Generative Artificial Intelligence in support of analytics: Copilot 365

Lappeenranta–Lahti University of Technology LUT

Master’s thesis in Industrial Engineering and Management

2024

Jonas Koskula

Examiners: Post-doctoral researcher, docent Lasse Metso

University Lecturer Leena Tynninen

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Engineering Sciences

Industrial Engineering and Management

Jonas Koskula

Generative Artificial Intelligence in support of analytics: Copilot 365

Master's thesis

2024

81 pages, 7 figures, 14 tables and 1 appendix

Examiners: Post-doctoral researcher, docent Lasse Metso and University Lecturer Leena Tynninen

Keywords: Artificial Intelligence, Generative Artificial Intelligence, Large Language Model, Analytics, Copilot 365

The rapid advancement of Generative Artificial Intelligence (Gen AI) has significant implications for enhancing analytics processes within organizations. This research investigates the utility of Large Language Models (LLMs), particularly Microsoft Copilot 365, in supporting and improving analytics workflows within the case company. While theoretical perspectives suggest that Gen AI can elevate analytical capabilities, it is evident from the literature that human analysts remain indispensable.

The study comprises both theoretical and empirical components. The theoretical section explores Gen AI's functionalities and potential benefits to analytics and reviews relevant methodologies. The empirical section employs the Design Science Research (DSR) methodology, integrating insights from interviews with case company employees to develop a tailored framework aimed at maximizing the benefits of using Copilot 365 in analytics workflows.

Findings indicate that while Copilot 365 cannot currently independently generate data insights, it excels in supporting data preprocessing tasks such as formula creation, data visualization, and pivoted table generation. The research concludes that effective implementation of Gen AI tools requires a comprehensive understanding of their capabilities and limitations, alongside a supportive organizational culture that fosters innovation and adaptability. A key limitation of this study is the inability to quantify the overall impact of the proposed framework on analytics workflows.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUTin insinööritieteiden tiedekunta

Tuotantotalous

Jonas Koskula

Generatiivinen tekoäly analytiikan tukena: Copilot 365

Tuotantotalouden diplomityö

2024

81 sivua, 7 kuvaa, 14 taulukkoa ja 1 liite

Tarkastajat: Tutkijatohtori, dosentti Lasse Metso ja Yliopisto-opettaja Leena Tynninen

Avainsanat: Tekoäly, Generatiivinen tekoäly, Laaja kielimalli, Analytiikka, Copilot 365

Generatiivisen tekoälyn nopealla kehityksellä on merkittäviä vaikutuksia organisaatioiden analytiikkaprosessien tehostamisessa. Tässä työssä tutkitaan suurten kielimallien (LLM), erityisesti Microsoft Copilot 365:n, hyödynnettävyyttä analytiikan työnkulun tukemisessa ja parantamisessa case-yrityksessä. Vaikka teoreettiset näkökulmat viittaavat siihen, että Generatiivinen tekoäly (Gen AI) voi parantaa analyyttisiä valmiuksia, niin kirjallisuudesta ilmenee, että ihmisanalyytikot ovat edelleen välttämättömiä analytiikassa.

Tutkimus sisältää sekä teoreettisen että empiirisen osan. Teoreettisessa osassa tarkastellaan Gen AI:n toiminnallisuuksia ja sen potentiaalisia hyötyjä analytiikoille sekä käydään läpi asiaankuuluvia menetelmiä. Empiirisessä osiossa käytetään suunnittelututkimus (DSR) menetelmää, johon on integroitu case-yrityksen työntekijöiden haastatteluista saatuja näkemyksiä, jotta voidaan kehittää räätälöity toimintamalli, joka pyrkii maksimoimaan Copilot 365:n käytön hyödyt analytiikan työnkuluissa.

Tulokset osoittavat, että vaikka Copilot 365 ei tällä hetkellä kykene itsenäisesti tuottamaan havaintoja datasta, niin se on erinomainen tukemaan datan esikäsittelytehtäviä, kuten kaavojen luomista, datan visualisointia ja pivot-taulukoiden luomista. Tutkimuksen johtopäätöksenä on, että tekoälytyökalujen tehokas käyttöönotto edellyttää kattavaa ymmärrystä niiden ominaisuuksista ja rajoituksista sekä sitä tukevaa organisaatiokulttuuria, joka edistää innovointia ja sopeutumiskykyä. Tämän tutkimuksen keskeinen puute on se, että ehdotetun kehyksen kokonaisvaikutusta analytiikan työnkulkuihin ei pystytä määrittelemään.

ACKNOWLEDGEMENTS

This thesis marks the conclusion of my studies at LUT University. I am extremely grateful to the friends I've met along the way, both at my home university and during my exchange, who have made this journey memorable. I will always cherish the experiences we've shared and look forward to creating more in the years ahead.

I also want to give a big thank you to my family and my girlfriend, whose support and encouragement have been invaluable throughout this journey.

Helsinki 28.11.2024

Jonas Koskula

ABBREVIATIONS

AI	Artificial Intelligence
BPE	Byte Pair Encoding
CoT	Chain of Thought
DSR	Design Science Research
Gen AI	Generative Artificial Intelligence
GPT	Generative Pretrained Transformer
ICL	In-Context Learning
LLM	Large Language Model
NLP	Natural Language Processing
RNN	Recurrent Neural Network

Table of contents

Appendices

Figures

Tables

1 Introduction

Generative Artificial Intelligence (Gen AI) has caused many changes in how organizations handle data analytics, thanks to transformative tools now able to create and not just analyze the content. Ever since their release in the early 2020s, OpenAI's Generative Pretrained Transformer (GPT) models GPT-3 and GPT-4 popularity has grown very fast. These steps propelled Gen AI into the pinnacle of modern technology, which has compelled organizations to execute decisions, process information, and perform many other tasks using Gen AI. In recent years Microsoft announced Copilot 365, a Gen AI assistant integrated into the software company's Office suite. This, in turn, has provided innovative ways to complete everyday tasks, including Excel workflows with a Gen AI assistant. However, these new technologies come forth with challenges on how to make effective use of them by being very much aware of their limitations. These limitations involve all sorts of problems such as hallucinations, where the model outputs imaginary or misleading information. Limitations like these call for retaining human oversight when Gen AI is used in critical domains, including those related to finance and healthcare, where precision is essential.

1.1 Objectives, Research Questions, and Scope

The primary objective of this research is to evaluate how Gen AI, with a particular focus on Large Language Models (LLMs) like Microsoft Copilot 365, can be utilized to support and enhance analytics workflows. Advantages brought by LLM driven tools are examined in detail for different kinds of workloads, and challenges associated with their limitations, which need to be understood by organizations. The research questions guiding this thesis are the following.

How can Large Language Models, particularly Copilot 365, be effectively used to support analytics?

How can Large Language Models be applied to different workloads, and what are the potential benefits and challenges?

What challenges and limitations may arise in analytics applications based on generative AI?

1.1 Research Methodology

The study incorporates a theoretical framework that examines the evolution of LLMs and their capabilities, as well as their application in analytics. Empirical insights are gathered through interviews with the case company's employees, providing real-world context for how Microsoft Copilot 365 could enhance their workflows.

The methodology follows the Design Science Research (DSR) approach, emphasizing the creation of an artifact, which in this case is a framework designed to enhance the effectiveness of the case company's employees' use of Copilot 365.

1.2 Structure of the Thesis and Use of LLM's in the study

This paper consists of five chapters, as seen in figure 1 below. Chapter 1 is the introduction chapter. Chapter 2 introduces Gen AI and its potential impact on data analytics, covering LLMs, their mechanisms, and applications. Chapter 3 explores how Gen AI enhances human expertise in finance and healthcare for better decision-making. Chapter 4 reviews a case study on Microsoft Copilot 365's effect on task automation in Excel. Chapter 5 summarizes the findings of the study, answering the research questions.

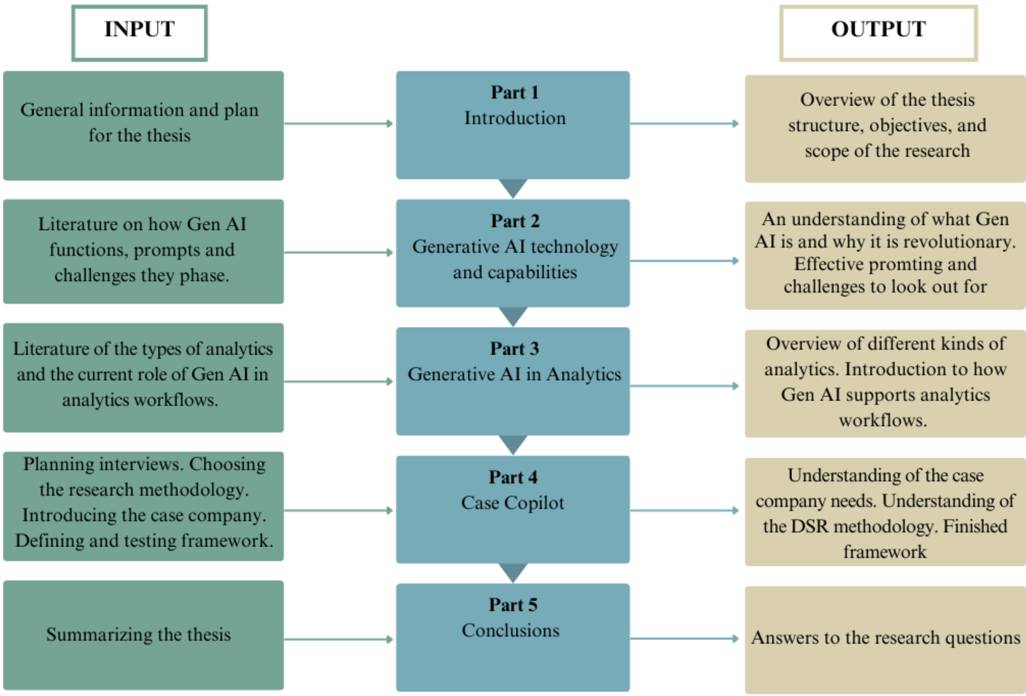

Figure 1: In and out figure of the structure of the thesis

LLM’s have been used in this study to enhance language and text refinement to improve the flow and overall readability of the study. LLM’s have also been used in the designing of the thesis structure and the framework created in chapter 4.

2 Generative AI technology and capabilities

Chapter 2 explores the development of Gen AI, with a focus on LLMs. It traces the evolution from traditional Artificial Intelligence (AI) algorithms to transformer-based models, highlighting how LLMs generate human-like text through techniques like self-attention. The chapter also addresses prompt engineering and challenges such as hallucinations, biases, and handling structured data. It concludes by discussing the ethical and organizational implications of integrating LLMs into various industries.

2.1 The Development of AI

The development of AI algorithms has progressed significantly from early frameworks, which primarily focused on supervised and unsupervised learning. Supervised learning involves a so-called supervisor, which provides labeled datasets to train the algorithm, which learns to map inputs to the correct outputs. (Zhou et al., 2018, pp.267-268) Un-supervised learning, by contrast, lacks a predefined outcome, and the algorithm must find patterns within the data on its own (Panesar, 2021). These foundational approaches are inspired by biological systems and natural processes, address complex, data-heavy problems while evolving alongside technological advancements. (Kar et al., 2023)

A major leap in AI innovation came with the rise of Gen AI. As seen in figure 2 below Gen AI is a subset of AI that focuses on creating new, original content, rather than simply analyzing or classifying existing data. Traditional machine learning algorithms have long been effective for text classification tasks, but advancements in natural language processing (NLP), particularly through Gen AI models like ChatGPT, have dramatically enhanced these capabilities. (Miró Maestre et al., 2024) As seen in figure 2, NLP is a branch of AI that enables computers to understand, interpret, and generate human language in both written and spoken forms. NLP has evolved to incorporate statistical and machine learning approaches, allowing machines to learn from large datasets and predict language patterns. Key applications of NLP include speech recognition, language translation, and text analysis. (Jain et al., 2018; Nadkarni et al., 2018)

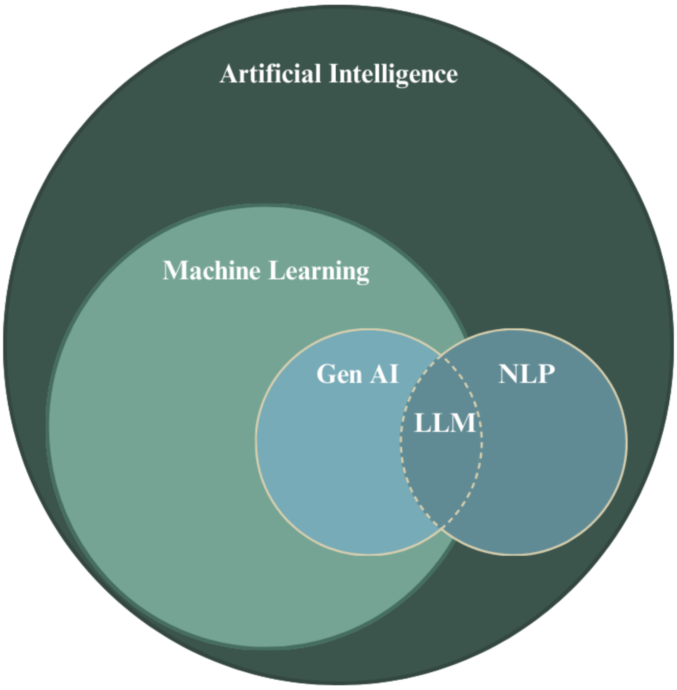

Figure 2: Visual representation of the concepts of AI (adapted from Amaratunga, T., 2023)

LLMs are a specific application of NLPs, as These models are designed to manage the complexities of language by learning from vast amounts of different textual data (Ooi et al., 2023). LLMs excel at tasks such as classification, summarization, and text generation by identifying complicated patterns in data and utilizing their findings to generate contextually relevant outputs (Devanny et al., 2023).

Their adaptability and ability to fine-tune for specific tasks make them particularly valuable for a wide range of applications, as they are used anywhere from customer service automation to advanced content creation (Saivasan et al., 2023). At their core, LLMs consist of three key components, a data processing pipeline, the user interface, and the core model that drives intelligent language generation (Malacaria et al., 2023; Saivasan et al., 2023).

Gen AI models leverage vast amounts of existing training data to create new content, improving performance as the volume of data increases, which mirrors human learning, where multiple possibilities are evaluated before reaching a conclusion (Ooi et al., 2023).

After pre-training, the model usually undergoes fine-tuning for a more specific task or domains, further refining its performance by training on labeled data, which is relevant to a particular application, such as language translation or sentiment analysis (Aggarwal, 2023). Gen AI 's versatility is evident in its ability to produce diverse outputs, including text, images, audio, and video, making it valuable across a wide range of applications. (Ooi et al., 2023; Dhoni, 2023; Corchado et al., 2023; Kar et al., 2023)

While Gen AI can produce various types of content, this study focuses mostly on LLMs which focus specifically on text generation. LLMs are trained to generate complete sentences, paragraphs, and even entire documents that often resemble human writing, positioning them as a vital subset within the broader field of Gen AI (Corchado et al., 2023; Dhoni, 2023; Corchado et al., 2023; Kar et al., 2023).

2.2 Large Language Models functionalities

At the heart of many LLMs is the transformer architecture, a neural network framework designed to handle data sequentially, making it ideal for tasks such as text generation (Ooi et al., 2023; Corchado et al., 2023). Neural networks are designed to mimic aspects of the human brain's structure and function, specifically in how humans learn from data (Singh et al., 2018). These networks use mathematical models inspired by human reasoning to process large amounts of data for analysis (Tarallo et al., 2019)

One of the most well-known transformer models is ChatGPT, where the GPT stands for Generative Pre-trained Transformer. GPT is a type of machine learning algorithm that uses deep learning and a large database of training text in order to generate new text in response to a user's prompt. GPT models have even outperformed traditional NLP techniques, particularly due to their ability to handle vast datasets and because they have been trained on such a vast array of texts, including academic articles and web content thus producing high-quality, human-like text. (Jiang, 2024)

Transformers have become the dominant architecture the development of LLMs, mainly because of their self-attention mechanism (Corchado et al., 2023; Zhao et al., 2024a). This mechanism revolutionized text processing by allowing the model to consider relationships between all words in a sentence, not just adjacent ones like their predecessors, which was the key improvement from the earlier models, as it enhances both the training and inference stages, which significantly improves the efficiency (Jiang, 2024; Korinek, 2023; Zhao et al., 2024a; Vaswani et al., 2017).

By utilizing the more sophisticated self-attention mechanism, transformers enable LLM models to efficiently handle larger datasets and capture long-range dependencies, which makes transformers better suited for handling complex sentence structures and long-form text generation, as well helping in understanding and generating text that closely mimics human communication, thus also improving consistency in conversational AI (Vaswani et al., 2017; Jiang, 2024; Korinek, 2023; Radford et al., 2019).

In the self-attention system, the input text is first transformed into a vector in a high-dimensional latent space, which the model processes to generate output tokens. Each word in the input is represented as three vectors, queries (Q), keys (K), and values (V). These vectors are used to calculate attention scores by comparing each query vector with all key vectors, determining how much focus each word should receive in relation to the others. (Korinek, 2023; Vaswani et al., 2017) This mechanism allows the model to weigh the importance of different words dynamically, enhancing its ability to understand and generate coherent text (Vaswani et al., 2017).

A key aspect of this prediction process is the softmax layer, which converts the model's output logits into a probability distribution over the entire vocabulary. The model then selects tokens based on these probabilities, either by choosing the most likely token or using more advanced techniques to refine the choice. (Zhao et al., 2024a; Heinsen, 2024). These scores are then used to weigh the value vectors, allowing the model to produce an output that effectively allows the model to build a richer, more nuanced understanding of context, leading to more coherent and accurate outputs. (Zhao et al., 2024a; Jiang, 2024)

The architecture of a transformer is built around two key components, which are called the Encoder and the Decoder. Both elements contain multiple layers of self-attention mechanisms and feed-forward neural networks, which are crucial for processing and addressing long-range dependencies within sequences of data (Zhao et al., 2024a; Kar et al., 2023; Che et al., 2023). Decoder-only models, such as the ChatGPT series, outclass the Decoder model in generating human-like contextually accurate and coherent text, which makes them ideal for tasks like creative writing or automated reporting (Zhao et al., 2024a).

Models that utilize both the Encoder and the Decoder are called Encoder-Decoder models, which are better suited for tasks like machine translation, where input text from one language is translated into another. These models process input text bidirectionally to maintain a semantic structure, which in turn ensures a high-quality output, important in tasks that require complex language understanding. In such models, each token in the input attends to only the preceding tokens, making them particularly effective for text generation tasks. (Zhao et al., 2024a; Cho et al., 2014)

2.3 Large Language Model prompt handling

LLMs are also adept at handling complex queries, processing long sentences, and interpreting complex questions. Their capacity for in-context learning (ICL) allows models like ChatGPT4 to excel in tasks that require common-sense reasoning and problem-solving. ICL is used to describe the inner mechanism of meta-learning, which can be further refined to describe “zero-shot”, “one-shot”, or “few-shot” based on the number of demonstrations available during reasoning. (Brown et al., 2020)

This flexibility in handling minimal training data underscores how the prompt structure and ICL can amplify a LLM’s ability to generalize tasks that it has not been explicitly trained for and be able to solve complex reasoning tasks (Brown et al., 2020; Sui et al., 2024). Their performance improves significantly when given structured, sequential prompts that break down complex, multi-step problems, which in turn leads to more enhanced outcomes in areas such as arithmetic, deductive reasoning, and common-sense understanding tasks (Zhao et al., 2024b; Zhao et al., 2024a).

Pre-trained LLMs can enhance their reasoning capabilities without relying on supervised data, by utilizing Chain-of-Thought (CoT) prompts, as LLMs can break down reasoning into logical steps, improving accuracy for similar future inputs through a process of filtering and using high-confidence predictions for further training (Ooi et al., 2023). Wei et al. (2022) further notes that CoT prompting brings major improvement in LLMs' ability to solve multi-step reasoning tasks as seen in figure 3 below, which illustrates the differences between standard and CoT prompting. Standard prompting provides a 'black box' result, which also happens to be incorrect in the example in figure 3. In contrast to standard prompting, CoT prompting offers a step-by-step solution, allowing the LLM to process each calculation step and reach the correct answer.

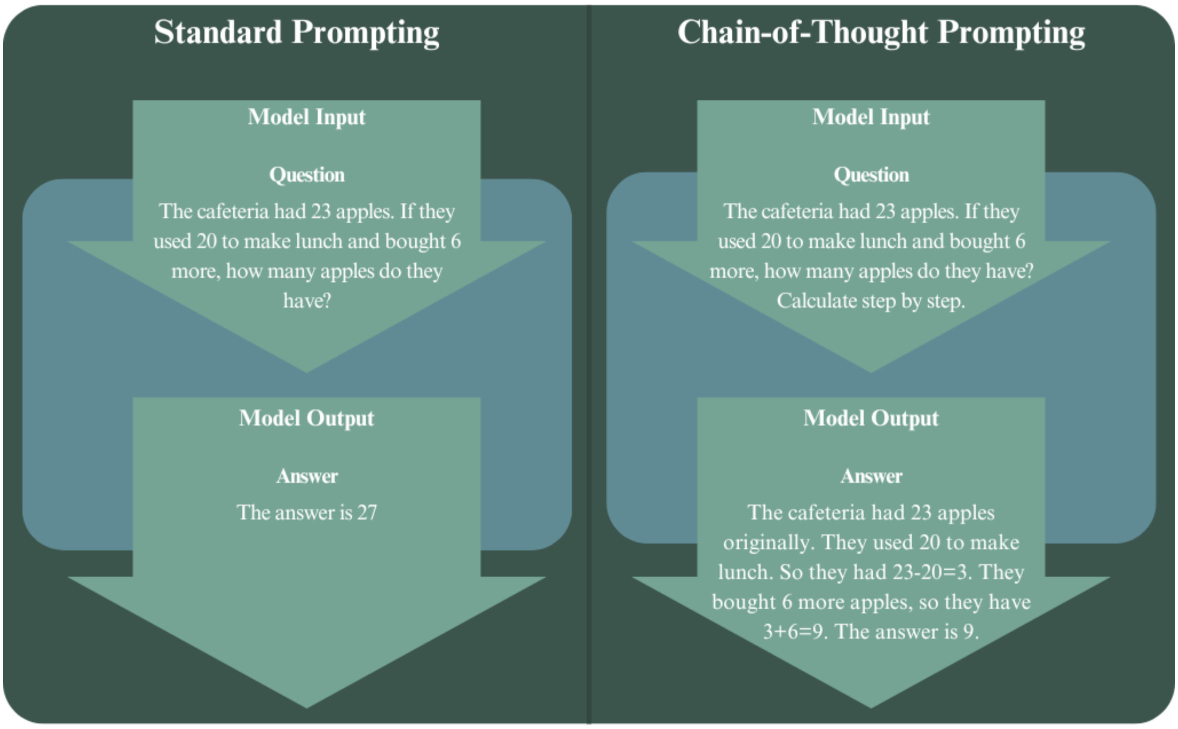

Figure 3: Representation of LLM prompting techniques (adapted from Wei et al., 2022)

CoT prompting also provides an explainable window into the behavior of the model, suggesting how it might have arrived at a particular answer and providing opportunities to debug where the reasoning path went wrong (Wei et al., 2022). However, Korinek (2023) notes that LLM outputs can still differ greatly depending on the phrasing and sequence of prompts, even though CoT is being used, which is similar to how human decisions can vary in different situations.

ChatGPT-4 exemplifies the capabilities of modern LLMs by supporting a variety of file formats, although the number of characters that can be processed varies depending on the file type. One of the clearer limitations of models like ChatGPT is their context window, which defines the maximum number of tokens the model can consider simultaneously when generating responses. (Korinek, 2023; OpenAI, 2023) For instance, ChatGPT-3 operates with a 2048-token window, which is equivalent to approximately 450-500 words, while ChatGPT-4 increases this capacity to 8192 tokens. This, however, is further surpassed by Claude 2.0, which can process an impressive 100 000 tokens at once, showcasing differences and the rapid advancements in LLMs. (Korinek, 2023; Anjos et al., 2024).

Buechel et al. (2018) highlighted that NLPs often struggle with understanding emotions like sarcasm, anger, or disappointment, resulting in misunderstandings, for example, during chatbot dialogues. This observation still holds up according to more recent findings by Jiang (2024) and Kar et al. (2023). Rashkin et al. (2019) explored the challenges of designing dialogue systems that can understand and respond empathetically to human emotions, noting that LLMs still struggle to provide adequate emotional support in real-world conversations. Ghosh et al. (2020) further noted the difficulty of detecting sarcasm, showcasing that even advanced pretrained language models struggle with this due to the subtlety and context-dependency of sarcastic language.

To manage vocabulary size and handle rare or morphologically complex words, LLMs typically employ techniques such as Byte-Pair Encoding (BPE) (Jiang, 2024). This method breaks down words into smaller sub word units and then transforms them into numerical representations that the model can process, enabling the model to understand both common and rare words more effectively (Jiang, 2024; Aggarwal, 2023). This step is called input encoding, which ensures that the model can manage diverse inputs by associating each token with a unique numerical identifier (Aggarwal, 2023). Kudo et al. (2018) introduced SentencePiece, which is an alternative to BPE, which provides language-independent tokenization, further aiding LLMs to handle rare words across various languages and dialects. Sennrich et al. (2016) emphasized the role of BPE in neural machine translation, where managing vast and diverse vocabularies is crucial.

ChatGPT also employs a technique called contextualization, which involves encoding both the input text and the surrounding context into a single vector. This enables the model to generate more relevant and accurate responses by considering the broader context of the conversation. (Briganti, 2024) The final step is decoding, where the model generates a sequence of words or sub-words based on the encoded input and conversation context, ensuring a coherent and contextually appropriate response (Aggarwal, 2023).

The process of crafting prompts, known as prompt engineering, is essentially a form of natural language programming. To effectively utilize the high-dimensional representation of input data, it is recommended to provide LLMs with detailed context and clear instructions when prompting them to generate content. (Wang et al., 2024; Korinek, 2023) Some LLMs even offer interfaces that allow users to set a "system message" that applies to all future interactions, which is particularly useful in scenarios requiring consistent application of a specific set of rules. LLMs are also increasingly improving in understanding user intent, even when initial prompts are vague or unclear. (Korinek, 2023)

2.4 Challenges of Generative AI

Integrating Gen AI into an organization goes beyond just a technological upgrade as it's a transformative shift that demands thoughtful management of both organizational and cultural changes. To succeed, organizations must establish strict governance frameworks to ensure both transparency and accountability, that also guide the ethical and effective use of

AI, while also nurturing a culture that values data-driven decisions and ongoing innovation. (Malacaria et al., 2023)

The use of LLMs within organizations offers considerable benefits, especially when it comes to automating and improving data-driven decision-making. However, a major challenge is the tendency of these models to produce hallucinating outputs that appear plausible. Bender et al. (2021) claims that these hallucinations appear because LLMs generate text by predicting patterns in the data, rather than focusing on factual accuracy. To maintain the reliability of LLM outputs, it's essential to cross-check facts with original data sources (Perkins et al., 2024; Zhou et al., 2024a; Devanny et al., 2023; Anjos et al., 2024). Hallucinations are particularly concerning in fields where precise and accurate data is crucial, such as financial forecasting (Devanny et al., 2023; Korinek, 2023).

LLMs, is that they tend to rely heavily on the data they were trained on, which can sometimes limit their accuracy and effectiveness. Since Gen AI tools are inherently unpredictable, they also lack consistency, meaning that repeating the same prompts doesn't always produce the same results. Marcus et al., (2020) further mention that this unpredictable nature of LLMs makes them unreliable for things like medical diagnoses or legal decisions, where you need accuracy and consistency. (Marcus et al., 2020) The situation gets more complicated with regular updates to models like GPT-4, which aren't always clearly communicated, leading to changes in the results they generate (Perkins et al., 2024).

Bender et al. (2021) explain that LLMs struggle to tell the difference between correct and incorrect data, which is a major factor in their tendency to produce inaccurate or biased content. Anjos et al. (2024) also point out that LLMs cannot process an entire document all at once, which causes them to focus on earlier parts of prompts and overlook key details that come up later, such as related ideas or the passage of time. It is crucial to have strong validation in place for AI systems to ensure their results are reliable, especially when used in important tasks that affect people directly (Basole et al., 2024).

One way to tackle the problem of hallucinations in LLMs is to bring in outside data or live updates to help improve their performance (Strubell et al., 2019). This can make the models more accurate, but it also means using more computing power, which can increase both costs and environmental impact, similar to training neural networks (Strubell et al., 2019). Using outside data also raises questions about bias, fairness, and how transparent the data is (Ooi et al., 2023). Anjos et al. (2024) suggest using prompt engineering to make sure LLM outputs align better with expert knowledge, while also keeping in mind the limits of the model's context window. Rudin (2019) points out that the 'black box' nature of many machine learning models is a problem because users often can't figure out where the mistakes come from. This is similar to the issue with LLMs, where it's hard to trace the source of errors like hallucinations.

Beyond hallucinations, using LLMs in organizations also brings up big concerns about data privacy and security. Since LLMs use both public and private data to generate responses, there's a risk that sensitive information could be exposed (Malacaria et al., 2023; Ooi et al., 2023). Korinek (2023) notes that LLMs can unintentionally produce outputs that breach privacy, and in some cases, the data entered by users could be stored and used for future training, which makes privacy concerns even bigger.

From a practical point of view, while LLMs sometimes make mistakes and show inconsistencies, their ability to give quick responses can be very useful for automating routine tasks. However, as Korinek (2023) points out, the level of mistakes that might be acceptable in an AI tool wouldn't be tolerated in a human assistant, which highlights the need for careful supervision when using these models.

One major challenge with LLMs is that they can pick up and repeat biases, harmful language, or prejudices from their training data. While providers are making efforts to reduce these biases, it's still important for users to stay aware and make sure these models are used responsibly and ethically. (Korinek, 2023) Bolukbasi et al. (2016) showed that gender and racial biases are often built into the word embeddings used in machine learning models, including LLMs, which makes careful debiasing important for ethical AI use. This is also tied to the challenge LLMs face in balancing memorization and generalization. While memorization helps the model answer specific questions accurately, overfitting on noisy data can lead to confusing answers. Generalization, however, helps the model deal with new inputs more effectively, meaning that finding the right balance between these two is key, because too much memorization can make the model rigid and less able to handle real-world tasks. (Naveed et al., 2024)

3 Generative AI in Analytics

This chapter examines the critical role data analytics plays in modern business, emphasizing how it drives decision-making and operational efficiency. It explains the four key types of analytics descriptive, diagnostic, predictive, and prescriptive and how businesses leverage these techniques to gain insights and optimize their processes. The chapter also explores the integration of Gen AI, particularly LLMs, highlighting their role in enhancing, rather than replacing, human expertise in data analysis. Specific use cases in industries like finance and healthcare are also discussed, showcasing the transformative potential of AI in these sectors.

3.1 Role and Process of Data Analytics in Modern Business

The modern business process generates enormous volumes of data, which becomes vital for optimization and thus makes the process more efficient. Decision-making has progressively become data driven, utilizing large datasets and advanced analytics methods that intelligently interpret the data. Brynjolfsson et al. (2014, pp.134-136) emphasizes leveraging data effectively is crucial for achieving operational efficiency, which enables businesses to transform their decision-making from being reactive to predictive, and ultimately prescriptive. According to Manyika et al. (2011) companies that are able to harness data analytics enjoy a significant competitive edge and can use insights from these tools to innovate or optimize their operations far better.

McAfee et al. (2012) highlight that organizations leveraging big data analytics are positioned to not only enhance customer engagement but also outperform competitors by fostering innovation and creating new market opportunities. Data analytics is typically divided into four distinct types descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics as seen in below table 1 (Yan et al., 2023; Liebowitz, 2021). Descriptive analytics provides insights into past events, predictive analytics forecasts potential future outcomes, and prescriptive analytics aims to identify the best possible course of action among various options (Shubho et al., 2022).

According to Davenport et al. (2017, pp. 7–30) organizations that successfully integrate all four types of analytics can derive strategic value by linking insights across descriptive, diagnostic, predictive, and prescriptive dimensions, which ultimately improves decision-making and operational efficiency.

Table 1: Four Types of Analytics (based on data from Yan et al., 2023; Liebowitz, 2021)

Type	Description	Question	Advantages
Descriptive	Uses data aggregation and visualization to understand how behaviors have evolved over time.	What is happening?	Makes complex data accessible to all, ensuring insights are available organization wide.
Diagnostic	Employs data mining and correlation analysis to uncover why certain trends and relationships occurred.	Why is it happening?	Identifies patterns, trends, and correlations, facilitating the recognition of opportunities and challenges.
Predictive	Combines historical data with statistical models to estimate future trends and behaviors.	What is likely to happen?	Enables rapid data assessment, supporting swift decision-making.
Prescriptive	Utilizes optimization techniques to determine the best method for achieving desired objectives.	What should we do?	Provides strategic recommendations, applicable to decision-making across various time horizons.

Descriptive analytics addresses the question, "What happened?" by analyzing historical data to help businesses compare past and present operations and even predict future outcomes (Wolniak, 2023; Shubho et al., 2022). According to Shmueli et al. (2017, p.366), descriptive analytics plays a critical role in business intelligence, allowing businesses to understand past performance and use these insights for strategic planning. By examining historical data, businesses can identify underperforming areas and take corrective actions, such as spotting a decline in sales in a specific region and making necessary adjustments (Wolniak, 2023). Descriptive analytics involves collecting, summarizing, and presenting historical data, enabling businesses to understand patterns, trends, and relationships (Shubho et al., 2022). As Manyika et al. (2011) highlight, this type of analytics helps businesses stay competitive by identifying shifts in market behavior quickly. This type of analytics allows analysts to spot data patterns and display them in real-time, enabling immediate insights and real-time decision-making (Wolniak, 2023; Shubho et al., 2022).

Bertsimas et al. (2020) argue that prescriptive analytics bridges the gap between prediction and action by providing tailored recommendations, which are especially valuable in dynamic industries like finance and healthcare. According to Shubho et al. (2022), it involves using optimization techniques to determine the most effective method of achieving a desired objective, whether that objective involves minimizing cost or maximizing outcomes. This process relies on a predictive model combined with actionable data and a feedback system to monitor the outcomes of actions taken. (Shubho et al., 2022) Davenport et al. (2017, pp. 7–30) point out that prescriptive analytics enables organizations to move beyond reactive responses, offering strategic guidance that cuts down trial-and-error decision-making.

Predictive analytics is a function of machine learning and AI, combined with historical data, to predict what will happen in the future. It involves creating a mathematical model from the patterns and trends in historical data, applying that to current data, and making accurate predictions about risks and opportunities by detecting and leveraging patterns in data. Siegel (2013, pp. 15-17, 155-158) also emphasizes that predictive analytics empowers organizations to anticipate customer behaviors and market trends, significantly improving forecasting accuracy. Predictive models come in two major types, classification models, where the forecasted outcome is categorical, and regression models, which forecast continuous outcomes. (Siegel (2013, pp. 15-17, 155-158). Waller et al. (2013) also highlights the importance of predictive analytics in supply chain management, allowing businesses to mitigate risks and enhance operational efficiency.

3.2 Enhancing, Not Replacing, Human Expertise in Data Analysis

Shneiderman (2020) emphasizes that effective AI systems must be human-centered and therefore designed to enhance human capabilities rather than replace them as they currently cannot replace the unique human ability to interpret, contextualize, and provide depth to findings. Human involvement remains essential in research, with Gen AI tools serving as transformative co-pilots that enhance and complement human capabilities rather than replace them (Perkins et al., 2024).

As we found out in chapter 2, their primary strength lies in their NLP abilities, leveraging their deep understanding of language semantics and structures (Korinek, 2023; Zhang et al., 2023). Advanced LLMs even allow users to upload files as input, enabling them to perform various data processing tasks, including complex analyses like regressions and file conversions (Korinek, 2023). As Brynjolfsson et al. (2014, pp. 82-84) have pointed out, these advancements in AI are indicative of a broader trend in which AI tools enhance human productivity across a range of tasks, allowing analysts to focus on more complex, high-value activities.

Dhoni (2023) states that leveraging LLM tools boosts efficiency in almost any field, including analytics, where they can transform productivity by providing rapid insights and speeding up analytical workflows. LLMs can quickly interpret large datasets by processing unstructured data, generating reports, providing verbal insights, optimizing operations, and identifying growth opportunities (Rane et al., 2023). They are also capable of performing tasks like text-to-SQL, which involves converting natural language questions into SQL queries (Cheng et al., 2023).

Cheng et al. (2023) research compared the performance of GPT-4 with human analysts at different expertise levels, as seen in Table 2 below, which showcases that GPT-4 completes tasks much faster than human analysts, requiring only 40-59 seconds compared to the 324-648 seconds for senior, junior, and intern analysts. (Cheng et al., 2023)

Table 2: Comparison of GPT-4 and Analysts' Performance (based on Cheng et al., 2023)

	Figure					Data Analysis				
Annotator	**Samples**	**Correctness**	**Chart Type**	**Aesthetics**	**Time (s)**	**Correctness**	**Complexity**	**Alignment**	**Fluency**	**Time (s)**
Senior	30	0,79	0,96	2,96	472	0.98	2.01	0.98	2.29	324
GPT 4	30	0,73	0,96	2,41	59	0.82	2.18	1	3	40
Junior	30	0,66	0,96	2,66	645	0.95	1.98	0.86	3	388
GPT 4	30	0,71	0,98	2,75	50	0.94	2.32	1	3	34
Intern	30	0,74	0,91	2,4	648	0.86	1.59	1	3	173
GPT 4	30	0,73	0,97	2,45	55	0.91	2.28	1	3	33

In terms of correctness, which is calculated from 0 to 1, while the aesthetics score is on a scale of 0 to 3, GPT-4's performance was similar to that of interns and junior analysts, achieving scores of 0.73-0.91, though slightly lower than senior analysts, who scored up to 0.98. GPT-4 excelled in handling complexity, outperforming both interns and juniors, with scores of 2.18-2.32, while its aesthetics scores were also high but marginally behind junior analysts. In areas like alignment and fluency, GPT-4 matched or surpassed human performance, achieving perfect scores, indicating that it produced grammatically correct, coherent, and well-aligned analyses. However, while GPT-4's correctness was comparable to less experienced analysts, it still lagged behind senior analysts in figure accuracy. This suggests that while GPT-4 is highly capable, especially in terms of speed and complexity, it does not yet fully replace the expertise of senior human analysts. As with any evaluation, caution is needed, and further testing is required to assess its performance in real-world business scenarios. (Cheng et al., 2023)

While human experts currently surpass LLMs in depth of knowledge, LLMs serve as excellent complementary tools, especially for brainstorming in areas where users lack expertise (Korinek, 2023; Devanny et al., 2023). This is in line with findings from Brynjolfsson et al. (2023), who observed that AI tools tend to benefit lower-skilled workers more significantly than highly skilled workers by augmenting their capabilities and boosting productivity.

Miller (2019) notes that, while AI systems can generate explanations and insights, their lack of contextual awareness limits their ability to match the depth of human reasoning and interpretation. Professional data analysts bring a unique integration of background knowledge into their work, prioritizing the avoidance of mistakes and ensuring confidence in their assumptions before offering insights, which contrasts with LLMs, which, while capable of providing quick suggestions, often do so without the same level of assurance or contextual understanding (Cheng et al., 2023). As such, LLMs are unlikely to replace human analysts in the near future. Instead, they should be viewed as complementary tools that enhance efficiency and effectiveness by assisting with tasks such as performing preliminary data analysis and proofreading reports.

Riemer et al. (2023) notes that the effectiveness of Gen AI diminishes when humans attempt to handle tasks that fall outside the AI's expertise, underscoring the importance of a symbiotic relationship between human judgment and AI-generated insights. The depth of contextual awareness and experience that humans have to offer, makes their judgment essential for interpreting and contextualizing AI-generated insights (Anjos et al., 2024; Dahal, 2023; Cheng et al., 2023). While GPT-4 is faster and cheaper than hiring human data analysts, its limitations in providing nuanced analysis mean that human oversight remains crucial (Cheng et al., 2023). The stakes in critical fields remain too high to consider LLMs as viable replacements at this stage in time (Devanny et al., 2023). Further caution is advised against over-reliance on LLMs due to their tendency to amplify biases embedded in their training data, which can result in misleading or harmful outputs. This reinforces the necessity for human oversight in critical decision-making processes. (Bender et al., 2021)

3.3 Large Language Models use cases in Analytic work

The impact of Gen AI has been profound, transforming sectors like healthcare, finance, research, and data analytics by enabling machines to not just interpret but also generate content (Malacaria et al., 2023; Rane et al., 2023). In healthcare, for example, Esteva et al. (2019) describe how Gen AI systems are now used to analyze medical images, predict disease risks, and provide diagnostic support and even claim that they have achieved physician-level accuracy in diagnostics, such as identifying tumors in radiographs and classifying skin lesions.

Ali et al. (2023) believe that the financial industry, where data-driven decision-making and customer interactions are paramount, LLMs will revolutionize these areas by leveraging vast amounts of high-quality, customer-centric data, leading to more informed strategic decisions and improved customer experiences.

Tools such as ChatGPT can serve as robust knowledge bases for financial regulations, helping financial institutions ensure compliance with applicable laws and regulations in their daily operations (Jiang, 2024). Bloomberg, for example, has developed its own AI model, BloombergGPT, which has been trained on Bloomberg's extensive financial data archives and public sources, excelling in financial tasks while maintaining proficiency in general LLM tasks (Ooi, 2023). BloombergGPT, as a specialized model, focuses on finance-specific tasks and excels at handling financial jargon, thus improving accuracy in predictions and risk assessments (Ooi, 2023).

Gen AI is able to meet real-time demands quickly, replacing manual processes with smart automation, which can significantly accelerate processes (Lin et al., 2023). These AI models can rapidly assess an organization's current state, providing advanced tools to analyze and interpret large, varied data sources, summarizing and evaluating unstructured text to address complex business questions (Lin et al., 2023; Shubho et al., 2022). As the prospects for Gen AI continue to grow, financial institutions should explore how to further integrate these technologies into their operations (Ooi, 2023).

Retail Banking is also very well-positioned to benefit from Gen AI applications due to its vast reserves of customer-centric data. This transactional data offers precise insights into customer behaviors, preferences, and risks. In retail banking, AI is increasingly used to forecast and tailor product offerings based on customer needs and behaviors, a practice that is quickly becoming standard in the industry. (Riemer et al., 2023)

Figure 4 illustrates how data from established banks, fintech companies, and regulators is processed through the Gen AI model, leading to outcomes that impact staff, consumers, and overall business operations. Figure 4 further highlights the importance that LLMs are trained with domain-specific data, ensuring they can effectively process and interpret the complex and specialized information required for tasks, in this example banking and finance, to be able to assess a company's financial performance, liquidity, stability, and other key metrics. (Sui et al., 2024; Aggarwal, 2023).

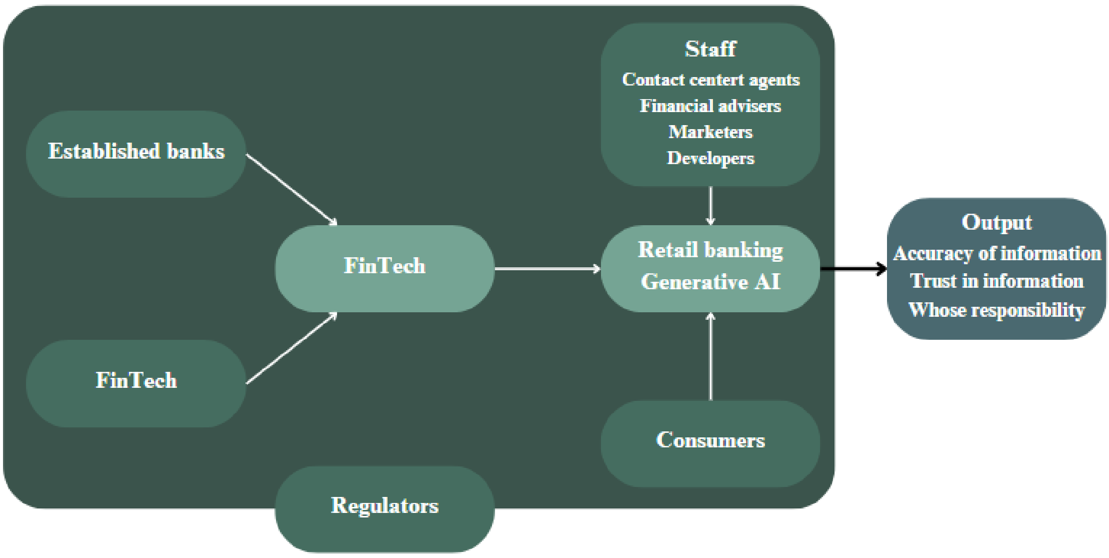

Figure 4: Example of Data flow in Retail Banking (Ooi, 2023)

LLMs significantly enhance this process by analyzing and processing inhumane amounts of financial data, including anything from stock prices to economic indicators, company announcements, and even analyst reports (Aggarwal, 2023; Jiang, 2024). LLMs offer valuable insights for predicting market trends and supporting informed decisions. They help investors spot potential risks and implement strategies like portfolio diversification to reduce losses (Huang et al., 2023, pp. 170-174).

LLMs are also being used to analyze consumer behavior and market trends helping organizations develop effective marketing strategies to identify new opportunities, and even detect suspicious activities in real-time, which reduces the risk of fraud and financial losses (Aggarwal, 2023; Jiang, 2024; Kar et al., , 2023). ChatGPT can also generate realistic simulations and scenarios that enhance financial models, improving their reliability and robustness (Che et al., 2023).

LLMs are gaining recognition for their role in predictive analytics, as they can easily aid organizations with demand planning, inventory management, and maintaining product availability. LLMs can uncover patterns and trends that might be overlooked by traditional market analysis methods enabling organizations to forecast market shifts, customer behaviors, and potential risks. The key advantage of Gen AI in predictive analytics is its capacity for continuous learning, which enhances accuracy as more data is processed. (Dahal, 2023) However, as Jiang (2024) notes, many financial research reports, expert summaries, and performance reviews are not part of ChatGPT's training set. Fine-tuning the model with specific financial data can address this gap, improving its relevance and accuracy for financial applications.

Dahal (2023) points out that these tools can sometimes still miss more complex, dynamic patterns in financial markets, underscoring the continued importance of human expertise in interpreting AI-generated predictions. Basole et al. (2024) further highlight challenges in quantitative data analytics, noting that LLMs face difficulties in creating accurate predictive models from structured data. These limitations suggest that LLMs are most effective when used in conjunction with other analytical tools. This limitation underscores why staff expertise remains crucial, as LLMs provide predictions that analysts must validate. The risk of hallucination in LLM outputs is also significant, requiring continuous monitoring of AI-generated content for accuracy.

To handle complex computations, Gen AI models can be augmented by converting natural language queries into structured code for execution by other processes. However, this integration introduces challenges, such as maintaining context and ensuring seamless communication between systems (Basole et al., 2024). While this integration can improve query formulation and streamline reporting, it requires careful management to ensure accuracy and effectiveness (Zhou et al., 2024b).

4 Case Copilot

This chapter introduces an overview of the research methodology used in this study, which is the called DSR. It first provides the theoretical basis of DSR and describes how DSR proceeds from problem identification to the design and evaluation of novel artifacts. Further on, this chapter details interviews conducted with the case company’s employees, laying the ground for the development of a framework for integrating Microsoft Copilot 365 into the case company’s existing workflows.

4.1 Introduction to research methodology

DSR, being a scientific problem-solving paradigm, paves the way for the development of human knowledge by creating innovative artifacts designed to solve problems occurring in the real world (Hevner et al., 2004). This paradigm is highly important in the context of Information Systems since it not only tries to solve the pressing business problems but also allows a growing body of design knowledge with the generation of constructs, models, methods, and instantiations (Vom Brocke et al., 2020, pp 2-5).

Hevner et al. (2004) offered practical guidelines on the ways DSR is to be conducted within the context of Information Systems, emphasizing the essence of DSR in developing an artifact capable of efficiently resolving the identified problem. Development of such an artifact is to be conducted taking into consideration substantial business issues that thus far were not resolved. It is mainly this focus on applicability within concrete settings that turns DSR different from other research approaches. (Hevner et al., 2004)

As Vom Brocke et al. (2020, pp. 2-5) indicate, DSR yields knowledge about the design of new solutions for crucial problems. More generally, such knowledge is referred to as design knowledge and it specifies the relationships between problem and solution spaces in means-ends terms. Therefore, design knowledge is about articulating how "entities can and should be designed with the intent to enhance the abilities of human beings and organizations through innovative artifacts" (Hevner et al., 2004; Gregor et al., 2013).

DSR is identified with a strict design and evaluation of the artifacts created to address identified problems. The outcome of these activities contributes in various manners to the scholarly research as well as to the practical knowledge bases (Hevner et al., 2004). The artefacts are modelled in several forms: methods, instantiations, models, social innovations, or new properties for technical, social, or informational resources - which are the inherent outputs of DSR (Järvinen, 2007; Aken, 2004).

That is, Hevner et al. (2004) defined it as a conceptual framework wherein environment and knowledge base become the two major influencers for DSR; and environment defines or demarcates the problem space comprising of people, organizations, and technologies; and the knowledge base provides the theoretical and methodological underpinning that is so essential both for creating and evaluating an artifact. The use of such knowledge ensures that the research is conducted upon solid theoretical and methodological ground dissolving rigor within the work undertaken. (Hevner et al., 2004)

Probably the most complete methodology related to DSR was proposed by Peffers et al. (2007), which established an orderly yet at the same time flexible process that guides the researcher from problem identification to the communication of results. The steps in this process can be found below in table 3 are generally carried out sequentially, although iteration between them may occur as the research progresses and the artefact evolves.

Table 3: Methodological Framework in Design Science Research. (Peffers et al., 2007)

Phase	Description
1. Problem Identification & Motivation	Defining the research problem and justifying the need for a solution.
2. Define Solution Objectives	Setting objectives that the solution must achieve, both quantitatively and qualitatively.
3. Design & Development	Creating the artifact that embodies the research contribution.
4. Demonstration	Demonstrating how the artifact effectively solves the problem.
5. Evaluation	Assessing the artifact's performance in meeting the defined objectives.
6. Communication	Communicating the problem, the artifact, and the results to relevant stakeholders.

Phase 1 starts with the identification and motivation of a certain research problem, which is an extremely important first step wherein, besides recognizing a gap that exists in knowledge or practice, one must identify the importance of filling such a gap, i.e., the importance of solving the identified problem. In other words, justifying the significance of the solution of a problem detected must be explained comprehensibly so that the research is relevant and effective. Establishing this foundation ensures that all the subsequent activities are carried out in a deliberate and well understood context. (Peffers et al., 2007)

The second phase is to outline what the requirements are that the solution needs to meet. These goals are directly obtained from the definition of the problem. Objectives can also be quantitative and qualitative in nature. They are considered to be quantitative because most of them are measurable targets of performance, whilst on the other hand, qualitative prescribes descriptive goals about how the new solution will solve problems that were hitherto not solved by prior solutions. Well thought objectives give clear guidelines on how the artifact is going to be designed and developed. (Peffers et al., 2007)

After the definition of objectives, the focal point is to ensure the design and development of the artifact-which is Phase 3. Through this phase, the researcher constructs a solution that after that personifies the contribution of the research. The artifact, which may range from any software to conceptual models, needs to integrate innovative elements therein that respond directly to the problem stated in the research. (Peffers et al., 2007)

Phase 4 starts once the artifact is developed. The artifacts effectiveness is determined by applying the artifact into one or more instances of the problem using such methods as experiments, simulations, or case studies. The demonstration phase is hence very critical because from here comes empirical evidence to prove that the artifact can indeed solve the identified problem and hence validates the design. (Peffers et al., 2007)

It is the assessment of the artifact which follows Phase 5, in which the performance of the artifact in the pursuit of meeting its objective derived in the previous phases. Now is the time of comparing the expected and actual outcomes based on what the deployment of the artifact is witnessing. These findings will help further refine the artifact if needed, or the research proceeds to the last phase. (Peffers et al., 2007)

Communication of the research is the last stage in the framework Phase 6. The findings with regard to anything pertinent to the problem, artifact and the results of its evaluation are made public now by the researcher to the interested parties. It will be maintained that good communication will at least have the impact of the research appreciated and the information already available to those that can use or develop the benefit of the artifact further. Whereas Peffers et al. (2007) described the process as a logical sequence, it is not one that should be seen as inflexible. (Peffers et al., 2007)

As figure 5 shows, this study will only be using the 5 first phases listed in table 3 as we won’t be able to communicate the actual results of the artifact. Phase 4 and 5 have been joined together to make the study flow better.

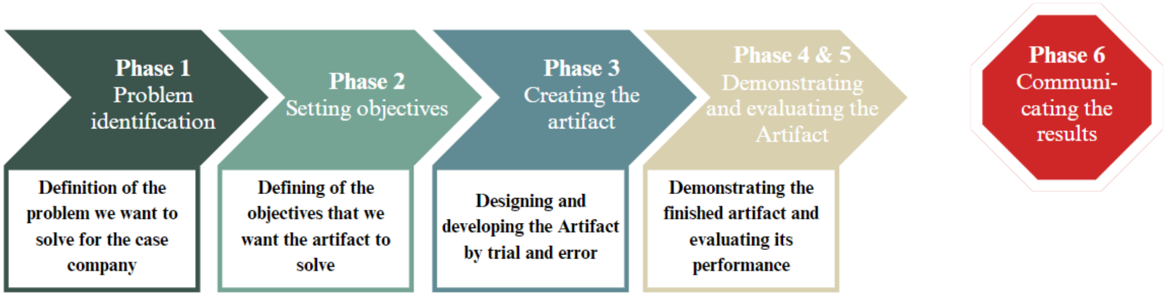

Figure 5: Demonstration of the phases used in the study (Peffers et al., 2007)

4.2 Interviews

The interviews were conducted in July 2024 and around 30 minutes was reserved per interview. There were a total of 6 interviews as seen in table 4. The interview questions were consistent for all participants and conducted remotely in Finnish via Microsoft Teams. They are available in Appendix 1. The answers were transcribed during the interview to Microsoft Forms. Transcriptions were then summarized and translated into English.

Table 4: Interviewee information

Role	Relevant experience (years)	Interviewed on
Team Lead	19	04.07.24
Sourcing Specialist	10	05.07.24
Development Manager	28	08.07.24
Jr. Analyst	1	08.07.24
Jr. Analyst	2	09.07.24
Controller	7	06.08.24

This interview aimed to gain a detailed understanding of how professionals at various levels within the case company's use Excel and other data tools as part of their day-to-day analytics workflows. It was done with the intention of mapping common workflows, finding pain points, and understanding the feasibility and scope related to the automation and introduction of AI-powered tools for better speed and accuracy in data processing, analysis, and reporting.

This is important, as the case company's has grown to be very data centric, relying on the availability of data tools to execute responsibilities and roles effectively. These interviews were conducted with people who already use Excel in their workflows before the core analysis phase, such as cleaning up data, and creating visualizations. In this respect, we tried to find where current processes could be further optimized, by adopting automation and AI solutions like Microsoft Copilot. Interview questions and the summaries of the answers are below:

Which business area do you belong to, and what are the main responsibilities of your role?

The interviewees represented a range of business areas within the case company, each contributing distinct responsibilities. Their roles encompassed data analytics, project management, and support functions. Specifically, responsibilities included maintaining financial reports, conducting group-level financial analysis for upper management, and compiling management-level reports by summarizing datasets. Some of the interviewees main responsibilities focused more on process automation to enhance data accuracy, while others were involved in project development, working to design data-driven solutions that align with business needs.

How often do you use Excel in your daily tasks, and for what purposes?

The interviews highlighted Excel as a widely used tool across distinct roles within the organization, serving various purposes. Finance and administrative professionals frequently employ Excel for generating financial reports, performing data analysis, budgeting, and forecasting cash flow, all of which are essential for informed decision-making and accurate reporting to key stakeholders. Many interviewees use Excel to combine and assess data from multiple various sources, which is particularly critical for conducting more macro-level evaluations.

For many, Excel played a crucial role in data cleaning processes, validating data precision is key before utilizing Excel's advanced tools, like macros, for complex tasks. Interviewees noted using Excel to combine data from several different systems, which is why data pre-processing is crucial to their workflows.

Additionally, a professional in a supporting role mentioned the importance of maintaining supplier data within Excel and managing P2P data to track expenses. Another respondent, responsible for residential rental services, reported daily use of Excel for monitoring rental payments, generating reports, and analyzing key metrics, including occupancy rates and maintenance schedules.

What type of data do you usually work with?

The interviewees primarily handle financial and real estate data, including customer portfolios, budget figures, cash flow statements, procure-to-pay (P2P) processes, billing, and supplier information. Additionally, they work with a variety of data types such as accounting transactions, property values, and tenant information. Interviewees once again highlighted the challenge of dealing with varied data formats, stressing the importance of standardizing data for accurate forecasting.

Are there any repetitive tasks in the data pre-processing phase that you wish you could automate or facilitate?

Interviewees frequently expressed a wish to streamline repetitive tasks in data pre-processing through automation. Key tasks for automation include the automation of combining Excel files, with common functions like V-lookups and also repetitive cleaning processes. These activities, which involve checking for errors, validating data, deleting duplicates, correcting mistakes, and unifying data formats, are currently performed manually. Respondents emphasized that these tasks are not only time-consuming but also prone to human error.

Respondents expressed that automating these repetitive tasks would allow them to shift their focus toward more value-added activities, such as reporting and extracting insights from the data.

What challenges do you face when analyzing data?

The interviewees identified several challenges in the data analysis process, with common issues once again revolving around data accuracy, consistency, and the sheer volume of work involved. A primary challenge mentioned was ensuring that data from different sources aligns correctly with the master data. Discrepancies between sources often lead to flawed analyses, and this issue is compounded when working with large datasets, where maintaining data accuracy and completeness becomes increasingly difficult.

Another significant challenge lies in the time-consuming nature of reporting, particularly when producing evaluation reports for clients or preparing presentations in PowerPoint. Updating existing data with new values requires substantial labor, background research, and the generation of charts. Such tasks are repetitive and could benefit from automation, as manual processes are not only time-intensive but also inclined to errors.

Furthermore, obtaining meaningful insights from complex or incompatible datasets adds another layer of difficulty. One interviewee mentioned how presenting data in a way that stakeholders can easily understand is another challenge, as clear representation is critical for conveying insights effectively.

What kind of reports or summaries do you most often create from the data?

Respondents indicated that the most frequently produced reports are financial and operational reports, which are crucial for decision-making and strategic planning. These reports include financial summaries, budget reports, cash flow forecasts, and detailed analyses of financial performance within the real estate sector. The purpose of these reports is to provide insights that guide management and stakeholders in making informed decisions.

Several interviewees also mentioned the use of pivot tables to summarize large datasets quickly, enabling dynamic analysis of key performance indicators (KPIs), trend analyses, and financial summaries. These pivot tables are often transformed into visual reports, such as charts and graphs, to present data in a more accessible format for non-technical stakeholders.

In addition to financial reports, respondents produce specialized reports, such as transaction analyses, budget comparisons, and visual dashboards that offer high level views of KPIs. These reports are essential for strategic planning, performance monitoring, and operational decision making.

The repetitive nature of manual data entry into templates was another pain point highlighted. An interviewee expressed a desire for automation tools, such as Microsoft Copilot, which could automatically derive data, upload it into premade presentations or reports, and significantly speed up the reporting process.

Other commonly produced reports track various operational metrics, including financial performance, real estate occupancy and maintenance, supplier performance, and the efficiency of processes such as billing and procure-to-pay (P2P). These reports are integral to the smooth running of day-to-day operations and ensure timely decision-making across various aspects of the enterprise.

What are the most time-consuming aspects of creating reports and visualizations?

The interviewees identified several time-consuming aspects associated with the creation of reports and visualizations, again particularly related to data preparation and ensuring the accuracy of the final output. The most frequently cited issue was the time spent on cleaning, manipulating, and formatting data. An interviewee pointed out how these tasks become especially demanding when dealing with large, complex datasets and when combining information from multiple sources.

Another significant time investment is required in creating visualizations that effectively communicate insights by ensuring that these visuals are clear, accurate, and insightful. Interviewees described that the process often entails multiple iterations, careful formatting, and repeated data checks to produce visualizations that enable stakeholders to easily derive insights and make informed decisions. Additionally, respondents once again noted that identifying patterns and anomalies in the data can be difficult when the dataset gets bigger through combining tables.

Are there any other particular aspects of your job that you think Copilot could improve or streamline?

Interviewees identified several areas where Gen AI tools, such as Microsoft Copilot, could save substantial time and improve productivity. A key area that came up has been a common pain point, which is data preparation. Interviewees believe that automation in this area would significantly reduce the time spent on manual data manipulation, especially when identifying inconsistencies or errors that require correction.

Respondents expressed a strong interest in automating parts of the analysis process. For example, tools that generate reports directly from the data, highlight discrepancies, or identify trends and patterns could significantly simplify their work while improving both efficiency and accuracy in data analysis.

Another area where automation could have a significant impact is in the development of complex formulas and lookups. Automating the creation of formulas difficult formulas such as calculations or lookups.

Furthermore, respondents highlighted the potential for AI tools to expedite the integration of data insights directly into PowerPoint presentations. This would enable users to perform preliminary analyses and create visualizations much faster, providing a solid foundation for more detailed reporting. By automating the initial steps of report creation, users could enhance the quality and speed of their work.

Additionally, automation could greatly assist in the development of advanced visualizations and dashboards, which typically require extensive formatting and iterative adjustments. A tool like Copilot could handle much of this process autonomously, making reporting and presentation tasks significantly easier and faster.

Have you used other automation or AI tools in your work? If so, how have they helped you?

Interviewees reported using a range of automation tools, from basic Excel macros and VBA scripting to more advanced AI solutions such as ChatGPT. These tools were employed in various capacities to improve efficiency and reduce manual effort.

Macros and VBA scripting were particularly useful for automating repetitive tasks related to data import, calculation, and reporting within Excel. These tools have already helped in cutting down on manual work and enhanced productivity, although noted that further automation could be implemented, particularly in the areas of data preparation, cleaning, and initial analysis.

ChatGPT has been widely used for a variety of tasks. These include translating documents, generating content for announcements, and drafting frameworks for different documents.

Several respondents also highlighted the tool's effectiveness in quickly translating Excel-formatted tables into Finnish, which saved time and simplified language-dependent tasks.

Most respondents agreed that while ChatGPT was highly useful for solving Excel formula queries and providing coding support, its utility was somewhat limited for tasks outside of these areas. Nevertheless, it has offered innovative ideas and solutions for complex problems and has been particularly helpful in explaining Excel formulas and functions, making them easier to understand and implement.

Do you have any other comments or thoughts on using Copilot 365 to support analytics?

Respondents suggested that Copilot 365 could play a crucial role in automating data quality checks, flagging inconsistencies, and offering suggestions for standardizing data formats across different departments or sources.

Another suggestion was the potential for Copilot to enable more intuitive, natural language queries within Excel. Instead of relying on complex formulas, users could ask questions in plain language, and Copilot would automatically generate the necessary data analysis and visualizations. Furthermore, respondents felt that Copilot could assist in customizing dashboards by automatically updating visualizations or recommending the best chart types based on the data structure, making reports more tailored and insightful.

One example frequently mentioned was the importing and normalizing of master data in Excel, which currently requires substantial manual adjustments. Automating these processes using procedures like a very definitive long prompt that would greatly reduce the time and effort involved, while also minimizing variability caused by manual handling. The Prompt would be made for that specific report and with one specialized prompt you could finish all the normalization and formatting tasks. These prompts could be then shared with the team.

The automation of VBA (Visual Basic for Applications) coding and debugging was also highlighted as a potential time-saving feature. Respondents suggested that Copilot's ability to automatically create and debug code could significantly reduce the time spent on these activities, enabling the development of more sophisticated automation solutions without the need for extensive manual coding. This, in turn, could lead to greater productivity and faster delivery of automation results.

Respondents emphasized the benefit of real-time collaboration, where Copilot could automatically update and reconcile changes made by different users, ensuring data consistency across shared workbooks. This could reduce errors from manual updates and streamline collaborative efforts, especially when multiple users are inputting data simultaneously.

Another area of interest was if Copilot 365 has the ability to integrate with external data sources beyond Excel, such as databases, cloud platforms, or APIs. Respondents expressed a desire to see Copilot automate the process of pulling in relevant external data for analysis, which would further enhance the data processing capabilities and reduce reliance on manual imports. The potential for seamless integration with platforms like Power BI or Azure for advanced analytics and reporting was also highlighted.

Problems that can possibly be solved with Copilot 365 are in table 5 below, along with a explanation of the possible solution of how Copilot 365 could be utilized in solving the problem or need.

Table 5: Objectives of implementing Microsoft Copilot 365

Problem	Solution
High manual effort in data cleaning and preprocessing	Automate repetitive tasks like data cleaning, normalization, and formatting, reducing time spent and minimizing human error.
Complex and time-consuming formula creation (e.g., V-lookups)	Automate formula generation and provide easy-to-understand examples and explanations, streamlining complex calculations.
Challenges in managing and analyzing large datasets from multiple sources	Use AI-driven insights to quickly identify patterns, trends, and anomalies, and automate data merging and standardization.
Manual report creation, including repetitive data entry and updating	Automate report generation, including creating pivot tables and visualizations, directly from data, saving time and improving accuracy.
Difficulty in creating effective visualizations that communicate insights	Automate the generation of advanced visualizations and dashboards, ensuring clarity and impact with less manual effort.
Repetitive tasks in routine processes (compiling reports)	Automate routine tasks and provide ready-made templates for common reports, freeing up time for more valuable analysis work.

4.3 Introduction to company and Case

The case company is a Finnish company that provides a wide range of real estate and facility management services. It focuses on property management, technical services, and facility maintenance, helping businesses and residential communities optimize their real-estate operations and is always looking for innovative ways to leverage new technologies to enhance its services for both employees and customers. Currently through this study the case company is evaluating whether Microsoft Copilot 365 can be a valuable investment for its employees to use, aiming to streamline their analytics workflows.

Our investigations into the interviews revealed some consistent opportunities for improvement that could be found across the different departments. Cleaning, normalizing, and preparing the data are pretty mundane and routine tasks that are very repetitive, hence workers make human errors in them. All these inefficiencies result in loss in quality and wasted time, which could otherwise be used for strategy or analyzing purposes.

The integration of Copilot 365 into Excel clearly has the potential to bring substantive benefits to the table for the case company. Regular automation of the data pre-processing and report generation by Copilot 365 could save working hours for its employees. This enables them to work on more strategic and analytical aspects hence, productivity with quality will increase due to less manual work.

4.4 Defining the objectives

This section discusses some of the major aims of implementing Microsoft Copilot 365 in Excel, taking into account the problem identified and the scope of the present study. This should particularly consider how Copilot 365 would enhance the present Excel-based workflow of the case company by automating repetitive work, reducing errors, and generally making the management and reporting of data more effective. This framework will not quantitatively measure the improvement of tasks, but we are going to estimate how well Copilot 365 does in supporting in analytics related tasks that we found out from the interviews. These tasks are summarized in Table 5, which was introduced at the end of chapter 4.2, which contains the main challenges identified from the interviews and how Copilot 365 could potentially address them.

The objective is to demonstrate how Copilot 365 could improve these workflows, focusing on the potential for increased efficiency and accuracy. By testing Copilot 365, we aim to explore its capabilities in real-world scenarios, such as automating time-consuming tasks like data cleaning, report generation, and managing large datasets. Although we won’t be measuring exact efficiency improvements, this study seeks to evaluate how well Copilot performs these tasks in practice.

The hypothesis is that automating mundane tasks in Excel through Copilot 365 will result in efficiency gains. Although precise measurements will not be part of this study, we will observe and evaluate how well Copilot 365 reduces the effort required for tasks like data cleaning, formula creation, and report generation. Testing Copilot 365 will also help us to understand at what level it operates in real-world scenarios, and if it is a useful tool for the employees of the case company to have. The objective for the framework is to be able to solve as many problems listed in table 5 as possible in a clear and understandable way.

4.5 Design and Development of the artifact

In this section, we design and develop the artifact introduced in Chapter 4.1, which in this case is a Copilot 365 framework for the case company. This framework serves as a guide for employees who may be unfamiliar with the capabilities of Microsoft Copilot 365 in Excel.

To ensure functionality, Copilot 365 requires files to be stored in OneDrive or SharePoint, and it operates exclusively on Excel tables. Given its integration within Excel, this framework is designed to seamlessly align with the case company’s existing Excel workflows, enhancing productivity. Copilot 365 can be activated from the Home ribbon, positioned in the top-right corner of the interface.

To demonstrate its functionality, I have generated sample data. Each sub-chapter begins by addressing specific objectives of the framework, followed by an immediate test of its capabilities to ensure clarity and coherence. The complete framework will be presented in Chapter 4.7.

4.5.1 Testing Copilot’s Data Preprocessing capabilities

One of the three objectives of this platform is to automate some of the manual tasks involved in cleaning, normalizing, and formatting data in Excel. Microsoft has provided several prompts to demonstrate the key features of Copilot 365, and out of the 26 total prompts, 3 were selected to evaluate data preprocessing abilities (Microsoft 2024a). These are listed in table 6 below.

Table 6: Microsoft Copilot suggested prompts for data processing (Microsoft, 2024b)

Type of Prompt	What the Prompt does
Split text in a column	Splits the contents of a column based on a specified delimiter (e.g., a space, comma, or other character) and separates it into multiple columns. Useful for breaking down complex data into simpler segments.
Extract text from a column to a new column	Extracts a specific part of a text string from an existing column and moves it into a new column. This can be based on patterns, substrings, or positions within the text. For instance, extracting domain names from email addresses.
Combine the columns of your choice in a specified way	Merges or concatenates data from multiple columns into one column using a specific format or delimiter. This can be helpful when combining first and last names or creating full addresses from separate columns.

In order to assess the quality of automation that Copilot 365 performs for data preprocessing, several tests were executed in the form of common scenarios corresponding to challenges indicated in previous interviews. The test prompts, results, and the specific formulas generated by Copilot 365 are presented in table 7 below.

Table 7: Results of Copilot 365 data preprocessing

Prompt	Key Outcomes	Formula Generated
Split the Invoice_Number column in the Invoices table into two columns	Successfully split invoice number into two columns	=SUBSTITUTE([@[Invoice_Number]], LEFT([@[Invoice_Number]], SEARCH("000", [@[Invoice_Number]]) - 1), "")
Extract the domain name from email addresses	Successfully extracted the domain from email addresses	=TEXTBEFORE(TEXTAFTER([@Email], "@"), ".com")
Combine columns A, B, and C using a '-' delimiter	Successfully combined specified columns together	=[@Department] & "-" & [@[Expense_Type]] & "-" & [@[Budget_Amount]]

The first part was to split the Invoice_Number column into two. In this regard, Copilot 365 offered a formula that can pull the numeric part of the invoice number effectively and efficiently. For example, the formula for this step was the following.

= SUBSTITUTE([@[Invoice_Number]], LEFT([@[Invoice_Number]], SEARCH("000", [@[Invoice_Number]])-1), "").

This is a formula that will eliminate the characters before the "000, " being a successful splitting of the invoice number into two different columns. This is considered accomplished in the sense that it actually showed Copilot can automate something which, otherwise, would have been done by hand.

The second task aimed to extract the domain name from the e-mail column. Copilot 365 did that by formulating the following formula.

=TEXTBEFORE(TEXTAFTER([@Email],"@"),"")

This was quite an efficient formula in that it extracted the text after the "@" symbol and ".com", therefore isolating the domain name of an e-mail address. Its functionalities come in handy in a task like customer segmentation, for which standardized email data is quite essential.

The third test was to take columns A, B, and C and combine them into one string, hyphenated. Copilot built the following formula.

=[@Department] & "-" & [@[Expense_Type]] & "-" & [@[Budget_Amount]]

The formula successfully merged the contents of the designated columns, demonstrating once more how proficient Copilot is at optimizing data manipulation operations. When working on big datasets, this is rather helpful as, manually done, this would be quite time-consuming and prone with human mistakes. It does result in nice data put together in a consistent form and automates that operation in Copilot very well, making it fast and dependable.

4.5.2 Testing Copilot’s formula creation capabilities

Creating complex formulas in Excel can at times become difficult, particularly when handling complex equations requiring accuracy and a thorough comprehension of the underlying data. By automating this task, one may save a lot of time and greatly lower the mistake risk. Table 8 below shows a variety of prompts meant to help with formula generating offered by Microsoft Copilot 365 (Microsoft, 2024b).

Table 8: Microsoft Copilot suggested prompts for formula generation (Microsoft, 2024b)

Type of Prompt	What the prompt does
Calculation based on other tables	Enables complex calculations by leveraging data from multiple tables, allowing for more advanced analysis and insights that go beyond a single dataset. This is particularly useful for tasks such as financial reporting, where data from various sources need to be aggregated or compared.
Lookup data	Retrieves and integrates external data into your table by referencing specific values from other tables or sources. This process enriches the current dataset and can help with tasks like filling in missing data, standardizing information, or cross-referencing key fields.
Compare columns in different tables	Compares columns from different tables to ensure data consistency and integrity. This feature is useful for validating data accuracy between datasets, detecting discrepancies, and performing quality checks during data integration or migration.
Get help with formulas	Offers intelligent suggestions for formulas based on the context of your task. This helps streamline the process of creating complex formulas, automating calculations, or solving specific data manipulation challenges without needing advanced formula knowledge.

To evaluate Copilot's ability to generate effective formulas, a series of scenarios were tested, each reflecting typical tasks encountered in data analytics as well as those highlighted during interviews with users. Examples of the prompts used in these tests, along with their outcomes and the specific formulas generated by Copilot 365, are again summarized in table 9 below.

Table 9: Results of Copilot 365 formula generation capabilities

Prompt	Key Outcomes	Formula Used
Calculate the difference between PO_Amount in the Purchase Orders table and the corresponding Amount in the Invoices table.	Successfully calculated the difference between the purchase order amount and the corresponding invoice amount.	=[@[PO_Amount]] - IFERROR(XLOOKUP([@[Invoice_Number]], Table1[Invoice_Number], Table1[Amount]), 0)
Perform a lookup to enrich the Invoices table with Payment_Date from the Payments table based on Invoice_Number.	Successfully retrieved and merged data from different tables.	=XLOOKUP([@[Invoice_Number]], Table3[Invoice_Number], Table3[Payment_Date])
Compare the Invoice_Number in the Invoices table with those in the Payments table to flag unpaid invoices.	Successfully identified and flagged unpaid invoices.	=IF(COUNTIF(Table3[Invoice_Number], [@[Invoice_Number]])=0, "Unpaid", "Paid")
Generate a formula to calculate the percentage of Amount_Paid compared to Amount in the Invoices table.	Successfully calculated the percentage of the amount paid.	=[@[Amount_Paid]] / XLOOKUP([@[Invoice_Number]], Table1[Invoice_Number], Table1[Amount])

First, it was required to calculate the variance between PO_Amount from the table Purchase Orders and respective Amount from the table Invoices. Copilot created the following formula.

=[@[PO_Amount]] - IFERROR(XLOOKUP([@[Invoice_Number]], Table1[Invoice_Number], Table1[Amount]), 0)

The difference had to be calculated by subtracting the amount of the invoice from the amount specified in the purchase order and in cases where no matching invoice could be found should return zero. This will mainly be useful for users who have to deal with heavy reconciliation between tables, as this is a highly error-prone and tedious operation if conducted manually.

The goal of the Copilot was to enhance the Invoices table by including the Payment_Date from the Payments table where the Invoice_Number matched. Essentially, this task is effectively accomplished by the formula.

=XLOOKUP([@[Invoice_Number]], Table3[Invoice_Number], Table3[Payment_Date])

This formula ensures that every invoice reflects the correct payment date. This accuracy is crucial for reliable financial reporting, and one of the frequent requests emphasized in these discussions is dependable and consistent data enrichment.

The third task was to compare the Invoice_Number in the Invoices table with those in the Payments table in order to determine whether there are unpaid invoices. Copilot returned the formula.

=IF(COUNTIF(Table3[Invoice_Number], [@[Invoice_Number]])=0, "Unpaid", "Paid")

which determines if the invoice number finds a similar duplicate in the payment table and describes that fact as "Unpaid". This kind of functionality proves extremely useful where manual errors, as would be described by the respondents, run very high in fields that require correct billing and tracking of payment.

The fourth test checked if Copilot could come up with a formula to find the percentage of Amount_Paid to the total Amount for each row in the Invoices table. Copilot used the formula:

=[@[Amount_Paid]] / XLOOKUP([@[Invoice_Number]], Table1[Invoice_Number], Table1[Amount])

and provided it as the % of amount in the invoice that is paid. This measure is very crucial in financial analysis and reporting. What is done here with automation is a real example of how Copilot can simplify complicated calculations, which, according to interviewees, are quite time-consuming.

4.5.3 Testing Copilot's Data Analysis and Visualization capabilities

As stated in the analytics chapter (Shubho et al., 2022), grasping and visualizing data is basic in contemporary analytics as it helps companies to understand prior trends, discover patterns, and make educated decisions. But given the complexity and amount of the data they are dealing with, many users find it difficult to spot trends or anomalies in big datasets, the interviews found.

Microsoft Copilot 365 provides a variety of prompts shown in table 10 designed to automate important facets of data analysis and visualization in order to meet these difficulties. These queries are meant to speed up data analysis, therefore enabling users to concentrate more on decision-making than on the technical aspects of data manipulation (Microsoft, 2024b).

Table 10: Suggested prompts for data analysis and visualization (Microsoft, 2024b)

Type of Prompt	What the Prompt does
Spot trends	Identifies patterns or trends over a specified time period or category
Provide a summary and analysis of data.	Provides a summary and/or detailed explanation of key metrics in your dataset
Get a count	Returns the number of entries, occurrences, or specific items in your dataset
See your data	Displays raw data or a selected subset in a readable format
Visualize relative values	Creates charts or graphs showing the proportion or comparison between data points
Find insights	Analyzes the data to reveal meaningful insights, correlations, or anomalies
Check for outliers	Detects values that are significantly different from the rest of the dataset

Again, all the Prompts were tested with several tests to evaluate the effectiveness of Copilot 365 in automating data analysis and visualization tasks. These tests were designed to mirror real-world scenarios. The results are once again in the below table 11, this time without formulas, as most results were either visualizations or Pivot charts.

Table 11: Summary of Data Visualization and insight task results

Prompt	Key Outcomes
Identify trends in Actual_Revenue over different time periods from the Time_Periods column in the Revenue table, with Time_Periods on the X-axis and Actual_Revenue on the Y-axis.	Generated a line chart, but the trend analysis was superficial.
Provide a summary of the Invoices data, including key metrics.	Successfully created a pivot table showcasing key metrics of the data. Written summary is missing
Create a bar chart comparing Actual_Revenue across different departments.	Successfully created a bar chart comparing actual revenue across departments.
Find insights within a dataset.	Insights varied in quality, sometimes failing to capture nuances in the data.
Check for outliers in the PO_Amount field within the Purchase Orders table.	Varyingly could identify and highlight outliers.

Using data from the Time_Periods column in the Revenue database, Copilot was assigned in the first test to spot patterns in Actual_Revenue over several time periods. Copilot created a line chart with Actual_Revenue on the Y-axis and Time_Periods on the X-axis. Although the chart was produced quickly, Copilot's examination was quite brief. Copilots assesment said, "Actual_Revenue decreases over time," even though the data was made to illustrate clear seasonal fluctuations. This result shows that although Copilot can efficiently create visuals, the depth of the insights it offers still needs more particular cues or refining to fully reflect the complexity of the data.

The second exam concentrated on summarizing the Invoices data’s key metrics. Copilot produced a pivot table summarizing metrics such as, the sum, average, minimum, and maximum of invoice amounts. The success of this prompt shows Copilot's fast and accurate summarizing of datasets, therefore offering important insights required by stakeholders for prompt decision-making. Copilot 365 did not give any textualized insights of the data.

Copilot 365 was instructed in the third scenario to produce a bar chart contrasting Actual_Revenue among many departments. Copilot was able to produce the wanted chart quickly, without needing to spend time manually developing and formatting charts. The ability to automate the development of visualizations directly answers a demand expressed in the interviews about automating visualization creation.

Asking Copilot to uncover insights from a dataset was the fourth test. Copilot produced insights, but the outcomes were unreliable and occasionally failed to effectively capture significant data details. Copilot, for example, occasionally overlooked particular patterns that were vital for spotting insights or trends. This restriction implies that even if Copilot is helpful for producing first insights, more user analysis is required to guarantee that all significant information is gathered.

The fifth test focused on Copilot 365s ability to find outliers from data. This was tested by having revenue fields with significantly larger revenue and also with data in the wrong format. Copilot 365 was able to find outliers from revenue when the revenue was significantly larger than the others, with no clear limit of what is considered an outlier. Out of place data formats or strings in revenues were not considered outliers by Copilot 365.

4.6 Final framework with evaluations

In this chapter, we evaluate the framework designed to assess how well Microsoft Copilot 365 aligns with the expectations of the case company’s employees that we interviewed. By integrating Copilot into Excel, the goal is to enhance productivity through the automation of Excel formulas, visualizations and insights. This assessment will highlight how employees can effectively use Copilot’s features in their daily tasks and adopt the best practices to maximize its potential.

This framework serves as a guide for the employees at the case company to effectively leverage Microsoft Copilot 365 in their Excel workflows.

Although Copilot 365 integrates with other Microsoft software, this framework focuses on its role in Excel to reduce manual work and minimize data processing errors. Automating repetitive tasks, such as data cleanup and formula creation, frees employees to focus on more strategic and analytical work. Copilot can be accessed directly within Excel, primarily through the Home ribbon. To use it, users must have an Excel workbook saved on OneDrive or SharePoint. Once opened, the Copilot button, located on the far right of the Home ribbon opens the Copilot 365 interface as seen in figure 6 below.

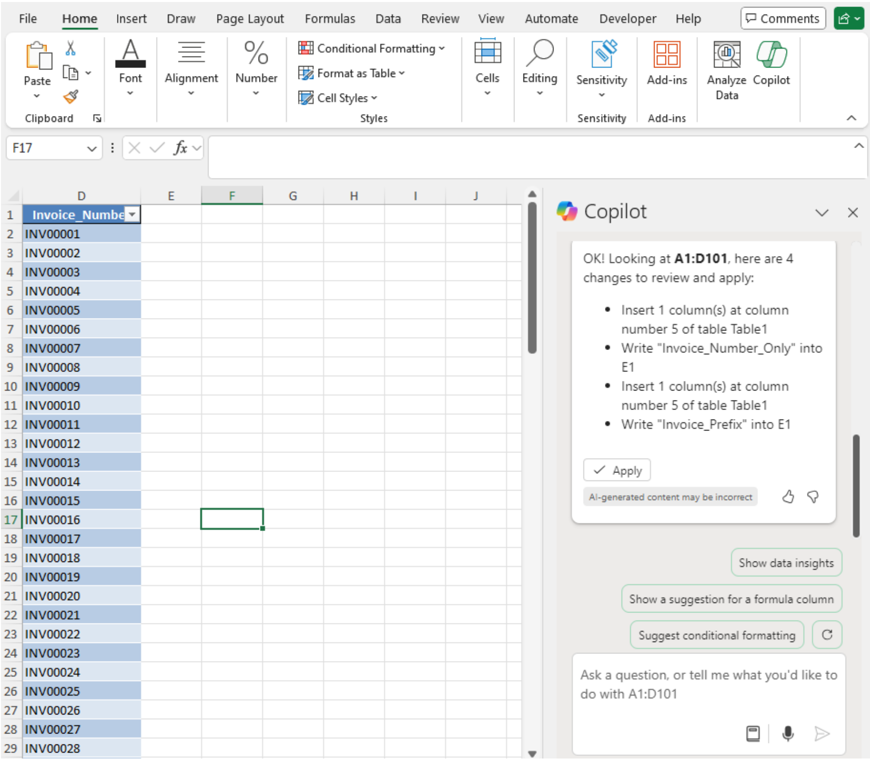

Figure 6: Showcasing Copilot 365 interface within Excel

The effectiveness of Copilot 365's responses depend largely on the clarity and specificity of the prompts provided by the user. When formulating prompts, be as specific as possible about the task you want to automate. For example, instead of asking Copilot to "clean the data," specify that you want to "split the Invoice_Number column into two separate columns based on a delimiter." Sometimes a working prompt might take some trial and error, it is a good idea to create a repository of common prompts within the team to ease reporting on recurring reports. This repository can serve as a reference for employees, helping them learn from each other's experiences and build on successful use cases, as copilot can take a handful of instructions at once, being able to preprocess data at the same time as doing lookups. This could save time in instances where the same changes need to be made to data frequently.

As Copilot continues to evolve, this framework should be revisited and updated to incorporate new features and best practices, ensuring that the case company’s employees remain at the forefront of technological innovation in data management and analysis.

4.6.1 Data preprocessing

The following section describes key capabilities of Copilot 365 with a focus on data manipulation, as one of the major capabilities of Copilot 365 rests in its ability to automate cleaning and preprocessing tasks with data. Employees spend a good amount of time splitting columns, normalizing data, and formatting text. Copilot handles data preprocessing tasks swiftly and with ease, automating key activities like splitting, extracting, and combining data effortlessly. It significantly reduces manual effort while ensuring accuracy in data handling.

Copilot 365 provides various prompts aimed at streamlining these processes. Each of these prompt’s functions by generating a calculated column with a formula designed to meet the requirements stated in your prompt. Table 12 below illustrates examples of these prompts in action.

Table 12: Showcasing of Data Manipulation Prompts and Formulas

Type of Prompt	What It Does	Example Usage	Formula Generated
Split text	Splits a column by a chosen delimiter	Split the Invoice_Number into two columns separating the number portion	=SUBSTITUTE([@[Invoice_Number]], LEFT([@[Invoice_Number]], SEARCH("000", [@[Invoice_Number]]) - 1), "")
Extract text	Extracts a specific part of a text string	Extract the domain name from email addresses	=TEXTBEFORE(TEXTAFTER([@Email], "@"), ".com")
Combine columns	Combines multiple columns into one	Combine columns, A, B, and C using a '-' delimiter	=[@Department] & "-" & [@[Expense_Type]] & "-" & [@[Budget_Amount]]

Each of these prompts automates the creation of formulas that handle common data preprocessing tasks. As seen in figure 6, after a prompt, Copilot 365 explains the changes it is going to make after the user presses the button to apply the changes. For instance, the "Split text" prompt can be used to separate numerical and text components within a single column into two columns, which is particularly useful when dealing with inconsistent data formats. The opposite "Combine columns" is also possible, being able to combine any columns of your choice into a single column with the delimiters you choose, which can come in handy when for example combining first names to last names, or for example adding area codes to phone numbers. The "Extract text" prompt can quickly isolate key information from text, such as domain names from email addresses.

4.6.2 Formula generation

Copilot 365 can simplify formula generation in Excel by automating the process and providing clear examples and explanations. Instead of manually searching for the right formula, it can be quicker to ask Copilot 365 to do what you want for you, especially for

Excel novices or when dealing with complex tasks, as Copilot allows users to describe what they need and generates the appropriate formula.

Copilot 365 handles complex formula generation tasks with great efficiency. For example, when tasked with calculating the variance between purchase orders and invoices or performing lookups to enrich data, Copilot consistently generated accurate and functional formulas. In testing it became clear that Copilot 365 effectively automates even the more complex calculations, significantly improving both the speed and accuracy of data analysis getting rid of tedious and error-prone formula testing. This capability is a valuable tool especially for professionals working with data-intensive tasks.

These capabilities are brought to life through specific examples where Copilot helps to generate and apply formulas tailored to various Excel tasks. Below are some practical scenarios demonstrating how Copilot can be used to address common challenges in Excel. Copilot 365 offers various prompts to assist with formula generation, as outlined in Table 13.

Table 13: Showcasing Data Calculation and Lookup Prompts and Formulas

Type of prompt	Prompt	What it does	Formula Generated
Calculation	Calculate the difference between PO_Amount and Amount in Invoices	Creates a calculated column that calculates the difference for each row.	=[@[PO_Amount]] - IFERROR(XLOOKUP([@[Invoice_Number]], Table1[Invoice_Number], Table1[Amount]), 0)
Lookup	Perform a VLOOKUP to enrich the Invoices table with Payment_Date	Retrieves and merges data from different tables using for example IDs in data	=XLOOKUP([@[Invoice_Number]], Table3[Invoice_Number], Table3[Payment_Date])
Comparison	Compare the Invoice_Number in Invoices with those in Payments table	Created a column that compared table values to see which invoices are still open.	=IF(COUNTIF(Table3[Invoice_Number], [@[Invoice_Number]])=0, "Unpaid", "Paid")
Calculation	Calculate the percentage of Amount_Paid compared to Amount in Invoices	Created a calculated column that shows in %, how much from the total invoices are already paid per row.	=[@[Amount_Paid]] / XLOOKUP([@[Invoice_Number]], Table1[Invoice_Number], Table1[Amount])

These examples illustrate how Copilot 365 can automate complex formula. The "Lookup Data" prompt, for instance, can automatically pull relevant data from other tables, saving time and reducing the potential for errors in manual data entry.

The "Calculate Based on Other Tables" prompt simplifies the process of performing calculations that involve multiple datasets, such as comparing financial data across different tables. It is also essential for data reconciliation, ensuring consistency across multiple sources, such as matching invoice numbers with corresponding payment records. Moreover, this automation streamlines the calculation of performance metrics, where key indicators are automatically computed based on aggregated data.

4.6.3 Visualizations and data insights

Copilot offers a variety of prompts to help analyze and visualize tasks, allowing users to spot trends, identify outliers, and create visual representations that clearly communicate insights. Copilot 365 is still in Preview mode, and especially the prompts relating to spotting trends, outliers or getting insights from the data do not function as well as advertised. The prompts are outlined in table 14 below.

Table 14: Showcasing Data Analysis Prompts and Their Outcomes

Type of Prompt	Prompt	Key Outcomes
Trend Identification	Identify trends in Actual_Revenue over different time periods	Generates a line chart with a superficial trend analysis.
Summary	Provide a summary of the Invoices data, including key metrics	Creates a pivot table displaying key metrics, without a written analysis of the data
Visualization	Create a bar chart comparing Actual_Revenue across different departments	Successfully created a bar chart comparing actual revenue across departments
Insights	Find insights within a dataset	Insights varied in quality, sometimes missing nuanced aspects of the data
Outlier Detection	Check for outliers in the PO_Amount field	Can identify and highlight outliers, when they are obvious enough

The summarization prompts enable quick insights from the dataset, creating a pivot table, with the metrics of your choice, for example sum, average, minimum, and maximum values, working as intended. Testing Copilot 365's Pivot capabilities showcased that its Pivot capabilities work as advertised. For example, Copilot can quickly generate pivot tables that summarize key metrics, such as total sales by region or average invoice amounts.

Creating visualizations also works quickly and as advertised by Microsoft, as seen from figure 7. When added to a new sheet, Copilot 365 creates a pivot table that is then visualized.

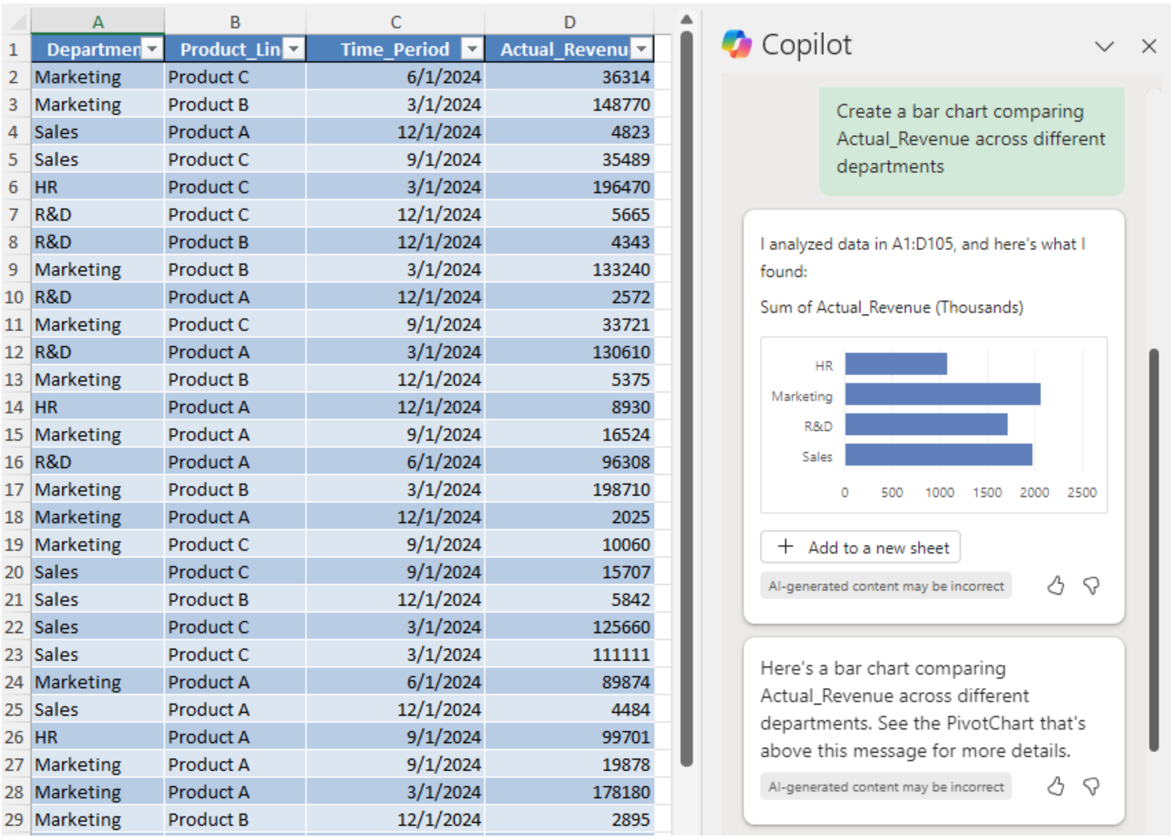

Figure 7: Showcases Copilot 365 creating a visualization in Excel

Trend identification prompt doesn’t work as intended as it misses clear patterns and requires further manual review by the user to gain more in-depth insights, as automated suggestions don't always reflect the full complexity of their data. While testing this capability it failed, for example, in noticing seasonal fluctuations in revenue, giving an insight into the fact that revenue is decreasing over time.

Outlier detection had similar problems, but it was able to point out outliers when they were obvious enough. When Copilot 365 was asked "What is considered an outlier in data?", it responded that Copilot 365 uses the Standard Deviation Method, which is effective when the data follows a normal distribution. In this approach, data points that deviate significantly from the mean, specifically those that are more than three standard deviations away, are flagged as outliers. This method assumes that most data points cluster around the mean, with only a few falling far from it.

5 Conclusions

The purpose of this thesis was to investigate if Gen AI, especially LLMs, can support analytics. We looked at the potential and uses of Gen AI to address important concerns about its place in analytics, its advantages and disadvantages, and the deployment constraints. The research questions are listed and answered below.

How can Large Language Models, particularly Copilot 365, be effectively used to support analytics?

The results of the research show that Gen AI can in fact assist analytics very successfully. As the literature in chapters 2 and 3 shows Gen AI, particularly LLMs have shown their capacity to absorb and evaluate vast amounts of data, spot trends, and create insights guiding decisions. It has been demonstrated to enhance the speed and accuracy of analytics operations in industries such as finance and banking by complementing human analysis through the automation of repetitive tasks, including data preparation and report generation. This allows for real-time, data-driven decision making that would be challenging or impossible to achieve manually. Generative AI has proven to be a valuable tool for this purpose.

Microsoft Copilot 365's integration with Excel at the case company offers a convincing argument for how Gen AI may efficiently assist in analytics. Copilot 365 greatly improves accuracy and efficiency by automating repetitive tasks like data preprocessing, formula development, visualization creation, therefore freeing staff members to concentrate on more strategic, high-value work. Some restrictions and difficulties related to Copilot 365's efficacy based mostly on the clarity of the user's prompts as more complicated or nuanced prompts call for improvement and manual control. Regarding Copilot 365 doing real-time data analysis, it is still quite lacking as its insights into data were superficial.

The lacking insight from data means that Copilot 365 users have to approach the integration of these solutions carefully, blending automation with human supervision, without mindlessly depending on the insights.

How can Large Language Models be applied to different workloads, and what are the potential benefits and challenges?

Gen AI has proven to be a transformative tool across various sectors, particularly in enhancing operational efficiency and supporting complex analytics tasks. Chapter 3.2 emphasized that while GenAI tools can augment workloads by automating routine tasks, they do not replace human expertise. For example, LLMs excel in NLP, data extraction, classification, and the generation of reports. However, they still require human oversight to ensure accuracy and mitigate the risks of hallucinations and biases. The ability of these tools to automate repetitive processes, such as data cleaning or formula generation, allows professionals to focus on more strategic tasks.

In Chapter 3.3, the use cases of LLMs in finance, healthcare, and banking highlight their growing role in predictive analytics, risk management, and fraud detection. These sectors are increasingly adopting AI to streamline processes like demand forecasting and credit risk analysis. The benefits of integrating Gen AI include significant time savings, enhanced accuracy, and the ability to derive insights from vast amounts of unstructured data. However, challenges such as the potential for over-reliance on Gen AI-generated outputs and the need for high-quality data inputs remain critical concerns. While GenAI boosts productivity, it's not a one-size-fits-all solution, particularly for complex analytic tasks requiring human judgment and specialized knowledge.

The interviews conducted with the case company employees revealed a strong desire to incorporate automation into their day-to-day workflows. Many respondents, especially in data analytics, project management, and support functions, expressed that repetitive tasks such as data cleaning, V-lookups, and formula generation were time-consuming. They saw

Microsoft Copilot 365 as a promising tool to alleviate these burdens, allowing them to refocus on value-added tasks. Sadly, as mentioned earlier, Copilot 365 is not yet capable of creating insightful reports from data.

The framework delivered for case company's employees, focusing on the integration of Microsoft Copilot 365 into Excel, demonstrates a practical application of Gen AI to different workloads enables employees to spend less time on manual labor and focus on more strategic, analytic, and decision-making activities. Additionally, while Copilot automates many tasks, its current limitations such as its inability to always recognize subtle patterns or nuances in data mean that human expertise remains crucial, particularly in higher-level analysis.

What challenges and limitations may arise in analytics applications based on generative AI?

LLMs have demonstrated value in automating and accelerating analytical tasks. However, as discussed in the theoretical chapters 2 and 3, key challenges persist, most notably in the areas of accuracy, privacy, and bias. The issue of hallucinations, where LLMs generate outputs that do not correspond to the input data, is a critical concern. Hallucinations can lead to incorrect or misleading analyses, particularly in high-stakes fields like finance and healthcare where precision is important. This limitation means that any analytical output generated by LLMs still requires cross-verification with primary data to ensure accuracy.

In addition to hallucinations, the risk of data privacy violations remains a serious concern. LLMs often rely on large datasets, some of which may inadvertently contain sensitive or proprietary information. The potential for LLMs to expose such data poses legal and ethical risks, especially in the context of regulations like GDPR.

The concern of bias propagation in AI models also stands out as a major limitation. Since LLMs are trained on vast datasets that may include biased or prejudiced information, they risk amplifying those biases in their outputs. This is especially dangerous in applications like hiring, customer service, or even financial decision-making, where biased outcomes could lead to unfair or discriminatory practices.

Future research could explore methods to enhance the depth of insights provided by Copilot 365, potentially by integrating it with other analytics tools or developing more advanced prompts. Further study on user experience, particularly concerning the onboarding process and long-term adoption, could help in designing more effective training programs and support systems. Testing the framework in a controlled environment or through a pilot program would also provide valuable insights into its real-world effectiveness, allowing for further refinement based on practical feedback.

References

Aggarwal, S., 2023. A review of ChatGPT and its impact in different domains. *International Journal of Applied Engineering Research,* 18(2), pp.119-123.

Aken, J.E.V., 2004. Management research based on the paradigm of the design sciences: the quest for field‐tested and grounded technological rules. *Journal of management studies*, *41*(2), pp.219-246.

Ali, H. and Aysan, A.F., 2024. Decoding Future of Generative AI in Finance: A Machine Learning Exploration of Academic and Grey Corpus. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4787211. [Accessed 27 July 2024].

Anjos, J.R., de Souza, M.G., de Andrade Neto, A.S. and de Souza, B.C., 2024. An analysis of the generative AI use as analyst in qualitative research in science education. *Revista Pesquisa Qualitativa*, *12*(30), pp.01-29.

Amaratunga, T., 2023. *Understanding large language models: Learning their underlying concepts and technologies*. Apress, p. 7.

Basole, R.C. and Major, T., 2024. Generative AI for Visualization: Opportunities and Challenges. *IEEE Computer Graphics and Applications*, *44*(2), pp.55-64. Available at: https://www.computer.org/csdl/magazine/cg/2024/02/10478355/1VBAA66c7Uk. [Accessed 2 May 2024].

Bertsimas, D. and Kallus, N., 2020. From predictive to prescriptive analytics. *Management Science*, *66*(3), pp.1025-1044.

Bender, E.M., Gebru, T., McMillan-Major, A. & Shmitchell, S., 2021. On the dangers of stochastic parrots: can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT)*. Available at: https://dl.acm.org/doi/pdf/10.1145/3442188.3445922. [Accessed 24 June 2024].

Bolukbasi, T., Chang, K.W., Zou, J.Y., Saligrama, V. & Kalai, A.T., 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. Available at: https://proceedings.neurips.cc/paper_files/paper/2016/file/a486cd07e4ac3d270571622f4f316ec5-Paper.pdf. [Accessed 21 June 2024].

Briganti, G., 2024. How ChatGPT works: a mini review. *European Archives of Oto-Rhino-Laryngology*, *281*(3), pp.1565-1569.

Brown, T. et al., 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems 33. NeurIPS* . Available at: https://papers.nips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html. [Accessed 16 June 2024].

Brynjolfsson, E. and McAfee, A., 2014. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & company. pp. 82-84, 134-136.

Brynjolfsson, E., Li, D. and Raymond, L.R., 2023. Generative AI at work (No. w31161). National Bureau of Economic Research.

Buechel, S. and Hahn, U., 2017. *EmoBank: Studying the Impact of Annotation Perspective and Representation Format on Dimensional Emotion Analysis*. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pp. 578-585. Available at: https://aclanthology.org/E17-2092. [Accessed 15 June 2024].

Che, C., Huang, Z., Li, C., Zheng, H. and Tian, X., 2024. *Integrating generative AI into financial market prediction for improved decision making*. In *Proceedings of the 6th International Conference on Computing and Data Science, Applied and Computational Engineering*, Vol. 64, pp. 154-160. Available at: https://www.ewadirect.com/proceedings/ace/article/view/12528. [Accessed 27 July 2024].

Cheng, L., Li, X., and Bing, L., 2023. *Is GPT-4 a Good Data Analyst?* In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 9496-9514. Available at: *https://aclanthology.org/2023.findings-emnlp.637.pdf*. [Accessed 11 May 2024].

Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y., 2014. *On the Properties of Neural Machine Translation: Encoder–Decoder Approaches*. In *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, pp. 103-111. Available at: https://aclanthology.org/W14-4012. [Accessed 25 August. 2024].

Corchado, J.M., López, S., Garcia, R. and Chamoso, P., 2023. Generative artificial intelligence: fundamentals. *ADCAIJ: advances in distributed computing and artificial intelligence journal*, *12*(1), pp.e31704-e31704.

Dahal, S.B., 2023. Utilizing Generative AI for Real-Time financial market analysis opportunities and challenges. *Advances in Intelligent Information Systems*, *8*(4), pp.1-11.

Davenport, T.H. and Harris, J.G., 2007. Competing on analytics: the new science of Winning. *Harvard business review press*, *Language*, *15*(217), pp. 7–30.

Devanny, J., Dylan, H. and Grossfeld, E., 2023. Generative ai and intelligence assessment. *The RUSI Journal*, *168*(7), pp.16-25.

Dhoni, P., 2023. Exploring the synergy between generative AI, data and analytics in the modern age. *Authorea Preprints*.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S. and Dean, J., 2019. A guide to deep learning in healthcare. *Nature medicine*, *25*(1), pp.24-29.

Ghosh, D., Vajpayee, A., and Muresan, S., 2020. *A Report on the 2020 Sarcasm Detection Shared Task*. In *Proceedings of the Second Workshop on Figurative Language Processing*, pp. 1-11. Available at: https://aclanthology.org/2020.figlang-1.1. [Accessed 25 August 2024].

Gregor, S. and Hevner, A.R., 2013. Positioning and presenting design science research for maximum impact. *MIS quarterly*, pp.337-355.

Heinsen, F.A., 2024. Softmax Attention with Constant Cost per Token. arXiv preprint arXiv:2404.05843.

Hevner, A.R., March, S.T., Park, J. and Ram, S., 2004. Design science in information systems research. *MIS quarterly*, pp.75-105.

Huang, K., Wang, Y., Zhu, F., Chen, X. and Xing, C., 2024. *Beyond AI: ChatGPT, Web3, and the business landscape of tomorrow*. Springer, pp. 170-174.

Jain, A., Kulkarni, G. & Shah, V., 2018. Natural language processing. International Journal of Computer Sciences and Engineering, 6(1).

Jiang, Z., 2024. Transforming the Finance Industry in China with ChatGPT. *Frontiers in Business, Economics and Management*, *13*(1), pp.80-83.

Järvinen, P., 2007. Action research is similar to design science. *Quality & quantity*, *41*, pp.37-54.

Kar, A.K., Varsha, P.S. and Rajan, S., 2023. Unravelling the impact of generative artificial intelligence (GAI) in industrial applications: A review of scientific and grey literature. *Global Journal of Flexible Systems Management*, *24*(4), pp.659-689.

Korinek, A., 2023. Generative AI for economic research: Use cases and implications for economists. Journal of Economic Literature, 61(4), pp.1281-1317.

Kudo, T., and Richardson, J., 2018. *SentencePiece: A simple and language independent subword tokenizer and detokenizer for Neural Text Processing*. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 66-71. Available at: https://aclanthology.org/D18-2012/. [Accessed 15 June 2024].

Liebowitz, J., 2021. Data analytics and AI. ProQuest Ebook Central, pp.12-13. Available at: https://ebookcentral.proquest.com/lib/lut/reader.action?docID=6264228. [Accessed 2 May 2024].

Lin, Y. and Yu, J., 2023. Study on the Modernization of Supply Chain Management of the Luxury Industry in the Context of the Digital Economy. *Academic Journal of Management and Social Sciences*, *4*(1), pp.5-11.

Malacaria, S., Grimaldi, M., Greco, M. and De Mauro, A., 2023. Business Talk: Harnessing Generative AI with Data Analytics Maturity. *International Journal on Cybernetics & Informatics (IJCI)*, *12*(7).

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A.H., 2011. Big data: the next frontier for innovation, competition, and productivity. *McKinsey Global Institute*. Available at: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation. [Accessed 1 July 2024].

Marcus, G. & Davis, E., 2020. GPT-3, bloviator: OpenAI's language generator has no idea what it's talking about. *MIT Technology Review*. Available at: https://www.technologyreview.com/2020/08/22/1007539/gpt3-openai-language-generator-artificial-intelligence-ai-opinion/. [Accessed 19 July 2024].

McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J. and Barton, D., 2012. Big data: the management revolution. *Harvard business review*, *90*(10), pp.60-68. Available at: https://hbr.org/2012/10/big-data-the-management-revolution. [Accessed 7 August 2024].

Microsoft, 2024a. Excel for Microsoft 365: Copilot for Microsoft 365. Available at: https://support.microsoft.com/en-us/office/identify-insights-with-copilot-in-excel-52d97339-86c0-431c-b46c-e7b07b2898dd. [Accessed 25 August 2024].

Microsoft, 2024b. Copilot Lab. Available at: https://copilot.cloud.microsoft/en-US/prompts/all?products%2Fname=Excel&createdBy=CreatedByAll. [Accessed 25 August 2024].

Miller, T., 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence*, *267*, pp.1-38.

Miró Maestre, M., Martínez-Murillo, I., Martin, T.J., Navarro Colorado, B., Ferrández, A., Suárez Cueto, A. and Lloret, E., 2024. Beyond Generative Artificial Intelligence: Roadmap for Natural Language Generation. Available at: https://www.proquest.com/docview/3081472602?accountid=27292&parentSessionId=l9FoqX3dls39yoR3WMur6RpRN%2Fx%2F4rIVSNvsUxLPMPQ%3D&pq-origsite=primo&sourcetype=Working%20Papers. [Accessed 25 September 2024].

Nadkarni, P.M., Ohno-Machado, L. and Chapman, W.W., 2011. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, *18*(5), pp.544-551.

Naveed, H., Khan, A.U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N. and Mian, A., 2023. A comprehensive overview of large language models. *arXiv preprint arXiv:2307.06435*.

OpenAI, et al., 2023. GPT-4 technical report. Available at: https://www.proquest.com/docview/2787735742?accountid=27292&parentSessionId=Is9WHv5yb%2BvgZHcarGL2rW6KIBuzmUFawSM%2FdrMY%2Fgs%3D&pq-origsite=primo&sourcetype=Working%20Papers. [Accessed 25 September 2024].

Ooi, K.-B. et al., 2023. The potential of generative artificial intelligence across disciplines: Perspectives and future directions. *Journal of Computer Information Systems*, pp.1-32.

Panesar, A., 2021. Machine learning and AI for healthcare, pp.72-75.

Peffers, K., Tuunanen, T., Rothenberger, M.A. and Chatterjee, S., 2007. A design science research methodology for information systems research. *Journal of management information systems*, *24*(3), pp.45-77.

Perkins, M. and Roe, J., 2024. The use of Generative AI in qualitative analysis: Inductive thematic analysis with ChatGPT. *Journal of Applied Learning and Teaching*, *7*(1).

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I., 2019. Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), p.9. Available at: https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf. [Accessed 27 June 2024].

Rane, N.L., Tawde, A., Choudhary, S.P. and Rane, J., 2023. Contribution and performance of ChatGPT and other Large Language Models (LLM) for scientific and research advancements: a double-edged sword. *International Research Journal of Modernization in Engineering, Technology and Science*, 5(10), pp. 875-884.

Rashkin, H., Smith, E.M., Li, M., and Boureau, Y-L., 2019. *Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset*. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 5370-5381. Available at: https://aclanthology.org/P19-1534/. [Accessed 24 June 2024].

Riemer, S., et al., 2023. A generative AI roadmap for financial institutions. *BCG Global.* Available at: https://www.bcg.com/publications/2023/a-genai-roadmap-for-fis. [Accessed 27 July 2024].

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, *1*(5), pp.206-215.

Saivasan, R. and Lokhande, M., 2023. Exploring Use Cases of Generative AI and Metaverse in Financial Analytics: Unveiling the Synergies of Advanced Technologies. *International Journal of Global Business and Competitiveness*, *18*(Suppl 1), pp.77-86.

Sennrich, R., Haddow, B., and Birch, A., 2016. *Neural Machine Translation of Rare Words with Subword Units*. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Volume 1: Long Papers*, pp. 1715-1725. Available at: https://aclanthology.org/P16-1162. [Accessed 15 May 2024].

Shmueli, G., Bruce, P.C., Yahav, I., Patel, N.R. and Lichtendahl Jr, K.C., 2017. *Data mining for business analytics: concepts, techniques, and applications in R*. John Wiley & Sons, pp. 366-387.

Shneiderman, B., 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), pp.495-504.

Siegel, E., 2013. *Predictive analytics: The power to predict who will click, buy, lie, or die*. John Wiley & Sons, pp. 15-17, 155-158. Available at: https://ebookcentral.proquest.com/lib/lut/reader.action?docID=4334745. [Accessed 5 August 2024]

Singh, A., Shukla, N. and Mishra, N., 2018. Social media data analytics to improve supply chain management in food industries. *Transportation Research Part E: Logistics and Transportation Review*, *114*, pp.398-415. Available at: https://www-sciencedirect-com.ezproxy.cc.lut.fi/science/article/pii/S1366554516303817. [Accessed 27 September 2024].

Strubell, E., Ganesh, A. and McCallum, A., 2020, April. Energy and policy considerations for modern deep learning research. In *Proceedings of the AAAI conference on artificial intelligence* Vol. 34, No. 09, pp. 13693-13696.

Sui, Y., Zhou, M., Zhou, M., Han, S. and Zhang, D., 2024, March. Table meets llm: Can large language models understand structured table data? a benchmark and empirical study. In Proceedings of the 17th ACM International Conference on Web Search and Data Mining, pp. 645-654. Available at: https://doi.org/10.1145/3616855.3635752. [Accessed 16 June 2024].

Tarallo, E., Akabane, G. K., Shimabukuro, C. I., Mello, J., & Amancio, D. (2019). Machine learning in predicting demand for fast-moving consumer goods: An exploratory research. IFAC-PapersOnLine, 52(13), 737-742. Available at: https://www.sciencedirect.com/science/article/pii/S240589631931153X?via%3Dihub. [Accessed 23.03.2022].

Shubho, O.Q., Tumpa, Z.N., Dipto, W.I.R. and Alam, M.R., 2022. Real-time data visualization using business intelligence techniques in small and medium enterprises for making a faster decision on sales data. *Decision Intelligence Analytics and the Implementation of Strategic Business Management*, pp.189-198.

Vaswani, A., 2017. Attention is all you need. *Advances in Neural Information Processing Systems*. Available at: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html. [Accessed 25 July 2024].

Veale, M. and Binns, R., 2017. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society, 4*(2), p.2053951717743530.

Vom Brocke, J., Hevner, A. and Maedche, A., 2020. Introduction to design science research. *Design science research. Cases*, pp. 2-5.

Waller, M.A. and Fawcett, S.E., 2013. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business logistics*, *34*(2), pp.77-84.

Wang, T., Zhou, N. and Chen, Z., 2024. Enhancing computer programming education with llms: A study on effective prompt engineering for python code generation. *arXiv preprint arXiv:2407.05437*.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q.V. and Zhou, D., 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, pp.24824-24837. Available at: https://proceedings.neurips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html. [Accessed 17 July 2024].

Wolniak, T., 2023. The concept of descriptive analytics. *Scientific Papers of Silesian University of Technology – Organization and Management Series – Issue No. 172*. Available at: https://managementpapers.polsl.pl/wp-content/uploads/2023/06/172-Wolniak-3.pdf. [Accessed 2 May 2024].

Yan, L., Martinez-Maldonado, R. & Gašević, D., 2023. Yan, L., Martinez-Maldonado, R. and Gasevic, D., 2024, March. Generative artificial intelligence in learning analytics: Contextualising opportunities and challenges through the learning analytics cycle. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pp. 101-111. Available at: https://dl.acm.org/doi/10.1145/3636555.3636856. [Accessed 2 May 2024].

Zhang, H., Dong, Y., Xiao, C. and Oyamada, M., 2023. Large language models as data preprocessors. *arXiv preprint arXiv:2308.16361*.

Zhao, H. et al., 2024a. Revolutionizing finance with llms: An overview of applications and insights. *arXiv preprint arXiv:2401.11641*.

Zhao, S. et al., 2024b. Retrieval Augmented Generation (RAG) and Beyond: A Comprehensive Survey on How to Make your LLMs use External Data More Wisely. *arXiv preprint arXiv:2409.14924*.

Zhou, J. and Chen, F. eds., 2018. *Human and machine learning*. Cham: Springer International Publishing, pp.267-268.

Zhou, M., 2024a. Improving LLM understanding of structured data and exploring advanced prompting methods. Available at: https://www.microsoft.com/en-us/research/blog/improving-llm-understanding-of-structured-data-and-exploring-advanced-prompting-methods/. [Accessed 25 September 2024].

Zhou, X., Zhao, X. and Li, G., 2024b. LLM-Enhanced Data Management. *arXiv preprint arXiv:2402.02643*.

Appendix 1: Interview structure Starting questions

Starting Questions

1. Which business area do you belong to, and what are the main responsibilities of your role?

Excel Usage and Data Processing

2. How often do you use Excel in your daily tasks, and for what purposes?
3. What type of data do you usually work with?
4. Are there any repetitive tasks in the data pre-processing phase that you wish you could automate or facilitate?

Challenges and Reporting

5. What challenges do you face when analyzing data?
6. What kind of reports or summaries do you most often create from the data?
7. What are the most time-consuming aspects of creating reports and visualizations?

Automation and AI Tools

8. Are there any other particular aspects of your job that you think Copilot could improve or streamline?
9. Have you used other automation or AI tools in your work? If so, how have they helped you?

Ending Question

10. Do you have any other comments or thoughts on using Copilot 365 to support analytics?