the United States. The primary decision studied was resource allocation, specifically determining which restaurants inspectors should prioritize for inspection. An AI system providing recommendations on which restaurants are more likely to have violations was introduced, outperforming the prioritizations of the inspectors significantly. However, only two-thirds of the recommendations were followed, as the authors argue that while AI improved inspection predictability, it did not take into account secondary objectives like minimizing travel distance and prioritizing popular restaurants. As a result, the AI system's superior predictive capabilities did not necessarily lead to improved decisions. The inspectors hence exercised their *decision authority*, to balance other organizational objectives in the decision-making process, leading to superior outcomes than solely relying on the AI recommendations.

Furthermore, the opaque nature of AI, as explained by Janiesch et al. (2021), has a significant impact on communicating AI generated knowledge. For instance, Lee (2018) found that in human tasks, AI decisions were perceived as more unfair and less trustworthy compared to human ones. This is elaborated upon in a 31-month ethnographic study by Waardenburg et al. (2022), which examines how knowledge brokers translate AI predictions of crime in various geographical areas with the introduction of an opaque, black-boxed ML system at a Dutch police station. The knowledge brokers' intended role was to translate insights from an ML-community to police management. However, due to the black-boxed nature of the system, the brokers struggled with explaining the predictions, leading them to adopt a role as *curator*. This meant that the brokers substituted the predictions of the system with their own judgements, as they argued that they were experts in their field and knew better than the predictions made by the system.

While it is important to understand how AI creates knowledge and how it is implemented in organizations, it is also essential to examine the effects of AI introduction on tasks. Therefore, the next section explores two key concepts: automation and augmentation of tasks through AI.

### 2.2.3 AI for Automation and Augmentation of Tasks

AI implementation in organizations is signified through either automation or augmentation. *Automation* is defined as a human task being carried out by machines, while *augmentation* means humans collaborating with machines to carry out tasks (Raisch & Krakowski, 2021). Automation serves as an effective approach to codifiable, rule-based tasks, while augmentation is better fit for more complex tasks (Raisch & Krakowski, 2021). Spring et al. (2020) explore the application of automation and augmentation in PSFs through a multiple case study involving two law firms and two accounting firms. Their research examines how ML-based AI systems are used to automate and augment tasks within these professional services. Their findings suggest that automation is frequent for repetitive tasks, while more complex tasks are augmented, in line with the findings of Raisch & Krakowski (2021). Furthermore, automation and augmentation enable professionals to free up time from client interaction, due to increased service delivery through AI-systems, as clients are enabled to self-service. This enables the professionals, combined with the AI systems ability to capture their knowledge, to provide improved quality or speed in the delivered services. Furthermore, the use of AI-systems is found to restructure the processes of service delivery, as opposed to replacing the professionals in the previous delivery process. Due to the increased analytics that AI systems provide, the offered services are extended, some even pertaining to realms outside of accounting and law (Spring et al., 2020).

While these concepts differ and should be applied in differing contexts, Raisch & Krakowski (2021) emphasize a paradoxical relationship between automation and augmentation of AI in management, suggesting that these two, seemingly distinct concepts to be interdependent, over time and space. This is echoed by Shollo et al. (2022), who suggest that knowledge creation can move to augmented processes, where AI technologies are increasingly used by professionals, while they still remain the decision-makers. This can then move to automated processes, where the AI systems turn into independently operating agents, excluding human intervention in the process (Shollo et al., 2022; Raisch & Krakowski, 2021). Furthermore, Raisch & Krakowski (2021) warn organizations in overemphasizing either automation or augmentation, arguing that this could lead to deterioration in overall performance.

In augmentation, the inherent challenge of codifying tacit knowledge necessitates *mutual learning* mechanisms. Raisch & Krakowski (2021, p.195) describe this as a *"coevolutionary process during which humans learn from machines and machines learn from humans."* This is accomplished by including domain experts in the development and deployment of augmented systems, and keeping these experts involved through *"human in the loop"* configurations (Raisch & Krakowski, 2021). This mutual learning mechanism is evident in the two-year long ethnographic study by van den Broek et al. (2021) on the development of an ML-based AI system for hiring. While the engineers initially intended to exclude domain experts from the creation process, they found themselves oscillating between excluding and including them in the development. This phenomenon is explained by the interaction between the technology and the domain specialists, which prompted a reconsideration of how knowledge was produced.

Grønsund & Aanestad (2020) further exemplify this with their two-year long case study on the implementation of an ML-based AI system for automating data analysis and predictive analytics in an international maritime trade organization. They found that the organization underwent various configurations of processes, resulting in augmentation, characterized by a "human in the loop". The authors argue that the development is augmented, as two essential processes emerged, namely, *auditing* and *altering* of the AI system. This meant that the system had to be continuously audited, through human supervision and assessment of the system's performance. It also had to be altered to perform more accurately, which meant improving the input data or presentation of the resulting output.

However, Grønsund & Aanestad (2020) highlight that auditing and altering of the system was feasible since identifiable errors could be corrected. In the context of GenAI, identifying errors may be more challenging (Zhou et al., 2024), necessitating innovative approaches to audit and modify these models. Cook et al. (2024) propose using feedback loops to enhance GenAI performance within PSFs. This is exemplified by Cook et al. (2024) with the task of creating initial drafts of documents such as research reports, a common task within professional services. The revision of these drafts, performed by the employees, could be used as a means of collecting feedback on the AI generated draft, to generate better drafts in the future. With professionals increasingly relying on GenAI to draft their work, feedback from the revision made by the professionals could not only boost the performance of the GenAI tool but also improve the quality of the knowledge created (Cook et al., 2024).

A common theme in the above literature is the increasing convergence of use and development of AI. Waardenburg and Huysman (2024), in their study of five empirical cases of ML-based AI systems in organizational contexts, argue this point, noting that AI's independent learning capabilities blur the boundaries between use and development. These blurred boundaries could also be extended to the popular game of chess, through the lens of how it affects competitive advantages. As opposed to traditional chess, in centaur chess, humans select, tune, and use chess engines to play the game. Krakowski et al. (2023) examine centaur chess competitions in terms of competitive advantage and the complementary effects of human-AI collaboration. They illustrate their findings with the 2005 Centaur Chess World Championship, where amateur players defeated grandmasters, showing that success in this context depended not on chess skill alone, but on the ability to select and tune chess engines effectively. Hence, as chess engines have surpassed and substituted human chess capabilities, additional non-chess-related competencies emerge as significant competitive advantages. Consequently, Krakowski et al. (2023) suggest that AI's augmentative potential to reshape competitive advantages can be applied to business contexts, shifting the focus toward complementary capabilities rather than solely relying on domain-specific skills.

As AI increasingly substitutes human capabilities in PSFs (Kronblad, 2020), similarly to the centaur chess case of Krakowski et al. (2023), its introduction in organizations can significantly impact workers' role identities, something that is evident in the adoption of GenAI (Sack et al., 2024). In a qualitative single case study by Strich et al. (2021), the implementation of an AI-based system[*] was examined at a loan consulting company. Experienced loan consultants, who previously played a crucial role in the loan application decision-making process, found their roles largely automated, and transformed to merely entering customer data into the AI system and communicating the decisions it produced. This shift challenged their role identities, particularly as customer support professionals were promoted to their position of loan consultants, a position viewed as prestigious. As a result, the professionals employed mechanisms to reinforce their role identities, some even attempting to influence the AI's decisions through data manipulation.

While challenges to role identity can prompt professionals to adopt mechanisms for self-preservation, the introduction of AI can also lead to conformity as individuals seek to protect their knowledge and expertise. Pachidi et al. (2021) conduct a case study on the introduction of an AI system in a sales department through the lens of regime of knowing. Account managers, responsible for client communication and sales, exhibited symbolic conformity to the new technology. This is as they pretended to comply with the changes, which enabled the data scientists to present its positive aspects to management, as a means of symbolically advocating for the technology. This dynamic relationship reinforced the account managers' symbolic conformity, which, in turn, enabled the data scientists to engage in further symbolic advocacy, resulting in a cyclical process where both parties demonstrated symbolic actions. Consequently, this process led to a radical change in the work of the account managers, as their regime of knowing shifted from client communication and relationship management to model-based predictions, ultimately resulting in layoffs and the outsourcing of the account managers' roles, even though the technology was initially intended to augment the work of the account managers.

GenAI is predicted to automate tasks but is suggested to mainly augment capabilities of workers[†] (Boulus-Rødje et al., 2024; Reinhard, 2024), especially in knowledge work (Alavi & Westerman, 2023; Sack et al., 2024). Hence, understanding how this technology transforms the work of knowledge workers becomes relevant. Retkowsky et al. (2024) examine the use of ChatGPT in knowledge work, highlighting that the introduction of GenAI is bottom-up as opposed to top-down as in previous AI-systems introduction in organizational contexts (Waardenburg et al. 2021 and Strich et al. 2021 are two such examples).

Through interviews with early adopters, Retkowsky et al. (2024) identified a process that users experience when adopting the technology. In the pre-use phase, adopters express emotions of curiosity about trying something perceived as highly valuable, while simultaneously experiencing anxiety as they view the technology as an existential threat to their jobs. The use phase is initially characterized by playful *"tinkering"*, where users experiment with the tool without assigning value to its potential applications or benefits in their work. However, upon discovering the AI's capabilities, they transition into the next phase, known as work-related "tinkering" with its work-related benefits in mind. In this phase, users begin to experiment with the tool in a work-related context and relate it to their knowledge-related limitations, such as a lack of time and creativity.

---

[*] The system is referred to as an Algorithmic Decision-Making system in Strich et al. (2021). However, it is considered an AI system, as its replacement of the human function, approving loan applications, indicates that it is likely equipped with AI capabilities.

[†] GenAI agents are an exception but not examined.

This realization prompts them to see how the technology can assist in their daily work, leading to what the researchers refer to as *"intertwinement."* In this phase, workers integrate the tool into their processes and daily knowledge work, making it an integral part of their tasks and expressing significant dependence on it. Interestingly, the experimental nature of the "tinkering" phases suggests similarities to Wensink's (2024) findings, which highlights a slow and time-consuming process adoption process within management consulting, that requires extensive experimentation with the technology. Furthermore, the concept of "intertwinement" presented by Retkowsky et al. (2024) resembles augmentation but places greater emphasis on the close integration between humans and AI, and the reliance of humans on the technology.

Retkowsky et al. (2024) further explore the affordances of using ChatGPT in knowledge work. They identify six use cases: searching for information, structuring content, brainstorming ideas, creating an initial draft, improving text, and proofreading work. In contrast, Wensink (2024) identifies information searching as a shared use case, while also highlighting notetaking, summarizing, transcribing, and creating PowerPoint presentations. Retkowsky et al. (2024) also describe the emerging use types in response to limitations related to knowledge work. For instance, they find that to address their limitations in information and inspiration, workers use ChatGPT to search for and summarize information as well as to brainstorm ideas. Additionally, to mitigate their constraints of limited time and competency, they employ ChatGPT for tasks such as drafting initial content and validating self-written text in terms of grammar. Hence, they suggest that the emerging use types are directly related to mitigating the limitations that knowledge workers inhibit.

Additionally, they examine three side effects from a managerial perspective. The affordances provided by the tool encourage workers to turn to it for assistance, validation, and brainstorming, rather than relying on peers, which suggests an increase in individualization in performing knowledge work. The second side effect noted is that ChatGPT may introduce lower levels of quality knowledge into organizational processes. The difficulty in distinguishing between AI-generated and human-generated work complicates efforts to maintain quality control in client deliverables. The final side effect pertains to role configurations; for example, rather than creating output manually, the worker's role shifts to editing AI-generated content. This shift creates challenges in assessing employee performance due to the intertwined nature of AI and the workers (Retkowsky et al., 2024).

### 2.2.4 AI Adoption Factors

There is no scarcity in literature examining factors influencing AI adoption, as a wide range of literature explores this context in various domains. Yang et al. (2024, p.1), however, questions their transferability due to AI's advancing essence and PSFs' *"distinct characteristics, such as knowledge intensity and regulatory environment"*. Hence, I deem there are four main adoption factors influencing the adoption of GenAI in knowledge-intensive organizations such as PSFs[*]. These factors, generally influenced by the size of the organizations, consist of social influence (Blazquez et al., 2024; Nesemeier, 2024), regulatory environment, competitive environment and organizational resources (Yang et al., 2024).

Yang et al. (2024) utilize Technology Organization Environment (TOE), while Blazquez et al. (2024) and Nesemeier (2024) use Unified Theory of Acceptance and Use of Technology (UTAUT). These

[*] This is a synthesis of literature addressing AI adoption in PSFs, and due to the specific nature of PSFs (Yang et al., 2024; Von Nordenflycht, 2010), and the technical discrepancies of GenAI as compared to traditional AI technologies (Feuerriegel et al., 2023; Zhou et al., 2024; Janiesch et al., 2021), only the literature deemed relevant to the context of this study are presented here. Those excluded deserving further attention in future research, are outlined in the Limitations section.

frameworks are salient in studying AI (and GenAI) adoption within PSFs and therefore deserve an overview and how they are applied in the works.

The TOE framework, pioneered by Tornatzky et al. (1990), analyzes the adoption of innovations from technological, organizational, and environmental perspectives. Yang et al. (2024) argue that it is a suitable framework for exploratory studies within AI due to its holistic and flexible nature, while also acknowledging critiques regarding its broadness. To address these concerns, Yang et al. (2024) propose incorporating the concepts of affordances and constraints within the technological dimension, as well as strategic factors typically overlooked in the traditional TOE framework.

Yang et al. (2024) utilize the TOE framework employing a multiple case study on three auditing firms of different sizes. Yang et al.'s (2024) findings suggest that within the regulatory environment factor, concerns regarding compliance with regulatory standards increased in importance with the size of the firm, as the large PSF expressed the most concerns, followed by the middle sized and then the small firm. Furthermore, the resources are suggested to influence adoption, as large firms have ample resources and capabilities, medium-sized firms have moderate, and small firms have limited. Finally, the factor of competitive environment was in the large firm driven by the competitors' adoption, while for the medium sized firm it was expressed through pricing competitiveness, and finally small firms experience the factor through intensive competition and customer acquisition (Yang et al., 2024).

UTAUT is on the other hand a rather new framework, introduced by Venkatesh et al. in 2003. This was done by examining the application of eight previous user acceptance and adoption behavior theories across four organizations, studying the introduction of various information systems during a six-month period. Their findings suggest that four key constructs outperformed the eight previous theories in the user acceptance of information systems, consisting of social influence, effort expectancy, performance expectancy and facilitating conditions[*]. These are defined as:

- Social Influence: is the technology's ability to deliver advantages and improve user performance based on their expectations (Venkatesh et al., 2003; Momami, 2020).

- Effort Expectancy: is the user expectations regarding the ease of using the technology (Venkatesh et al., 2003; Momami, 2020).

- Performance Expectancy: is the anticipated impact of others on the user's decision to initiate and maintain the use of the technology (Venkatesh et al., 2003; Momami, 2020).

- Facilitating Conditions: is the anticipated extent of organizational and technical infrastructure capable of supporting the use of the technology (Venkatesh et al., 2003; Momami, 2020).

Blazquez et al. (2024) examine GenAI adoption in South Korea through surveys and the UTAUT framework. Their study found that social influence and support from coworkers, particularly supervisors and peers, are crucial in facilitating the adoption process, indicating a common thread of social dynamics impacting GenAI acceptance. Nesemeier (2024) also utilized the UTAUT framework but concentrated on professional services within management consulting. His findings reaffirmed the significance of social influence, while emphasizing that peer support often outweighed management support. This aligns with Blazquez et al.'s results, suggesting that social

[*] It should be noted that UTAUT takes into account how certain aspects such as age and experience affect these constructs, however, since these factors are out of scope in the context of this study, they are omitted.

dynamics play a critical role in the adoption of GenAI technologies across different organizational contexts.

### 2.2.5 Ethical AI Adoption

As AI comes with new capabilities, it becomes crucial to ensure sustainable use and deployment. Monod et al. (2024) examine the introduction of an AI tool for salespeople at a large Chinese technology services company. The intended purpose of the tool was to enhance the salespeople's work by augmenting their processes, which it initially achieved by helping them better understand their customers' needs. This allowed the salespeople to serve customers more effectively through individualized approaches while freeing them from time-consuming repetitive tasks. However, these benefits were gradually overshadowed by negative consequences when managers recognized the value of utilizing AI to monitor and control the salespeople to their advantage. Firstly, while the salespeople were engaged in training the AI tool, managers attempted to reduce the workforce without explicitly communicating this intention to them. Secondly, managers increased surveillance of the salespeople to track performance and behavior. Thirdly, inequity and opacity were employed by the managers, who, to maintain the subordination of the salespeople, ensured that the latter had access to less information and fewer details than the AI or the managers themselves.

AI can also lead to injustice as employing black-box models can lead to discriminatory and biased decisions made. One such instance is discussed by Kronblad et al. (2024b), where the Public School Administration (PSA) in the city of Gothenburg implemented an AI-based system* to automate the previously manual task of assigning children to schools based on proximity, in accordance with Swedish and city regulation. However, the system failed to adhere to these regulations, as errors made by the system resulted in thousands of children being placed in wrong schools, as well as most of them being left uncorrected. Kronblad et al. (2024b) argue that the PSA knew about the erroneous system, as school principals and the software firm providing the system raised concerns. However, the PSA did neither address any of the warnings, attempt to understand the functionality nor assess the result. Hence, the black-box concept is extended to the organizational and social context, as the authors argue that PSA blinded itself through *intraorganizational black-boxing*.

Considering that both local and city laws were breached and that each case required individual legal contestation, children from resourceful families were more likely to appeal the decision. Therefore, the authors argue that the consequences of the system's failure, coupled with the public entity's inability to rectify its mistake, lead to structural injustice. Consequently, one of the authors filed a lawsuit against the PSA. However, the case was ultimately ruled in favor of the PSA. This outcome was influenced by several factors: the burden of proof was placed on the author rather than the PSA, the PSA withheld technical information about the system, and each decision was legally required to be addressed individually. The authors argue that the court system failed to adequately assess the legality of the situation, suggesting that the legal system is not equipped to handle the illegal implications of AI in decision-making systems.

Kronblad et al.'s (2024) case emphasize the importance of implementing AI in an ethical manner. This is addressed by Asatiani et al. (2021), who conducted a case study on the Danish Business Authority (DBA), a government entity utilizing AI and ML technologies in its operations. This study highlights DBA's commitment to ethical procedures and serves as an intriguing case for examining the concept of *envelopment*, defined as *"establishing clear boundaries within which the AI is to interact with its surroundings, choosing and curating the training data well, and appropriately*

---

* Similar to Strich et al. (2021), the system is referred to as an Algorithmic Decision-Making system in Kronblad et al. (2024b). However, again, it is considered an AI system, as its replacement of the human function of assigning children to appropriate schools, indicates that it is likely equipped with AI capabilities.

*managing input and output sources"* (Asatiani et al., 2021, p.325). Their findings indicate that DBA effectively employs envelopment methods in its automated decision-making processes. Moreover, envelopment encompasses not only technical factors but also the interaction between technical and social factors, and operationalizing this interaction helps DBA to balance the tradeoff between explainability and performance.

# 3 Method

During a five-month period, a qualitative single case study was conducted within a large PSF to explore the adoption of GenAI tools. This section begins with a detailed description of the empirical setting, outlining the specific GenAI tools utilized by the firm and the firm's background and operations. Following this, the data collection methods and data analysis employed in the study are covered.

## 3.1 Empirical Setting

The adoption of AI technologies in PSFs is an emerging phenomenon that necessitates in-depth exploration within its real-life context. The case study approach is particularly suitable for this study as it facilitates a comprehensive examination of phenomena as they naturally occur in their real-world settings, *"leading to rich, empirical descriptions"* (Saunders et al., 2015, p.185). Additionally, case studies are useful for investigating contemporary events where the *"boundaries between phenomenon and context are not clearly evident"* (Yin, 2003, p.13). This is particularly relevant for understanding the adoption of GenAI within PSFs, as highlighted by Yang et al. (2024), given that distinct contextual factors such as the regulatory environment significantly impact the adoption process. Therefore, utilizing a case study approach allows for an in-depth analysis of this phenomenon and the specific context in which it operates, capturing the intricate interactions and influences shaping GenAI integration into PSFs.

The subject of this case study is a prominent multinational professional services firm, widely recognized for its leading position in accounting and as one of the *Big Four* accounting firms. While the firm operates independently in Sweden, it remains part of a global network and has offices throughout the country. To maintain anonymity, this organization is anonymized and referred to as *Professional Services Global (PSG)*, and any identifiable information such as the names of the interviewees are anonymized.

PSG's organizational structure is divided into two primary divisions: Line of Services and Internal Firm Services. As the names suggest, Line of Services provide client services and comprises Assurance, Tax, and Advisory. In contrast, Internal Firm Services are responsible for internal support functions, including IT, Human Resources, and Finance. Each Line of Service can be further categorized into business units, who work specifically within a realm within its Line of Service. Figure 2 provides a comprehensive overview of this organizational structure.

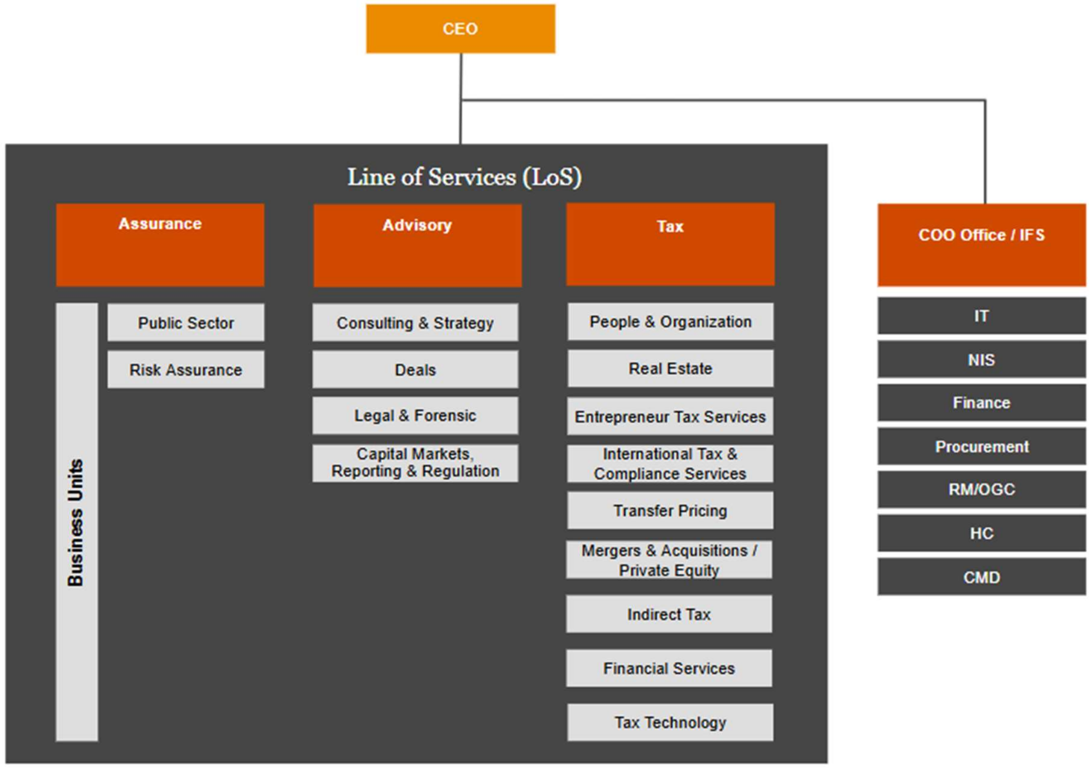


**Figure 2:** **Organizational Overview of PSG**

The firm's structure, typical of large PSFs, is hierarchical. The general promotional ladder includes the following titles: Associate, Senior Associate, Manager, Senior Manager, Director, and Partner.

This thesis focuses on the Tax Line of Service due to its engagement with three GenAI technologies, unlike the rest of the organization, which only employs one. Hence, the tax focus offers deeper insights into how varying technologies affect adoption. Additionally, the Tax Line of Service spearheads technological transformations within the firm, both locally and globally, making it an intriguing subject for studying new innovations within professional services.

During the five-month period of this thesis, I was based full-time at PSG's Stockholm office. In addition to writing this thesis, I worked part-time within the Tax Technology business unit. This team is characterized as a technology unit supporting the tax division and its clients with data and automation solutions. My role facilitated frequent interactions with a broad range of tax stakeholders, allowing me to gain insights into their daily operations.

## 3.2 Overview of Generative AI at Tax Line of Service

This case study explores the adoption of GenAI within the Tax Line of Service, which has implemented three GenAI technologies, unlike other divisions that generally only utilize one. Currently, the Tax Line of Service has three GenAI tools in use, although another one is under development with uncertain prospects for launch. Table 1 provides an overview of all these tools, including their status, key characteristics and if they are examined in this study.

**Table 1: Overview of GenAI technologies at PSG**

| Tools | Status of Launch | Stakeholder access | Characteristics | Included in this study |
|---|---|---|---|---|
| ChatPSG | Launched Spring 2023 | All throughout PSG | LLM chatbot based on ChatGPT 4o model | Yes |
| Harvey AI | Launched Spring 2023 | 200 at Tax, 140 in advisory (Legal & Deals) | Law specific RAG chatbot from Harvey AI | Yes |
| Transfer Pricing Artificial Intelligence (TPAI) | Launched Spring 2024 | Transfer Pricing | Unknown, globally launched | No |
| Microsoft Copilot | Under development | Under development | Integrated AI copilot in PSG's internal systems | No |

PSG has made significant strides in adopting GenAI technologies, including a partnership with OpenAI to deploy ChatGPT for internal use. This collaboration ensures that employees can securely input sensitive data, such as client information. ChatPSG (see Figure 3), a general-purpose chatbot based on OpenAI's ChatGPT-4 model, has been adapted for internal use following a thorough technology risk assessment to ensure the secure handling of confidential information. This means that PSG professionals can input client specific information, without breaking confidentiality.

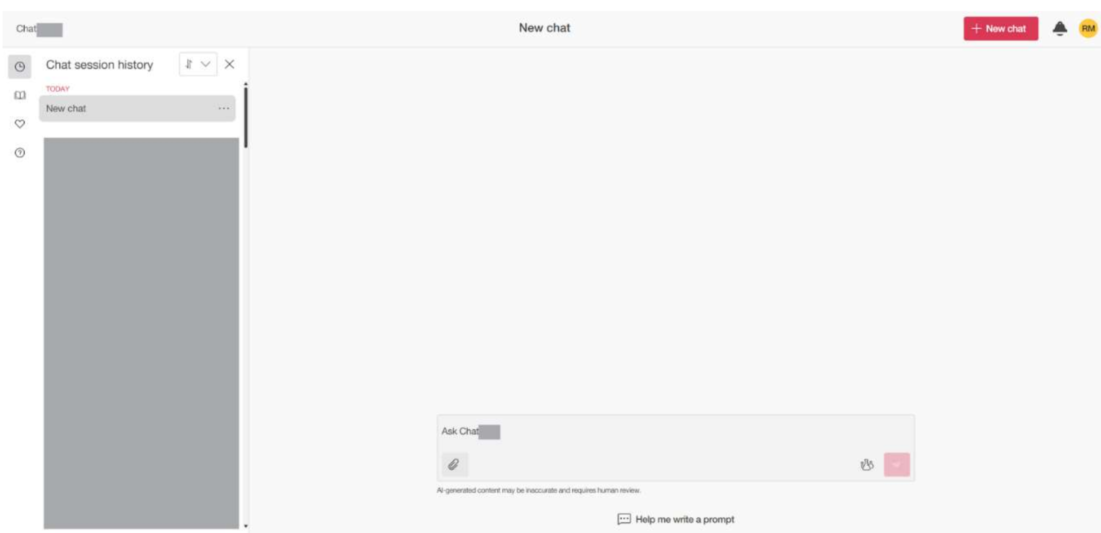


**Figure 3: User interface of GenAI tool ChatPSG**

As of March 28th, 2025, ChatPSG has seen a launch of a new version, with extended functionality as well as a new user interface. The new version allows the user to switch between different models, which include: OpenAI's GPT-4o and GPT-4o mini, Anthropic's Claude 3-5 Sonnet and Claude 3-5 Haiku, and Google's Gemini 1-5 pro and Gemini 1-5 Flash. It also has a feature where the user can create an assistant, where saved prompts can be used in combination with uploaded files to create outputs. Considering the recency in this launch, data collection has not been conducted for this version, and hence, the new version has not been subject for analysis.

In addition, the firm has partnered with a LegalTech startup offering a RAG model specifically designed for legal work. Harvey AI, shown in Figure 4, is primarily accessible to tax professionals and was, during my time at PSG, undergoing an internal compliance review to ensure the secure handling of confidential client data without breaching regulations. As the General Data Protection Regulation (GDPR) mandates that customer data cannot be accessed by third parties without the customer's consent, PSG professionals were unable to input client data into the United-States-based startup's system. To address this limitation, PSG was collaborating with Harvey to relocate the data processing servers to Europe, thereby ensuring that data is processed within the European Union (EU) and in compliance with GDPR. Additionally, Vault, a feature provided by Harvey AI, was also undergoing a compliance review at PSG, which means professionals did not have access to it during data collection. Both processes were completed towards the very end of this study; therefore, no data collection regarding these aspects was conducted after their completion.

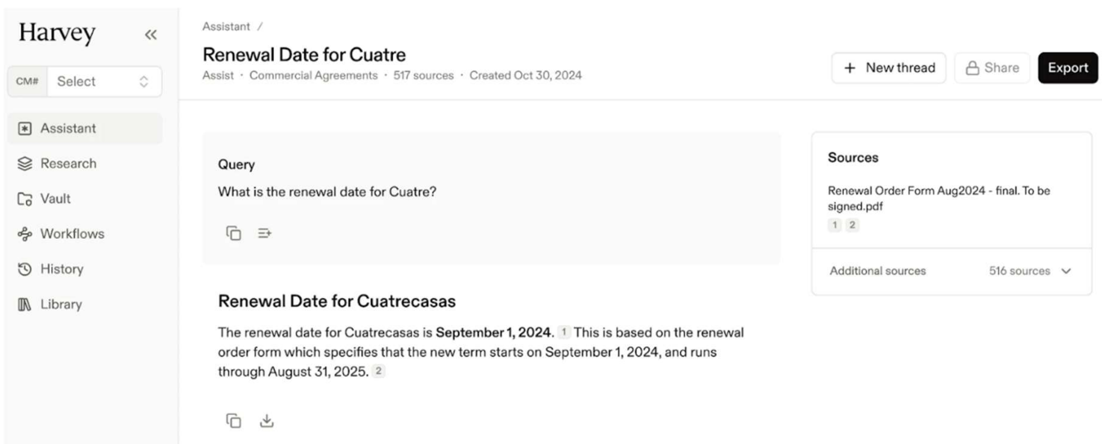


**Figure 4:** **User interface of GenAI tool Harvey AI, adopted from LaBarre (2024)**

Within the Transfer Pricing (TP) business unit, Transfer Pricing Artificial Intelligence (TPAI), a GenAI tool provided by the global PSG network, is used when performing benchmarking, a central task for TP associates. Contrary to traditional LLMs, this GenAI, based on prompts, filters benchmarking firms by analyzing their websites, and selecting firms that meet the criteria for benchmarking to the tested party. This service is provided by the global network of PSG, meaning that the architecture and technical details of this software are unknown to the local entity of PSG and subject to the global firm.

The firm is also in the midst of a transition from Google services to Microsoft 365, a process ongoing for approximately two years. As part of this transition, the firm is evaluating and developing Microsoft's Copilot for integration into the new environment. However, the future of this initiative remains undecided.

This study focuses on ChatPSG and Harvey, as they generate similar textual outputs as opposed to TPAI, for which there is less accessible information due to its global scope. Additionally, ChatPSG and Harvey are utilized across various tax areas, making their applicability broader within the tax profession. In contrast, TPAI is specifically designed for a single task in TP, namely, benchmarking. This broader applicability of ChatPSG and Harvey also enhances the transferability of the findings to PSG's other service streams, but also to PSFs in general.

## 3.3 Data Collection

To comprehend the internal processes and the use of GenAI tools, interviews with various stakeholders within the firm were conducted. The interviews were conducted with PSG tax professionals, as well as AI transformation leaders. Furthermore, observations also contributed as a source for data collection, as well as secondary data provided by PSG.

### 3.3.1 Interview Context

Given the extensive organizational structure, some interviews aimed to provide an overview of the GenAI transformation and its evolution within the firm. For this purpose, Respondent 1, a partner in Tax responsible for the GenAI program (including Harvey AI and ChatPSG) organization-wide, was interviewed. This offered a comprehensive view of the firm's GenAI transformation, top management initiatives, support structures, and strategic outlook. Additionally, interviews were conducted with Tax Director Respondent 2, who oversees Harvey AI across the organizational units utilizing the technology.

- *Respondent 1: Tax partner responsible for the AI transformation locally (Harvey AI and ChatPSG)*
- *Respondent 2: Tax director responsible for Harvey AI locally*

These two interview subjects served as a means to understanding the overall background of the transformation process, and provided insights into what has happened, and what is currently happening at the firm. Furthermore, it enabled me to guide myself in the organization and find cases and interview subjects interesting to my study. One such instance was learning that the TP team were at the forefront of adopting this technology, and that they had employed effective strategies at utilizing this technology. Considering this factor, I chose to examine the TP professionals in detail.

Since this thesis focuses on how TP professionals adopt the technology in their work, it is worth understanding what their work entails. TP refers to the pricing of transactions between associated enterprises, typically within a multinational enterprise (MNE) group. It ensures that transactions between related entities are conducted at arm's length, meaning the prices are consistent with those that would be charged between independent enterprises. This is important for tax purposes to ensure that profits are taxed where economic activities generating the profits are performed and where value is created. Furthermore, external comparables, which are other firms that should be benchmarked to the tested party, need to be identified and assessed based on their comparability. Through performing adjustments to comparables or other variables, the arms-length range remuneration is determined (OECD, 2022).

Furthermore, there are two key components of transfer pricing documentation: the master file and the local file. The master file provides a high-level overview of the MNE group's global business operations and transfer pricing policies. It is intended to give tax administrations a comprehensive understanding of the MNE's global structure and activities. The local file instead provides detailed information on specific intercompany transactions that are material in the context of the local jurisdiction's tax system. It supplements the master file by focusing on the local affiliate's transactions with associated enterprises in different jurisdictions (OECD, 2022).

Understanding the factors driving and hindering this adoption, and how the technology has transformed their work, offers valuable insights into the impact of AI on tax professionals' roles. Consequently, interviews were conducted with four TP associates, one TP manager, and one TP senior manager. To provide a balanced perspective, interviews were also held with two AI-skeptical and hesitant adopters: a partner in Entrepreneurial Tax Services (ETS) and a senior manager in Real Estate Tax (RE).

Interestingly, certain professionals within TP have pivoted into roles where they spend half their time in tax related projects, while the other half is dedicated to working with GenAI and other technologies related to it. Respondents 5 and 7 are such professionals, and their interviews provide further material on how their work has transformed with officially dedicated roles to work with the technology.

For clarity, interviewees are categorized as either adopters or AI transformation leaders. Table 2 summarizes the interviewees' characteristics and provides an identifier for each responder.

**Table 2: Summary of respondent's characteristics**

| Number | Title | Category | Business Unit | Stage 1 | Stage 2 |
|---|---|---|---|---|---|
| 1 | Partner | AI Transformation Leader | Internal Tax & Compliance Services (ITCS)[*] | Conducted | |
| 2 | Director | AI Transformation Leader | Tax Technology | Conducted | Conducted |
| 3 | Partner | Adopter | Entrepreneurial Tax Services (ETS) | Conducted | |
| 4 | Senior Manager | Adopter | Real Estate (RE) | Conducted | |
| 5 | Senior Manager | Adopter | TP (50% tech role) | | Conducted |
| 6 | Manager | Adopter | TP | | Conducted |
| 7 | Associate | Adopter | TP (50% tech role) | Conducted | Conducted |
| 8 | Associate | Adopter | TP | | Conducted |
| 9 | Associate | Adopter | TP | | Conducted |
| 10 | Associate | Adopter | TP | | Conducted |

[*] While Respondent 1 is part of ITCS, he focuses on the transformation and less on tax related activities in this domain.

A two-stage interview process has been conducted, where the first stage served as an exploratory stage, employing open-ended questions. This stage utilized responders from different business units, and transformation leaders, to inform and guide the study. As TP was chosen for closer examination, gathering data in the second stage was more structured, consisting of semi-structured interviews.

Given the exploratory nature of this study, the use of unstructured and semi-structured interviews is appropriate (Saunders et al., 2015, p. 392). Unstructured interviews provide in-depth insights into ongoing processes and their contexts (Saunders et al., 2015, p. 393). Since the context of PSFs is distinct (Von Nordenflycht, 2010), unstructured interviews were used to establish an understanding of the nature of the PSF and the technological transformation brought about by AI at PSG. With an enhanced understanding of these aspects, semi-structured interviews were employed to explore the relationships between the PSF context and the AI transformation. This approach is suitable, as semi-structured interviews can effectively examine relationships between variables (Saunders et al., 2015, p. 393).

### 3.3.2 Observations and Secondary Data

As mentioned, I spent five months working in the tax technology team, and this enabled me to observe tax professionals in their daily work but also gain insights into how various stakeholders (including my team) drive and implement technical solutions, often data-related or automation scripts, to different Tax business units. Furthermore, through discussions with various stakeholders, I was presented with demonstrations of digital technologies, including AI technologies and their specific features.

A wide range of secondary data was utilized, including internal documents, intranet content, the AI technologies employed by PSG, along with documentation related to the firm's GenAI solutions and change work. Furthermore, I incorporated ChatPSG into my own work, to increase my knowledge in the technology but also to get familiar with the AI technologies at the firm. Moreover, to increase my understanding of the different AI technologies and how they work, I spent time researching their technical architecture and limitations.

This approach not only provided a deeper understanding of internal recommendations and guidelines but also contributed to data triangulation, complementing the insights gathered from interviews. By employing triangulation, the overall validity and reliability of the methodology were strengthened (Saunders et al., 2015, p.320).

## 3.4 Data Analysis

The interviews were voice recorded, and transcripts were generated automatically using an AI transcription feature within Microsoft Word. The transcripts were then manually corrected by listening to the recording and editing it through removal of mistakes and adding statements that were not captured by the auto-generated transcripts. The reason for employing AI for transcribing the recordings was simply to save time in doing so, as listening through the recordings and checking the transcript, made sure they were accurate.

The method utilized for data analysis cannot be classified under any specific qualitative data analysis method. However, parallels can be drawn to Thematic Analysis, as the findings from the interviews and other collected data were re-analyzed and classified into themes that can be generalized over all collected data, which are central practices when utilizing Thematic Analysis (Saunders et al., 2015, p.579). However, where it differs from said method is that none of the data is coded using a systematic approach, where data units, consisting of separate words or short phrases, yielded from interviews are labeled into themes that summarize its meaning (Saunders et al., 2015, p.580).

Instead, an iterative and reflexive approach inspired by Wensink (2024) and Engwall et al. (2021) was employed. By abstaining from coding, and iteratively revisiting the data throughout the study, the meaning and its context to the case could be better understood, as I constantly, through a reflexive method, was able to challenge my earlier assumptions and understanding of the data. This was made possible by consistently aligning my analysis with observations and secondary data through triangulation. Furthermore, through conversations with supervisors and stakeholders at PSG, I was able to refine and reassess the meaning of my data to more accurately reflect the context. This resulted in a non-linear and reflexive approach, which is fitting to the qualitative nature of this study (Alvesson & Sköldberg, 2017).

Furthermore, participant validation was utilized in this study, which is a method for increasing the validity of findings by allowing participants to review and validate the research results (Saunders et al., 2015, p.207). This process involved holding a presentation to showcase the findings at PSG, during which participants could comment and correct any inconsistencies, enhancing my overall understanding of the analyzed data and the case examined.

# 4 Empirical Findings

At PSG, GenAI has currently been integrated primarily as an augmentation to the work of professionals. Initially, I detail how this augmentation has shaped perceptions of the implementation process, highlighting a perceived slow pace of adoption and how fears of automation over time evolved into a more positive outlook. Furthermore, as GenAI is a novel technology, I, similarly to Retkowsky et al. (2024), explore its use cases to understand how it transforms the professionals' work.

However, the augmentation leads to effectiveness of this technology being contingent on how it is used, as increased disparities in performance are reported compared to prior to the technology's introduction. Given that the outcomes heavily depend on usage proficiency, it becomes crucial to investigate the challenges that the professionals perceive in their augmented use, as this affects employed mechanisms. Thus, I present the relevant obstacles perceived by professionals in adopting this technology.

In light of these challenges, I examine how AI is integrated into the professionals' workflows, leading to the emergence of a method-based approach. This approach comprises specific methods aimed at either enhancing work performance or facilitating the learning processes related to their implementation. Accordingly, I delineate two conceptualizations of approaches: those I refer to as *implementation-methods* and those I refer to as *learning-methods*.

Finally, I provide a forward-looking perspective on the adoption of GenAI at PSG, where expertise surrounding GenAI adoption is initially developed through internal learning before being transferred to clients. However, senior professionals express concerns that reliance on this technology may ultimately result in knowledge loss among junior professionals, potentially hindering their ability to effectively contextualize AI-generated insights when they eventually move on to senior positions.

## 4.1 Current Stage of AI Adoption

The current stage of AI adoption is signified as slow, with promised transformation of the business yet to be delivered. Furthermore, the adoption is mainly characterized by augmenting the processes of the professionals, as opposed to automating them. However, this leads to potential increases in performance disparities, as those being more proficient in augmenting their work reap the benefits to larger extents compared to before the introduction of AI.

### 4.1.1 Use Cases

There is a proposed slowness in the transformation process that is quite evident over the organization, and that the change promised when this technology was first introduced has not yet been fulfilled. But that does not only pertain to the firm specifically, but also something that relates to the market in general. *“Two years ago, maybe when we got this program and started, it was like this. In two years, the market will be completely different. Now two years have passed. And it is different, but not very different"* (Respondent 3). This is also reflected upon by Respondent 2 who is leading the introduction of Harvey, when he states, *“we thought that the whole world would be turned upside down within six months, but that hasn't happened yet.”*

This perceived slowness is explained as a reason for why professionals, who exclaimed themselves to previously have been highly concerned about the technology, to now adopt a more positive outlook on it. Two perspectives are presented by Respondent 3, “*Either it's a fantastic tool that we are in* *the process of learning to use, which, with a bit of hindsight, might be what it leads to, but there is also the other side that, [...] we are in the process of rationalizing ourselves, our profession, away*". Emphasis is being put on the current stage of the technology's effects and is viewed as a tool that helps the professionals become more efficient. While being critical at first, Respondent 4 expressed the introduction of AI as having turned out *"better than expected",* while Respondent 3 explains: *"this dystopian view has perhaps calmed down a bit, [...] we should [instead] probably see this as a very good tool."* This switch in view is also evident among the junior professionals. Respondent 8 describes his work as being quite manual and initially being concerned that AI would reduce his work. *"More than reduced I think it just has streamlined it a bit."* Hence, an initial dystopian outlook on the technology seems to have shifted towards a more utopian one, where AI serves as a complement to the professionals' work, as opposed to being perceived as a threat to replace them. As Respondent 5 puts it *"those who get it understand—this is not magic; currently, it's just an assistant."*

The identified use cases suggest that most of them serve to augment the work of professionals rather than replace their core functions. As illustrated in Table 3, these use cases provide complementary support that enhances productivity and broadens the professionals' capabilities.

**Table 4: Overview of use cases and descriptions**

| Use Cases | Descriptions |
|---|---|
| Brainstorming | Supports professionals in the brainstorming process by providing alternative perspectives, encouraging broader consideration of options when approaching tasks. |
| Drafting documentation (local and master file) | Assists in drafting local and master files, by helping professionals integrate more comprehensive and relevant content. |
| Drafting email | Enhances email drafting by assisting with structure, grammar, proofreading, and overall clarity. |
| Data extraction & analysis | Extracts key information from documents and analyzes it according to the specific requirements of a task. |
| Summarizations & translations | Facilitates the summarization and translation of legal documents and cases, making them accessible to non-Swedish-speaking professionals. |
| Rapid research | Enables professionals to quickly gather and integrate relevant information on a topic into their work, streamlining the research process. |

A commonly utilized use case among the professionals is brainstorming, *"AI helps us in brainstorming, say, what could be more that you can think about? And why not this?"* (Respondent 6). This is especially useful in TP practices, as Respondent 6 explains: *"In TP, there's a lot of room for, you know, debating, and there's room for different argumentation. So, in that sense, AI helps us transfer pricing specialists to get a different point of view."* This demonstrates how AI enhances the professionals' ability to broaden their perspectives and fosters more effective discussions, illustrating its role in augmenting their expertise.

Professionals also utilize Harvey to draft local and master files, as these documents must be provided to tax authorities by the client firm. Respondent 8 describes Harvey as an excellent tool for drafting documentation in unfamiliar industries specifically. He cites the medical software industry as an example, noting that when *"drafting something, you might overlook certain aspects. A medical software provider does more than just sell software; it also offers after-sales services, which I might not consider because I'm not involved in the industry."* Hence, utilizing AI when drafting documentation, he explains, *"makes my knowledge broader."* Again, the use case extends the accessibility to knowledge, augmenting the professionals in their writing.

Both Harvey and ChatPSG are extensively used to write emails, and more specifically they are used to formulate emails, correct grammar, proofread and translate writing. As Respondent 4 puts it:

> *"A lot of what might otherwise take a lot of time where you're sitting and pondering a sentence and how to phrase it can now somehow be written as it comes to mind, even though it's something you'd never send to the client. But then it will still give me a suggestion that turns into something. So, the time from jotting down the first thought to having a complete draft becomes much shorter because you don't have to formulate every word individually".*

Again, the professionals display working with the AI in writing the emails, suggesting augmentation of the process. Respondent 7 further explains that he opts for ChatPSG for simple tasks like email-writing, since he knows that Harvey is more costly, and that any of the GenAI tools are good enough for this purpose. However, Respondent 6 uses Harvey to write emails because she likes the structure and style of the text it provides. These indicate differing approaches for the same task, which is a theme prevalent throughout the use cases.

Data extraction and analysis is also a common use case among professionals, where most use cases emerge in extracting unstructured data, as structured data is extracted using software automation scripts. As Respondent 7 explains *"extracting data from large documents, say, 100 to 150 pages, when you need a specific value or text on a page is made easier with Harvey."* Respondent 9 also analyzes annual reports and extracts key information, stating, *"I can ask, for example, where to find the operating profit or how to calculate the EBITDA [(Earnings Before Interest, Taxes, Depreciation, and Amortization)], and Harvey provides the information."*

Extraction and analysis of data is used in a wide range of cases, and one common use case among senior professionals is conducting a so-called *red-flag analysis*, where tax professionals identify potential tax-related issues in client firms by uploading their annual report. This is done in multiple tax domains, and as Respondent 5 illustrates through an example *"Now I have a client meeting with Company X in 5 minutes [...] so here are these 20 red flags in the tax area—that I get just by uploading the annual report."* He further elaborates on how this saves him time and can make him quickly prepare for the meeting: *"[...] within 3 minutes, I know, okay, these are the things I need to highlight, or these are the people I need to include in the meeting with this client because they might have issues in these areas."* Hence, the task of data extraction and analysis is partly automated, however, the ability of analyzing and advising clients is augmented.

AI can also aid in summarizing and translating information. Respondent 5 continues, *"We have the ability to gather more data and compile it, giving us a better overview. We can now do things that were previously impossible, such as creating a database of legal cases with summaries and translations that everyone on our team can access, not just those who speak Swedish."* Both Respondents 6 and 9, who are English speakers and not proficient in Swedish, agree that Harvey significantly improves their access to Swedish legal resources efficiently and accurately. Respondent 6 emphasizes the ability of Harvey to summarize Swedish case laws, as she sometimes does not like the word-by-word translation of it into English. Summarization is also valuable for providing an overview of meetings, as key takeaways from recorded sessions are summarized and presented intuitively to the professionals.

Tax professionals also benefit from AI when rapid research is needed. Respondent 8 illustrates this with an example of a client-meeting he had, *"Someone might ask, 'In Sweden, what are the thresholds for providing transfer pricing documentation?' Instead of extensive Googling, I ask Harvey for an explanation of debt capacity and specifics on Swedish thresholds, and it provides the needed information."* Respondent 6 expands on this by recalling a case involving a question that was outside her primary expertise (TP) and included a corporate tax component. *"Before consulting a corporate tax colleague for the answer, I thought to do some initial research myself."* She utilized Harvey and reviewed relevant sources to draft a response. She then approached her corporate tax colleague and shared her findings, to which her colleague responded, *"yes, it makes sense."*

The outlined use cases demonstrate that they primarily enhance professional capabilities by automating simpler tasks. Activities like brainstorming and drafting documentation allow professionals to gain diverse perspectives, enriching their insights. Moreover, use cases focusing on data extraction, analysis, summarization, and translation significantly improve information accessibility, resulting in what Respondent 5 describes as a *"better overview."* Although tasks such as summarization and translation are automated, they are typically minor in scope, ultimately enabling professionals to focus more effectively on their core responsibilities. Thus, at this stage, the technology plays a crucial role in augmenting the abilities of professionals

While there is a perception of slowness in the augmented work among tax professionals, all professionals within TP consistently report spending less time on tasks since the introduction of AI. For instance, Respondent 10 observed, *"Just in one year, so much has changed,"* while Respondent 7 remarked, *"The work really goes much faster, simply put."* Additionally, in my scoping interview with Respondent 1, TP emerged as a compelling example of successful AI adoption. Initially, some junior staff were given access to Harvey, leading managers to report efficiency increases and improvement in the quality of work. However, there are indications that these outcomes depend on the "correct" use of the technology and may increase the disparity in performance between professionals.

### 4.1.2 Use-dependent Outcome

Due to the augmented transformation of the tax professionals, the effect of using AI is determined by their skills in utilizing it. Professionals often develop slightly different methods for utilizing AI technologies, validated by Respondent 6 who notes, *"Everyone has their own personal best use cases [...] It just depends on how comfortable you get using it."* Given this, Respondent 5 emphasized that efficiency gains vary greatly based on how AI is utilized, stating, *"It's not everyone that becomes more efficient by using AI. Only those who know how to do it, who try, and who have been involved from the beginning—those who find it fun—they become extremely efficient."* Respondent 7 in accordance, describes AI as a superpower *"for those who know how to use it."* This proposes that the outcome depends on the use of the technology.

Furthermore, Respondent 5 suggests that a growing disparity in efficiency is emerging between those skilled with the technology and those less adept.

> *"In the past, you could have one person doing the work for you, and it might take, well, either six or eight hours for the same job. [...] Now, the difference is so much greater—it can be like this: one person might do it in one hour, while for another, it might take 25 hours for the same work, and it depends on how they use the tools."* (Respondent 5)

This has led Respondent 5 to collaborate more with junior professionals skilled in AI use to perform tasks. Furthermore, he expresses frustration with those less proficient, often choosing to complete the work by himself: *"I'd rather do it by myself, instead of giving it to a junior—it's so quick anyways. Why should I wait two weeks for them to do it?"* He attributes this to the iterative process of editing the juniors' work, providing feedback, and issuing new instructions. While the general consensus among interviewees is that AI will not replace tax professionals but rather that tax professionals embracing AI will replace those who don't, proficiency seems to remain essential. Reflecting this sentiment, Respondent 5 states*, "honestly, I feel more and more that I just want to replace most of the juniors with AI."* This proposed disparity in performance calls for understanding of how AI is used among the professionals. However, since the use is highly influenced by the challenges posed by augmented workflows and the organizational context, identified challenges are covered next.

### 4.1.3 Challenges in Augmented Workflows

To better understand how professionals adopt the technology effectively, it is important to first examine the limitations they perceive and how they are navigated. This section therefore outlines technical and organizational challenges observed among PSG tax professionals.

#### *4.1.3.1 Hallucinations*

As respondent 3 explains *"during these two years, it has been the case that when AI cannot find an answer, it lies."* Indeed, the phenomenon of hallucinations in AI-generated outputs presents a unique challenge for the professionals. Respondent 6 highlights an intriguing distinction between a junior professional and AI, noting that *"a junior [professional] does not hallucinate."* Respondent 8 points out that hallucinations are more likely in particular situations. *"Sometimes you want a market analysis skewed in a certain direction. For example, if an industry is struggling, you might want to create the impression that it is not performing well. In this case, Harvey might pull sources but also hallucinate information."* Nevertheless, hallucination is evident as a challenge among the professionals.

#### *4.1.3.2 Potential Data Breaches*

As one of the largest traditional audit firms, PSG's main business has been, and remains, helping clients comply with regulations. Due to PSG's client-heavy work and the handling of substantial amounts of confidential data and insights, both the services provided, and PSG itself must adhere strictly to regulations. This adherence is crucial not only to avoid regulatory breaches but also to maintain client trust.

All client specific information that is not publicly shared, needs to be anonymized before being put into Harvey to avoid data breaches. Respondent 2 explains that this is the main challenge that adoption of Harvey faces within the organization, as anonymization can be difficult for certain tasks. The professionals do the anonymization manually, where they remove or replace any client-specific information before inputting it into the AI. Harvey is the desired choice for domain specific tasks, however, sometimes the extensive need for anonymization makes them resort to ChatPSG. As Respondent 9 says, *"sometimes I go to ChatPSG when it's a lot of client sensitive information, because the purpose is to save time right?"*

The risk of data breach is a fundamental reason for Respondent 4, who is from the Real Estate Tax stream, to completely refrain from adopting Harvey, as she states, *“It feels safer when you don't have to worry about whether you're handing out something you're not allowed to, as that's one of the biggest risks we have, sending out something you're not allowed to disclose. If you don't need to feel that worry, it becomes easier to use that tool without having to think about it."*

Respondents 2, 5, 6 and 7 all state that there can be more use cases unveiled with the ability to input client information. Furthermore, Respondent 7 argues that the possibility of removing client specific data will develop new use cases for other tax streams specifically. Respondent 6 conveys that the limitation of not being able to use client specific information is not as detrimental in TP, as opposed to other tax streams, as *“they have legal documents, which are confidential, so they cannot input it there.”*

#### *4.1.3.3 Time Allocation*

Throughout the organization, time allocation poses a significant challenge for stakeholders, as described by Respondent 1: *"It is a professional organization, we are very ambitious, [and] we have an insane amount of work to do every day."* This issue is further supported by the observations of Respondents 2 and 4, who note that tax professionals need dedicated time to learn and effectively utilize AI technologies. Respondent 2 explains, *“People need to set aside time in order to actually save time but find it difficult to see the immediate benefit and, in a stressful everyday life, try to set aside time to try to incorporate this."* Respondent 4 echoes this sentiment by emphasizing that client delivery is her top priority, leaving little room for experimenting with AI tools.

## 4.2 Enhancing Outcomes: Methods for Implementation and Learning

In the midst of the augmented use cases and the challenges presented, certain methods employed by the professionals, conducive to successful implementation, emerge. These either serve as counteracting the challenges, improving the outcome of the work or both as these are not mutually exclusive. The methods are conceptualized as either *implementation-methods,* meaning methods implemented by professionals when using AI in their daily tasks, or *learning-methods*, concerning how to learn and develop the implemented methods and use of the technology.

### 4.2.1 Implementation-Methods

Respondent 5 articulates quite well what it means to work with tax deliveries, explaining that working with AI is essentially equivalent to how one would previously engage with their own work. He puts it this way:

> *"[...] if you are a good lawyer and tax advisor who wants to deliver a quality product, it is never the first thing you write that is the final product. No, you set it aside for a bit, you come back, make some adjustments, read it over again, add and remove. That’s how you should work, and in exactly the same way, you should work with AI here. You can’t just take the first product and be satisfied with it, no, you have to go in and work with it."*

In line with this notion, the implementation-methods developed by professionals primarily revolve around two key actions: *revising* and *validating*. These actions specifically target the AI's input and output and are often non-linear in nature. Professionals perform tasks to revise the input before it is processed by the AI. For the output, they validate the results generated by the AI while also engaging in additional tasks to revise the output as necessary. Given that there is a necessity to *work* with the input and output, and that this is *implemented* trough methods such as revision and validation, this section explores these critical, non-linear implementation-methods in detail.

#### 4.2.1.1 Revising Input: Anonymizing & Prompt Engineering

To mitigate the risk of data breaches, TP professionals employ manual anonymization. Respondents 8, 9, and 10 indicated that, although this process is manual, it is manageable because they can quickly remove identifiable information from the input data. This involves manually eliminating client-specific or personally identifiable information, such as names and social security numbers, ensuring that the data is no longer linked to the individual or organization it pertains to.

Respondent 5, who dedicates 50% of his role to technology, focuses on the development, testing, and refinement of input prompts for AI tools like Harvey and ChatPSG. He notes that for prompts extracting data, the initial accuracy is around 55%. To improve this, he tests prompts with various inputs, exemplifying this with the development of red-flag analysis using annual reports. By experimenting with different currencies, languages, and formats, he analyzes the outputs and adjusts the prompts accordingly. *"Usually around the 17th or 18th version, the accuracy is around 98-99%, and that's when it is completed"*, he explains. He also mentions that feedback from other professionals using his prompts helps him make further enhancements. Respondent 5 has extensively worked on prompt development, and with the assistance of Respondent 7 and other tech enthusiasts, he has created a prompt library for Harvey, which not only serves TP but also shares use cases with other tax streams.

#### 4.2.1.2 Validating Output: Source Checking, Independent Research & Feedback

By conducting independent research and verifying sources provided by Harvey, professionals can mitigate the risks associated with hallucinations. Respondent 8 illustrates this with an example involving a client in the financial sector: *"If I ask Harvey for an industry and market outlook for 2023, 2024, and 2025, it provides a rough text with linked sources, allowing me to verify them and refine the analysis."* Respondent 9 echoes this approach, emphasizing the importance of source checking: *"You cannot just blindly trust the tool. That's not the purpose of it. So, you have to check it, but it saves a lot of time because you just don't have to write that yourself. You can just edit it and make it better."* Furthermore, Respondent 8 explains that reviewing sources isn't sufficient, adding, *"you also need to conduct independent research to understand the industry—its growth and current trends. If this doesn't align with Harvey's output, you need to edit it."*

Moreover, Respondents 5 and 10 provide feedback to Harvey by rating the generated outputs through a built-in feedback feature. Respondent 10 notes that this feedback is *"very quickly implemented,"* as he has observed the startup integrating his input into the system, leading to an increased frequency of his desired outputs. Respondent 5 has especially taken on this role, as he works extensively with developing prompts, but also due to his seniority, knows what the desired output should be.

#### 4.2.1.3 Revising Output: Humanizing, Contextualizing & Continuous Prompting

Revising the AI output based on the context and specific client needs is a common practice among professionals. Respondent 7 emphasizes that many make the mistake of having AI write text for managers without human revision, noting, *"it's immediately noticeable that it's AI-written and doesn't meet the needs. If they had spent even half as much time doing it manually, it would have been of better quality and more professional."* He also highlights the importance of aligning the AI-generated content with the firm's style, saying, *"I have become better at using AI. In the beginning,* *I sometimes copied and pasted the output from AI into the result, but now I adapt it to the [PSG] style."*

Respondent 5 emphasizes the importance of humanizing the output, explaining that a skilled tax professional continually works on refining the text. The same approach should be applied when using AI: *"You get a base from AI, then you make the text human with small adjustments, and I think that's how you use AI correctly"* (Respondent 5). He further explains that using AI to generate only certain parts of a text can result in incoherence with the rest, which is undesirable.

While manual revision is a common method for editing output, using AI for this purpose is also prevalent. This is characterized by continuously prompting the AI iteratively until the desired output is achieved. Respondent 6 illustrates this by describing her process for drafting emails to clients, where she asks Harvey to either enhance her existing email or generate a new one. By providing feedback through prompts to the AI, on elements like desired heading titles and where the text should be bulleted, she guides the AI to produce her preferred output.

### 4.2.2 Learning-Methods

Professionals have identified use cases and developed implementation-methods through a combination of experiment-based learning and collaborative efforts. This section explores how professionals acquire this knowledge and how it is collaboratively shared throughout the organization.

#### 4.2.2.1 Experiment-based Learning

Experimentation is viewed as crucial for learning to work with AI. Respondent 5 underlines the importance of experimentation, stating that it is instrumental in discovering new use cases. He explains, *"sometimes, in certain projects, it's not immediately clear whether using AI will be beneficial. But if you don't try, you won't uncover its potential effectiveness in these projects."*

Responding 6 explains the importance of evaluating the possibility of integrating Harvey into a task: *"you have to constantly remind yourself, okay, when I'm doing something, can I use Harvey for it?"* TP professionals commonly remarked that they experienced management support in experimenting with the technology, as they are encouraged to spend time working with the AI technologies. This is highlighted by some professionals within TP who have been given 50% dedicated tech-roles (Respondent 5 and 7 in this study). Respondent 7 highlights the firm's willingness to this cause, as *"[management] is comfortable with me working with AI, and they want me to help others use AI, and they want us to build applications with AI."*

#### 4.2.2.2 Collaborative Learning

While the organization has introduced a series of online lectures on AI usage, the professionals attribute collaborative efforts as a means of learning. When initially given access to the AI tools, TP professionals initiated a series of internal training sessions focusing on the technology, to share use cases and ideas. While this initiative was temporary and only at TP, tax professionals value "Harvey Community Calls", regular meetings open to all Harvey users, where use cases and updates are shared. Respondent 8 highlights their cross-tax benefit, noting, *"you kind of see what other people are doing and how they are engaging with Harvey, and you think, 'How could this translate to TP?'"* Respondent 7 appreciates these sessions for their educational value and the opportunity they provide to share his findings with the organization and educate others on AI use.

Furthermore, At TP's Stockholm office, a segment at the end of the weekly Monday meetings is dedicated to sharing AI insights, news, and use cases. In contrast, TP's Gothenburg office emphasizes technical sessions such as lunches and "fikas," where tax professionals, primarily from

TP but also from other tax streams, are encouraged to share their experiences working with technologies like AI. *"We always tell everyone, giving them examples of the use cases and telling them that you know, you just have to think about how you can use it."* (Respondent 6)

Respondent 7 explains that he and Respondent 5 actively inspire each other to work with AI: *"we often encourage each other, saying things like, 'have you seen this? Have you tried this? Great, now we can do this.' This way, we can explore it together, which I find makes it more fun."* This collaborative spirit is exemplified by Respondent 7, who, along with Respondent 5, developed a specific prompt for red-flag analysis and shared it throughout the tax organization. Respondent 7 explains:

> *"In the beginning, we did it for TP, then we showed it to ITCS [(Internal Tax & Compliance Services)], and they said that they must have it too. When ITCS was interested, we said, okay, we should also implement it for VAT [(Value Added Tax)]. So we went to VAT, and when the others saw it, they wanted their own as well. Now it has turned into a domino effect overall."*

Interestingly, Respondent 4 explains that her refrain from using Harvey partly stems from her lack of inspiration, and the proximity to help. *"First, I have to contact a colleague who refers me to the next one. Then I don't do it, but instead, I do it the old way because it goes faster."* She also explains that she would prefer to ask questions to someone she knows well, as opposed to someone she doesn't, as she expresses discomfort in asking *"dumb questions"*. This suggests that Respondent 4, who is in Real Estate and not TP, could be lacking collaborative efforts in her adoption of the technology.

## 4.3 Future Prospects of AI Adoption

Looking forward, PSG aims to position itself competitively in the AI playing field, as it is looking to commercialize the adoption of GenAI. However, the increasing reliance on AI technologies in client deliverables might have an adverse effect on the learning of the junior professionals, posing future issues in contextualizing AI generated output.

### 4.3.1 Leading by Example

Respondents 1 and 2 explain that there is a bigger purpose to adopting GenAI at the firm, namely, to commercialize it. Respondent 3 notes that PSG will need to embrace the technology due to competitive pressures and the evolving nature of AI, stating that they are, *"in a way, [...] forced to be involved and try to be at the forefront."* Respondent 2 explains that AI can help PSG in improving its service offerings in terms of quality and efficiency. However, both Respondents 1 and 2 emphasize how AI can turn into a commodity, selling it directly to customers. Interestingly, Respondent 2's duties are not only in developing and adopting Harvey at PSG, but also to bring it to clients, through selling subscriptions of the AI system. Throughout my time at the organization, Respondent 2 has been meeting with representatives from some of Sweden's largest firms, to pitch the AI tool and showcase its capabilities.

Furthermore, Respondent 1 brings up an initiative internally, where several partners and directors have been invited to in total three *"AI Labs"*, where the technology is demystified, and the best use cases are presented to the directors and partners. Respondent 1 states that the intention of these meetings is to make the professionals identify potential use cases among clients.

> *"[...] if a client describes a problem, they should think, 'wait a minute, they mentioned something about this at the lab I attended.' Then they can say, 'I think we have something in AI that could help you.'"* (Respondent 1)

Respondent 1 highlights that it is not only about providing AI solutions to clients but also to help them in their transformation. *"We are trying to drive this transformation partly internally. But we also want to help our clients with their transformation too."* Hence, the client focus makes the firm have another incentive apart from improving the delivered services, as they want to adopt GenAI successfully, so that they can aid their clients in doing the same. This is further expanded upon by Respondent 2 who explains: "*we do this to remain relevant, but we are also doing it for commercial purposes; the skills we acquire should be utilized and capitalized on, [...], through the offering of a service."* Furthermore, he explains that while PSG is a distinctly different organization compared to the clients it serves, the departments the tax stream work closely with *"have similar workflows and processes."* This means that developed knowledge in how to effectively utilize AI can still be applied as *"the implementation of AI and using AI in cognitive tasks is similar."* (Respondent 2)

### 4.3.2 Potential Unintended Consequence: Loss of Domain Knowledge

While junior professionals are more proficient in using AI, they may risk losing domain knowledge due to their growing reliance on AI technologies. Respondent 5, being a manager, explains that he can tell right from wrong in the deliverables, explaining: *"I often know when something is right or wrong because I've done it myself [...] and so I know what I want to write about and what I think is right to write."* Hence, seniors can and quickly assess the accuracy of the work produced, whereas juniors need to learn this through experience and research. In fact, this is a common practice as junior professionals' hand over the work to seniors for review, as Respondent 8 explains, *"you're going to have to give it to a senior to review, so they immediately know if it is incorrect."*

The way the senior professionals have acquired this knowledge is through traditional methods when doing their work, prior to the introduction of AI, through extensive manual research, involving extensive reading and writing of legal documents. As Respondent 5 explains, *"during that process, you learn a lot yourself. And then you learn to see if someone else is doing it the same way, if what you're doing is right and you learn what is right and wrong."* Further, a dilemma emerges among senior professionals, in how today's junior professionals should learn, when the traditional methods are increasingly being replaced by AI assisted ones. As Respondent 5 puts it: *"[...] now you can take very quick shortcuts if you know how to use AI correctly, but you might not learn as much."*

Furthermore, managers have reasons beyond helping juniors learn to prefer traditional methods, namely, that they can ensure that the AI generated output is correct. Respondent 5 explains: *"[...] so far you absolutely cannot trust any answer at all. [...] how is the new generation, who don't know and haven't learned in a traditional way, supposed to assess whether AI is right or wrong?"* Respondent 3 further elaborates on this, expressing concerns over a possible future where junior professionals risk obsolescence and those replacing seniors lack the ability to contextualize AI-generated knowledge. *"Then who will put it in context and explain it to the client, and see if this answer is correct?"* (Respondent 3)

There are, generally, two different approaches used by the junior professionals when working with AI, possibly affecting acquisition of domain knowledge: generating the output followed by research or conducting research before generating the output. These approaches are not mutually exclusive, as some professionals utilize both. Respondent 6 prefers juniors to first generate output using Harvey to guide subsequent research: *"You still have to go to Google, but then you will have something in your head when you're going to Google to ask."* She believes this approach structures both the output and the research process. Respondent 3 sees no importance in the sequence of using Harvey versus traditional research, emphasizing the need to develop both skills. However, he encourages juniors to perform sanity checks and prefers traditional research methods, such as analyzing legal cases independently. He warns that if tax associates only write prompts, their law degrees might become irrelevant: *"Then we could just hire people from high school and train them in writing prompts."*

Respondent 5 compares two junior professionals to illustrate differing approaches and how they affect learning. Associate A is quick and efficient with AI, while Associate B takes more time to research and learn, allowing him to better justify his outputs. Interestingly, A does the generation, and later research, While B instead does research before and during generation of AI output. Despite A's efficiency, Respondent 5 values B's approach for its learning benefit, stating a preference for B's approach *"for their own sake."*

Respondent 2, who is leading the introduction of Harvey, dismisses the concerns of domain knowledge loss by drawing an analogy, stating that *"Nobel prize winners today are not dumber because they have more tools at their disposal"*. However, he acknowledges that the role of tax professionals will increasingly revolve around understanding and communicating AI-generated knowledge while automating current tasks, allowing them to focus on more emerging complex responsibilities. As a result, he points out the future challenges of training new hires. Consequently, he is collaborating with representatives from the other Big Four accounting firms and academic institutions in Sweden to help shape the future of law degrees.

# 5 Discussion

The emerging AI augmented work is discussed from a mutual learning perspective, focusing on the learning dynamics involved and their implications for professionals in PSFs. Next, the acquisition of knowledge is discussed, relating it to literature. Following this, the future of the tax profession is discussed, given concerns and opportunities expressed by professionals and management.

## 5.1 Augmented AI Adoption

The identified use cases have similarities to both Wensink (2024) and Retkowsky (2024), such as drafting text and brainstorming, however, the tax professionals seem to also be extending the use to collecting and analyzing data. Similarly to Retkowsky et al. (2024), the use of the AI technologies mitigates constraints faced by workers. Thereby facilitating the augmentation of knowledge in their profession. The interplay between the knowledge augmentation provided by AI and the implementation-methods employed by professionals suggests the emergence of a mutual learning dynamic characterized by a "human in the loop" element, consistent with Raisch & Krakowski (2021). As suggested by Henriksen & Bechman (2020), a mutual learning mechanism could eventually lead to a redefinition of domain truth and facilitate the discovery of knowledge that is truer (Agarwal & Dhar, 2014). This perspective resonates with the findings, as the quality of output is suggested to improve through the augmented use of the technology.

In alignment with Retkowsky et al.'s (2024) findings, there appears to be a convergence (or intertwinement like Retkowsky et al. (2024) calls it) between professionals and the technology they utilize. As Waardenburg & Huysman (2024) suggest, the boundaries between use and development of AI are becoming increasingly blurred. While professionals do not typically collaborate directly with AI developers, they engage in identifying beneficial use cases and creating implementation-methods that help them navigate technical limitations. This contrasts with the collaborative approach described by van den Broek et al. (2021) and Grønsund & Aanestad (2020), where domain specialists work together to uncover truths through technical advancements. Instead, domain experts in this context tend to operate independently, refining how the technology is integrated into their practice*. This may be due to the nature of the technology; the outputs produced are neither inherently correct nor incorrect and up for discussion. Another, perhaps more reasonable explanation is the partnership with Harvey, as the development is completely left up to the start-up. How this translates to ChatPSG should be further explored, however, there was seemingly no interaction between professionals and developers of the technology at the Swedish branch.

However, the use of AI at PSG generates feedback loops where professionals assess the quality of the generated output, within the concept of RLHF, thereby excluding AI developers from direct involvement in this process. Interestingly, the feedback loops seem to differ to the ones proposed by Cook et al. (2024), as the manual revision of the generated AI output is not explicitly considered in the feedback. However, revision through continuous prompting could serve as a means of providing feedback to the AI, if this is captured by Harvey and ChatPSG developers. Instead, currently, the main utilized method for feedback seems to be RLHF, where the professionals rank the usefulness of the generated output. Furthermore, considering that senior professionals are the ones who know what the correct output is, they are more valuable in providing feedback to the AI developers.

Utilizing the insights from Raisch & Krakowski (2021) and Spring et al. (2020), reasons behind the phenomenon of augmentation within the tax profession can be explored. This profession is

* Harvey AI utilizes legal expertise in its development, however, the legal professionals at PSG and Harvey seems to remain disconnected apart from occasional feedback provision

characterized by complex tasks that are challenging to codify, largely due to its indeterministic nature. The augmented implementation-methods of professionals in collaboration with AI, can be contextualized through the concepts of auditing and altering outlined by Grønsund & Aanestad (2020). Professionals in the field employ auditing mechanisms to validate AI-generated output through various processes, such as source checking, independent research, and feedback provision. Additionally, they exhibit altering behaviors by adjusting input prompts through techniques like prompt engineering and anonymization, and by refining the generated output through humanization, contextualization, and ongoing prompting. Hence, what I refer to as validating could be viewed as auditing, while what I refer to as revision could be corresponding to altering.

Amidst these augmented developments, professionals' perceptions seem to have shifted from a dystopian viewpoint to a more utopian one. Initial concerns about automation potentially threatening their job security echo the findings of Strich et al. (2021) regarding loan consultants, as professionals could have felt that their role identity was being threatened. As the professionals increasingly recognized the necessity of augmentation instead of outright automation, their attitudes became more positive. This aligns with Retkowsky et al. (2024), where initial fears of automation gave way to a more optimistic perspective on the technology as a facilitator of augmentation. Another factor potentially explaining this shift may be the existence of symbolic conformity among some professionals, while others exhibit symbolic advocacy, as discussed in Pachidi et al. (2021). The group of professionals in the tax domain appears to have successfully adopted the technology, but those who remain skeptical may conform symbolically to the prevailing trends to safeguard their profession. This conforms to a broader management strategy that is determined to implement the technology, leading skeptical professionals to possibly react defensively.

The practice of humanizing AI-generated output has emerged as a compelling altering mechanism to enhance the quality of the outputs produced by AI. The need for this practice may stem from the observation that the professional characteristics of PSFs, as described by Von Nordenflycht (2010), diminish when relying on AI for knowledge creation—a finding supported by Kronblad (2020). To preserve professionalism in their output, professionals manually edit the AI-generated results. Thus, the drive to humanize AI output could be attributed to the client-centric focus of the profession, as human-generated outputs are often preferred over their algorithmically produced counterparts, in knowledge-related tasks (Lee, 2018). The significance of professional delivery in services may also be viewed as an organizational objective; resonating with Kim et al.'s (2024) case study on restaurant inspectors, who made better decisions by considering such secondary organizational objectives in their decisions. In the context of tax professionals, the absence of an organizational objective aimed at ensuring the professionalism of AI-generated output could compel professionals to intervene and implement humanizing practices.

As highlighted in the findings, there is no universally accurate output in TP practices due to their indeterminate nature; thus, what is deemed correct is influenced by the judgment of senior professionals. If these professionals do not consider the AI-generated output to be accurate, particularly if it does not align with their own knowledge, they may feel compelled to alter it. As Retkowsky et al. (2024) highlight a management concern in quality control, lower quality knowledge risks being integrated in the knowledge generation process. Perhaps this could be another explanation as to why the professionals utilize altering mechanisms. By intervening, potential lower quality knowledge might be identified and omitted or recreated into more quality one.

This phenomenon is related to Strich et al.'s (2021) study of loan consultants, where the need for output manipulation served as a protective mechanism for their professional identities. However, caution should be exercised when extending these findings to explore altering mechanisms in the context of this study, which focuses on augmentation rather than automation, particularly given the technological challenges such as hallucinations and potential data breaches, marking the necessity of both altering and auditing mechanisms.

Interestingly the performance-related outcomes of this technology are heterogeneous and seems to have shifted to technical abilities in utilizing AI, manifested through the auditing and altering mechanisms employed. This could be explained by the findings from the competitive advantages in centaur chess, presented by Krakowski et al. (2023). As AI increasingly substitutes knowledge generation in the tax professionals' work, the competitive advantage could be shifting to the utilization of AI, through auditing and altering practices. Hence, those who develop AI utilization capabilities could be the ones yielding better performance, explaining the increasing disparities in performance.

Further, the convergence between human and AI created output, suggests that when evaluating performance among professionals, the technical abilities in utilizing AI might emerge as more important than domain-specific ones. Similar notions are expressed in Retkowsky et al. (2024), as managers have issues in evaluating performance of junior workers. This does not, somehow, translate extensively to the case of TP, due to its traditionally, tech-driven nature. However, other tax streams that to a larger extent place value on domain knowledge might perceive an issue in this very context.

The technical limitations of hallucinations and potential data breaches are, currently, factors necessitating an augmented approach. However, bias did not emerge as a concern among the professionals, nor among the transformation leaders. This could be explained partly by Zhou et al.'s (2024) findings, as the architectural design of GenAI, limits the detection of bias in its output. As tax professionals' work does not generally revolve around a precise and sole correct answer, this could make the work more compatible with the technology, however, at the same time making it difficult to detect bias. Hence, an intriguing situation seems to emerge, where the underlying bias is not an explicit concern in the professionals' delivery of services to clients, potentially risking integrating AI bias into client-deliverables unknowingly. This suggests that professionals may be less inclined to raise concerns about bias, as it did not emerge as a technological challenge, unlike the findings reported by Yang et al. (2024). Since Yang et al. (2024) focused on AI technologies in general, rather than GenAI specifically, this may indicate that the perceived technological limitations of GenAI within PSFs differs from that of earlier AI technologies.

## 5.2 Learning Focus

The findings indicate that experimentation is an effective learning-method for utilizing AI and plays a crucial role in the adoption and development of use cases. This aligns with the conclusions of Wensink (2024) and Retkowsky et al. (2024), which suggest that experimentation is a central factor in the diffusion of GenAI. By sharing use cases through mediums such as regular meetings and workshops, a collaborative learning environment is fostered, inspiring the workforce to experiment with various applications within their tax stream. Thus, social support emerges as a key factor for effective adoption, aligning with the findings of Nesemeier (2024) and Blazquez et al. (2024).

While Retkowsky et al.'s (2024) study presents the introduction of GenAI tools from a bottom-up perspective, this case highlights a top-down approach, wherein the technology is implemented strategically by the firm. Nevertheless, the findings reveal that the creation of use cases and implementation-methods occurs through a bottom-up process led by the knowledge-workers. This may explain why experimentation proves to be an effective method: it encourages professionals to determine how the technology can be integrated into their work, rather than simply mandating its adoption.

The bottom-up approach could further be exemplified by upskilling of employees, as management-driven upskilling initiatives, such as online lectures, did not substantially empower professionals to develop their skills and knowledge in this area. This suggests that the introduction of GenAI requires localized upskilling, as use cases are often domain-specific, and experimentation occurs at an individual level in collaboration with peers. This was reflected in the findings, as the professionals attributed Harvey community calls and local upskilling initiatives as influencing their adoption.

The challenge of time allocation is also addressed through localized upskilling and collaborative sessions that set aside specific periods for working with the technology. This approach contrasts with the difficulty of finding time in an already busy schedule, where priorities often center on client service. Additionally, the reassignment of Respondents 5 and 7 to focus specifically on technology-related work further alleviates time constraints, facilitating the adoption of GenAI. Hence, TP expressed management support when they were given the opportunity to experiment with the technology, which they argue facilitated their learning journey. This can then further be attributed to the extensive resources large PSFs possess (Yang et al., 2024), facilitating adoption. So, while Nesemeier suggests social support outweighing management support, the management support expressed through resource reallocation, can effectively contribute to the adoption within this context.

Established players may be better positioned for AI adoption due to their abundant resources, which can be reallocated to support development and adoption efforts. However, their hierarchical organizational structures could pose a challenge. Effective bottom-up adoption requires initiative from junior professionals, which may be difficult if not supported by clear top-down communication. In other words, professionals accustomed to following directives from senior staff may struggle to implement AI into their workflows without explicit guidance. As a result, the hierarchical nature of these organizations may hinder the development of new implementation-methods as they should be driven by those directly engaged in the work.

## 5.3 The Future of Professional Services

The business model of PSG has yet not undergone any major changes due to the introduction of GenAI, corroborating the tentative findings of Kronblad et al. (2024a). However, the firm is positioning itself to transform its business model. Therefore, adopting the technology holds significant potential for PSG as it can capitalize on this transformation, potentially reducing the capital intensity of the business, aligning with trends in increased AI utilization (Kronblad, 2020). Thus, the firm views internal adoption not only as a means to enhance its service offering but also as a pathway to transform its business by offering GenAI transformation services to clients. Furthermore, the increased digital service offering is in line with Kronblad (2020), and that it pivots away from domain-specific offerings is also in line with Spring et al. (2020).

The firm-level incentive for adoption of the technology can also be explained using the perspective of competitive environment (Yang et al., 2024), as competitors' advancements in the use of the technology forces PSG to adopt it to retain its market-position. Moreover, the regulatory environment, as expressed by Yang et al. (2024), evident in the importance of compliance in the professionals’ work, shows itself in the findings mainly through the challenges of potential data breaches, as it is handling large amounts of client-specific data. Since both perspectives in Yang et al. (2024) are those of large PSFs, the findings suggest that PSG aligns with these perspectives.

Furthermore, concerns about future knowledge loss highlight that, despite the augmentation of knowledge, the learning process among professionals may suffer as a result. This is as traditional implementation-methods are substituted for AI-assisted ones, and due to the competitive advantage shifting as exemplified by Krakowski et al. (2023), the AI assisted ones will likely be preferred by professionals due to incentive in performance gains.

However, augmentation through research is proposed as a strategy to mitigate the effects of potential knowledge loss, and professionals at TP seem to employ this approach. Nonetheless, if knowledge is first produced by Harvey and subsequently serves as a guide for further research, as suggested by Respondent 6, it may limit professionals' exploration of the domain. Nevertheless, AI assisted work may lead to future difficulties in understanding, contextualizing and communicating AI generated knowledge, as increasingly human knowledge creation is passed on to AI. The comprehension and communication aspects are specifically important in PSFs due to their client centric nature and hence could create future issues.

Since this section discusses the future context, I believe the implications of a potential knowledge loss should be explored further and put in relation to the “human in the loop” perspective by Raisch & Krakowski (2021). If professionals experience domain knowledge loss while serving as humans in the loop, and if AI surpasses them in its knowledge creation capabilities, their augmented contribution could arguably be rendered obsolete, or nearly so. This could then lead to increased automation of the process, which is feasible considering the interdependent nature of automation and augmentation (Shollo et al., 2022; Raisch & Krakowski, 2021).

### 5.4 Limitations

It is difficult to study a technology that is at its infancy, and therefore there are several limitations to this study. Firstly, the adoption factors reviewed in the literature are limited, as only those considered relevant to the context of the study are explored. Hence, in Prasad’s (2023) findings when analyzing data from 108 organizations in India concerning GenAI adoption, only organizational size is related to the findings. Prasad (2023) suggests six determinants; complexity and regularity support acting as barriers to adoption, while compatibility, environmental uncertainty, organizational size and competition intensity act as enablers. Furthermore, Yang et al.'s (2024) findings suggest that there are six main factors influencing AI adoption, of them competitive environment and regulatory environment being the only ones taken into consideration in this study. Technology affordances, technology constraints, the firm's innovation management approaches and AI readiness are left up to future studies to examine in relation to GenAI in PSFs.

Due to the risks associated with professionals sharing unauthorized information, a limitation is that they may be unable to fully express their views on AI. For instance, the firm's Generative AI Business Rules, an internal document outlining the guidelines for AI usage, explicitly instructs that as a PSG professional to *"not suggest publicly [...] that you are actively using Generative AI to improve the quality of [PSG] deliverables."* Even though informed about the anonymization of the data collection in this study, this restriction may have impacted data collection, as some interviewees might have been hesitant to discuss their experiences with AI and its quality related implications. Another factor limiting the data collection is a potential pro-AI culture at PSG, which may lead to critical professionals refraining from voicing their concerns. For example, one of the respondents explained, *“Saying you are skeptical about AI is kind of saying that you are dumb; nobody wants to do it.”* This suggests the emergence of a culture where those wishing to criticize AI may be perceived as lacking understanding, thereby refraining from expressing their doubts or concerns.

# 6 Conclusions

This work explores how GenAI is adopted within an established PSF, through a single case study approach within its tax operations. By focusing on the practices employed by a team that is deemed successful in the adoption of the technology, the study addresses the need for understanding how GenAI is adopted in knowledge work. While AI evolves rapidly, its actual impact in this context appears slower and more gradual than expected. The anticipated disruption has not yet fully materialized, leading to a shift in perception from fear of automation to optimism about augmentation. Professionals are increasingly viewing GenAI as a tool that enhances, rather than replaces, their work.

In terms of use cases, I find that the introduction of GenAI at PSG has led to automation of some menial tasks, however, primarily serves as augmenting the capabilities of tax professionals. Both professionals and managers report improvements in efficiency and quality. However, these benefits are unevenly distributed and highly dependent on how the technology is used. This heterogeneity in performance highlights that GenAI's value does not lie in the tool itself, but in the practices that surround its use.

Furthermore, several challenges have emerged in the integration of the technology. One prominent technical issue is hallucinations, a phenomenon specific to GenAI, identified as a recurring concern among professionals. In addition, organizational challenges are also evident. For instance, the risk of data breaches significantly affects adoption, as professionals handle sensitive client information and are therefore at risk of compromising confidentiality. Time allocation poses another challenge; with a primary focus on client deliveries, professionals often struggle to dedicate time to developing effective uses of the technology. These challenges highlight key contextual factors influencing GenAI adoption in PSFs. Importantly, they impact performance, which appears to be closely tied to the professionals' proficiency in using the technology.

Amid these challenges, a method-based conceptualization is used to understand two distinct strategies that professionals use to navigate the challenges and enhance performance. These strategies consist of implementation-methods, which refer to the procedures for integrating AI into their tasks, and learning-methods, which encompass the ways in which professionals acquire this knowledge. The implementation-methods are characterized by auditing and altering practices of the AI input and output, while the learning-methods explain how the methods for use are developed and learned, through extensive experimentation and collaboration, suggesting importance of social support in the adoption process. In light of these methods, and the heterogeneity of the performance, the competitive advantage seems to be shifting from traditional domain-specific skills to AI utilization skills.

Lastly, the adoption of AI at the firm level is seen not only as a way to enhance client deliveries but also as a valuable learning opportunity. This knowledge can subsequently assist clients in adopting the technology, enabling PSG to provide novel service offerings. However, there are growing concerns among senior professionals about acquiring domain knowledge through AI implementation methods, and that junior professionals' acquisition of domain knowledge may suffer. Coupled with the observed shift in competitive advantage from domain-specific expertise to AI proficiency, professionals may become increasingly incentivized to prioritize learning how to use AI over building foundational subject-matter knowledge. These dynamics may hinder the next generation's ability to comprehend, contextualize, and effectively communicate AI-generated outputs as they move into more senior roles.

## 6.1 Managerial Recommendations

The adoption process of GenAI is different compared to traditional AI technologies in that it necessitates extensive learning efforts due to its domain-specific nature and personalized use. Managers should therefore understand that the adoption is bottom-up, and heavily dependent on experimentation and collaborative efforts from professionals. Therefore, facilitating conditions through creating an environment encouraging experimentation and collaboration, could serve as effective means for adoption. This could be done by freeing up time and giving the professionals the freedom to explore the technology themselves, as opposed to solely being assigned to client projects. This pertains mainly to large PSFs as they have vast resources that could be utilized in this manner, as was done in the TP team. Moreover, local upskilling proves to be highly effective given the domain-specific use cases of the technology and collaborative nature. Consequently, management should allocate more resources to enable teams to engage in this process, if they want to scale up GenAI adoption.

Furthermore, there are evident concerns among managers that the introduction of AI might lead to domain knowledge loss. While collaborative efforts between industry and academia could create an educational program more suitable for today's tax professionals, it is essential to not lose domain specific knowledge. This is to protect the augmented configuration of a human in the loop, as technical challenges still necessitate this. Hence, what I refer to as implementation-methods in this work, should be designed in a way to promote learning among the professionals, and become an established practice that is consequently with the development of the technology evaluated and improved upon. In the development of these, there should be senior professionals involved, considering that they can help tell right from wrong, and facilitate the development of implementation-methods that enable the professionals in delivering the correct output, but also engaging with the work in a way that helps them learn right from wrong themselves.

As Retkowsky et al. (2024) point out, challenges in performance review of junior professionals may arise, making it potentially beneficial to explore new methods for evaluating performance. Furthermore, the goal should not place excessive emphasis on efficiency but to also recognize and reward learning and the acquisition of domain knowledge, thereby retaining expertise within the staff by providing incentive to do so.

## 6.2 Future Work

Future work should further explore the adoption of GenAI in knowledge work and examine other tax domains, as they differ in the expertise they provide. Additionally, the study of GenAI adoption in PSFs through a case-study approach should be extended to other domains, such as accounting and technology consultancy services. Furthermore, the educational and learning aspects need further investigation to understand how the emergence of GenAI affects knowledge acquisition among professionals. Strategies to mitigate potential knowledge loss should also be explored.

This research suggests that performance disparities are increasing, likely due to the development of new implementation-methods, with those who embrace these changes gaining a competitive advantage. The relationship between this emerging competitive advantage and traditional competitive advantage in the context of knowledge work should be better understood in future research.

While the competitive advantages seem to shift, the business model remains reliant on hourly billing. The financial incentives for professionals to invest time in experimenting with the technology should be further explored, as such efforts may be seen as less profitable in both the short and long term, as increased efficiency would reduce billable hours. As a result, the business model transformation anticipated by Cook et al. (2024) may be hindered by the very incentives embedded in the existing structure. Future research should examine how these incentive structures related to the current business model influence AI adoption in PSFs.

The bottom-up approach should be further examined, particularly how its development differs from earlier forms of AI. This should be contextualized within the hierarchical organizational structures of established PSFs and its impact on technology adoption. Moreover, the implications of these findings for medium-sized and smaller firms with fewer resources should be explored, specifically whether they are better positioned in terms of organizational structure to adopt the technology.

Although bias was not identified as an issue in this research, it was noted in Yang et al. (2024). Future research should investigate the reasons for this discrepancy and explore whether factors beyond the technology's architecture contribute to difficulties in detecting bias. Additionally, organizations should be provided with strategies for addressing more challenging identifiable biases. Ethical implementation means should also be explored, taking into account long-term effects. Furthermore, given that PSFs conduct a significant amount of client-facing work, the black-box nature of GenAI and its potential impact on communication warrant further exploration.

## References


Agarwal, R., & Dhar, V. (2014). Editorial—Big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research, 25*(3), 443-448.

Alavi, M., & Westerman, G. (2023, November). *How generative AI will transform knowledge work*. Harvard Business Review. https://hbr.org/2023/11/how-generative-ai-will-transform-knowledge-work (Accessed 3rd of June 2025).

Alvesson, M., & Sköldberg, K. (2017). *Reflexive methodology: New vistas for qualitative research*. Sage.

Asatiani, A., Pekka, M., Nagbøl Rådberg, P., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems*, 22(2). DOI: 10.17705/1jais.00664.

Blazquez, V., Kim, Y., & Oh, T. (2024). Determinants of generative AI system adoption and usage behavior in Korean companies: Applying the UTAUT model. *Behavioral Sciences, 14*(11). https://doi.org/10.3390/bs14111035

Boulus-Rødje, N., Cranefield, J., Doyle, C., & Fleron, B. (2024). GenAI and me: the hidden work of building and maintaining an augmentative partnership. *Pers Ubiquit Comput 28*, 861–874. https://doi.org/10.1007/s00779-024-01810-y

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W.W. Norton.

Brynjolfsson, E., Rock, D., & Syverson, C. (2017). Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics [White paper]. *National Bureau of Economic Research.*

Carlsson, B. (2004). The Digital Economy: what is new and what is not? *Structural Change and Economic Dynamics, 15*(3), 245-264.

Cook, S., Hagiu, A., & Wright, J. (2024). Turn generative AI from an existential threat into a competitive advantage. *Harvard Business Review.*

Copeland, B. J. (2025, April 19). *History of artificial intelligence (AI)*. Encyclopaedia Britannica. https://www.britannica.com/science/history-of-artificial-intelligence#ref379822 (Accessed 3rd of June 2025).

Dell'Acqua, F., McFowland III, E., Mollick, E., Lifshitz-Assaf, H., Kellogg, C. K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, R. K. (2023). Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality. *Harvard Business School Working Paper*, No. 24-013.

Emmerich, S. (2023, June 30). Artificially unintelligent: Attorneys sanctioned for misuse of ChatGPT. *Proskauer*. https://www.mindingyourbusinesslitigation.com/2023/06/artificially-unintelligent-attorneys-sanctioned-for-misuse-of-chatgpt/ (Accessed 3rd of June 2025).

Engwall, M., Kaulio, M., Karakaya, E., Miterev, M., & Berlin, D. (2021). Experimental networks for business model innovation: A way for incumbents to navigate sustainability transitions? *Technovation, 108*. https://doi.org/10.1016/j.technovation.2021.102330

Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2023). RAGAS: Automated evaluation of retrieval augmented generation.

Ferrara, E. (2024). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci, 6*(1). https://doi.org/10.3390/sci6010003

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. Business & *Information Systems Engineering, 66*, 111–126. https://doi.org/10.1007/s12599-023-00834-7

Goldman Sachs (2023). *Generative AI could raise global GDP by 7%*. https://www.goldmansachs.com/insights/pages/generative-ai-could-raise-global-gdp-by-7-percent.html (Accessed 3rd of June 2025).

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems, 29*(2).

Henriksen, A., & Bechmann, A. (2020). Building truths in AI: Making predictive algorithms doable in healthcare. *Information, Communication & Society*, *23*(6), 802–816. https://doi.org/10.1080/1369118X.2020.1751866.

Institute for Management Development (IMD). (2024). IMD World Digital Competitiveness Ranking 2024. *Institute for Management Development*. Lausanne, Switzerland.

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

Jones, C.R., & Bergen, B.K., (2025). Large Language Models Pass the Turing Test.

Kim, H., Glaeser, E. L., Hillis, A., Kominers, S. D., & Luca, M. (2024). Decision authority and the returns to algorithms. *Strategic Management Journal, 45*(4), 619–648. https://doi.org/10.1002/smj.3569

Krakowski, S., Luger, J., & Raisch, S. (2023). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal, 44*(6), 1425–1452. https://doi.org/10.1002/smj.3387.

Kronblad, C., Jonsson, A., & Pemer, F. (2024a). Generative AI beyond the hype—new technologies in the face of organizing and organizations. *Organization Studies, 60*(4). https://doi.org/10.1177/00218863241285454.

Kronblad, C., Essén, A., & Mähring, M. (2024b). When justice is blind to algorithms: Multilayered blackboxing of algorithmic decision making in the public sector. *MIS Quarterly*, *48*(4), 1637–1662. https://doi.org/10.25300/MISQ/2024/18251.

Kronblad, C. (2020). How digitalization changes our understanding of professional service firms. *Academy of Management Discoveries, 6*(3), 436-454.

LaBarre, J., (2024, November 4). *How Harvey Uses Harvey*. Harvey Internal. https://www.harvey.ai/blog/harvey-internal (Accessed 3rd of June 2025).

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society, 5*(1). https://doi.org/10.1177/2053951718756684.

Maleki, N., Padmanabhan, B., & Dutta, K. (2024). AI hallucinations: A misnomer worth clarifying. In 2024 *IEEE Conference on Artificial Intelligence (CAI)*, Singapore, Singapore, 133-138. DOI: 10.1109/CAI59869.2024.00033.

Monod, E., Mayer, A., Straub, D., Joyce, E. & Qi, J. (2024). From worker empowerment to managerial control: The devolution of AI tools' intended positive implementation to their negative consequences. *Information and Organization, 34*(1).

Momani, A. (2020). The Unified Theory of Acceptance and Use of Technology: A New Approach in Technology Acceptance. *International Journal of Sociotechnology and Knowledge Development*, *12*, 79-98.

Muller, E., & Doloreux, D. (2009). What we should know about knowledge-intensive business services. *Technology in Society, 31*(1), 64-72.

Nesemeier, D. (2024). GenAI acceptance in professional services: The case of management consulting. *Proceedings of the International Conference on AI Research, ICAIR 2024, 4*(1). https://doi.org/10.34190/icair.4.1.3021.

OECD (2022). OECD Transfer Pricing Guidelines for Multinational Enterprises and Tax Administrations 2022. OECD Publishing, Paris. https://doi.org/10.1787/0e6.

Pachidi, S., Berends, H., Faraj, S., & Huysman, M. (2021) Make Way for the Algorithms: Symbolic Actions and Change in a Regime of Knowing, *Organization Science, 32*(1), 18–41. DOI:10.1287/orsc.2020.1377.

Prasad Agrawal, K. (2023). Towards adoption of generative AI in organizational settings. *Journal of Computer Information Systems, 64*(5), 636–651. https://doi.org/10.1080/08874417.2023.2240744.

PwC. (2024). *PwC's Global Workforce Hopes and Fears Survey 2024.*

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *The Academy of Management Review*, 46(1), 192–210. https://doi.org/10.5465/AMR.2018.0072.

Reinhard, Philipp (2024): Augmentation through Generative AI: Exploring the Effects of Human-AI Interaction and Explainable AI on Service Performance. *Mensch und Computer 2024 - Workshopband*. DOI: 10.18420/muc2024-mci-dc-360.

Retkowsky, J., Hafermalz, E., & Huysman, M. (2023). From playmate to assistant; User experiences of integrating ChatGPT into knowledge work. In *ICIS 2023 Proceedings: Tech and the Future of Work, 16*, 1-9.

Retkowsky, J., Hafermalz, E. & Huysman, M. (2024). Managing a ChatGPT-empowered workforce: Understanding its affordances and side effects. *Business Horizons, 67*(5), 511-523.

Rogers, M. E. (1962). *Diffusion of innovations* (1st ed.). The Free Press.

Sack, D., Krayer, L., Wiles, E., Abbadi, M., Awasthi, U., Kennedy, R., Arnolds, C., & Candelon, F. (2024). *GenAI doesn't just increase productivity. It expands capabilities. Boston Consulting Group.* https://www.bcg.com/publications/2024/gen-ai-increases-productivity-and-expands-capabilities (Accessed 3rd of June 2025).

Saunders, M., Lewis, P., & Thornhill, A. (2015). *Research methods for business students* (7th ed.). Pearson.

Shollo, A., Hopf, K., Thiess, T., & Müller, O. (2022). Shifting ML value creation mechanisms: A process model of ML value creation. *Journal of Strategic Information Systems, 31*(3). https://doi.org/10.1016/j.jsis.2022.101734.

Shuster, K., Poff, S., Chen, M., Kiela, D., & Weston, J. (2021). Retrieval augmentation reduces hallucination in conversation.

Spring, M., Faulconbridge, J., & Sarwar, A. (2022). How information technology automates and augments processes: Insights from artificial-intelligence-based systems in professional service operations. *Journal of Operations Management, 68*, 592-618.

Strich, F., Mayer, A. S., & Fiedler, M. (2021). What Do I Do in a World of Artificial Intelligence? Investigating the Impact of Substitutive Decision-Making AI Systems on Employees' Professional Role Identity. *Journal of the Association for Information Systems, 22*(2).

The Finance Story. (2024, June 4). *Big 4 firms scramble to win the consulting race: Investing over $4B in AI*. https://thefinancestory.com/big-4-invest-over-usd-4-bn-in-ai (Accessed 3rd of June 2025).

Tornatzky, L., Fleischer, M., & Chakrabarti, A. (1990). *The processes of technological innovation*. Lexington Books.

Turing, A. (1948). *Intelligent Machinery*.

Van den Broek, E., Sergeeva, A., & Huysman, M. (2021). When the machine meets the expert: An ethnography of developing AI for hiring. *MIS Quarterly, 45*(3), 1557-1580.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998-6008.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: *Toward a Unified View*. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540.

Von Nordenflycht, A. (2010). What is a professional service firm? Toward a theory and taxonomy of knowledge-intensive firms. *Academy of Management Review, 35*(1), 155–174.

Waardenburg, L., Huysman, M., & Sergeeva, A. V. (2022). In the land of the blind, the one-eyed man is king: Knowledge brokerage in the age of learning algorithms. *Organization Science, 33*(1), 59–82.

Waardenburg, L., & Huysman, M. (2022). From coexistence to co-creation: Blurring boundaries in the age of AI. *Information and Organization, 32*(4).

Weiser, B. (2023, May 27). Here's what happens when your lawyer uses ChatGPT. *The New York Times*. [https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html](https://www.nytimes.com/2023/05/27/nyregion/avianca-airline-lawsuit-chatgpt.html) (Accessed 3rd of June 2025).

Wensink, L. (2024). Early Adoption Stage of AI Technologies in a Small-Scale Management Consulting (Dissertation).

Yang, J., Blount, Y., & Amrollahi A. (2024). Artificial intelligence adoption in a professional service industry: A multiple case study. *Technological Forecasting and Social Change*, *201*. DOI: 10.1016/j.techfore.2024.123251

Yin, R. (2003). *Case study research design and methods* (3rd ed.). Sage.

Zhou, M., Abhishek, V., Derdenger, T., Kim, J., & Srinivasan, K. (2024). Bias in generative AI.

Ziegler, D. M., Stiennon, N., Wu, J., Brown, T. B., Radford, A., Amodei, D., Christiano, P., & Irving, G. (2019). Fine-tuning language models from human preferences.