Junior Management Science

www.jums.academy
ISSN: 2942-1861

Success Factors and Development Areas for the Implementation of Generative AI in Companies

Julian Anton Meyer

Technical University of Munich

Abstract

With the significant increase in public interest in ChatGPT since its breakthrough following the public release in November 2022, an expanding array of application possibilities is being discovered. This heightened interest is also reflected in economic contexts and for businesses. These Generative AI (GenAI) models are believed to have the potential to contribute trillions of dollars in value to the global economy. Now, pioneering companies face the challenge of successfully leveraging this Generative AI technology to their advantage, positioning themselves successfully at the forefront of AI. The adoption of Generative AI proves to be neither straightforward nor simple for companies and is associated with various challenges. Within this thesis, these challenges will be identified by conducting a multiple-case study involving expert interviews. Practical insights will be obtained to identify the decisive factors for the successful adoption of Generative AI, and these insights will be translated into a hands-on implementation framework for companies.

Keywords: ChatGPT Enterprise; Generative AI; GenAI; GenAI adoption; GenAI framework

1. Introduction

1.1. ChatGPT versus Turing Test

> *"I PROPOSE to consider the question, 'Can machines think?'"*
>
> *Turing (1950, p. 433)*

To date, a question one could argue to be answered with yes, at least partially. However, how one argues this is partly a matter of interpretation and definition. Especially since Turing, in this context, is using "intelligence" as a synonym for "the capacity to think" and human rational (Hanna, 2023, p. 2). Intelligence may be defined by others very differently, hence the subjectivity in this argumentation. The 'Turing Test' or the 'Imitation game' was proposed by Turing in 1950 to assess the intelligence of computers based on generating human-like responses (Turing, 1950, pp. 433-434). Despite widespread criticism of the test, it represents a central, thought-provoking idea. An idea that would still concern people to this date: can computers have real human-like intelligence, or do they already have it, and what are the consequences?

The test involves a covert interrogation game with three players, one of whom is replaced by a computer. The computer would pass the test not based on the correctness of the answers but rather if the human interrogator could not reliably distinguish the covert written responses of the human from those of the computer (Turing, 1950, pp. 433-441). Since the test relies on written human-like responses, it is especially interesting in this context regarding the current developments in the field of Generative AI (GenAI) models such as ChatGPT, which can indeed produce human-like texts.

I wish to express my gratitude to those who contributed to the realization of this bachelor thesis. Special acknowledgment goes to my supervisor, Benedikt Blümelhuber, for continuous support, valuable guidance, and especially consistent trust throughout the research process. I am also grateful to the department and Univ.-Prof. Dr. Dr. h. c. mult. Horst Wildemann for facilitating the exploration of the chosen research topic. I extend my appreciation to the interview participants, which cannot be explicitly mentioned to ensure anonymity. Their willingness to share insights and perspectives significantly influenced the formulation and development of the presented theoretical framework.

DOI: https://doi.org/10.5282/jums/v10i1pp1-23

Can a state-of-the-art chatbot like ChatGPT 4.0 pass such a test? There is no clear answer to that, it depends on the type of test. An example approach to answer this question resulted in answering with “No, it cannot pass the Turing Test” (Hanna, 2023, p. 8). However, it is important to note that the original Turing Test from 1950 did not provide specific guidelines for conducting the test; it was more of a thought experiment by Turing. Nonetheless, current GenAI models cannot pass various machine-human-like tests and benchmarks, yet they come dangerously close, or should one rather say, unsurprisingly close?

One could argue that current GenAI models are indeed capable, as proposed by Turing in 1950, of "understanding and speaking English." This is unquestionably a reality. How should one navigate this reality? How does one manage the associated risks? How can GenAI be leveraged? These questions will be explored within the scope of this thesis from the perspective of businesses. Due to the novelty of this technology, companies still face numerous challenges in the successful adoption of this technology. A framework for integrating these GenAI technologies will be developed, enabling companies to effectively harness this reality to their advantage.

1.2. GenAI’s economic potential

From an economic perspective, why should companies even consider incorporating GenAI into their business in any form? What economic potential does this technology represent?

An estimate by McKinsey, identifying 63 GenAI use cases across 16 business functions, suggests that GenAI could have an annual total contribution of $2.6 trillion to $4.4 trillion. 75 percent of this value comes from four use cases: Customer operations, marketing and sales, software engineering, and research and development. Furthermore, after accounting for overlaps and considering additional beneficial impacts on knowledge workers’ activities, the total sum could grow to $6.1 trillion to $7.9 trillion annually. (Chui et al., 2023, pp. 1-10)

The potential impact is enormous, but it remains a potential. To gradually unlock this potential, companies must now evolve towards incorporating GenAI technologies into their practices, simultaneously engaging in and managing risks, and overcoming challenges associated with it. However, there is still no unified approach or generally known blueprint on how companies can perform this transformation. This will be explored in the framework of this thesis by identifying the strategic, organizational, and technical factors that influence the successful deployment of GenAI in companies. The focus lies on Large Language Models such as ChatGPT and their implementation and application within the framework of corporate processes. Particular attention is given to gathering the utmost practical insights to ensure high relevance in the results.

2. Literature Review

2.1. Overview of current literature on GenAI

The following section provides a brief overview of the current research status regarding GenAI, specifically ChatGPT.

A recent study by Liu et al. from 2023 analyzed the current development of ChatGPT-related research. Figure 1 illustrates the monthly publication counts of ChatGPT-related papers as well as the cumulative daily submitted papers on arXiv (an online archive of scholarly articles) from 2022 to April 2023. It demonstrates a significant upward trend, indicating the growing and sustained interest in ChatGPT-related research. (Liu et al., 2023, p. 2)

As this demonstrates only general results regarding the topic of ChatGPT, it provides a good indicator of the overall interest in GenAI, but not specifically in the subtopics. Liu further categorized her analysis results based on the types of publications related to ChatGPT. This allows for a more detailed insight into the distribution across various fields.

As Figure 2 depicts, Computation and Language is by far the largest field in this database on GenAI. Machine Learning, Computers and Society, and AI are the next largest fields. There are many other areas, but they are relatively underrepresented. Particularly, the field of Applications, which is of interest to this research paper for analyzing the use cases and integration strategies for companies, is only weakly represented. (Liu et al., 2023, pp. 2-3)

The consequences concluded from these findings are presented in the subsequent methodology chapter in the introductory part. To address this research gap, the Methodology chapter will elaborate on how and through which methods this thesis aims to contribute.

2.2. Insight into underlying principles of Large Language Models

For a thorough analysis of how companies can successfully integrate GenAI models into their business processes, it is crucial to understand the underlying architecture of the most prevalent type of GenAI models: text-based models such as ChatGPT. They are based on Large Language Models (LLMs) combined with an easy-to-use interface through which humans can prompt requests. The primary goal in the research and development of Language Models (LMs), especially LLMs, was to improve their effectiveness in handling Natural Language Processing (NLP) tasks. (Chang et al., 2023, p. 7)

Defining LLMs works best by first explaining what a model and an LM is. A model, which in its base is a set of rules or math equations, in this specific case a LM, is designed to understand, replicate, and generate human language, achieved by computing probabilities of subsequent word series (Chang et al., 2023, p. 4). Foregoing text is analyzed by the model, most commonly by the n-gram model (Brown et al., 1992, pp. 467–480), and represents the input to a distribution model to obtain an output value that most accurately predicts the next words of a given sequence.

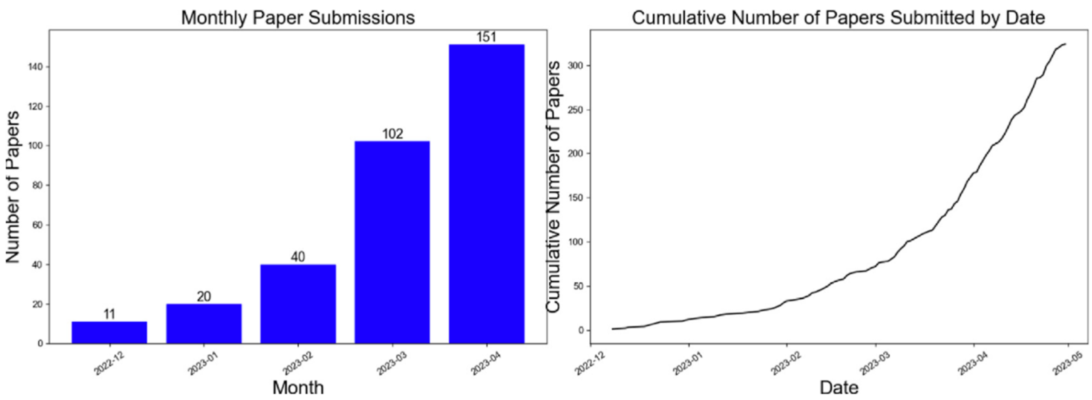

Figure 1: Graphical representation of ChatGPT-related research (Liu et al., 2023, p. 2)

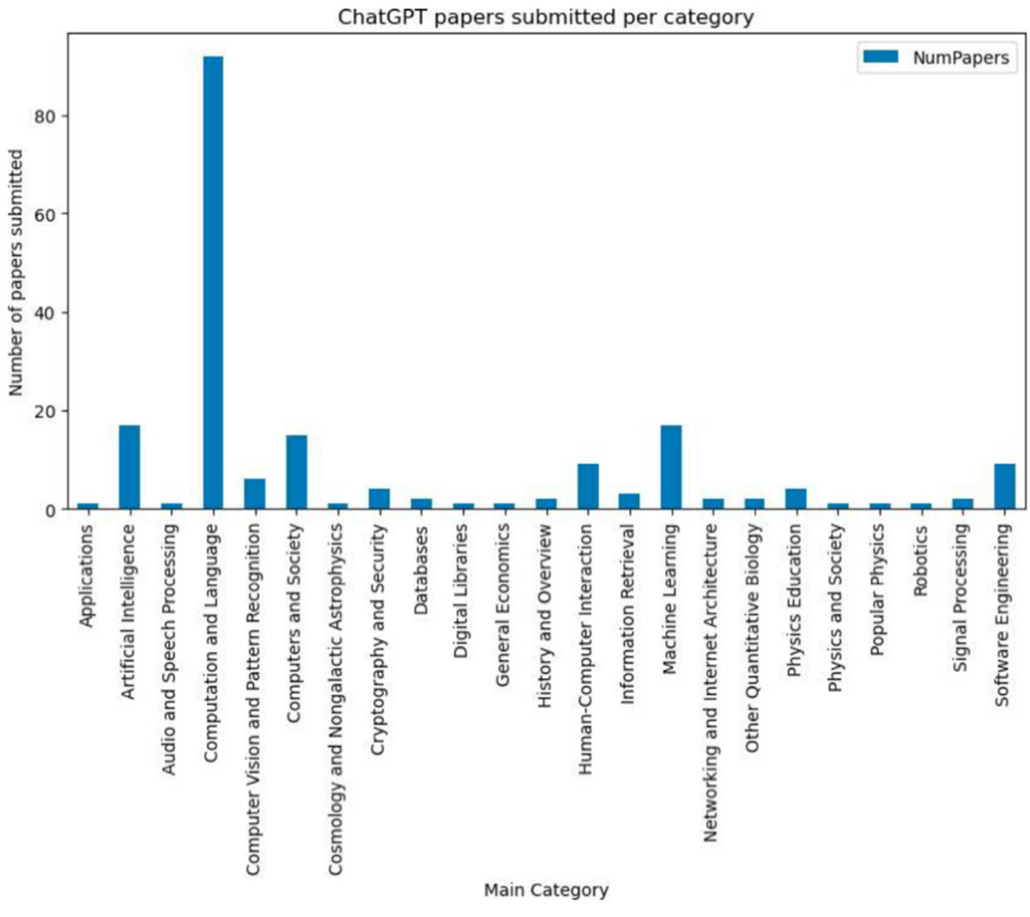

Figure 2: The distribution of ChatGPT papers submitted across various fields (Liu et al., 2023, p. 3)

Problems faced by basic LMs are phenomena such as compositions of complex linguistics, rare words, and overfitting issues. (Chang et al., 2023, p. 4)

A far more advanced version of basic language modeling that not only remarkably improves on these weaknesses, are LLMs. Trained on a massive amount of data, hence the word "Large", as well as utilizing derivations of the 2017 newly introduced transformer architecture (Ozdemir, 2024, pp. 36-43), LLMs are capable of far more complex and versatile tasks and deliver significantly more sophisticated results. They are based on many different principles, with one principle standing out: 'Attention', which enables them to recognize context in language. No matter how the different LLMs are derived from the original transformer architecture, they all operate on a similar working method: **Distributing attention to previously assigned tokens.**

To break that apart, first, it is important to understand the terminology of the token. "A token is the smallest unit of semantic meaning, which is created by breaking down a sentence or piece of text into smaller units; it is the basic input for an LLM. Tokens can be words but also can be "subwords," [...]. Some readers may be familiar with the term "n-gram," which refers to a sequence of n consecutive tokens" (Ozdemir, 2024, p. 37).

The objective is to comprehend and correctly interpret the relationship between the tokens assigned to the input text. The key is to combine and weigh as precisely as possible the semantic meaning (basically, word definitions) and context (with the surrounding tokens) to generate the most contextually rich token embeddings conceivable. This is where the power of transformers comes into play. They utilize attention calculations to achieve this complex counterplay, shifting attention between word definitions and contextual correctness (Ozdemir, 2024, p. 43); Something usually only intuitively achievable by humans.

As seen in the example in Figure 3, contextual awareness is essential for understanding Natural Language (NL). In the first sentence, "ruler" is intended to stand for a measurement tool, compared to the second sentence, where it is in place of a leader or commander. Distributing attention between the semantic meaning of words and the context of surrounding tokens allows the LLM to differentiate between these two sentences, which is critical for advanced Natural Language Understanding (NLU).

2.3. Large Language Model providers

There are many different GenAI models built upon LLMs, such as GPT-4, Bard, LLaMA 2, PaLM 2, BloomberGPT, and Claude, to name a few. The chronological release of these is shown in Figure 4. The variety of LLMs is extensive, giving a good overview of recent developments. However, an LLM can manifest itself through multiple Application Programming Interfaces (APIs). For example, the Azure OpenAI service and the OpenAI ChatGPT Enterprise offer run on the same underlying model. They are based on OpenAI's GPT models. As we aim to stay as practical as possible within the scope of this thesis, the focus is not on model diversity but rather on the integration into companies. Therefore, as part of the qualitative investigation, it will later become apparent that these are the common models used in practice, and thus, the focus lies on the above mentioned solutions.

Models on the upper half of the timeline are open-sourced, while the ones below are closed-source models, additionally, dark blue rectangles stand for 'instruction-tuned' models, and light blue rectangles for 'pre-trained' models (Naveed et al., 2023, p. 2).

2.4. Fields of GenAI

Following the previously provided overview of the functionality and underlying principles of GenAI and its various providers and models, the next step involves categorizing it. GenAI encompasses all scenarios in which an AI generates something 'new'; this can take on various entities. Subsequently, GenAI will be categorized based on these output entities. Given the absence of a uniformly established definition of GenAI and its categorization, an appropriate approach is to observe its scientific context. A literature review conducted by García-Peñalvo and Vázquez-Ingelmo in 2023 illustrates how the topic of GenAI and its entities is reflected in scientific publications. One of their studies displays the publications of papers from 2020 to May 2023 on GenAI categorized by the output generated (García-Peñalvo & Vázquez-Ingelmo, 2023, p. 13).

As Figure 5 demonstrates, the thematic scope of GenAI publications includes generated content in the forms of Images, Data, Text, Video, Videogame assets, Code, 3D, and Audio (García-Peñalvo & Vázquez-Ingelmo, 2023, p. 13). For further consolidation, considering the context of this work focusing on introducing GenAI and its use cases into businesses, some of these fields can be amalgamated. With a focus on practical applications, three overarching categories emerge: Media (Video, Images, Audio), Text, and Code (Code, 3D, and Videogame Assets). To clarify the category "Code", besides secondary aspects such as Videogame Assets and 3D, it includes producing numerical algorithmic code across various programming languages, code debugging, addressing missing segments in numerical code, and translating existing code into different programming languages (Kashefi & Mukerji, 2023, p. 1). Moreover, GenAI models can be utilized to directly convert human input text into code, a process known as text-to-code (Gozalo-Brizuela & Garrido-Merchan, 2023, p. 19).

Ultimately, the categorization is a matter of interpretation, but in this context, such segmentation into Text, Media, and Code, is advantageous for clarity and overview. As will be evident later, these categories are good distinguishers for different applications and business use cases, with text-based solutions being the main one. As this research paper focuses on LLMs specifically and consequently on all text-based solutions, the categories Media and Code will only be addressed secondarily. Additionally, as mentioned earlier, the research results from Figure 5 include only data until May of 2023. Certainly, there have been shifts in interest and, consequently, in the distribution and number of publications since then, especially due to the increase in accessibility and utilization of GenAI and its expanding range of applications. However, these changes have no impact on the categorization.

2.5. Status quo of GenAI in companies

2.5.1. Overview of application areas and use cases

After outlining the fundamental functionalities of GenAI and its classification and categorization, this section briefly shows some potential applications of GenAI in corporate environments. The focus later lies on the actual factors affecting the implementation rather than primarily on the use cases. Nevertheless, the following provides a couple of examples of different application areas in order to establish a comprehensive overview of the application possibilities for the subsequent discussion.

GenAI can be deployed in various types of applications across different industries, for example, customer support,

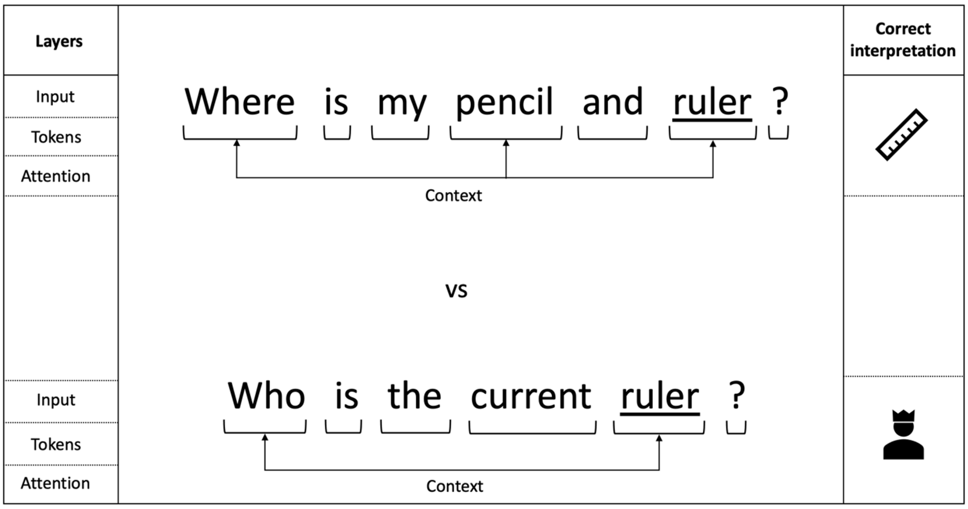

Figure 3: Example of contextual differences in meaning

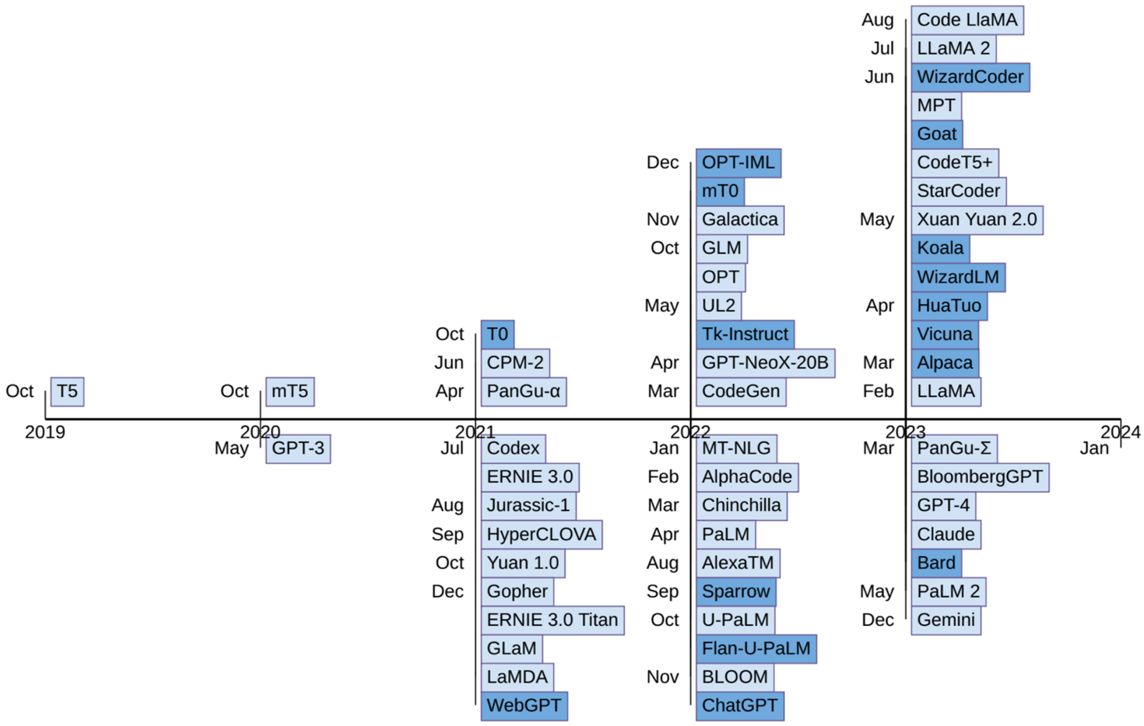

Figure 4: Chronological display of LLM releases [. . .] (Naveed et al., 2023, p. 2)

knowledge management, marketing and sales, education, health, recruitment, social media, and translation to name a few.

In customer support scenarios, GenAI can be used in the form of chatbots to provide users with tailored responses, thereby enhancing customer interactions across diverse sectors such as customer service, marketing, and e-commerce. The technology's versatility allows for customization to meet specific business needs, and therefore potentially contributing to improved user satisfaction. GenAI's proficiency extends to machine translation applications, where its neural network architecture is used to train on large multilingual text datasets. This enables GenAI to comprehend linguistic structures and semantic relationships, as described in chapter 2.2 by NLU and attention distribution, across different languages, resulting in translations of great accuracy and naturalness compared to traditional rule-based systems. The possibility to fine-tune a model further allows the adaptation to specific translation tasks for specific business settings, such as medicine or engineering. Here it can be used to translate technical terminologies into every-day language. For content writing, GenAI seems to be very valuable for generating high-

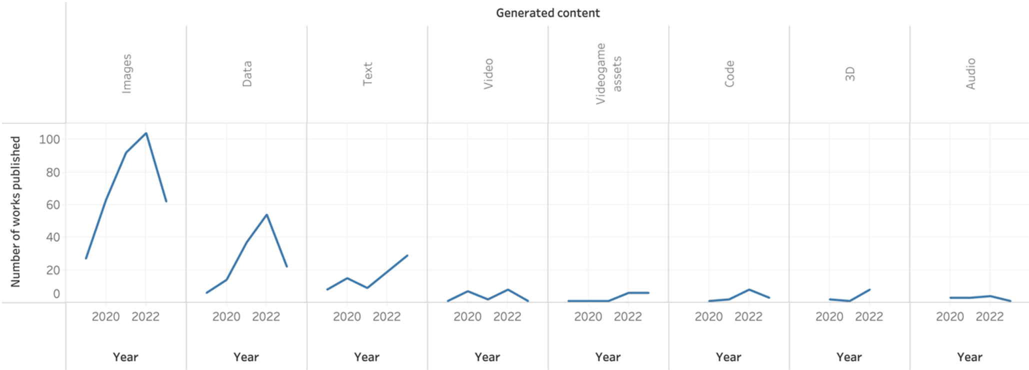

Figure 5: Number of works published over the years grouped by generated content type (García-Peñalvo & Vázquez-Ingelmo, 2023, p. 13)

quality NL text. Its ability to learn from vast datasets distinguishes it from rule- or template-based methods, making it particularly interesting for businesses requiring scalable content production, such as product descriptions, blog posts, reports, and text summaries. Another example for fine-tuning is that a journalist can efficiently generate article ideas or catchy headlines by providing a brief topic description and streamlining the content creation process. GenAI can also be used for information retrieval through its capability to answer diverse queries posed in NL. Users can obtain accurate and detailed responses by utilizing a chatbot interface with a pre-trained model. These models can also be trained and limited to specific datasets to control what data your answers will be relying on. Adjusting parameters like sampling temperature and noise level allows users to control the creativity of responses, ensuring flexibility in tailoring outputs to meet specific needs. The chapter demonstrates GenAI's transformative impact on NLU and NLP, presenting opportunities for improved efficiency and user experiences across different industry application areas. (Sarrion, 2023, pp. 32-43)

To gain insights into the most relevant advantages of GenAI applications in the corporate context, a study by Raj et al. from 2023 is referenced. Utilizing the Preference Selection Index and Complex Proportional Assessment approaches, the potential benefits of ChatGPT were weighted and prioritized in a tabular format based on their actual utility in businesses. (Raj et al., 2023, p. 1)

Table 1 shows the various advantages that can arise from the implementation of ChatGPT in the corporate context. The main categories are Cost Saving (CS) factors for businesses, Enhanced Customer Engagement (ECE) aspects, and advantages by Generating High-Quality Content (GHC) (Raj et al., 2023, p. 3). These are further detailed into their respective sub-categories. Additionally, this also provides a first useful overview of which potential advantages generally exist. The results for the overarching categories CS, ECE, and GHC are as follows: Generating High-Quality Content is by far the most valuable advantage, followed by Enhanced Customer Engagement and Cost Savings ranking third (Raj et al., 2023, pp. 5-6). It is noteworthy that CS and ECE are relatively close together in the evaluation.

Furthermore, the following results in Figure 6 offer an initial indication of which sub-fields are most relevant for the subsequent discussion. This is crucial as we aim to stay as practical as possible in the research for this work, focusing on the introduction of GenAI use cases in companies. Accordingly, the most relevant fields that offer the most benefits in the corporate setting are the most interesting for implementation.

Besides the results of the overarching categories, the evaluation of the top five specific sub-benefits is more intriguing for this research. The complete results can be seen in Figure 6. "Providing quick, informative, and more natural responses" (ECE1) under the category of "Enhanced Customer Experience" (ECE) is the most advantageous application of ChatGPT for business operations, according to the study. "Personalize customer interactions and tailor responses based on the customer's preferences" (GHC3) and the "Ability to generate human-like text" (GHC2) from the category "Generate High Quality Content" are very closely ranked, securing the second and third positions, respectively. "Automate repetitive tasks such as answering frequently asked questions" (CS3) and "Leads to a more positive experience for the customer" (ECE2) constitute the last two of the top five use cases from this analysis. Thus, these top five use cases, along with the remaining ones from Figure 6, provide a solid foundation and overview of the applications that will be further discussed in the context of the successful integration of GenAI into businesses. (Raj et al., 2023, p. 8)

2.5.2. Impact of GenAI use in corporate context

The actual effectiveness of GenAI in its variety of entities in its deployment in the corporate context was demonstrated by a very recent field experiment conducted by Dell'Acqua et al. at Harvard Business School of Technology & Operations in September of 2023. They investigated the impact of GenAI using GPT-4 on performance on realistic, complex, and knowledge-intensive tasks in collaboration with the Boston Consulting Group, a global management consulting firm.

The field experiment involved 758 consultants, with an initial establishment of a performance baseline on similar typical tasks. Benchmarking the initial situation is crucial for

Table 1: Explanation of different aspects of benefits and their sub-benefits [Adapted by author from Raj et al. (2023, p. 3)]

Benefits	CS	Cost Savings	ECE	Enhanced Customer Engagement	GHC	Generate High Quality Content
Sub-Benefits	CS1	Increased efficiency within a business	ECE1	Providing quick, informative, and more natural responses	GHC1	Save businesses time and resources for content creation
	CS2	Improved accuracy within a business	ECE2	Leads to a more positive experience for the customer	GHC2	Ability to generate human-like text
	CS3	Automate repetitive tasks such as answering frequently asked questions	ECE3	Increased customer satisfaction and loyalty	GHC3	Personalize customer interactions and tailor responses based on the customer's preferences

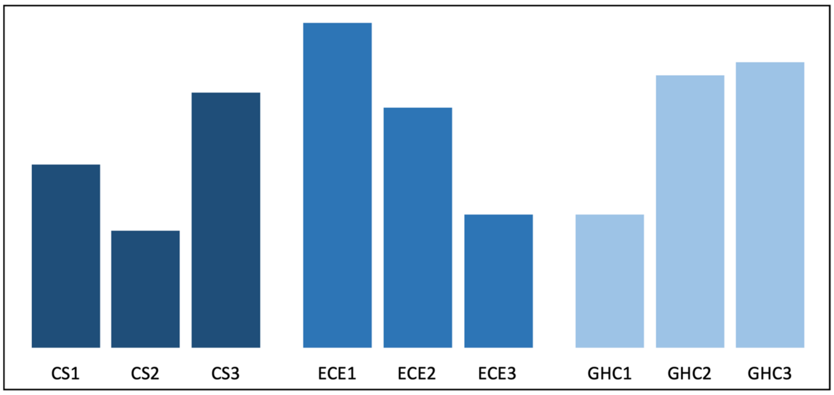

Figure 6: Illustration of sub-benefits parameters scores [Adapted by author from Raj et al. (2023, p. 7)]

conducting a meaningful comparative analysis later on. The consultants were allocated to three groups. The first group had no AI access, the second group had GPT-4 AI access, and finally, the third group had GPT-4 AI access along with an additional prompt engineering overview. For certain of the 18 realistic consulting tasks that were investigated, the results showed that consultants who used AI to handle these were overall significantly more productive. They processed the tasks 25.1% faster. Additionally, they were able to complete an average of 12.2% more tasks and demonstrated 40% higher quality results compared to the control group. However, these figures are based on a specific subset of the 18 tasks. According to the Harvard Business School study, these tasks are referred to as "inside the jagged technological frontier," meaning they have proven to be suitable tasks for the application of GenAI. For the remaining tasks outside this technological frontier, there was a 19 percentage points less likely outcome to produce correct solutions by the consultants compared to those without AI. (Dell'Acqua et al., 2023, pp. 1-19)

To conclude on the relevancy of this field experiment for this research paper, a particularly relevant insight gained from it is that there are tasks and use cases where the introduction of GenAI technology proves extremely advantageous, while in some cases, tasks might be better solved by a human alone. These cases, termed "inside the jagged technological frontier" of GenAI, need to be differentiated from tasks "outside the jagged technological frontier" of GenAI. Only through this differentiation can a successful, sustainable, and particularly rewarding implementation of GenAI in companies be ensured. The identification of these use cases thus emerges as one of the cornerstones and prerequisites for the introduction of GenAI. The empirical part of the study, in the form of expert interviews, aims to elaborate on how this identification can be achieved in practice and emphasizes its importance.

2.6. Critical success factors for the integration of GenAI

Many companies aspire to leverage the aforementioned potential benefits by integrating GenAI into their business processes, such as chatbots, advanced translations, and knowledge management support, among others. Realizing that there are many as yet unexplored difficulties along the way of implementing GenAI models into their company and establishing best practices, the topic of how to successfully adopt GenAI is becoming more apparent. Given the short time span from the public release of ChatGPT in November of 2022 (Gordijn & Have, 2023, p. 1) to the beginning of 2024, there have been only a few to no publications addressing this exact topic. The following is intended to highlight the current research findings regarding the successful implementation of GenAI in companies.

2.6.1. Influencing factors toward adoption

A study conducted by Prasad Agrawal in 2023 yielded some interesting results regarding the influencing factors for the adoption of GenAI in organizations. The study followed a framework that considered three overarching categories of influencing factors: Technology, Organization, and Environment. These unfolded into several subcategories, as shown in Figure 7. Results indicated that compatibility, competition intensity, organizational size, and environmental uncertainty were positively associated with the likelihood of corporate adoption of GenAI technology. On the contrary, complexity and regulatory support had a negative association with the likelihood of the adoption of GenAI technologies in companies. Absorptive capacity, relative advantage, and technological resource proficiency showed no statistically significant relationship to distinguish between adoption or no adaptation. This provides a good initial impression of which factors may have an impact. (Prasad Agrawal, 2023, pp. 1-11)

The key findings of this analysis are described below. As mentioned, absorptive capacity, relative advantage, and technological resource proficiency showed no clear indication and will not be further interpreted in this step. The negative association of complexity acts as a barrier to GenAI adoption due to the technology's immaturity, the absence of widely accepted standards, and the complexities involved in trying to adopt this new technology. There was no clear interpretation of the negative influence of regulatory support. However, it was noted that a certain environment with the right supportive regulations and policies must be created by policymakers to support GenAI adoption for companies. In this field, there may be changes soon with the enforcement of the EU AI Act (Schuett, 2023, pp. 1-4). Compatibility has a positive impact. For example, if companies have previous experiences aligning information systems with GenAI applications, and these turn out to be compatible with their existing information infrastructure, it indicates that future GenAI adoptions are more likely to succeed. Competition intensity, as a driving force in the business environment, was found to positively impact firms by encouraging receptiveness to GenAI technologies. The adoption of GenAI becomes more pronounced when competitors use it strategically, leading to increased concerns about competitive differentiation among adopters, surpassing those of non-adopters. Environmental uncertainty plays a crucial role in promoting the adoption of GenAI. Companies operating in environments marked by higher levels of uncertainty in their relationships with trading partners are more inclined to perceive opportunities, possibly contributing to the adoption of GenAI. The size of organizations plays a significant role in the adoption of GenAI, with larger firms being more likely to embrace this technology. Companies that have initiated the use of GenAI show fewer concerns about expenses related to acquisition, replacement, and ongoing costs compared to organizations that have not adopted the technology. The obstacles to GenAI adoption include software and hardware costs, consultancy support costs, as well as challenges related to installation and integration. Larger organizations with greater resources are positively influenced towards adopting GenAI. (Prasad Agrawal, 2023, pp. 11-14)

2.6.2. Adopting GenAI in organizational settings

Another perspective on possible adaptation paths is provided by the Harvard Business Review in Technology and Analytics by Davenport and Alavi from 2023. The article titled "How to Train Generative AI Using Your Company's Data" offers an interesting initial insight into how GenAI can be integrated into companies by training a GenAI model with their own corporate data. Particularly interesting for companies is the potential leverage of GenAI capabilities in the field of knowledge management to express complex topics in articulate language (Davenport & Alavi, 2023, p. 2).

Davenport and Alavi explain, that in the pursuit of customizing LLMs for specific domains, three main approaches can be employed. Firstly, training a domain-specific LLM from scratch is a rare, cost and resource-intensive method, as it demands vast amounts of high-quality data, which most companies don't have, considerable computing power, and expert data science talent. Bloomberg's creation of BloombergGPT for finance exemplifies this approach, utilizing over 40 years of financial data. BloombergGPT is an LLM the size of 50 billion parameters designed to achieve best-in-class results specific to financial benchmarks (Wu et al., 2023, p. 4). Secondly, the fine-tuning approach involves modifying an existing LLM and adding domain-specific content to a pre-trained model. Google's Med-PaLM2, tailored for medical knowledge, achieved notable success, answering 85% of U.S. medical licensing exam questions (Singhal et al., 2023, pp. 1-2). Despite its advantages in requiring less data and computing time, fine-tuning can be expensive and demands data science expertise. Thirdly, prompt-tuning, a common method for non-cloud vendor companies, involves freezing the original model and then modifying an LLM through prompts containing domain-specific knowledge. Morgan Stanley, for instance, utilized prompt tuning to train OpenAI's GPT-4 for financial advising (Ayoub et al., 2023, p. 7; Morgan Stanley, 2023). This approach is computationally efficient and doesn't require extensive training data. However, it presents challenges in handling large and unstructured text data, often necessitating the use of vector embeddings (Li et al., 2023, p. 1350). These are three popular approaches to incorporate proprietary data into GenAI models. (Davenport & Alavi, 2023, pp. 3-6)

To draw a first conclusion from these various approaches available for companies, developing an own LLM is often financially unrealistic for most organizations due to the high cost involved. While this might be a viable option for the largest enterprises and governments, it still demands a substantial time commitment. If the decision is made not to pursue in-house development, organizations can choose more cost-effective approaches. These approaches enable the adaptation and integration of an off-the-shelf model or LLM service using proprietary enterprise data. Several techniques can be employed to link data to an LLM, including few-shot

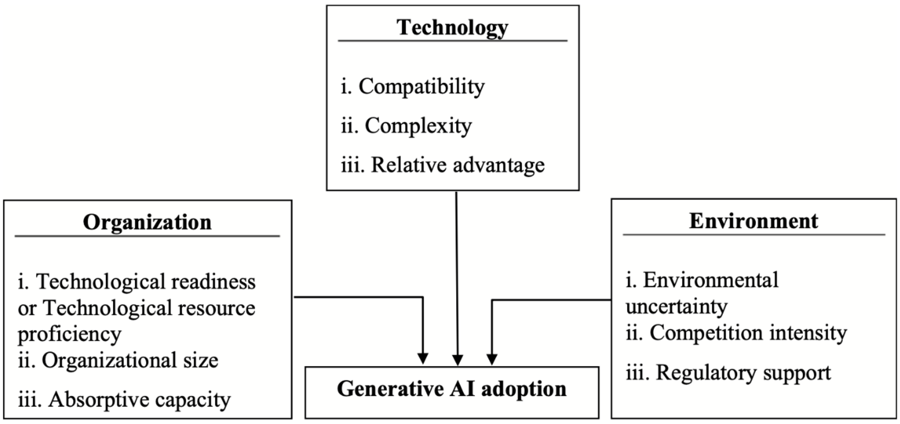

Figure 7: Proposed research model (Prasad Agrawal, 2023, p. 4)

prompting, the previously mentioned prompt-tuning and fine-tuning, and the Retrieval Augmented Generation (RAG) method. (Sweenor & Ramanathan, 2023, pp. 3-31)

The literature presented so far provides an initial theoretical insight into how companies can leverage, utilize, and partially integrate GenAI. However, there is a gap in comprehensive publications specifically addressing the success factors in the integration of GenAI into businesses. The current research in the field of AI, GenAI, and LLMs primarily focuses on functionality and, to some extent, on application possibilities but lacks describing detailed strategies on how companies can effectively integrate them into their business processes.

3. Methodology

3.1. Introduction

GenAI technologies only had their big breakthrough very recently, such as ChatGPT's public release in November 2022 (Gordijn & Have, 2023, p. 1), and since then a wave of research has been triggered. There have been many publications in the past on AI in general (Maslej et al., 2023, p. 24), but since the release of ChatGPT, the number of papers specific to GenAI has been increasing rapidly (Liu et al., 2023, p. 2). While conducting this research, it quickly became apparent that there are predominantly publications on the functioning of LLMs and general concerns of GenAI and, to some extent, application areas of GenAI. However, there are only very few publications that address the critical factors for a successful adaptation of these technologies for businesses, as described previously. In order to approach this meagerly researched topic, the following qualitative research approach is particularly suitable for detaching oneself from existing theories, misconceptions, and assumptions and inductively developing new theories (Schwaiger & Meyer, 2009, p. 413). For this reason, expert interviews were conducted in addition to the literature review to gain real-world insights from business perspectives. These are intended to identify additional important factors that the respective interviewee has experienced in their corporate environment, factors that have either hindered or supported the integration of GenAI models inside companies. As a result, the following section of this paper introduces the methodology and qualitative approach for this empirical research.

3.2. Data collection and methodology

Due to the limited literature on the introduction of GenAI into companies and the subsequent conduction of expert interviews, an inductive research approach is being pursued (Eisenhardt, 1989, pp. 532, 537; Gioia et al., 2013, pp. 17, 21). The goal is to create a new theory and framework based on these new insights, which companies can use as a guide for the successful implementation of GenAI models in their own businesses. The data source for the qualitative analysis consists of four semi-structured, in-depth expert interviews and the resulting transcriptions. The data analysis was conducted using the Gioia method on three levels (Gioia et al., 2013, p. 21).

3.2.1. Semi-structured interviews

Semi-structured interviews offer the advantage of providing the interviewee the necessary freedom in conversation, responses, and individual verbal expressions while still ensuring a guide during the interview that both the interviewer and interviewee can follow. Additionally, the interviewer has the flexibility to ask follow-up questions and delve deeper in case of any uncertainties. This ensures the necessary flexibility during the interviews. (Adams, 2015, pp. 493-494)

This is especially important in this case, as the research question being investigated is a very new and untouched area, making guidance during the interview essential. For this reason, the semi-structured interviews are closer to conventional structured interviews. Pre-formulated questions were prepared to obtain the most detailed answers, given the novelty of the topic of GenAI integration and the lack of established common practices in companies. Consequently, terms and phrases that would typically be known for describing observations within the company in this context are generally less prevalent, requiring more precise support in the questions and more guidance.

3.2.2. Expert interviews

The individuals interviewed in these conversations are experts in the sense described by Gläser and Laudel in 2010: An expert, in this context, refers to the specific role of the interviewee as a source of specialized knowledge about the social phenomena under investigation. Expert interviews serve as a method to tap into this knowledge (Gläser & Laudel, 2010, p. 12). In the context of this research, experts refer to individuals who have direct experience with the integration or hindrance of GenAI models in companies. This particularly involves individuals in managerial and implementation roles related to internal IT processes or those who may already oversee GenAI applications within the company. The 'insider' knowledge gained from these experts, derived from real-world scenarios, is of extremely high significance to gain additional insights beyond the literature review. For this reason, identifying a smaller, more selectively chosen group of experts was of higher significance rather than settling for potentially more but semi-experts who might not contribute relevant and valuable new insights to this very recent GenAI integration topic during the interviews.

3.2.3. Interview guide

In the methodology section of this study, there are recurrent references to an 'Interview Guide' (Appendix C). However, it is important to note that there were two different 'versions' of the Interview Guide: the initial version sent to potential interview candidates and the second version. These two versions were used interchangeably, which makes a precise differentiation in wording secondary in this context. The first version, titled 'Information & Interview Guidelines', included a cover letter, an explanation of the research background, an overview of the essential interview process details, the consent form, and a summary of the thematic questions. The second version, employed during the interviews, differed only in including fully formulated questions. These questions are additionally provided in Appendix D. They were designed to be used in a dynamic style, allowing for spontaneous adjustments and weighting of questions during the interview. The questions were structured into an introduction section, the main questions comprising three parts, and the outroduction part. The questionnaire is extensive, with questions precisely formulated and categorized into three main themes: 'Strategic', 'Organizational', and 'Technical' factors influencing the integration of GenAI in companies. With 37 questions, which would exceed an appropriate length of an expert interview, the Interview Guide is designed to enable both prior and live dynamic adjustments of order and weighting of questions during the interview, based on the interviewee's personal expertise in these three areas. This information was partially ascertained during the preliminary acquaintance phase or revealed during the interview itself. In other words, if the interviewee's expertise focuses on technical factors, the questioning was weighted towards that aspect, allowing the interview partner to provide as detailed knowledge as possible. The remaining questions related to other aspects are then addressed secondarily. This approach aims to achieve the highest resolution and relevance in the shared knowledge and responses.

3.2.4. Sampling method and target population

After extensive preparation and the completion of the detailed Interview Guide, as described in the previous chapter, several potential interview candidates were identified. The interviews were offered to be conducted in English and alternatively in German to reduce potential language barriers that might discourage candidates from participating. The researcher's personal professional network, a leading IT consulting firm, was used as the starting point for case selection. Additional industries from which the interviewees were drawn include software and the insurance sector. As a result of brief preliminary discussions about the interview questions and personal experience in the relevant field, the selection was narrowed down. Following the iterative approach of conducting interviews, as described later, four interview partners resulted. With their specific expertise and personal experiences from their work environment on the subject of GenAI, including its use cases and introduction into companies along with the associated challenges, they were excellently suited as interview candidates. They held positions in their respective companies that involved significant knowledge and responsibility, as seen in Table 2. Among them was a manager from one of the three leading global business consulting firms, bringing experiences from client projects where GenAI use cases were identified and implemented, as well as insights from internal applications within the company. Additionally, there were two individuals from a major software provider. One is serving as the Head of Brand Creative and Communication with experience in GenAI marketing applications, and another follow-up interview with the company's Chief Technology Officer, who oversees the entire technology aspects of the company. The last interviewed person is from a large insurance company, which, in contrast to the other companies, does not yet have a fully integrated GenAI solution internally but is currently in the process of implementing one. Therefore, contributing valuable experiences concerning difficulties during GenAI integration efforts.

Comprising this knowledge and individual background, these individuals served as appropriate experts, as mentioned in chapter 3.2.2 according to Gläser and Laudel (2010), to support answering the central research question.

3.2.5. Data collection and approach

The interviews were conducted over a period of three weeks. The average interview length was 43 minutes with a range of 37 to 56 minutes, to ensure comparability and a similar level of granularity in the interviews. Theoretical saturation (Glaser & Strauss, 1967) was reached after the fourth interview. All expert interviews were conducted in remote online meetings, recorded via Zoom, and transcribed with the "f4x Audiotranskription" software to prevent context and information loss (Gläser & Laudel, 2010, pp. 152-158). A declaration of consent for data recording and processing was obtained through a consent form, which is the last section

Table 2: Overview of the interviewed experts

Industry	Description	Title	Status of GenAI Use in Company	Impact
Business consulting	Digital & data transformation	Manager	Several working company solutions and client implementation projects	High
Software development	CPM platform for finance	Chief Technology Officer	Working company solutions	High
Software development	CPM platform for finance	Head of Brand Creative and Communication	Working company solutions	Medium
Insurance	Diverse insurance products	Manager Information Security	Solution currently in development	Medium

of the Interview Guide. The transcripts were then manually reviewed, cross-checked with the recordings, and carefully corrected in the sense of falsely transcribed text by the software and other minor errors. Two interviews were held in German. To uphold traceability as well as originality, the original transcripts are in Appendix B. For the analysis part of this research, the translations were done as precisely and contextually correct as possible. To prevent information loss or data distortion caused by the translation, the translated transcripts, as well as the original transcripts, were sent to the interviewees for review and approval after completion and prior analysis.

Regarding the inclusion and exclusion criteria of the collected data, none of the interviews yielded unusable or irrelevant information. No technical disruptions occurred, and there were no data losses during the conduct or transcription processes. Due to the deliberate pre-selection of interview candidates, none of them subsequently proved to be unsuitable. Consequently, all conducted interviews could be included, and there was no need for exclusions.

3.3. Methods of analysis

Details regarding the approach to the qualitative data collection process and analysis have already been discussed. Following, a brief theoretical background will follow, describing the chosen method of analysis by Gioia et al. 2013, before moving on to the description of the coding and synthesis of the collected data.

Gioia's method for analyzing qualitative interview results consists of five stages in which the interview data is coded along a certain analysis structure. In the first step, named 1st-order analysis, a large number of terms and phrases are directly gathered from the interview transcripts, often resulting in many categories and expressions in the code. In the second step, similarities and differences among the numerous categories are to be identified. This is referred to as the 2nd-order analysis. The goal is to reduce the number of 1st-order concepts to a more manageable number, somewhere typically around 25 or 30. During this phase, it is possible that the researcher identifies new connections in the data, resulting in possibly reiterating the questionnaire and adjusting it for further interviews. In the third step, the theoretical realm is reached. Here, the researcher aims to describe the observed phenomena with the identified themes and concepts. Special attention is given to concepts that are not well presented in the current literature. The theoretical saturation (Glaser & Strauss, 1967) is reached when the 1st-order and 2nd-order concepts and themes are sufficient for a comprehensive analysis and no further decisive new findings emerge. These are then further abstracted to the so-called aggregated dimensions. The fourth step involves building a static data structure that shows how the 1st and 2nd-order concepts and themes result in the main aggregate dimensions. This data structure is a key component for the theory that is to be developed. The final step in Gioia's approach consists of connecting all the above-mentioned concepts and data to derive a holistic theory from it. To support this, the data and its interconnections are to be visualized in a dynamic visualization. (Gioia et al., 2013, pp. 20-23)

3.4. Scientific quality criteria

The scientific quality criteria of transparency, scope, and intersubjectivity that apply to qualitative research were maintained in the context of this work. The description and disclosure of the procedure within this methodology chapter transparently and comprehensibly revealed how the data collection and analysis took place. By formulating and posing open-ended questions for discussion, maintaining intersubjectivity was ensured. A consistent set of questions and the Interview Guide were used, contributing to the reproducibility of results and, therefore, ensuring a sufficient scope. By elucidating the appropriateness of the chosen research approach and disclosing the empirical work and the developed theory, a sufficient level of comprehensibility is ensured (Schwaiger & Meyer, 2009, p. 408).

3.5. Data synthesis and coding

As previously described, the interview data were processed and coded using the Gioia method. From the transcripts in Appendix B, a total of 187 1st-order concepts were identified. In contrast to the classical formation of 1st-order concepts, according to Gioia, particular attention was paid here to incorporate statements very close to the original of the interviewees into the 1st-order concepts. This adjustment was made due to the novelty of the topic and the lack of comprehensive abstraction capabilities of interconnecting categories at the time of the interview analysis. Therefore, the approach was slightly modified in the first step to minimize data loss through early abstraction endeavors. These 1st-order concepts resulted in 23 2nd-order themes. Six overarching aggregate dimensions could be derived from those 2nd-order themes: Strategic Grounds, Awareness and Central Enablement, Feasible Framework and Tangible Use Cases, Technical Considerations, Risk Identification and Management, and Regulatory Measures. A static data structure of the 1st-order concepts, 2nd-order themes, and the six overarching aggregate dimensions is provided in Appendix A. These form the basis for the theory that is to be developed. The following will elaborate on these dimensions based on their respective coding structure and 2nd-order themes. (Gioia et al., 2013, pp. 19-21)

Strategic Grounds

Strategic Grounds define the necessary prerequisites for the strategic alignment of GenAI with the respective company that intends to incorporate GenAI models into its structures and business model. Additionally, competitive advantages are crucial in this context; they need to be identified to establish the feasibility of implementing GenAI. Adaptability to tech describes the capability and readiness of the respective company to adopt, based on its structure and employees.

Awareness and Central Enablement

Awareness and Central Enablement address the resistance and acceptance stance by employees towards the GenAI models within the company. It primarily revolves around forming a comprehensive understanding of the company's workforce to achieve a high adoption rate for the introduced GenAI models. This is driven by GenAI awareness within the company. GenAI-specific in-house expertise and awareness need to be nurtured through specialized trainings. The right executive support and leadership are crucial to ensuring a successful implementation. Detailed collaboration and communication within the organization are necessary, for example, to exchange success stories related to GenAI use cases.

Feasible Framework and Tangible Use Cases

A Feasible Framework and Tangible Use Cases are crucial to achieving a sustainable and for the employee's comprehensible introduction of GenAI technologies. Possible preidentification of clear use case definitions can help recognize the potential value gained and find the right approach. Through tangible use cases and best practices, it is also possible to create a very conceivable and practical image for the employees, who are ultimately the end-users, of how they can use the introduced technology in their day-to-day business. Additionally, various rollout approaches are available, which must be carefully chosen depending on the use cases and company specifics to ensure a proper and successful implementation.

Technical Considerations

There are various Technical Considerations that arise when planning to introduce GenAI in any form into a company. Different GenAI solution types and technical challenges must be taken into account. A thoughtful and considered decision must be made, considering various model selection criteria. Model explainability and interpretability are of essential importance in this regard to ensure a certain level of comprehensibility in its usage.

Risk Identification and Management

Risk Identification and Management initially involves the identification of typical risks and mitigation measures. A comprehensive risk assessment must be conducted to establish the foundations for effective risk management later. To successfully address the perceived risks, various concrete actions against data risks need to be undertaken.

Regulatory Measures

A multitude of new Regulatory Measures must be introduced. Initial regulatory guidelines need to be redefined or expanded to incorporate GenAI-specific requirements. If no specific governance of users is in place, at the minimum, thought must be given to potential use restrictions. Furthermore, continuous monitoring and optimization of the introduced GenAI model must be ensured, and ongoing monitoring and evaluations need to be conducted to respond adequately to emerging successes and issues.

The dynamic relationship and interplay of the 2nd-order themes and the six aggregate dimensions are depicted in the Figure 8. At the center is "Awareness and Central Enablement", ensuring the facilitation of the surrounding implementation efforts. The core messages of the interviewees' statements are illustrated around this center. Serving as the foundation and prerequisite are the "Strategic Grounds" around the model, addressing crucial preceding points.

Subsequently, a comprehensive theory will be derived from this coding evaluation. This dynamic model serves as the foundation, which will be modified and adapted after the subsequent final analysis.

4. Results and Discussion

4.1. Interview results and analysis

The following will present and evaluate the interview results along the six overarching aggregated dimensions.

4.1.1. Strategic Grounds

Respective companies need to clearly understand why they want to incorporate a specific GenAI model into their business processes. In principle, identifying business cases and determining business value is crucial, asking, for example, what competitive advantage they can gain. Productivity gains are the most typical benefits that come to mind, additionally, one can leverage the current market momentum, position themselves as pioneers in the field, and potentially even market the gained expertise.

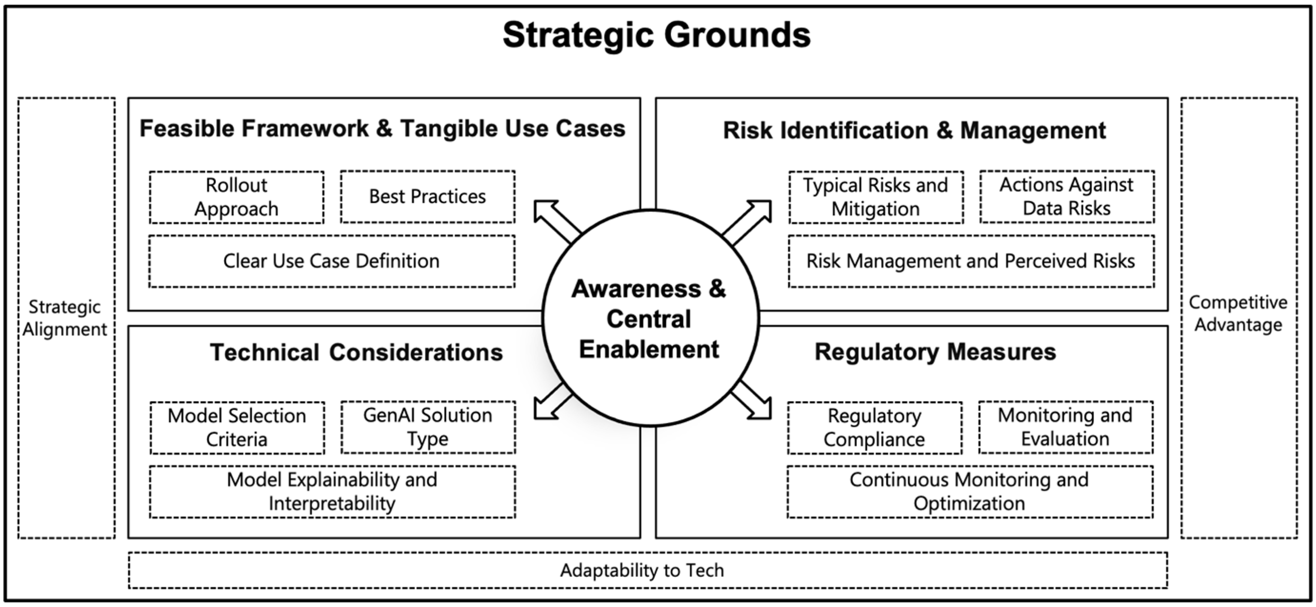

Figure 8: Developed dynamic model of the structural relationships

> *"You can currently take advantage of a very good momentum because there's a lot happening in the market."*

> *"[We] leverage generative AI to really give us productivity gains [. . .]"*

> *"So for the GitHub copilot [. . .] allegedly [it] gives a productivity boost to developers of 30 to 40%. Given its cost, which is like $20 per user per month. If that's true, then it's a no brainer from a commercial point of view, right."*

> *"[. . .] We will likely reach a point where we all use it relatively similarly. So, it's not an endless competitive advantage. But where it currently benefits us a lot is in offering consulting services related to GenAI because it not only affects us but also our customers. And if we are pioneers in that regard, it gives us a competitive advantage to offer consulting services."*

> *"I would say if you snooze, you lose, right? [. . .] People hesitate to move with technology. [. . .] So the early adopters will have the edge, but they have also the challenges because they have to do all the learning curve. [. . .] But if you don't do it, you're definitely going to fall behind because the speed with AI is incredible, right? [. . .]"*

Furthermore, it is essential to ensure a strong strategic fit between GenAI and the company's strategy and culture. This is what is referred to as Strategic Alignment in this context. This significantly enhances the likelihood of successful adoption. However, an imperfect fit should by no means be considered as an exclusion criterion. A clear advantage here is the immense boost in growth.

> *"[. . .] We are a company that's growing 50% a year and [GenAI] is helping us do it even faster. You have to know that we're also a private equity invested company with the intention to exit. So we're on this fast value creation program to improve the value of the company. So AI plays a key role there."*

> *"[. . .] It fits like a glove because our goal is, of course, growth. And as a consulting company, we are also a people business, and scaling essentially only happens through people. We could hire a lot more people, but that also comes with risks. And generative AI makes individual employees even more productive."*

In addition to the strategic fit, adaptability to technology is also a crucial consideration. Different companies have diverse backgrounds within the workforce, such as generational-related factors and varying technical expertise. These need to be considered during the implementation process to achieve the highest possible adoption rate of the introduced GenAI tools. The type, size, and structure of the company also play a role in this context.

> *"If you think about who our employees are, 50 plus or even older. They can use Google, but getting correct answers from ChatGPT is something else."*

> *"Are they ordinary clerks who have only been on the phone their whole lives and maybe occasionally use Google privately? Or are they people our age who have grown up with these issues, for whom using AI is natural?"*

> *"[. . .] Being that we are small or mid sized company, we are very agile and very quick. So we have*

quick, very quick, very flat hierarchy, very quick decision making processes."

"[. . .] As a consulting company, we are much more adaptable, and we will use such technology much more strongly and quickly. [. . .] [We] are used to constant changes in conditions, dealing with new customers, and new employees coming in."

Thus, Strategic Grounds emerge as one of the most crucial foundation elements and prerequisites in the consideration and implementation of GenAI integration initiatives.

4.1.2. Awareness and Central Enablement

To advance the topic of GenAI within the company, a certain foundation must be established. Employees need to be introduced to the topic of GenAI, considering various levels of knowledge and backgrounds. Option rooms and use cases for incorporating GenAI into existing processes must be explored and illustrated, and touchpoints need to be created, especially with employees who may not be tech-savvy. Questions such as "What can we do with this new tool?" or "Why do we need it?" and "How can I effectively use this in my day-to-day activities?" must be addressed. This is crucial to ensure that a later-introduced tool achieves a high adoption rate and does not go unnoticed. All these aspects fall under the theme of "Awareness and Central Enablement." For this, it is important to have dedicated GenAI teams. An illustrative example from the interviews that demonstrates how this can be transferred into practice is the introduction of so-called "GenAI Blackbelts," who act as GenAI experts in the company. They serve as points of contact and GenAI ambassadors.

"[. . .] I am also a so-called GenAI Blackbelt. This is our internal multiplier concept. We have set up a GenAI Lab centrally, which has set itself the mission of introducing GenAI in the Germany, Austria, Switzerland, or Central Europe region."

"Internally, the Blackbelt program is now being rolled out, where, as a Blackbelt, I am required to drive and promote the topic in my office. I will also receive material to set up learning sessions [. . .]"

"And you have the designated AI teams."

Next, one must proactively address the topic of resistance and acceptance by employees. A certain proportion of employees may not immediately resonate with the tool. They might not immediately recognize its benefits and may not know how to integrate it into their day-to-day tasks. Or, they may simply not invest the time for it. A certain 'activation energy' is required to overcome this initial hurdle, which, as later described, can be facilitated through practical explanations, tangible use cases, and trainings, enabling them to become familiar with the tool and build up a certain level of excitement. This proactive approach serves as a preventive measure against potential rejections towards the GenAI tool.

"But there's also another angle which is around adoption, because if you just give it to people, probably 30% will really use it and really get into it, but 60% or whatever, 70% will probably not even bother to use it. And so, you want to get the engagement up, you want to get people using it, build a certain level of excitement and get that adoption up so you really get the value from it."

"So it's just some people are too busy with their day to day to give it a shot."

"I couldn't prompt properly, or I didn't know how specific I could be. The results were somewhat mediocre, and I thought, well, in that time, I could have just done it quickly myself instead."

"I think it's more a little bit of fear and a little bit of "well I don't know how to start. I don't know how to do this.""

To address the lack of in-house expertise and awareness, various methods and communication channels can be employed. Dedicated Microsoft Teams or Slack channels and internal newsletters can be utilized to raise awareness and 'promote' the topic internally. This is where simple and easily understandable use cases should be demonstrated to engage people and spark interest. Specific training sessions and company-wide meetings can complement this, with a focus on comprehensibility and applicability. Moreover, there should be an in-depth exploration of prompt engineering in these sessions. Only through this approach can a highly effective environment be created, enabling users to derive significant value from the introduced GenAI tool. At the same time, attention should also be given to the described risks, such as distortions, hallucinations, biases, and how to handle them. More details on this will be elaborated later in the context of regulatory compliance.

"We have a few Slack channels, and they sort of formed by themself, and there are some GenAI distribution lists that someone set up, where a lot of knowledge is shared. Sometimes very specific things, sometimes very basic things."

"[. . .] Subscribe to certain newsletter distribution lists, and always screen what comes out of there."

"[. . .] [We had] company wide meetings where our CTO introduced it, showed how to use it, showed some of the benefits."

"[. . .] [We had] trainings in the company where people have been educated showing how it works and also on prompt engineering [. . .]"

"Especially with a strong focus on prompt engineering [. . .]"

"Be aware that the data, the responses, may not always be correct, that it may be simply made up."

Further examples of the necessary clear communication and enablement to ensure GenAI awareness throughout the company include the following:

> *"Enablement is crucial. Clearly communicate to employees what they can do with it, what they can do well with it, and what they should avoid."*
>
> *"[. . .] Showing people scenarios like that really helps because it just kind of fires up their imagination [. . .]"*
>
> *"[. . .] There's an element of showing people [what's possible]. And then encouraging them to use it as well."*
>
> *"Creating awareness and also showing examples, such as in one training where the random invention of information is shown."*
>
> *"[. . .] Enablement is crucial and to clearly tell the employees what the risks are and how they need to deal with them."*

Furthermore, it has proven to be highly beneficial to share success stories related to introduced GenAI tools between business units and working groups within a company. This is precisely where individuals with a more conservative attitude towards the topic can observe and identify with it, making it more conceivable for them.

> *"It's definitely really, really crucial because if one department [. . .] has success, then they need to share it with the other departments, right."*

Lastly, to effectively ensure central enablement, strong executive support and leadership are essential. They need to commit to the GenAI topic, not only ensuring the necessary resources for the actual model itself but also for training, dedicated teams, and all supporting functions mentioned. This is crucial to achieve a sustainable introduction and the highest possible adoption rate.

> *"I mean, the executive leadership our C-level is definitely the driving force of AI."*
>
> *"[. . .] Number one, leadership commitment. Clearly, someone from the top must drive this initiative. This person also needs to believe that it has an impact and should provide the necessary resources to drive it within the company."*
>
> *"[. . .] It is crucial that you have a leader and also a sufficiently high-ranking leader who drives it. In our case, even towards a global initiative, a lot was driven from Germany. So, our ***** Central Europe CEO personally took on the topic and founded the GenAI LAB, and he drove it through his peers and network [. . .]. [Our Managing Director and Senior Partner] specifically ensured that consultants were staffed on this internal topic to make it correspondingly large. So, super important."*

4.1.3. Feasible Framework and Tangible Use Cases

A feasible framework for the introduction and implementation of GenAI in a company is crucial for achieving effective and smooth integration of the technology. A central component of this framework is the rollout of the model. There are various rollout approaches that can be considered, and some companies may already have experience with the rollout of other introduced tools. Depending on the specific technology, company structure and size, and employee competence and expertise, the chosen rollout approach should be adjusted.

A notable example from the interviews was the observation of one interviewee who implemented a GenAI tool for a client. The client, an automotive company, exhibited significant differences in competence, receptiveness, and speed to adaption between technology-savvy departments and other 'normal' departments. The approach was quickly adapted to ensure earlier access to the introduced GenAI tools for these departments. This difference in technology affinity between departments can be leveraged. Due to their unique expertise, these departments can identify useful use cases more quickly and experience possible problem areas early on, and these can then be disseminated later through the chosen rollout approach and communication channels throughout the company.

> *"[. . .] [Regarding] decentralized implementation, we actually observed this at the automotive manufacturer. Technology-savvy departments tend to take matters into their own hands; they don't wait for IT to come around and say, "Look, we've developed something for you, and you can use it now." Therefore, from an IT perspective, it's crucial to provide tools quickly, such as a sandbox, so that data analysts, who are already abundant in vehicle development, logistics, and sales, can experiment and implement their own use cases."*

However, it is crucial to note that there is not one universally correct rollout approach. As described, it needs to be adaptively fitted to the situation. An example of this is provided by the following two statements, both from the same person and referring to two different tools that were rolled out within the same company.

> *"Specifically, with ChatGPT, we conducted a pilot over the past year and evaluated its success. The decision was made to roll it out, and it's happening in what I would call a rapid Big Bang, over a very short period. I believe it was rolled out within two to three weeks."*
>
> *"In other examples, if I take the Deckster tool, [. . .] it also went through a pilot, and we have dedicated Slack channels for each pilot with product owners. So, we do have iterative development as well."*

The first tool, ChatGPT Enterprise, was rolled out in a few weeks in a type of Big Bang, all at once. This allowed everyone across the company to access it directly and simultaneously. In contrast, the Deckster tool, which is a PowerPoint integration involving slide creation and providing content from knowledge management in the company style without the need for manual searching, was developed and rolled out in an iterative style involving various product owners.

One of the interviewees, the CTO of a software company, chose a very scientific approach regarding the rollout of the GitHub Copilot. Divided into two groups, comparable to a typical 'control group' and an 'experimental group', he could clearly benchmark the performance gained through the introduced GenAI tool using various metrics.

> *"So for the GitHub Copilot rollout, [. . .] what I didn't want to do was just: "Hear you go, developers take it, run with it, see if you like it". I wanted to take a more scientific approach. And so, what we've done is we've done effectively A / B testing, whereby we rolled it out to a group of engineers and we have engineering metrics, tooling that allows us to understand, the current performance of those engineers. So, we benchmark that performance, benchmark that against other teams. We've rolled it out and now we've been monitoring it for the last four months just to see the impact of that tool."*

In addition to the quantitative measures described, he also incorporated a complementary qualitative approach by conducting surveys.

> *"And then we also take a qualitative approach too, we survey them about every six weeks or so just to kind of ask questions along the lines of the impact of that tool, so on and so forth. So that's generally the approach and that's been really successful."*

That was an example of a rather scientific approach and a very detailed procedure, but it does not necessarily have to unfold exactly like that. In principle, the various rollout approaches, however, share some commonalities, which can be summarized as follows:

> *"Do a hackathon. Do a prototype. Figure out what is possible. Start to narrow down some of the scenarios where we think we can get impact and add value for the products. Then take those prototypes to the next level and then roll them out."*

The importance of use cases should be indisputable by now. Centrally providing tangible use cases is crucial to comprehensibly demonstrate to employees how and for what purposes they can use the introduced GenAI tool. A clear framework and guidelines must be provided to create a productive environment for the use of GenAI.

> *"[. . .] There are very clearly defined frameworks in which generative AI is incredibly helpful."*

> *"[. . .] The big impetus for generative AI was use case identification [. . .]."*

The approach to successfully identify these use cases can vary. A best practice from an interviewee's experience shows that gradually accumulating and filtering use cases proved to be sensible in order to ultimately identify the crucial so-called 'High Value Top Use Cases'.

> *"[. . .] We built a large use case funnel, basically built an idea list and then gradually filtered it to get to the so-called "High Value Top Use Cases.""*

4.1.4. Technical Considerations

Among the technical considerations, one of the most crucial decisions that needs to be made before the rollout is apparent is: 'Which model, in what form, from which provider do I choose to incorporate in my company?' For this, comprehensive knowledge about the corresponding model selection criteria is needed. An interviewee from a company highly advanced in the GenAI field, referring to their company as "the largest OpenAI customer there is", described the process of selecting the right approach as follows: "It's like navigating a multi-level decision tree, I would say." She divided the process into an initial make-or-buy decision under certain criteria, with various consequences, such as the potential training of LLMs or the use of pre-trained models and fine-tuning.

> *"Yes, I think at the core, the first decision you have to make is a make-or-buy decision. That essentially depends on whether there is an existing solution in the market that adequately covers your use case. It could be a ChatGPT Enterprise license if you have many small use cases where you just want a helper alongside. If such a solution doesn't exist, or if commercial off-the-shelf solutions don't meet these requirements, or if you have very specific data security requirements, then building one yourself may make sense. Then you move to the next level, where you also ask yourself: Do I train my own model, or do I use a pretrained model? And if pre-trained, do I want to fine-tune it, or do I use the pre-trained model as is and incorporate my proprietary data, for example, through a Retrieval Augmented Generation (RAG) approach?*

Especially the last point mentioned here, RAG, represents a very practical solution for companies, giving GenAI models the ability to access external data, particularly in complex and knowledge-intensive tasks. This way, effects like hallucination can be significantly reduced, leading to more factual consistency and reliability in the responses (Gao et al., 2023, pp. 1, 17).

To be as close to current practice and application reality as possible, practical and realistic methods for the introduction of GenAI in companies are to be developed. Therefore, the following observation from the same person mentioned previously, who has conducted various GenAI implementation projects in her company as part of client services, provides a good overview of the current common and functional practices.

> *"For most use cases I've seen in companies, they actually rely on a pre-trained large language model. Not even on some small open-source models, but most of them use either the GPT APIs from OpenAI through a ChatGPT Enterprise account or the Azure Open AI Service. But ultimately, it's the same GPT model behind it."*

If a company has very specific requirements and decides to develop its own solution, it is extremely important to develop a well-thought-out architecture, adjust 'freedom' parameters, try different 'temperature' settings as well as to test different token length sizes. Additionally, they have to carefully consider whether a very demanding and resource-intensive fine-tuning process should be rolled out or if the in comparison, less complex and proven RAG approach is sufficient (Gao et al., 2023, p. 17).

> *"For building your own solution, a well-thought-out architecture is essential, defining how much freedom you want to give to a large language model in your solution. That's the consideration. If the model is supposed to do something company-specific, do you really want to fine-tune it, or do you use the mentioned RAG approach."*

The next aspect that needs to be considered is model explainability and interpretability. It primarily revolves around the user's expectation of transparency when using GenAI models. An exemplary experience from the interviewed consulting firm illustrates this:

> *"I recently used ChatGPT to brainstorm GenAI use cases for a company [. . .] The first thing ChatGPT did was go to the [competitor's] website, from another consulting firm. I even provided in the prompt: "I am a ***** consultant." So, that shows that there is no logical reasoning behind it. They are very opportunistic, whatever they find quickly that might fit will be used."*

To meet these requirements of model explainability and interpretability measures such as providing sources by the model, as done in the 4.0 version of GPT, need to be taken. Moreover, users save time when they can extract useful information from the response without having to search for the source.

> *"[. . .] Providing sources is [. . .] an important point. ChatGPT does that quite well through the 4.0 version and the browser integration."*

> *"[. . .] Explainability is super important for us when we use output or information on customer projects."*

> *"I think it's super important because if an employee is unsure, they can look it up directly instead of searching for hours."*

The specific solution types that the companies of the interviewed individuals have in their practice include the previously mentioned ChatGPT Enterprise offering as well as the so-called OpenAI Playground. The crucial point with these solutions is that the input into the models is not used for further training of the models by the providers. This allows for the use of company-sensitive data for prompts and incorporation into the model.

> *"[. . .] OpenAI Playground, basically your data isn't used to train the model. It's just in your own domain."*

If you want to be on the safe side, there is the option of using a pre-trained model and feeding in your specific ground data to generate responses based on this limited dataset. Another client example illustrates this for a company in the legal context:

> *"If you take a pre-trained LLM or a regular LLM service and you input the relevant legal foundations, [. . .] [for example] a German Civil Code, into a vector database, basically embedding it as knowledge, and then you let the model work only with this provided knowledge and get the corresponding text passages. That works really well and is, I would say, always the safest solution when you have a manageable knowledge base that is relevant to you."*

In terms of technical challenges, no major issues were experienced. The challenges experienced were more towards prompt engineering, which has already been addressed in the prompt training and GenAI awareness section.

> *"[. . .] It's actually pretty simple to integrate and use these models via API."*

> *"The challenges are typically more on things like prompt engineering, data engineering, making sure your data is in the right format, it's clean, so on and so forth."*

A certain flexibility in setup and cost can be ensured through special Platform as a Service (PaaS) offerings. In this approach, resources such as computing power, storage, and network costs, are outsourced and made available on-demand. This consumption-based model allows for a concise cost overview in one single fee and a simplified setup of the GenAI model.

> *"'Platform as a Service', which basically means you [. . .] pay for what you use. So, all of that, that power usage, the storage, the network costs, that's all rolled up into a single fee that you pay based on the amount that you use the service. So that's how we will leverage it in a production environment. And that's actually how we're currently leveraging it even with Playground as well. It's a consumption model."*

4.1.5. Risk Identification and Management

As part of Risk Identification and Management, the perceived risks of the company, mainly concerning data leakage, are addressed. Possible other risks are very company-specific, depending on the type, structure, and IT landscape of the organization. Data leakage risks can occur, especially if adequate measures are not taken, such as not using a model that does not share the data with the GenAI model provider. Furthermore, there is the risk that users may use responses from GenAI models directly without verifying them, and in the worst case, unfiltered responses may be sent to customers, for example.

> *"Another risk, of course, is that employees simply adopt the answers one-to-one."*

Important for the GenAI introduction is directly communicating and confronting the employees with this topic. One initial way to manage these risks without directly resorting to regulatory measures is to be proactive. Employees must be made aware during internal training and workshops to treat certain data or tools with special care, aiming to prevent potential risks from arising in the first place. Following the principle:

> *"So, prevention is the best course of action [. . .]"*

It is important to clearly communicate to users, possibly based on use cases, how to categorize and handle different scenarios. An example of a concrete implementation in the form of a three-color stage traffic light system that could be later transferred into regulatory measures is:

> *"[The] AI policy [. . .] has basically green, amber and red use cases for how you can use generative AI. And basically those use cases [that are okay] to pass to Gen AI [green] and the types of data that you need explicit approval for, which is amber, and then the types of data that you explicitly cannot pass to any GenAI. Those are the red use cases."*

4.1.6. Regulatory Measures

As a central component of Regulatory Measures, the emerging regulatory compliance issues around the GenAI topic must be considered. The current most important aspect is the EU AI Act, which companies must adapt to accordingly. Use cases are classified into different risk classes; therefore, companies need to adjust their internal guidelines to comply with the new EU AI Act regulations. A practical example demonstrates how this can be implemented by adding another dimension to already existing risk processes:

> *"Foremost is the EU AI Act for all applications to be deployed in Europe. It regulates certain things but still allows enough room for companies to work. Ultimately, use cases are divided into risk classes, which is not much different from what companies already do with their applications in risk management. You're just adding another dimension. How is it solved? For the customer company, they have mainly considered additional risks posed by GenAI. These could be things like hallucinations or the fact that certain things in a model cannot be traced, as most models are built, yet the benefit outweighs the risks. In such cases, different control mechanisms are needed, and these criteria, risks, and corresponding mitigation measures are embedded in the existing risk process. No additional GenAI risk process is established."*

To further this, various usage restrictions can be implemented. Texts generated by the AI, for example, are marked as such and accompanied by a disclaimer. Apart from visibly marking the GenAI text, there can also be covert marking by employing the watermarking method, where algorithmically detectable statistical markers in the form of a short span of tokens are embedded into the text that are invisible to humans (Kirchenbauer et al., 2023, p. 1). This can be helpful if one wants to trace back GenAI content. Furthermore, the use of the introduced GenAI tools can be restricted until the respective employee has completed specific mandatory trainings and workshops. In addition, they must also agree to the newly established guidelines that describe issues such as factual incorrectness, hallucinations, distortions, and other biases. An example from one of the interview companies illustrates this through the establishment of so-called 'Responsible AI Guidelines':

> *"[. . .] To activate this license, must read our Responsible AI Guidelines."*
>
> *"[. . .] First unlock all the AI systems as soon as employees have completed mandatory training."*
>
> *"[. . .] We always clearly mark what is generated and what has been created by humans. This is to ensure clarity and includes providing a disclaimer."*
>
> *"[. . .] There are potential distortions, hallucinations, biases, every employee is called upon to check the output of the models and not to use it just like that."*

It is important to note that this proactive approach is much more sensible, as employees can resort to alternatives in the case of a potential complete ban. If websites like ChatGPT are blocked company-wide, employees could turn to so-called 'mirror sites' that replicate the content and functions of the original site. This could lead to even more dangerous data leaks. An interviewee working as an Information Security Manager described it as follows:

> *"[. . .] Then you could just have a mirror site that just mirrors ChatGPT and then it works anyway. Then it's not blocked. If employees want to use it, they can do it on their private computer in home office if necessary. There are always ways to bypass the issue."*

Another useful aspect regarding the governance of users is, for example, to gain insights by reading out the network logs of the respective website on how much a tool is being used. This allows measuring the adoption rate or getting a rough overview of how much and by whom it is used. Another aspect gained from that, as part of the initial consideration of introducing a GenAI tool, is that this could also be used to get an idea of the level of interest in GenAI. This helps to better assess how much ChatGPT, for example, is already covertly used before a rollout and whether an introduction is worthwhile.

> *"[. . .] Read the network logs and see how many calls, for example, ChatGPT has made. That was already a good indication for the automotive manufacturer that the technology is being utilized, even if they don't offer their own application."*

Lastly, to ensure continuous monitoring and optimization of the GenAI models and processes, incorporating feedback mechanisms is necessary. Besides already mentioned surveys, possibly setting up monitoring dashboards or simple things like evaluating thumbs up or down, or even the copy button as part of the chatbots, can provide valuable information to further track and develop the tool.

> *"With ChatGPT, for example, you can give a thumbs up or thumbs down, and they actually measure when you press the copy button too."*
>
> *"This may involve implementing a monitoring dashboard or some monitoring interface where a GenAI application owner can randomly check responses."*

This monitoring can be used to evaluate system-related KPIs (Key-Performance-Indicators) that concern topics such as latency or outages. Additionally, measuring gained benefits and converting these to classical KPIs should be considered to be able to track these for management.

The constant involvement of a human checking over the model output in the so-called 'human in the loop' process is important, especially in sensitive cases, to always ensure a high-quality standard.

> *"The point of "human in the loop" is definitely crucial, especially for critical use cases or cases where it's essential for the output quality to be high and for nothing untoward to happen."*

However, in the long-term, to achieve a significantly high level of scalability, one must gradually move away from this temporary solution of the 'human in the loop.' This is well described by an interviewee in the concluding quote:

> *"You start with a human in the loop as a gatekeeper and eventually reach a point where you trust the solution enough to only perform random checks. At some point, you could even use your GenAI models for verification. Haha, the AI checks the AI. But it actually works quite well. We need to evolve in that direction, in my opinion."*

4.2. Discussion

Subsequently, following this detailed analysis of the interview results, all insights are to be summarized and evaluated in conjunction with those from the literature research. These results are intended to be developed and synthesized within a framework that can be used by companies and other interested parties. Derived from the dynamic representation in Figure 8 from the interview results, the obtained results suggest the following proposal for a six-step step framework (Figure 9) that can be structured as follows:

At first, the Strategic Grounds (1) lay the foundation for the successful introduction of GenAI in companies. Secondly, a detailed Risk Identification and Assessment (2) procedure follows. Thirdly, based on that, first concrete Regulatory Measures (3) must be implemented. Consequently, as the fourth step, adequate Technical Considerations (4) must be made. Simultaneously, during these steps, comprehensive GenAI Awareness and Central Enablement (5) must be ensured to drive this process continually. The sixth step involves securing a Feasible Framework and Tangible Use Cases (6) for the rollout process.

The Strategic Grounds represent prerequisites and external influencing factors. Awareness and Central Enablement act as a central foundation, continuously driving the process through necessary facilitation. It is important to note that this constitutes a cyclical process, meaning it does not end at the last step but rather begins anew, continually moving, for example, to identify newly emerging risks, mitigate them, and continually optimize the model. Furthermore, it does not present a rigid sequence of steps; these can be handled iteratively depending on various external factors.

In the following, the final identified key success factors and development potentials for companies to adopt GenAI into their businesses will be detailed and summarized.

Number 1: Strategic Grounds

As part of the Strategic Grounds, companies are encouraged to first address the question of why they want to incorporate a specific GenAI model. It is essential to initially identify business cases and determine business value, such as generating human-like text, providing quick, informative, and more natural responses, and personalizing customer interactions or other potential cost savings factors, as described in Table 1 and Figure 6. Other side effects can also arise, such as potentially increasing the company's overall value and even marketing (commercializing) the gained expertise. A factor from the literature review that should be incorporated here is the observation regarding the competition intensity. This can act as a driving force in the business environment, increasing the receptiveness of these companies to adopt GenAI due to competitive pressure. Companies

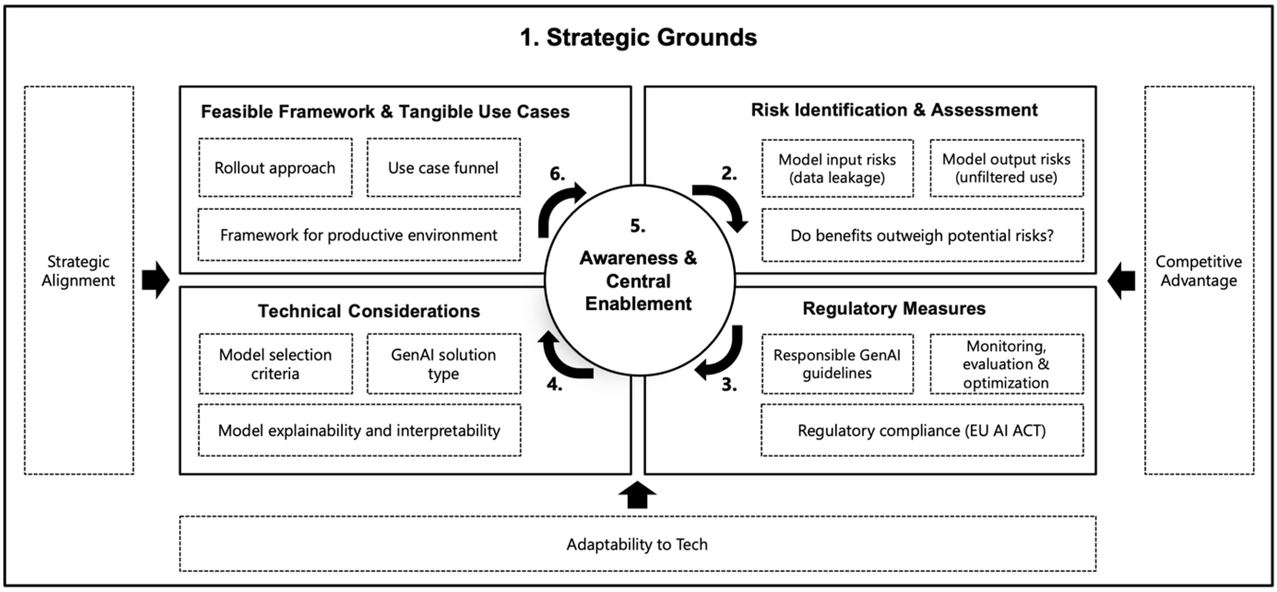

Figure 9: Proposed GenAI implementation framework for companies

in this situation should, therefore, consider GenAI integration more intensively (Prasad Agrawal, 2023, pp. 11-14). Leveraging the current market momentum is an option that should not be overlooked, positioning themselves as pioneers in the field of AI. It's important to consider that there is no endless competitive advantage in incorporating GenAI since competitors will follow; therefore, being on the forefront and having the edge as early adopters will be crucial. Furthermore, the strategic fit of the company's business model and principles to the GenAI technology must be considered, and adaptability to technology should be taken into account. The inclusion of diverse backgrounds within the workforce plays an important role, varying from company to company, in order to achieve the highest possible adoption rate later on.

Number 2: Risk Identification and Assessment

Extremely important is the identification of company-specific risks, primarily those related to data leakage. Mishandling or underestimating the importance of data security measures is one of the most significant risk factors. These risks, along with others, such as the one-to-one adaptation of GenAI output by users and its unfiltered reuse and distribution, must be identified beforehand. A comprehensive risk assessment must then be conducted to subsequently evaluate whether the benefits outweigh the perceived risks. Afterward, proactive measures, as explained in later steps through training and awareness, should be considered early on to ensure that these identified risks do not arise in the first place. In this case, prevention is always the best course of action.

Number 3: Regulatory Measures

Fundamentally, correct, and effective regulatory measures are decisive. This includes considering newly emerging regulatory compliance issues around the GenAI topic, with a central focus being on the EU AI Act. Companies need to adjust their internal guidelines. It is not necessarily required to establish an entirely new GenAI risk process; rather, an additional dimension must be added. GenAI-specific effects, such as hallucinations or partial untraceability, need to be integrated into this dimension, and cases must be classified into different risk classes. Implementing various usage restrictions, for example, through the introduction of 'Responsible AI Guidelines,' is crucial. For instance, the categorization of use cases could be realized through a color-coded 'traffic light' system. Red indicating explicitly prohibited use cases, yellow for approval-dependent cases, and green for free-to-go use cases. Furthermore, in some cases, GenAI-generated text should be marked as such and accompanied by a disclaimer. Employees should only gain access after successfully completing specific mandatory training and workshops, which explain effects like factual incorrectness, hallucinations, distortions, and other biases. It's worth noting that in the case of the prohibition of GenAI websites, employees could easily bypass restrictions by using so-called 'mirror sites'. Therefore, clear and proactive communication and implementation prove more sensible than a prohibition policy or gray zone strategy. Additionally, limited governance of users can be advantageous. Reading out the network logs of GenAI site calls can enable measuring the adoption rate or determining to what extent the introduction of GenAI is worthwhile by assessing how much it was covertly used beforehand. Additionally, a comprehensive 'Continuous Monitoring and Optimization' system should be established. Incorporating feedback mechanisms through surveys and measuring actions such as thumbs up and down, as well as the copy button, is important and should be captured through a monitoring dashboard. This allows for the measurement of system-related KPIs, such as latency or outages, and classical productivity KPIs for management. In the long term, to achieve a significantly high level of scalability, one must gradually move away from this temporary solution of the 'human in the loop' and consider methods such as "AI checking the AI."

Number 4: Technical Considerations

The first technical consideration is the correct model selection. Navigating a multi-level decision tree, the initial make-or-buy decision must be made depending on whether an off-the-shelf solution is sufficient or not. Observations from qualitative research have shown that for most companies, the GPT APIs from OpenAI through a ChatGPT Enterprise account or the Azure Open AI Service are satisfactory. This aligns with the approaches discussed in the literature section described by Davenport and Alavi in 2023, who particularly emphasize that self-developing company own GenAI models and fine-tuning is associated with significantly high costs. However, if companies have very specific requirements, potential training of their own LLMs or the use of pre-trained models and fine-tuning can be considered. Besides the input data, different parameters such as 'freedom', 'temperature', and token length size can then be fundamentally changed but require a well-thought-out architecture. These are associated with considerably more resource effort; however, methods like the RAG approach, which reduces hallucinations and increases factual consistency and reliability, can simplify this, especially for knowledge-intensive tasks. In contrast to the observation made in the literature (Chapter 2.6.1) that the complexity of the models and their associated immaturity is an obstacle, no significant problems were identified in the interviews. On the contrary, the wide range of offers and easy integration through the APIs make it 'relatively' easy. Next, attention must also be given to the model's explainability and interpretability. Providing sources, for example, through the 4.0 version of GPT, can contribute to creating transparency for the end user. Additionally, this saves time since they can trace back information from the response without having to search for the source. In the end, regardless of the chosen GenAI solution type, it is crucial that the input into the models is not used for further training of the public models; otherwise, data security is jeopardized. Alternatively, or additionally, one can use an 'offline' pre-trained model to which specific ground data has been fed in advance to generate responses based on this limited dataset. A particularly attractive consumption-based model for businesses is the PaaS, where computing power, storage, and network costs, are outsourced and made available on-demand. It provides a concise cost overview in one single fee and a simplified setup.

Number 5: Awareness and Central Enablement

The GenAI Awareness and Central Enablement shall act as a central driving force, consistently advancing the GenAI topic within the company and thus fueling this implementation process from the inside. Initially, attention must be paid to the different levels of knowledge backgrounds. Particularly, questions like 'how and for what purpose can I incorporate GenAI in my day-to-day activities' must be encountered through tangible use cases. By introducing dedicated GenAI teams and representatives, sufficient touchpoints must be created, continuously ensuring awareness of GenAI, to later achieve a high adoption rate. These internal GenAI ambassadors act as a multiplier concept, promoting the topic through, for example, learning sessions and driving it into the individual offices. A proactive approach considering potential resistance to adoption by employees must be taken. Some may not immediately resonate with the tool, and some may simply not take the necessary time during their workday to give it a shot. Building up a certain level of excitement, for example, by showcasing the capabilities through demos, can act as a catalyst to help overcome the initial 'activation energy' needed. To counteract the lack of in-house expertise, various countermeasures must be implemented. Through diverse communication channels and internal newsletters, the topic can be further promoted, drawing attention to specific training sessions and company-wide meetings that primarily focus on the comprehensibility and applicability of demonstrations and learnings. An in-depth exploration of 'Prompt Engineering', explaining effects such as distortions, hallucinations, and biases, and addressing potential ethical concerns arising should be a central part of this. Sharing GenAI success stories across business units and working groups can be highly advantageous, especially to reach and engage individuals with a rather reserved or conservative attitude towards GenAI. Strong executive support and leadership towards the introduction are the decisive factors to ensure, through this central enablement, that not only the necessary resources for the actual model itself are sufficient but also for the training and dedicated teams. Practice shows that if not enough employees are staffed on this topic, not much will come of it.

Number 6: Feasible Framework and Tangible Use Cases

As part of the Feasible Framework and Tangible Use Cases, the first step is selecting the appropriate rollout approach, which is based on various factors such as the chosen solution, company structure and size, and the employees' competencies and expertise. Possibly even a mixture of a rollout in waves and phases, pilot projects, iterative development, step-by-step, and big bang should be considered. The difference in technology affinity between departments does not necessarily have to comprise a disadvantage; it can be leveraged during the rollout approach to identify useful use cases more quickly and experience possible problem areas early on, for example, by favoring these tech-savvy departments and including them in pilot projects. A more effort-intensive but insightful approach is the scientific one, which splits into a 'Control Group' and an 'Experimental Group,' where GenAI users are benchmarked against conventional users. This allows for capturing detailed metrics. The clear definition of use cases is crucial. These must be vividly and comprehensibly communicated to the employees. A clear framework and guidelines must be provided to create a productive environment. Accumulating and filtering use cases along a use case funnel is essential to identify the 'High Value Top Use Cases.' These must be collected and made available for everyone to access.

5. Conclusion

In summary, it can be said that based on the insights gained from the interviews and their insights into companies with GenAI, a clear conclusion can be drawn. Companies that see an opportunity based on the proposed framework to integrate GenAI into their business processes should take this seriously. Not just take it seriously but also act on it as quickly as possible. Time is of the essence here. Being at the forefront, being a GenAI pioneer in a specific business field, will prove to be very advantageous not only in the long term but especially in the short term. In contrast to most investments companies make, which are usually planned on a mid- to long-term perspective (Zellweger, 2007, pp. 1-2), GenAI integration represents a contrary strategy. It can be rolled out relatively quickly through the correct rollout strategy outlined in the proposed framework and has the potential to disburse its benefits sooner.

The central research question and objective were to identify the success factors and development areas for companies for the integration of GenAI technologies. The end result is summarized in the developed framework. Based on the developed GenAI implementation framework, companies considering or already implementing GenAI can use it as guidance. It aims to be a practical and hands-on guide derived from real-world company insights, providing assistance and guidance on various success factors while pointing out risks and important considerations. In conclusion, considering the identified benefits throughout this thesis as well as the generally perceived risks by society, the following can be said:

GenAI will not necessarily replace people and their jobs in the short term, but individuals and companies leveraging GenAI will soon replace and surpass those who refuse.

5.1. Limitations

While conducting this qualitative work, there were several limitations. Firstly, it should be mentioned that there is not yet a comprehensive or substantially meaningful amount of research papers in this field, especially regarding the influencing factors of GenAI integration for companies. Therefore, the otherwise extensive and solid foundation of literature is not fully present in this case.

Furthermore, the qualitative analysis, conduct, and evaluation of the interviews are always subject to a certain degree of subjectivity. On the one hand, caused by the chosen sampling method, which included the selection of experts in the personnel network of the researcher, on the other hand, is limited due to the novelty of the topic, as only a small percentage of companies actually have valuable experiences in the GenAI integration process. Additionally, expert interviews and the resulting statements are naturally subject to an inherently subjective and personal influence and opinion of the interviewees. It is also worth noting that generalizing results from a small case study to a larger population is often challenging.

Various data losses and distortions due to conducting interviews online via Zoom cannot be completely ruled out. The best efforts were made to minimize this, as interviews were recorded, transcribed, partially translated, corrected, prepared, and subsequently rechecked and consulted with the interviewees.

Therefore, the proposed framework should be viewed only as a guide and reference for companies and interested parties, not as a comprehensive solution. Ultimately, it is up to companies to determine how they implement GenAI into their business processes. Indeed, at this earlier stage, forming a universally valid assessment or judgment about the plausibility and value contribution of the developed theory is problematic and would border on a philosophical discussion (Schwaiger & Meyer, 2009, p. 408). Instead, it is intended to encourage further research, as described below.

5.2. Further research opportunities

Moreover, the presented results are intended to encourage additional researchers to complement them through further investigations. By encompassing a broader range of perspectives, approaches, types of companies and industries, and various GenAI solutions, this model can be further developed. This ongoing refinement will enable a continuous elaboration of the developed theory over time to ensure a development towards a more robust theory. The following interesting observation encountered during the development of this thesis should encourage reflection and provide a possible impulse for further research.

As with any newly introduced technology, GenAI brings not only the mentioned benefits and risks but also disadvantages. One particularly interesting aspect should be briefly mentioned here. Companies introducing GenAI technologies might be exposed to certain side effects. The use of, for example, ChatGPT by employees could lead to a reduction in the diversity of ideas among the results produced with GenAI. Initially, one might expect that the use of ChatGPT would exclusively improve the diversity of results, given its extensive knowledge base. However, the study discussed in the literature review section from Harvard Business School by Dell'Acqua et al. from 2023 revealed a tendency toward a decrease in the variation of results compared to the control group that had not used ChatGPT. Consequently, companies could be exposed to this side effect and might expect, in some cases, less conceptual variation in the production of GenAI-supported results. (Dell'Acqua et al., 2023, pp. 53-54)

However, this phenomenon is not yet fully explored, as the production of results heavily depends on how and in what context GenAI is used. Additionally, the quality of the results is not taken into account here. For instance, while the semantic similarity of results may increase overall, the relevancy and quality of the results may differ compared to non-GenAI results. Overall, this point is intended to stimulate consideration and reflection and does not represent a fully substantiated statement. (Dell'Acqua et al., 2023, pp. 53-54)

References

Adams, W. C. (2015). Conducting semi-structured interviews. In *Handbook of Practical Program Evaluation*.

Ayoub, T. Y., Amami, A., & Hachaïchi, Y. (2023). Expectations of AI Economic Impact & Giant AI Experiments: Case of ChatGPT 3.5.

Brown, P. F., Della Pietra, V. J., Desouza, P. V., Lai, J. C., & Mercer, R. L. (1992). Class-based n-gram models of natural language. *Computational linguistics*, *18*(4), 467–480.

Chang, Y., Wang, X., Wang, J., Wu, Y., Zhu, K., Chen, H., Yang, L., Yi, X., Wang, C., & Wang, Y. (2023). A survey on evaluation of large language models.

Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., Yee, L., & Zemmel, R. (2023). The economic potential of generative AI.

Davenport, T., & Alavi, M. (2023). How to Train Generative AI Using Your Company's Data.

Dell'Acqua, F., McFowland, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality [Working paper: 24-013]. *Harvard Business School Technology & Operations*.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*(4), 532–550.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey.

García-Peñalvo, F., & Vázquez-Ingelmo, A. (2023). What do we mean by GenAI? A systematic mapping of the evolution, trends, and techniques involved in Generative AI. *International Journal of Interactive Multimedia and Artificial Intelligence*, *8*(4).

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology.

Glaser, B., & Strauss, A. (1967). *Discovery of Grounded Theory: Strategies for Qualitative Research*.

Gläser, J., & Laudel, G. (2010). *Experteninterviews und qualitative Inhaltsanalyse* (Vol. 4). Springer-Verlag.

Gordijn, B., & Have, H. t. (2023). ChatGPT: evolution or revolution. *Medicine, Health Care and Philosophy*, *26*(1), 1–2.

Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). ChatGPT is not all you need. A State of the Art Review of large Generative AI models.

Hanna, R. (2023). How and Why ChatGPT Failed The Turing Test.

Kashefi, A., & Mukerji, T. (2023). ChatGPT for programming numerical methods. *Journal of Machine Learning for Modeling and Computing*, *4*(2).

Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., & Goldstein, T. (2023). A watermark for large language models.

Li, L., Zhang, Y., & Chen, L. (2023). Prompt distillation for efficient LLM-based recommendation. *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*.

Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., & Liu, Z. (2023). Summary of ChatGPT-related research and perspective towards the future of large language models. *Meta-Radiology*.

Maslej, N., Fattorini, L., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Parli, V., Shoham, Y., Wald, W., R., Clark, J., & Perrault, R. (2023). The AI Index 2023 Annual Report.

Morgan Stanley. (2023). Morgan Stanley Wealth Management Announces Key Milestone in Innovation Journey with OpenAI.

Naveed, H., Khan, A. U., & Qiu, S. (2023). A Comprehensive Overview of Large Language Models.

Ozdemir, S. (2024). *Quick Start Guide to Large Language Models*. Addison-Wesley.

Prasad Agrawal, K. (2023). Towards Adoption of Generative AI in Organizational Settings. *Journal of Computer Information Systems*, 1–16.

Raj, R., Singh, A., Kumar, V., & Verma, P. (2023). Analyzing the potential benefits and use cases of ChatGPT as a tool for improving the efficiency and effectiveness of business operations. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *3*(3).

Sarrion, E. (2023). *Exploring the power of ChatGPT*. Apress.

Schuett, J. (2023). Risk Management in the Artificial Intelligence Act. *European Journal of Risk Regulation*, 1–19.

Schwaiger, M., & Meyer, A. (2009). *Theorien und Methoden der Betriebswirtschaft: Handbuch für Wissenschaftler und Studierende*. Vahlen.

Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Hou, L., Clark, K., Pfohl, S., Cole-Lewis, H., & Neal, D. (2023). Towards expert-level medical question answering with large language models.

Sweenor, D., & Ramanathan, K. (2023). *The CIO's Guide to Adopting Generative AI: Five Keys to Success*. TinyTechMedia LLC.

Turing, A. M. (1950). I.—Computing Machinery and Intelligence. *Mind*, *59*(236), 433–460.

Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., Kambadur, P., Rosenberg, D., & Mann, G. (2023). BloombergGPT: A large language model for finance.

Zellweger, T. (2007). Time horizon, costs of equity capital, and generic investment strategies of firms. *Family Business Review*, *20*(1), 1–15.