I. Introduction

The following section introduces the research topic and outlines the content of this thesis. The first subsection (I.1) highlights the relevance of the investigated research topic. Subsequently, the second subsection (I.2) identifies the research gaps of this thesis and derives relevant research questions. Consequently, subsection I.3 describes the structure of this thesis, followed by subsection I.4 which presents the research design and positioning. Finally, subsection I.5 provides a summary of the anticipated contributions of this thesis.

I.1 Motivation

> *"If a machine can think, it might think more intelligently than we do, and then where should we be?" (Turing, 2004 [1951], p. 485)*

In 1951, Alan Turing, a British mathematician and computer scientist, proposed the Turing Test to answer the question: Can machines think? (Turing, 2004 [1951]). The test's core idea was that a machine could be considered intelligent if it could convincingly mimic human behavior. In the Turing Test, a human participant engages in a natural language conversation with both a human and a machine, without knowing which is which. The premise was that if the human participant could not distinguish between the human and the machine during the conversation, the machine could be deemed intelligent. Although the test has faced various criticisms and has undergone multiple revisions over the years, its core idea remains powerful (French, 2000). Therefore, it can be concluded that the Turing Test foresaw future developments in AI that would extend the boundaries of machine intelligence toward human-like capabilities (Boden, 2016).

Building upon, and perhaps even expanding Turing's vision of AI, the present day has indeed witnessed a revolution in the capabilities of AI-based systems, which can generally be regarded as algorithms that reference human intelligence. This revolution has been driven by significant advancements in AI models, such as Generative Pretrained Transformers (GPT)-3 and GPT-4, which have not only passed variations of the Turing Test (Biever, 2023; Dwivedi et al., 2023) but have also consistently demonstrated human-like performance across various tasks (Fügener et al., 2021b). This has led to substantial advancements in various fields since AI can now conduct data-intensive and repetitive tasks in a fully automated manner and do so faster than humans (Dellermann et al., 2019; Rai et al., 2019; Raisch & Krakowski, 2021). For example, AI is being integrated into processes such as the delegation of transportation services to humans (e.g., Uber) (Stelmaszak et al., 2024), automated loan credibility checks (Strich et al., 2021), and advertisement bookings (Einola et al., 2024). Moreover, especially in the healthcare domain, AI has advanced processes such as drug development (Lou & Wu, 2021),

medical information extraction (Wang et al., 2018), and clinical decision-making (Lebovitz et al., 2022; Pumplun et al., 2023). These advances are a direct result of AI's inherent characteristics, which include its increasing autonomy, ability to learn, and inscrutability. Thereby, AI is transforming roles, processes, and structures within organizations (Berente et al., 2021).

While the inherent and distinctive characteristics of AI enable innovative applications, they also present significant sociotechnical challenges that necessitate effective integration within organizational contexts (Benbya et al., 2021). Consequently, one challenge stems from the statistical reliance of contemporary AI models. Unlike earlier rule-based systems, these modern models depend heavily on statistical methods, which can result in unpredictable behavior. This reliance highlights the need for human-AI collaboration across various tasks, particularly in organizational settings where accountability is crucial and the costs of errors can be significant (Benbya et al., 2021). In the healthcare domain, for instance, where misdiagnoses can have fatal consequences, human oversight becomes vital for effectively managing AI outcomes (Jussupow et al., 2021). Furthermore, as the AI is capable of learning, data is needed to adapt AI to local requirements or to incorporate special knowledge that the models have not been trained on. For example, AI models trained primarily on English data will not be useful or will be only minimally effective, when applied to other languages (Névéol et al., 2018).

Moreover, as AI continues to be integrated into organizational structures, several studies highlight the potential negative consequences if it is not properly integrated. This underscores the importance of careful planning and consideration to mitigate these risks. Research shows that if such challenges are not properly handled, introducing AI into organizations can even result in unintended consequences, including the erosion of organizational knowledge (Sturm et al., 2021), tensions, and manipulations of employees to their advantage (Strich et al., 2021) and a decrease in the performance of individuals and groups (Fügener et al., 2021). Therefore, successful integration of AI requires careful consideration of the challenges posed by both technological components and the organizational context to achieve a cohesive sociotechnical fit.

To effectively navigate these challenges and fully harness the benefits of AI, it is essential to deepen our understanding of this fit between AI and organizational context. In short, **'AI'** can be defined as referencing human intelligence (Berente et al., 2021), an advanced definition will be derived in section II.1.1. Drawing on a definition from the human-computer interaction literature, **(organizational) context** can be understood as the "certain setting […] that imposes constraints or significance for doing and completing the tasks" (Zhang & Li, 2004, p. 130) at hand. The process of fitting AI to meet the specific requirements of the organizational context will be referred to as **'AI alignment'** (see

Figure 1). This phenomenon of AI alignment sets the stage for the interaction between humans and AI in the context of their tasks, which will be referred to as **'human-AI collaboration'**.

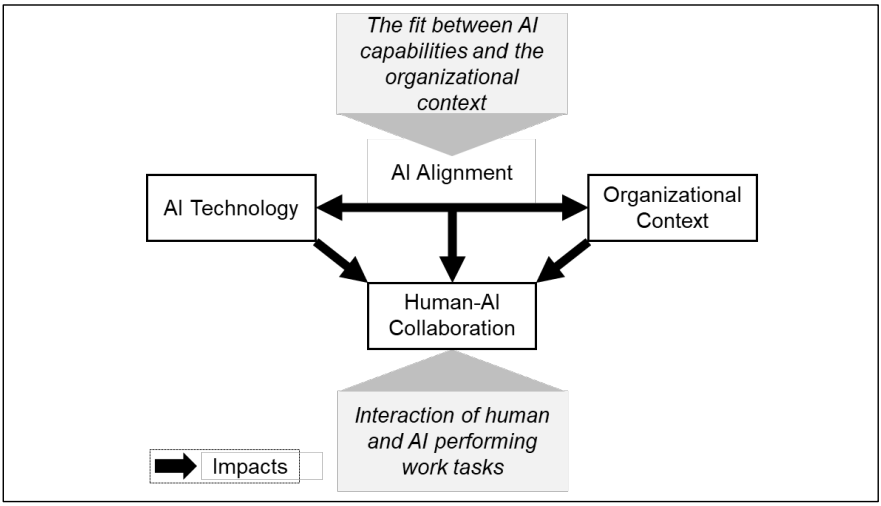

Figure 1. The Sociotechnical Components of AI Alignment

Furthermore, gaining a deeper understanding of AI alignment contributes to a better comprehension of the resulting human-AI collaboration, which is naturally dispersed across various organizational contexts, phenomena, and research streams. For instance, computer science focuses on algorithms and mechanisms, such as explainable AI (e.g., Yin et al., 2019), while psychology examines the cognitive, emotional, and social aspects of collaboration with AI systems (e.g., Yue & Li, 2023). Management research views AI as a strategic asset to be leveraged for organizational success (e.g., Raisch & Krakowski, 2021), whereas the IS domain adopts a more holistic approach aimed at investigating AI as a complex sociotechnical phenomenon (Baird & Maruping, 2021; Seeber, Bittner, et al., 2020). This fragmentation of research efforts highlights the need for a more comprehensive understanding of AI alignment, including its definition and the various aspects it encompasses, leading to human-AI collaboration.

To facilitate a thorough understanding of AI alignment, it is essential to investigate the alignment process within a specific context. The organizational context of clinical healthcare serves as a case study for this thesis. Clinical healthcare refers to the organizational structures, processes, and systems that enable or facilitate the delivery of medical care to patients, encompassing clinical decision-making, patient management, and administrative operations. This includes care from specialists at secondary and tertiary levels, where patients with specific conditions are referred to by primary care

providers to access highly specialized services available mainly in regional centers such as hospitals (Akbari et al., 2008; McWhinney & Freeman, 2009).

First, clinical healthcare is a unique organizational context due to the heterogeneous constraints that it imposes on AI. The general healthcare context covers issues ranging from “health promotion to disease prevention, treatment, rehabilitation, palliative care and more” (World Health Organization, 2019a, p. 1), and hence, is of utmost importance for society. Given the high-stakes nature of the healthcare context, where errors can have devastating and even fatal consequences (Holzinger et al., 2008), integrating the technology of AI is a substantial challenge for the organizational context. This necessitates the development of specialized methods to ensure alignment with the requirements of the healthcare domain (Rajpurkar et al., 2022).

Second, the information systems of clinical healthcare organizations are highly regulated. For instance, software needs to be certified if it is used for medical purposes (European Commission, 2012) and other software classes, such as smartphone apps for patients need to undergo special certification processes (Federal Institute for Drugs and Medical Devices, 2020).

Third, systems such as Health Information Technology (HIT) systems are highly centralized (Colicchio et al., 2019) and lack interfaces for interoperability (Kohli & Tan, 2016), which renders AI implementation in these proprietary systems particularly challenging. Additionally, the clinical healthcare sector is facing unprecedented challenges, including a shortage of healthcare professionals exacerbated by an aging population, which is contributing to a higher prevalence of chronic diseases (Klecun, 2016). AI is viewed as a potential solution to address this challenge and help alleviate the workload of healthcare professionals. For instance, AI can aid in streamlining medical documentation – a task, for which healthcare professionals spend approximately two hours for every hour of patient contact (Arndt et al., 2017). Consequently, the organizational context of clinical healthcare can not only be seen as being strongly regulated but also AI can and should be considered an integral part of its future.

AI has already had significant success in some primarily visual-based tasks in the clinical healthcare sector, such as diagnosing disease conditions through the analysis of radiographic imagery and lesion identification (e.g., Ardila et al., 2019) however, its potential extends far beyond. In particular, implementing AI in the domain of Electronic Health Records (EHRs) warrants exploration. The introduction of EHRs is seen as one of the core pillars of digitalization in healthcare that can address the upcoming challenges the sector will be facing (Berwick, 2002; Kohli & Tan, 2016; Mihailescu et al., 2017). By increasing information quantity and quality, EHR adoption aims to improve patient care. Reduced documentation times, higher information quality through implemented documentation policies, and a reduction in medication errors are some advantages that

are commonly mentioned (Campanella et al., 2016; Domaney et al., 2018). Despite this, many of the promised advantages have not yet materialized (Colicchio et al., 2019). Conversely, EHR adoption in healthcare introduced several unintended consequences for clinicians (Gephart et al., 2015), particularly related to increased documentation times for clinicians resulting from the additional required information and prolonged workflows. Hence, several authors pledge for innovative solutions to restructure EHRs to ultimately support clinicians in clinical documentation processes (Colicchio et al., 2019; Lin et al., 2018).

In this context, advancements in NLP offer a promising opportunity to leverage the vast, often unstructured textual data found in medical records. It is estimated that around 70% of medical information is stored in a free-text format, which is not directly processable by HIT systems (Ford et al., 2016). By enabling machines to process and generate human language in a meaningful way, NLP unlocks a wide range of potential benefits. For example, it can automate the extraction of relevant information from patient records (Y. Wang et al., 2018), and thereby significantly increasing efficiency and reducing manual effort in documentation processes (Biswas & Talukdar, 2024). Here, NLP has the potential to reduce the time spent by healthcare professionals on documentation tasks. However, while these potentials have been demonstrated in research contexts, they often fail to translate into everyday clinical practice due to various sociotechnical challenges, particularly regarding implementation, accountability, and fairness (Panch et al., 2019; Rajpurkar et al., 2022).

This thesis studies the phenomenon of AI alignment from an IS perspective. The research stream of AI in organizational contexts (e.g., Benbya et al., 2020, 2021; Berente et al., 2021) and the lens of IS alignment (e.g., Benbya & McKelvey, 2006; Gerow et al., 2014) provide a relevant framework for understanding the complexities of this phenomenon. However, as Benbya et al. (2020) note, the increasing complexity and change of sociotechnical systems over time reduce the durability of knowledge claims, making it necessary to revisit previous research. Given the continuous evolution of AI, existing knowledge on IS alignment may not be directly applicable to contemporary AI systems.

To explore this challenge further, this thesis specifically focuses on aligning NLP with the organizational context of clinical healthcare. To ensure practical relevance, parts of the research have been conducted within the context of the research project AuMEDa (Automated information and feature extraction from MEDical documents), which is positioned within the context of clinical healthcare and has implemented a functional NLP-based system into clinical healthcare. The project's goal of aligning an NLP system with medical documentation processes serves as a case study, providing valuable insights into the challenges and opportunities of integrating AI into complex organizational systems at different levels. By connecting the research discourse and practical value, this

thesis aims to contribute to a better understanding of how AI can be successfully aligned with organizational contexts.

I.2 Research Gap and Resulting Research Questions

This thesis aims to investigate the phenomenon of AI alignment by identifying the sociotechnical components that shape it, with particular emphasis on the resulting human-AI collaboration. As AI has evolved beyond the automation of routine tasks, it has become clear that more complex tasks, such as critical decision-making and other intricate workflows, often require collaboration between humans and AI (Benbya et al., 2021; Dellermann, Ebel, et al., 2019). To address this evolving landscape, this thesis is organized around three overarching research questions. These questions explore the phenomenon of AI alignment and the resulting dynamics of human-AI collaboration. In the following sections, each research question is derived from identifying a relevant gap.

AI is continuously advancing in terms of its performance and scope (Berente et al., 2021), with large language models being one of the latest developments that enable generative AI (GenAI). GenAI is characterized by its ability to produce new information, such as texts, images, or videos (Benbya et al., 2024). In this thesis, if not explicitly defined, GenAI will be regarded as a subset of AI. The rapid evolution of AI is transforming the nature of work and challenging traditional organizational structures, presenting both opportunities and challenges that result from the integration of AI into organizations (Eloundou et al., 2023).

With these increasing capabilities of AI, organizations are moving towards more dynamic collaborations between humans and AI (Dennis et al., 2023). Effectively aligning AI within these collaborations is crucial, as its unique characteristics, autonomy, learning, and inscrutability (i.e., incomprehensible AI behavior for humans) need to be addressed to realize the benefits of AI (Berente et al., 2021). Proper alignment not only facilitates routine interactions but also enhances complex decision-making and supports workflows that require human insight and oversight (Benbya et al., 2021; Dellermann, Ebel, et al., 2019). Thus, an exploration of how collaborative processes evolve with AI and how alignment mechanisms can address the associated challenges is essential. This context informs the subsequent research question, focusing on understanding the impact of AI technology, the organizational context, and its resulting human-AI collaboration:

> ***RQ1:*** *What is the status quo of human-AI collaboration in organizational contexts?*

To gain a deeper understanding of AI alignment, this thesis investigates this phenomenon in the organizational context of clinical healthcare, which offers a unique combination of societal importance and potential for AI. The clinical healthcare sector encompasses various disciplines, including radiology, pathology, and gastroenterology, and involves a

range of tasks that could be supported by AI, such as patient anamnesis, disease detection, and recommending and conducting therapies (Panch et al., 2019; Rajpurkar et al., 2022).

Notably, AI has already shown promise in various clinical settings (Rajpurkar et al., 2022; Topol, 2019). AI applications for visual tasks have been steadily progressing over time, with techniques such as whole-slide imaging and the development of deep learning algorithms fueling this advancement (Rajpurkar et al., 2022). For instance, AI-based image processing has significantly enhanced the visual detection of diseases in radiology, as evidenced by several works (e.g., Sorantin et al., 2022; Tadavarthi et al., 2020).

Yet, the potential for AI, particularly NLP, in non-image work processes has been neglected for a long time. However, it has recently gained momentum in text processing through technological advances such as the transformer architecture and foundation models (Bommasani et al., 2021). For example, NLP has shown potential in improving administrative task performance in clinical documentation processes (Chen et al., 2019; Pons et al., 2016; Y. Wang et al., 2018) and in risk prediction (Sterckx et al., 2020). Despite these advancements, research indicates that NLP's integration into clinical practice remains a significant challenge due to poor alignment, manifested in the design of EHR interfaces and the mismatch between free-text formats and structured data (Panch et al., 2019; Stone et al., 2016). This gap between practical implementation and research leads to the following research question:

> ***RQ2:*** *How can NLP systems be aligned with clinical healthcare?*

Effective integration of NLP systems into clinical practice requires a comprehensive understanding of their technical and functional characteristics. The complex ecosystem of everyday clinical practice presents challenges related to implementation, fairness, and accountability. These challenges are the result of the diversity and complexity of patient data, critical decision-making processes, and ethical considerations such as privacy and bias (Panch et al., 2019; Rajpurkar et al., 2022). In addition, NLP systems must be aligned with the organizational constraints of clinical healthcare. This leads to the following research question:

> ***RQ 2.1:*** *Which key system characteristics and organizational constraints must be considered when integrating NLP into the clinical healthcare context?*

To ensure the successful adoption and utilization of NLP systems, the specific requirements of healthcare professionals and the underlying tasks need to be considered on an individual level (Velupillai et al., 2018; Wen et al., 2019). These systems must emphasize transparency, explainability, and trustworthiness, as medical staff need to be able to understand the recommendations and insights generated by the NLP system (Lebovitz et al., 2022; Panch et al., 2019).

However, a gap between practical utility and demonstrated potential in research can be observed (Panch et al., 2019). On the one hand, retrospective studies on the digitalization of healthcare report that the design of EHR systems lacks usability and even increases documentation times for physicians (Domaney et al., 2018; Stone et al., 2016). On the other hand, research on NLP can potentially alleviate physicians' workload by extracting relevant information (Chen et al., 2019; Y. Wang et al., 2018), coding and structuring treatments (Bossen & Pine, 2023), or summarizing relevant biomedical literature (Q. Yang et al., 2023). However, such applications are often not transferred into practical settings. This leads to the following research question:

> ***RQ 2.2:*** *How should NLP systems be aligned with the special requirements of healthcare professionals?*

In work contexts, enhancing performance, which depends on the individual task (Maedche et al., 2019), is an important promise of human-AI collaboration because humans and AI can potentially utilize their complementary strengths to exceed the performance of either AI or humans alone (Dellermann, Ebel, et al., 2019). However, previous research also suggests that the integration of AI is not necessarily beneficial and can lead to unintended consequences, such as decreasing performance within the workforce (Benbya et al., 2021; Fügener et al., 2021b). Much of this research has focused on measuring human-AI collaboration performance by investigating single components of alignment (Bauer et al., 2023; Berente et al., 2021) often overlooking the contextual components that inevitably influence the findings. Hence, a holistic perspective is needed to understand the impact of AI alignment and its resulting contextualization on human-AI collaboration performance. This results in the following research question:

> ***RQ 3:*** *How does AI alignment affect human-AI collaboration performance in organizational contexts?*

Figure 2 illustrates the research framework, outlining the three research questions and their interrelations with each other.

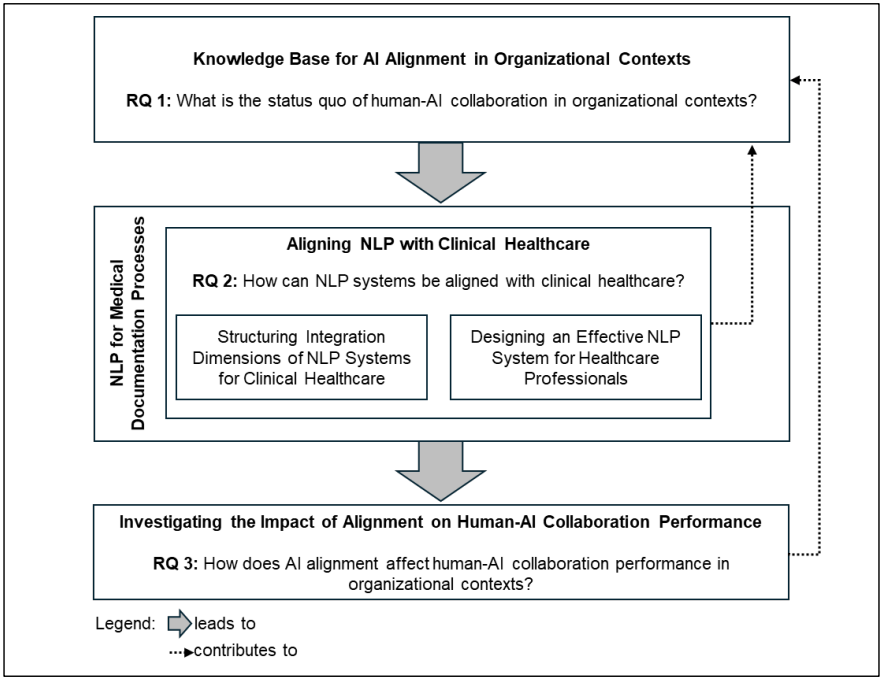

Figure 2. Research Overview

I.3 Structure of this Thesis

This thesis is structured cumulatively and presents four studies conducted on aligning AI with a deep dive into the healthcare context from the perspective of IS research. It is divided into three parts (see Figure 3).

The first part (A) of this thesis provides a foundation for the following studies by introducing the research endeavor (A.I) and outlining the relevant research background (A.II). The introduction section starts with the motivation of the research endeavor (A.I.1), identifies a research gap and resulting research questions (A.I.2), and provides an overview of the structure of this thesis (A.I.3). Afterwards, the conducted research is located within IS research (A.I.4) and the anticipated contribution is outlined (A.I.5). The research background of this thesis is described in section A.II. First, the foundations of AI alignment are presented in section A.II.1. Next, the practical context of NLP in clinical healthcare and its associated challenges are discussed, framing these issues as an alignment problem (A.II.2). Finally, the research project AuMEDa is described as a case study for this thesis (A.II.3).

The second part (B) comprises the four main studies of this thesis (see Table 1) that address the research questions formulated in A.I.2 and provide the basis for the third part (C).

First, the findings of the studies are utilized to answer the research questions (C.I.1-3) and synthesized in a holistic framework addressing the role of aligning AI with an organizational context (C.I.4). Next, based on the synthesized framework, the limitations of this work and future research directions are discussed. Additionally, the implications for both research and practice are considered (C.II). Finally, a conclusion summarizing the main findings of this thesis is presented (C.III).

Table 1. Main Studies of this Thesis

No.	Outlet	Status	Ranking*	Section	RQ	Contribution
1	In preparation for submission to a journal. A previous version was published in the Proceedings of the 44th International Conference on Information Systems (2023)	Preparation/ Published	A	B.I	1	Synthesis of current human-AI collaboration in organizational contexts. Identification of three distinct roles of AI within human-AI teams.
2	Proceedings of the 44th International Conference on Information Systems (2023)	Published	A	B.II	2(.1)	Identification and conceptualization of integration dimensions of NLP-based systems for clinical healthcare organizations.
3	Proceedings of the 17th International Conference on Design Science Research in Information Systems and Technology (2022)	Published	C	B.III	2(.2)	Design and validation of an NLP-based system for healthcare professionals.
4	Journal of Decision Systems (2023) A previous version was published in the Proceedings of the 30th European Conference on Information Systems (2022)	Published	B	B.IV	3	Understanding the impact of alignment, i.e., human-AI collaboration designs, on performance.

*JOURQUAL ranking at the time of publication (VHB-JOURQUAL 3)