Data Science & Artificial Intelligence

The Effect of GenAI-Tools on Knowledge Management, Workload and Productivity in the Salvation Army

Thijmen Bouman

First supervisor: Tyron Offerman
Second supervisor: Natalia Amat Lefort

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

09/07/2026

Abstract

Background & Related Work: Generative Artificial Intelligence (GenAI) tools are increasingly adopted in organizations for knowledge-intensive tasks such as documentation, summarization, and information retrieval. While existing research demonstrates productivity improvements in experimental and commercial settings, limited work examines the simultaneous impact of GenAI on knowledge management, workload, and productivity in real-world non-profit contexts.

Aim: This study investigates how GenAI transcription and summarization tools influence perceived knowledge management, workload, and productivity within the Dutch Salvation Army, a non-profit organization where administrative burden significantly reduces time available for client care.

Method: A single case study methodology was followed, combining semi-structured interviews with developers, a quantitative end-user survey ($n = 51$) measuring six constructs (Efficiency, Quality, Workload, Cognitive Load, Trust, Usability), and objective behavioral metrics extracted from Azure Application Insights log data using Kusto Query Language queries.

Results: Survey respondents reported positive perceptions across all constructs, with the highest scores for Usability ($M = 4.02$) and Efficiency ($M = 3.94$). Reliability analysis showed acceptable internal consistency for Efficiency ($\alpha = 0.78$), Quality ($\alpha = 0.76$), and Workload ($\alpha = 0.86$). Frequent users reported significantly higher perceived Efficiency than occasional users ($p = 0.014$). Log data analysis revealed that the user base grew from 125 to a peak of 465 weekly active users between April and July, with an AI-generated summary edit rate of only 1.49% and 27.1% of recordings transitioning to chat-based follow-up.

Conclusion: GenAI tools can effectively reduce administrative burden and improve documentation quality in non-profit care environments, without replacing professional judgement. The convergence of positive survey perceptions and objective behavioral metrics provides robust evidence that LuisterLinie supports knowledge management and productivity in a mission-driven organizational setting. Human oversight of AI-generated output remains essential.

Acknowledgements

I would like to thank Tyron Offerman, Niek Schreuder and Manuel Mol for their extensive assistance and supervision across my research within the Dutch Salvation Army, they have shown me many new ways of dealing with a corporate environment and have helped me improve across a multitude of different areas that will certainly help me in the future. Tyron Offerman has provided me with tremendous help on how to write scientific papers, conduct action research, communicate my own ideas and incentives, communicate ideas with others, analyze different data sources, and present my findings in an academic way. Niek Schreuder has been my main contact point on how to get things done within a larger corporate organization such as the Dutch Salvation Army, while never losing the main goal of helping others out of sight. He provided me with authorization, convenient introductions, and was always there if I needed him. Manuel Mol has shown me what it is like to work within a high-level programming team, using the most modern ways of working together and the future of software development within organizations. I would also like to thank the rest of the AI Innovation team, for welcoming me into their team, and allowing me to take their time and really go in depth on the research that I wanted to do. Furthermore, I would also like to thank the Dutch Salvation Army for having me as an intern to perform this research, and offering me a long-term position within the AI Innovation team.

Contents

Chapter 1

Introduction

Generative Artificial Intelligence (GenAI) is rapidly transforming the modern workplace, offering new capabilities for automating knowledge-intensive tasks such as documentation, summarization, and information retrieval. Since the release of large language models like ChatGPT, organizations across sectors have begun experimenting with GenAI tools to improve efficiency and streamline administrative processes [18, 10]. However, the broader organizational implications of implementing these systems remain insufficiently understood, particularly in non-profit contexts where productivity is measured not in profit margins but in the quality and quantity of care provided to society.

The Dutch Salvation Army (Leger des Heils) is one such organization. As a non-profit providing social care and support, its employees, caretakers and social workers, spend a substantial portion of their time on administrative tasks such as documenting client meetings, writing reports, and processing records. Internal assessments indicate that administrative work consumes approximately 30 to 40 percent of employee time, reducing the hours available for direct client care. To address this burden, the organization has developed two GenAI tools: LuisterLinie, an automated transcription and summarization application for client conversations, and ChatLDH, a conversational AI assistant for internal knowledge retrieval.

These tools promise to reduce administrative workload, improve the quality and consistency of documentation, and allow caretakers to focus more on their clients. Yet the actual impact of such tools on employee productivity, knowledge management, and perceived workload in a real-world non-profit environment has not been systematically evaluated. Existing research on GenAI in organizations has largely focused on commercial settings or controlled experimental studies [10, 13], leaving a gap in understanding how these technologies function within mission-driven, resource-constrained contexts.

1.1 Problem statement

The implementation of GenAI tools in organizational settings presents both opportunities and challenges. On one hand, automated transcription and summarization can reduce repetitive administrative work and improve the accessibility of organizational knowledge [14]. On the other hand, interacting with AI systems may introduce new forms of cognitive workload, as employees must verify and refine AI-generated output [19]. Furthermore, factors such as privacy concerns, organizational resistance to change, and varying levels of user proficiency can influence how these tools are adopted and experienced.

Within the Dutch Salvation Army, the development of LuisterLinie and ChatLDH has progressed from pilot phases to broader organizational rollout. However, no systematic research has been conducted to evaluate whether these tools achieve their intended effects on productivity, knowledge management, and workload, or to identify the factors that influence their successful adoption. Understanding these relationships is essential not only for the organization itself, but also for the broader field of GenAI implementation in non-profit and care-oriented settings.

1.2 Research questions

This study aims to investigate the impact of GenAI tools on knowledge management, workload, and productivity within the Dutch Salvation Army. The following research questions guide the investigation:

RQ1 How do GenAI tools influence the perceived productivity and efficiency of administrative work within the Dutch Salvation Army?

RQ2 What is the perceived impact of GenAI tools on knowledge management and documentation quality?

RQ3 How do GenAI tools affect perceived workload and cognitive load among employees?

RQ4 How do objective behavioral metrics from application log data relate to subjective user perceptions of GenAI tool usage?

1.3 Overview of the thesis

Chapter 2 reviews the existing literature on Generative AI in organizational contexts, knowledge management, and the relationship between productivity and workload. Chapter 3 describes the case study

methodology used, including the diagnosing interviews, survey development, and extraction of application log data. Chapter 4 presents the findings from the exploratory interviews. Chapter 5 reports the quantitative survey results and the behavioral metrics derived from application usage data. Chapter 6 interprets these findings in the context of existing literature and discusses the convergence of subjective and objective measurements. Finally, Chapter 7 summarizes the contributions of this study and outlines directions for future research.

Chapter 2

Background and Related Work

2.1 Generative AI in Organizations

Artificial Intelligence (AI) has become one of the most influential technological developments of the past decade, fundamentally changing how organizations process information, make decisions, and perform knowledge-intensive work. Although AI has existed as a research field since the 1950s, recent advances in computational power, data availability, and machine learning algorithms have enabled AI systems to perform increasingly complex cognitive tasks that were previously considered exclusive to human intelligence [21]. Today, AI technologies are widely applied across industries to automate repetitive processes, support decision-making, recognize patterns in large datasets, and augment human expertise.

Generative Artificial Intelligence (GenAI) represents a relatively new branch within the broader AI field. Unlike traditional AI systems that primarily classify, predict, or optimize existing information, GenAI is designed to generate new content such as text, images, computer code, or audio based on learned statistical patterns from training data [3]. The rapid public adoption of systems such as ChatGPT has significantly accelerated organizational interest in applying GenAI to support everyday knowledge work. Rather than replacing existing information systems, these technologies provide employees with interactive assistants capable of generating summaries, drafting reports, answering questions, and assisting in administrative tasks.

2.1.1 Large Language Models and Their Organizational Applications

The recent success of GenAI has largely been enabled by the development of Large Language Models (LLMs). LLMs are transformer-based neural networks trained on extremely large collections of textual data, enabling them to learn statistical relationships between words, sentences, and broader linguistic concepts [4]. Through this training process, LLMs are capable of understanding natural language prompts and generating coherent responses that resemble human-written text. Beyond simple text generation, these models are capable of summarization, translation, information extraction, question answering, and conversational interaction, making them particularly suitable for supporting knowledge-intensive organizational processes.

Within organizations, the application of GenAI extends beyond simple automation. Earlier generations of AI systems often relied on structured rule-based knowledge systems or expert systems that required manually encoded knowledge [17]. Contemporary GenAI systems instead enable employees to interact with organizational knowledge through natural language, allowing users to retrieve information, summarize documents, generate reports, and synthesize information from multiple sources with considerably less effort. Consequently, organizations increasingly position GenAI not only as an automation technology but also as a collaborative tool that augments human knowledge work rather than replacing it [9].

2.1.2 Empirical Evidence of Productivity Improvements

Recent empirical research provides growing evidence that GenAI can improve employee performance across a variety of knowledge-intensive tasks. In one of the largest controlled field experiments to date, Noy and Zhang [16] demonstrated that knowledge workers using ChatGPT completed professional writing tasks approximately 40% faster while simultaneously producing higher-quality output than participants without AI assistance. Similar findings were reported by Brynjolfsson et al. [5], who investigated the implementation of a conversational AI assistant in a customer support environment. Their study found significant improvements in employee productivity, particularly among less experienced workers, suggesting that GenAI may help reduce differences in expertise by providing immediate access to relevant knowledge and recommendations.

The productivity gains associated with GenAI appear to extend beyond writing tasks. Experimental studies have shown that employees using AI support are often able to complete analytical tasks more efficiently while maintaining or improving output quality [20]. Rather than performing tasks independently, GenAI appears to function most effectively as a collaborative partner in which human judgement and AI-generated suggestions complement one another. This concept of human–AI collaboration has received increasing attention within organizational research, where AI is viewed as a technology that augments human capabilities instead of replacing professional expertise [12].

2.1.3 Organizational Challenges in GenAI Implementation

Despite these promising developments, the implementation of GenAI also introduces several organizational challenges. One frequently discussed limitation is the occurrence of so-called *hallucinations*, in which language models generate plausible but factually incorrect information [3]. Because these outputs are often presented confidently, users may incorrectly assume that generated information is accurate. In professional environments where documentation quality, accountability, and decision-making are critical, incorrect AI-generated information can have significant consequences. Consequently, effective human verification remains an essential component of responsible GenAI adoption.

In addition to concerns regarding output quality, organizations must address issues related to privacy, governance, transparency, and responsible AI use. Many organizational applications process confidential or sensitive information, requiring compliance with legal frameworks such as the General Data Protection Regulation (GDPR). Furthermore, successful implementation depends not only on technical performance but also on employee acceptance, organizational support, and adequate training. Research indicates that users who understand both the capabilities and limitations of GenAI are generally more successful in integrating these systems into their daily work processes [9].

Taken together, the existing literature demonstrates that Generative Artificial Intelligence has considerable potential to improve organizational performance by supporting knowledge-intensive work, reducing repetitive administrative activities, and increasing employee productivity. However, realizing these benefits depends on effective integration into organizational workflows and continued human oversight. One organizational domain in which these characteristics become particularly relevant is knowledge management, where GenAI technologies are increasingly applied to capture, organize, retrieve, and generate organizational knowledge. The following section therefore examines the concept of knowledge management and the role of Generative AI within contemporary knowledge management practices.

2.2 Knowledge Management and GenAI

Knowledge has long been recognized as one of the most valuable organizational resources. Unlike physical assets, organizational knowledge enables employees to make informed decisions, solve complex problems, and continuously improve organizational performance. As organizations become increasingly knowledge-intensive, effectively managing knowledge has become essential for maintaining operational efficiency and service quality. Knowledge Management (KM) therefore focuses on the systematic creation, storage, sharing, and application of organizational knowledge to achieve strategic objectives [7].

One of the most widely accepted definitions of Knowledge Management is provided by Alavi and Leidner [1], who describe KM as a process through which organizations create, store, transfer, and apply knowledge to improve organizational effectiveness. Rather than considering knowledge as static information stored in documents or databases, KM views knowledge as a dynamic organizational asset that is continuously created and refined through interactions between employees, organizational processes, and technological systems. Effective knowledge management therefore depends not only on information technology but also on organizational culture, collaboration, and individual expertise.

2.2.1 The Tacit–Explicit Knowledge Distinction

A fundamental distinction within the Knowledge Management literature is the difference between tacit and explicit knowledge, originally introduced by Nonaka and Takeuchi [15]. Tacit knowledge refers to personal experience, intuition, professional judgement, and practical skills that are difficult to formalize or communicate. In healthcare and social care environments, tacit knowledge often includes interpersonal communication, professional observations, and contextual understanding developed through experience with clients. Explicit knowledge, by contrast, consists of information that has been documented and can be stored, transferred, and reused in the form of reports, procedures, guidelines, or databases.

According to the knowledge creation model proposed by Nonaka and Takeuchi [15], organizational learning occurs through continuous interaction between tacit and explicit knowledge. Their SECI model describes four recurring processes: socialization, externalization, combination, and internalization. During externalization, tacit knowledge is transformed into explicit documentation, allowing valuable professional expertise to become available to the wider organization. In care organizations, documentation of client interactions represents an important example of this process, as conversations and professional

observations are converted into structured reports that support continuity of care and organizational learning.

2.2.2 Core Processes in Knowledge Management

Building upon this perspective, Alavi and Leidner [1] identify four core Knowledge Management processes that together form the foundation of effective organizational knowledge management: knowledge creation, knowledge storage and retrieval, knowledge transfer, and knowledge application.

Knowledge creation refers to the generation of new organizational insights through experience, collaboration, or problem solving. Knowledge storage ensures that organizational knowledge remains available over time through documentation systems, databases, or digital repositories. Knowledge transfer focuses on sharing knowledge between employees and departments, while knowledge application concerns the effective use of available knowledge during daily work processes and decision-making. Information technologies have traditionally supported these processes through document management systems, knowledge bases, and enterprise information systems.

2.2.3 The Role of Generative AI in Knowledge Management

Recent developments in Generative Artificial Intelligence have significantly expanded the technological possibilities for supporting Knowledge Management. Unlike conventional information systems that primarily store and retrieve existing information, GenAI systems are capable of generating summaries, extracting key information, synthesizing knowledge from multiple sources, and responding to natural language questions. As a result, GenAI can actively support multiple stages of the Knowledge Management cycle rather than functioning solely as a passive repository of organizational information.

Several studies have identified document summarization as one of the most promising applications of Large Language Models within Knowledge Management. By automatically extracting the essential information from lengthy conversations or documents, GenAI reduces the effort required to produce structured documentation while simultaneously improving consistency across reports [6]. Similarly, conversational AI systems enable employees to retrieve organizational knowledge using natural language queries rather than manually searching through large collections of documents. This lowers the barrier to accessing existing organizational knowledge and may improve knowledge sharing across teams.

Another important contribution of GenAI concerns organizational memory. Organizational memory refers to the accumulated knowledge that remains available to an organization after individual employees leave or change roles [22]. High-quality documentation is therefore essential to ensure continuity of care, preserve institutional knowledge, and support future decision-making. Because GenAI systems can automatically generate structured reports from recorded conversations, they may contribute to more complete and standardized documentation, thereby strengthening the organization's collective memory.

Speech-to-text technology further expands the role of GenAI within Knowledge Management. Instead of requiring professionals to manually document conversations after client interactions, AI-based transcription systems can automatically convert spoken conversations into textual records. These records may subsequently be summarized, structured, and integrated into organizational information systems. This not only reduces administrative effort but also decreases the risk of information loss that may occur when documentation is delayed or based solely on memory. Previous research has shown that automated transcription technologies can improve documentation completeness while allowing professionals to devote greater attention to client interactions during conversations [24].

Despite these opportunities, the integration of GenAI into Knowledge Management also introduces important challenges. Automatically generated documentation remains dependent on the quality of the underlying language model and the completeness of recorded conversations. Incorrect summaries, omitted information, or hallucinated content may negatively affect documentation quality if generated output is accepted without sufficient human verification. Consequently, recent literature increasingly emphasizes the importance of maintaining human oversight when integrating Generative AI into organizational knowledge processes [9].

2.2.4 Knowledge Management at the Leger des Heils

Within the context of the Leger des Heils, Knowledge Management primarily concerns the documentation of client conversations and the preservation of accurate administrative records. Professionals routinely conduct conversations that contain valuable contextual information, observations, and agreements that must subsequently be documented within organizational information systems. Traditionally, this documentation process is both time-consuming and cognitively demanding, requiring employees to reconstruct conversations after they have taken place.

The LuisterLinie application addresses this challenge by automatically transcribing recorded conversations and generating structured summaries that can be used as a basis for administrative documentation. Rather than replacing professional judgement, the system supports the externalization of tacit knowledge by transforming spoken conversations into explicit documentation. In doing so, the application contributes to several Knowledge Management processes simultaneously, including knowledge creation through structured summaries, knowledge storage through standardized documentation, knowledge retrieval through searchable reports, and knowledge application by making client information more readily available during future care processes.

Consequently, Knowledge Management forms one of the central theoretical constructs of this research. The extent to which users perceive improvements in documentation quality, report completeness, and information accessibility provides an important indicator of the organizational value created by Generative AI within the Leger des Heils.

2.3 Productivity and Workload

One of the primary motivations for adopting Generative Artificial Intelligence in organizations is the expectation that it will improve employee productivity. Productivity is commonly defined as the relationship between outputs produced and the resources required to produce them, such as time, effort, or labor [8]. Within knowledge-intensive organizations, productivity extends beyond simply producing more work in less time. It also includes improvements in work quality, decision-making, information processing, and the efficient allocation of employees' cognitive resources.

2.3.1 Defining Productivity in Knowledge-Intensive Work

Measuring productivity in knowledge work presents unique challenges because many professional tasks involve problem solving, communication, and decision-making rather than easily quantifiable production outputs. As noted by Drucker [8], the productivity of knowledge workers depends not only on efficiency but also on the quality of their work, their ability to focus on value-adding activities, and the effectiveness with which knowledge is applied. Consequently, technologies that reduce administrative burden may increase productivity even when total working hours remain unchanged.

Recent research has demonstrated that Generative AI can substantially improve productivity across a wide range of professional tasks. In a controlled experiment, Noy and Zhang [16] found that knowledge workers using ChatGPT completed writing tasks approximately 40% faster while simultaneously producing higher-quality outputs. Similarly, Brynjolfsson et al. [5] observed productivity improvements following the introduction of an AI assistant in customer support environments, particularly among less experienced employees. These findings suggest that Generative AI can act as a performance-enhancing technology by reducing the time required to complete repetitive cognitive tasks while improving access to relevant knowledge.

Within administrative professions, one of the largest opportunities for productivity improvement lies in reducing documentation effort. Professionals in healthcare and social care frequently spend considerable time documenting client interactions, writing reports, and completing administrative records. AI-assisted transcription and summarization systems can automate substantial portions of these activities, allowing professionals to devote more attention to direct client interactions. Consequently, productivity improvements should not solely be interpreted as producing more documentation, but rather as enabling professionals to spend more time on activities that directly contribute to organizational objectives.

2.3.2 Cognitive Workload and the NASA-TLX Framework

Although productivity represents an important organizational outcome, it should not be evaluated independently from employee workload. Workload refers to the amount of physical and mental effort required to complete work-related tasks. While automation technologies often reduce physical or administrative workload, they may simultaneously introduce new forms of cognitive effort associated with supervising, interpreting, and verifying automated outputs.

Cognitive workload is generally defined as the mental effort required to process information, make decisions, and maintain attention while performing a task [11]. Unlike physical workload, cognitive workload is often influenced by factors such as task complexity, uncertainty, interruptions, and information processing demands. As work becomes increasingly knowledge-intensive, cognitive workload has become an important indicator of employee well-being and job performance.

One of the most widely used instruments for measuring perceived cognitive workload is the NASA Task Load Index (NASA-TLX), originally developed by Hart and Staveland [11]. The NASA-TLX conceptualizes workload as a multidimensional construct consisting of six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. Although originally developed for aviation and human factors research, the framework has since been widely applied across healthcare, information systems, and human-computer interaction research. The emphasis on perceived mental effort makes the NASA-TLX particularly relevant when evaluating AI-supported work environments.

2.3.3 The Interaction Between Productivity and Cognitive Workload

Generative AI introduces an interesting paradox with respect to workload. On one hand, AI systems reduce the amount of manual effort required to complete documentation and repetitive administrative tasks. On the other hand, users remain responsible for reviewing, correcting, and validating AI-generated output before it can be used in professional practice. Consequently, traditional workload may decrease while cognitive workload remains unchanged or even increases due to the additional responsibility of monitoring AI performance.

This phenomenon has received increasing attention within the literature on Human–AI Collaboration. Rather than replacing human professionals, contemporary AI systems are increasingly viewed as collaborative partners that augment human capabilities [12]. Human–AI collaboration combines the computational capabilities of AI systems with human judgement, contextual understanding, ethical reasoning, and domain expertise. Successful collaboration therefore depends on an appropriate distribution of responsibilities between humans and AI rather than full automation.

Research suggests that the effectiveness of Human–AI Collaboration is influenced by several factors, including system usability, transparency, trust, and the user's understanding of the AI system's capabilities and limitations [2]. Employees who understand when AI-generated output should be trusted and when additional verification is required are generally better able to integrate AI into their daily work processes. Consequently, effective collaboration requires not only technically accurate AI models but also appropriate organizational training and governance.

An important implication of Human–AI Collaboration is that productivity gains should not automatically be interpreted as reductions in overall effort. Instead, the nature of work changes. Manual documentation activities are partially replaced by reviewing AI-generated summaries, evaluating generated recommendations, and making professional decisions regarding the correctness of AI output. As a result, employees increasingly function as supervisors of AI systems rather than solely performing administrative work themselves.

This distinction is particularly relevant within healthcare and social care organizations, where documentation quality directly influences continuity of care and organizational accountability. Incorrect or incomplete AI-generated documentation may have significant consequences for both clients and professionals. Therefore, human verification remains essential despite the considerable efficiency gains offered by Generative AI.

The relationship between productivity and workload is therefore more nuanced than a simple inverse relationship. While increased productivity is often expected to reduce workload, Generative AI may simultaneously redistribute work from manual administrative activities towards cognitive evaluation and

quality assurance. Understanding this balance forms one of the central research questions of the present study.

Within the context of the Leger des Heils, LuisterLinie is intended to reduce the administrative burden associated with documenting client conversations while allowing professionals to devote greater attention to clients themselves. However, users remain responsible for verifying generated summaries before incorporating them into official client records. Consequently, this research examines both perceived productivity improvements and cognitive workload to obtain a comprehensive understanding of the organizational impact of AI-assisted documentation. By evaluating these constructs simultaneously, the study aims to determine whether the efficiency gains achieved through Generative AI outweigh the additional mental effort associated with human oversight.

2.4 Research Gap and Objectives

The existing literature demonstrates that Generative Artificial Intelligence has considerable potential to transform knowledge-intensive work by improving productivity, supporting knowledge management, and reducing repetitive administrative activities. However, several important research gaps remain regarding the implementation and organizational impact of GenAI in real-world settings.

2.4.1 Gaps in the Current Literature

First, much of the current empirical research on Generative AI focuses on controlled experiments, technology development, or commercial organizations. Studies such as Noy and Zhang [16] and Brynjolfsson et al. [5] provide valuable evidence that GenAI can improve task performance and employee productivity. However, these studies primarily examine short-term effects on specific tasks, such as writing or customer support, rather than the broader organizational consequences of integrating AI into existing workflows. Less is known about how GenAI affects organizational processes over time, how employees adapt to AI-supported work, and how these technologies influence knowledge creation and documentation practices.

Second, existing research has predominantly focused on productivity improvements while giving comparatively less attention to the relationship between productivity gains and changes in workload. Although GenAI can reduce repetitive administrative effort, the introduction of AI-supported workflows may also create new cognitive demands related to reviewing, validating, and interpreting generated output. Understanding this balance is particularly important in professional environments where employees remain responsible for the accuracy and ethical implications of documented information. Therefore, research is needed that considers both the benefits of increased efficiency and the potential cognitive costs associated with human oversight.

Third, although Knowledge Management literature has extensively studied the creation, storage, and transfer of organizational knowledge, the role of Generative AI within these processes remains an emerging research area. Existing Knowledge Management systems have traditionally focused on storing and retrieving explicit knowledge, whereas GenAI introduces new possibilities for transforming tacit knowledge into accessible organizational information. In particular, the use of AI-assisted transcription and summarization to convert conversations into structured documentation remains underexplored. This is especially relevant in organizations where valuable knowledge is generated through human interactions rather than traditional written documentation.

A further limitation in existing research is the limited attention given to non-profit organizations and public-service environments. Much of the current evidence regarding AI adoption originates from commercial organizations where productivity, efficiency, and financial performance are dominant objectives. However, non-profit organizations operate under different conditions, where social value, professional responsibility, ethical considerations, and service quality are equally important outcomes. In care-oriented environments, the objective of AI implementation is not necessarily to replace employees or maximize output, but rather to support professionals by reducing administrative burden and enabling more attention to clients.

The Leger des Heils provides an example of such a context. Professionals within the organization perform knowledge-intensive activities involving client conversations, documentation, and administrative reporting. These activities require both efficiency and careful preservation of contextual information.

Therefore, evaluating GenAI within this environment requires considering multiple dimensions simultaneously, including productivity, knowledge management, workload, cognitive effort, usability, and trust.

2.4.2 Methodological Rationale for Single Case Study

Due to the complexity of implementing GenAI within organizational practice, a single case study methodology is particularly suitable for addressing these research gaps. Unlike traditional experimental approaches that examine isolated effects, a case study allows for deeper investigation within the Dutch Salvation Army. This approach allows researchers and organizational stakeholders to collaboratively examine how AI technology is introduced, how employees experience its use, and how implementation strategies can be improved based on empirical findings.

Within this research, a single case study enables the combination of multiple data sources, including developer interviews, end-user surveys, application usage data, and AI interaction logs. This methodological combination provides a more comprehensive understanding of both the technical performance and organizational consequences of GenAI adoption. By combining quantitative measures of user perceptions with qualitative insights and behavioral usage data, this study aims to move beyond evaluating whether GenAI works and instead investigate how it can be effectively integrated into organizational practice.

2.4.3 Research Questions and Contribution of This Study

Based on the identified research gaps, this thesis addresses the following main research question:

> *To what extent can Generative Artificial Intelligence tools improve knowledge management, productivity, and workload management within a non-profit organizational context?*

To answer this question, the research investigates how employees perceive the influence of GenAI tools on productivity and administrative workload, and how these tools contribute to knowledge management processes, particularly the creation, structure, and retrieval of organizational documentation. Furthermore, the study examines what cognitive workload challenges arise when professionals use GenAI-assisted documentation systems and which factors influence user acceptance, trust, and effective adoption of GenAI tools within the organization. Finally, the research considers how a single case study approach can support the successful implementation and continuous improvement of GenAI applications in a non-profit organizational setting.

By addressing these questions, this research contributes to the growing body of literature on Generative AI in organizations by providing empirical insights into AI adoption within a non-profit care environment. Furthermore, it contributes to practice by identifying how GenAI systems can be implemented in a way that enhances administrative processes while preserving human expertise, professional judgement, and organizational values.

Chapter 3

Methodology

This study applies a single case study methodology to investigate how GenAI tools influence knowledge management, productivity and workload within the organizational context of the Dutch Salvation Army. A single case study is suitable for this research as it allows for an in-depth, contextualized analysis investigation of GenAI application within a real-world context [23]. Unlike experimental methods that isolate variables from their context, the case study approach enables the researcher to examine how GenAI tools are actually developed, deployed, and experienced within a specific organizational environment. This is especially important when studying a non-profit organization such as the Dutch Salvation Army, where the interplay between technological innovation, organizational structure, and mission-driven objectives creates a unique context that cannot be replicated in a laboratory setting.

The research is carried out within the AI Innovation team of the Dutch Salvation Army, where current caretakers summarize and document client meetings as part of their administrative responsibilities. These knowledge-intensive tasks require significant time and effort that could otherwise be spent on direct client care. The introduction of GenAI tools aims to address these challenges and enables the study of both efficiency gains and broader transformations in how knowledge is created, processed, and stored within the organization.

3.1 Case Study Design

The case study follows a structured approach consisting of four phases: exploratory interviews with developers to understand the organizational and technical context, survey development based on insights from both the interviews and the existing literature, data collection through survey distribution and application log extraction, and evaluation through statistical analysis and methodological triangulation. This structure allows for a comprehensive investigation that combines qualitative insights from developers with quantitative data from end-users and objective behavioral metrics from the applications themselves.

3.2 Exploratory Interviews

The first phase of the case study consists of exploratory semi-structured interviews with developers at the Dutch Salvation Army who are responsible for building and maintaining the GenAI tools. The purpose of these interviews is to develop a comprehensive understanding of the current situation within the organization, including the challenges that led to the development of the tools, the design intentions behind them, and the organizational context in which they are deployed. Key topics explored during these interviews include the current use of GenAI tools within the organization, the intent behind their development, existing documentation processes, perceived benefits and limitations from the organizational perspective, and differences between tools at varying maturity levels.

An important distinction made during these conversations is between the fully developed LuisterLinie transcription application and the newer conversational AI tool ChatLDH. LuisterLinie has been in use longer and has a clearly defined purpose, while ChatLDH is still in an earlier phase of development and adoption. This distinction enables a comparison of the challenges and opportunities that arise at different stages of GenAI tool maturity. During the interviews, the researcher follows a semi-structured question list covering system use and evolution, changes over time, technical constraints, organizational enablers and barriers, user feedback mechanisms, and future development priorities. The interview protocol is included in Appendix A.

In addition to the formal interviews, informal user discussions, observations of workflows, and analysis of internal documentation help contextualize the findings and provide a deeper understanding of the organization as the researcher dives deeper within the organization. These additional sources of insight make sure that the interview findings are grounded in the reality of the organization.

3.3 Survey Design

The insights gathered from the exploratory interviews are translated into a quantitative survey instrument designed to collect end-user perceptions of the GenAI tools. The survey focuses on six theoretical constructs: Efficiency, Quality, Workload, Cognitive Load, Trust, and Usability. These constructs are aligned with both the priorities identified by developers during the exploratory interviews and the relevant dimensions identified in the existing literature on GenAI adoption. Topics covered by the survey

include time savings, improvements in documentation quality, trust in AI output, and the effort required to verify and correct AI-generated content.

The survey is implemented directly within the LuisterLinie application, where a Microsoft Forms pop-up activates when users end a recording session. To collect additional answers, the survey could also be distributed through an e-mail. A similar survey is used for ChatLDH, tailored to the different functionality of that tool. The survey and its translated English version are included in Appendix B. Prior to distribution, the survey was reviewed by the development team to ensure clarity and relevance of the questions. The survey uses 5-point Likert scales for all closed questions, with response options ranging from strongly disagree to strongly agree. Three open-ended questions are included at the end of the survey to allow respondents to provide additional feedback and context to the given answers.

3.4 Data Collection

Data collection proceeds along two parallel tracks: survey responses from end-users and behavioral log data from the GenAI applications. The survey is distributed through the in-app pop-up, and users from different activity levels were sent an e-mail to answer the survey. Responses are collected over a period of several weeks to ensure an adequate sample size. The applications log data is extracted from Azure Application Insights using Kusto Query Language (KQL) queries executed through a Grafana dashboard. Six concrete metrics are extracted from the log data: weekly active users to track adoption growth, recording length distribution to understand usage patterns, chat turns per session to measure engagement depth with the chat feature, time from recording end to final action to understand workflow integration, chat-to-recording ratio to measure how often recordings transition into chat follow-up, and summary edit rate as a proxy for output quality.

All data collection follows strict ethical guidelines. All data is fully anonymized prior to analysis, no personally identifiable information is accessible to the researcher, it is not possible to trace survey responses back to individual users or link them to log data, and data is used exclusively for academic research purposes. Participants are informed about the purpose of the study, and data handling procedures comply with organizational and data protection regulations.

3.5 Evaluation

The evaluation combines analysis of the survey data with interpretation of the behavioral log metrics. The survey data is analyzed using a systematic multi-step approach. First, the data is imported from Microsoft Forms and cleaned of any personally identifiable metadata and incomplete responses. Likert-scale responses are converted to numerical values, with negatively phrased items reverse-coded to ensure consistent interpretation. Individual survey items are grouped into the six theoretical constructs, and construct scores are calculated as the arithmetic mean of the items belonging to each construct.

Descriptive statistics are calculated at both the item and construct levels, and internal consistency is assessed using Cronbach's alpha. Pearson correlation coefficients are calculated between all construct scores to examine relationships among the latent concepts. An exploratory factor analysis is performed to investigate whether the empirical survey items reflect the theoretically proposed measurement constructs. To investigate whether perceived benefits differ according to experience with the tool, respondents are divided into occasional and frequent user groups based on their reported usage frequency, and construct scores are compared using the Mann–Whitney U test.

The application log data provides objective behavioral indicators of actual tool usage that complement the subjective survey perceptions. While individual-level integration with survey data is not possible due to anonymization constraints, the combined use of behavioral logs and perceptual survey data enables a triangulated understanding of the impact of GenAI tools on organizational work. The evaluation interprets the convergence and divergence between these two data sources to provide a comprehensive picture of how the tools are used and perceived.

Chapter 4

Exploratory Interview Results

This section presents an analysis of the exploratory conversations conducted with five members of the AI team at the Dutch Salvation Army, referred to as P1 through P5 throughout this chapter. P1 is a developer of the GenAI applications LuisterLinie and ChatLDH, P2 is an AI engineer, and P3 focuses on DevOps and infrastructure. P4 holds a managerial role within the AI team, and P5 operates in an advisory capacity. The goal of this analysis is to identify key themes, challenges, and opportunities regarding the implementation of GenAI tools within the organization. These findings inform the development of a survey for end-users, which is included in Appendix B. The interview results show how GenAI tools influence knowledge management, productivity, and workload, as well as the organizational and technical context in which they are developed, allowing the survey to assess whether the tools are functioning as the development team intended.

The analysis follows several themes that emerged as recurrent patterns across the interviews. These include administrative workload as the primary driver for tool development, the impact on quality of work and knowledge management, human-AI interaction and cognitive workload, organizational barriers such as privacy and resistance, differences between tools at varying stages of maturity, and the importance of user-centered development.

4.1 Administrative workload as the primary driver

A central theme across all conversations is that the development of GenAI tools within the Dutch Salvation Army is driven by the need to reduce administrative workload. As P4 explained, administrative burdens consume a substantial portion of employees' time: care workers spend approximately 30 to 40 percent of their working hours on administrative tasks. This highlights the main issue at hand: caretakers are so occupied with administrative obligations, necessary though they may be for compliance and documentation, that they have considerably less time available for their primary mission of client care. P4 further elaborated on this by describing the administrative work as repetitive and burdensome.

A further concern raised during the conversations was that administration also intrudes upon interactions with clients. P4 illustrated this by pointing out that if a care worker is taking notes during a conversation, they are not able to give their full attention to the client. This illustrates how administrative requirements directly interfere with the quality of client interactions. Consequently, the introduction of tools like LuisterLinie is not solely about efficiency gains; it extends into fundamentally changing the way people work, where administration is minimized and the quality of interactions is maximized through GenAI-supported transcription and summarization.

4.1.1 Impact on quality of work and knowledge management

Beyond efficiency and administration, the developers expressed a strong interest in the quality of both the administration and the interactions themselves. The leading idea is that the quality of interactions increases while administration improves through the generation of more neutral, objective content. As P4 noted, care workers have more attention available for the client, which improves the conversation itself. P4 added that documentation becomes more neutral because a language model simply reports facts without emotional bias. P4 reported that several care workers find it very difficult and energy-intensive to write a neutral report, particularly when a client has relapsed or gone against advice; LLMs can assist by producing this content more neutrally and efficiently.

On the other hand, a common critique is that AI-generated output can feel impersonal. P4 noted that the generated text lacks tone and that one cannot sense the person behind the words. This highlights a tension between objectivity and context in documentation. While AI improves consistency, it may omit important contextual or emotional cues that a care worker would otherwise provide. To address this limitation, the concept of a human-in-the-loop is considered essential. As P4 put it, the AI output provides a foundation, and the care worker must still refine it into a proper report. This perspective reinforces that GenAI is not an all-solving solution but rather a tool that supports human judgement. Caretakers remain responsible for interpreting and refining the output to ensure that quality standards are met.

4.1.2 Human-AI interaction and cognitive workload

While GenAI is intended to reduce administrative work, the interviews also reveal that these tools can introduce new forms of cognitive workload. The developers expressed concerns about usability, user experience, and the cognitive effort required to interpret and refine AI-generated outputs. P2 noted that while chat-based interaction feels intuitive to technically oriented team members, many users open the chat and do not know where to go or what to do. This highlights that development must prioritize accessibility for all users, including those who are less technologically experienced. P2 added that for some users, the way in which an answer is structured and presented is very important. This suggests that cognitive workload is not solely determined by task complexity but also by how information is presented to the user. Poorly designed interfaces or responses can increase mental effort, even when the underlying task has been automated, an important consideration for the end-user survey.

4.1.3 Organizational barriers: Privacy, structure and resistance

Implementing GenAI tools within a larger organization presents several organizational challenges. The interviews point to three major barriers: privacy and compliance concerns, mismatches between development speed and organizational structure, and resistance to change.

Regarding privacy and compliance, this was consistently identified as the most time-consuming challenge. P4 noted that privacy and compliance procedures took approximately one and a half years to complete, illustrating how privacy processes can significantly slow down innovation in larger organizations. Furthermore, the same privacy restrictions that protect client data also limit the development team's ability to access usage data for system improvement. As P2 explained, being unable to look into the database as a developer slows down the feedback loop, even though these restrictions are necessary for data protection.

With respect to organizational structure, the development team experiences a mismatch between the fast pace of sprint-based development and the slower decision-making processes of the broader organization. P1 characterized this imbalance between a quickly moving development team and a slowly moving organization as an inherent challenge of digital transformation.

Resistance to change was also identified as a significant barrier. The Dutch Salvation Army employs many caretakers who are not technologically oriented and are often resistant to digital transformation. P4 recalled the initial skepticism, noting that people did not believe care workers would ever record their conversations. However, P4 observed that a year and a half later, attitudes had shifted considerably. This suggests that the adoption of GenAI tools is a gradual process that requires time and demonstrated value to overcome initial resistance.

4.1.4 Tools and stages of maturity

The interviews reveal clear differences between the two main tools studied: LuisterLinie and ChatAI. LuisterLinie is described as a more mature product focused on a specific use case, namely transcription and summarization. Its functionality is relatively clear and user adoption is accordingly more straightforward. P3 described it as a simple product that speaks for itself. In contrast, ChatAI is still evolving and has a broader, less clearly defined scope. As P3 noted, the development team is still exploring how to convince colleagues that this tool provides added value beyond what LuisterLinie already offers. This difference underscores the importance of clarity in value proposition: tools with a clear and immediate use case are easier to adopt, while more general tools require additional effort to demonstrate their usefulness. Development approaches also differ between the two tools: P3 noted that ChatAI is approached more experimentally, with features sometimes built in just two days, whereas LuisterLinie has transitioned into a more stable, production-oriented phase.

4.1.5 User-centered development and feedback

A final key theme is the importance of user involvement in the development process. P4 emphasized the need to test tools in real-world settings by finding people on the work floor who work with real cases. This reflects a strong alignment with action research principles, where iterative feedback is central to continuous improvement. Feedback mechanisms are being actively developed; P4 mentioned comprehensive logging systems that track which prompts are used and how frequently. Additionally, structured

processes such as product ownership and Scrum methodologies have been introduced to ensure that development aligns with user needs. Nevertheless, P4 acknowledged that user feedback is not always fully utilized, noting that they were not sure whether the team actively monitors all the feedback they receive. This indicates an opportunity for further improvement in feedback integration.

4.1.6 Summary

The analysis of the exploratory phase reveals that the implementation of GenAI tools is driven by a clear need to reduce administrative workload, but its impact extends far beyond efficiency gains. The tools influence the quality of work, reshape knowledge management practices, and introduce new forms of cognitive workload. At the same time, implementation is shaped by significant organizational and technical challenges, including privacy constraints, resistance to change, and the complexity of integrating innovation into existing structures. The findings also highlight the importance of tool maturity and clarity of use case, as well as the critical role of user-centered development. Overall, the results provide a nuanced understanding of how GenAI tools function within a real-world organizational context, forming a strong foundation for subsequent phases of the research.

Chapter 5

Results

5.1 Survey Results

5.1.1 Respondent Characteristics

A total of 51 completed questionnaires were included in the quantitative analysis after data cleaning. Responses containing excessive missing values were removed according to the procedure described in Chapter 3. The remaining dataset consisted of complete responses across 17 Likert-scale items and three open-ended questions.

Respondents represented different user groups within the Leger des Heils and reported varying frequencies of LuisterLinie usage. Based on reported usage frequency, respondents were categorized into two groups:

- Occasional users ($n = 30$);
- Frequent users ($n = 21$).

These groups were used for the comparative analyses presented later in this chapter.

5.1.2 Descriptive Statistics

Table 5.1 presents the descriptive statistics for each survey construct.

Table 5.1: Descriptive statistics of the survey constructs

Construct	Items	Mean	SD	SEM	Cronbach's α
Efficiency	3	3.94	0.74	0.10	0.78
Quality	4	3.60	0.74	0.10	0.76
Workload	2	3.73	0.93	0.13	0.86
Cognitive Load	3	3.47	0.71	0.10	0.55
Trust	2	3.77	0.62	0.09	0.17
Usability	2	4.02	0.69	0.10	0.29

Figure 5.1 visualizes the construct means together with their corresponding 95% confidence intervals.

Overall, respondents evaluated LuisterLinie positively across all measured constructs. The highest average score was observed for the Usability construct ($M = 4.02$), followed by Efficiency ($M = 3.94$). Quality and Workload also received positive average scores of 3.60 and 3.73, respectively.

Cognitive Load obtained the lowest construct mean ($M = 3.47$), indicating greater variation in respondents' perceptions regarding the mental effort required when using the application.

5.1.3 Item-Level Results

The highest-rated survey items were those related to administrative efficiency. Respondents reported strong agreement that LuisterLinie reduces the time spent on documentation ($M = 4.24$), accelerates administrative work ($M = 4.24$), and increases overall efficiency ($M = 4.08$).

Items relating to documentation quality received moderately positive scores. Respondents generally agreed that LuisterLinie improves the quality of documentation ($M = 3.71$), produces more neutral summaries ($M = 3.69$), improves report structure ($M = 3.55$), and adds additional detail to documentation ($M = 3.47$).

Regarding workload, respondents agreed that LuisterLinie reduces both administrative workload ($M = 3.82$) and work-related stress ($M = 3.63$).

Responses concerning cognitive workload were more varied. Participants moderately agreed that additional time is required to improve AI-generated output ($M = 3.22$), while lower agreement was observed for the need to think more during documentation ($M = 2.55$). The statement concerning additional pressure when working with AI received the lowest average score ($M = 1.82$).

Privacy concerns also received relatively low agreement ($M = 1.86$), while ease of use ($M = 3.98$) and improved focus during conversations ($M = 4.06$) were rated positively.

The complete distribution of responses for each survey item is illustrated in Figure 5.2.

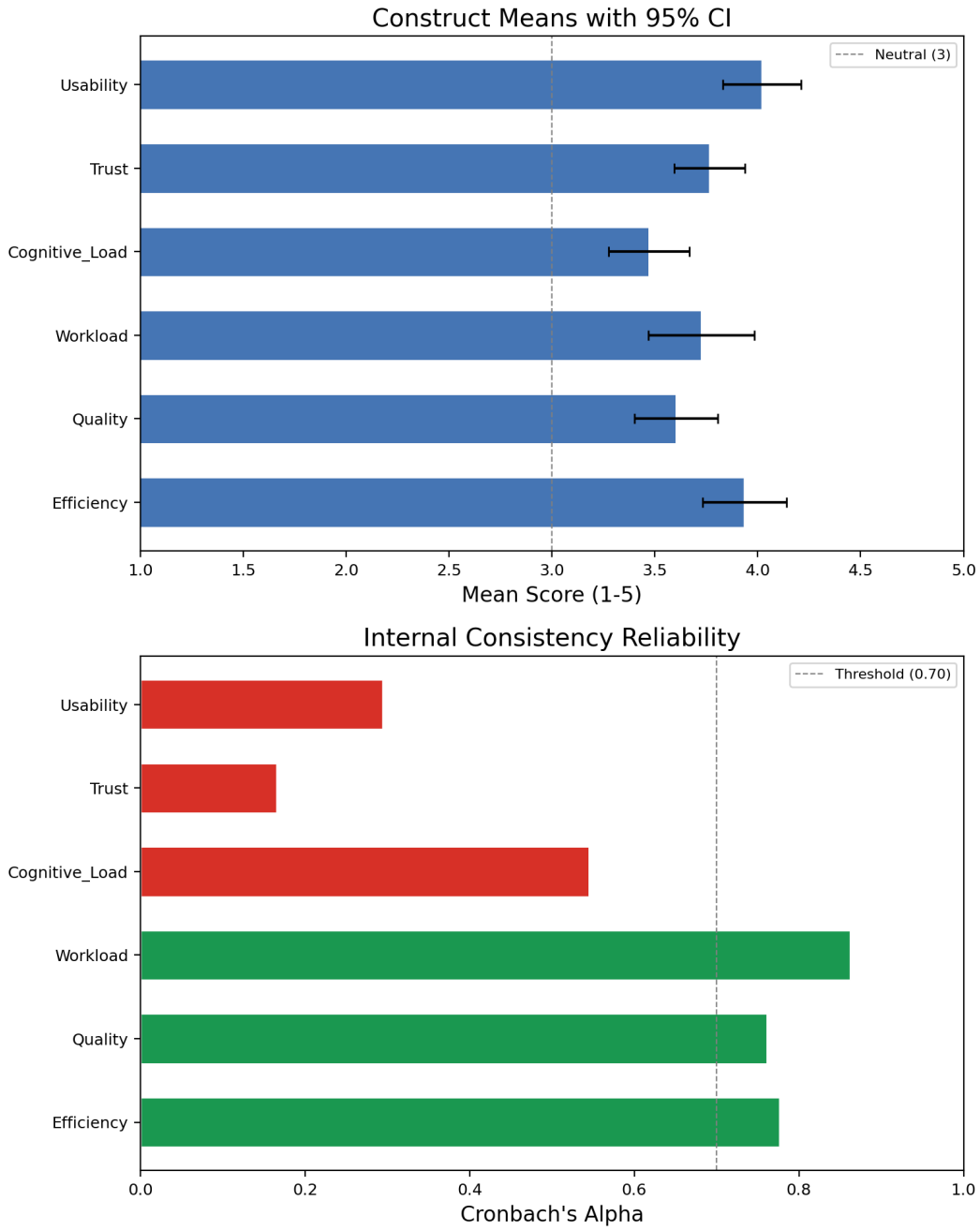

Figure 5.1: Construct means with 95% confidence intervals. Error bars indicate the standard error of the mean. Cronbach's alpha values are displayed above each construct.

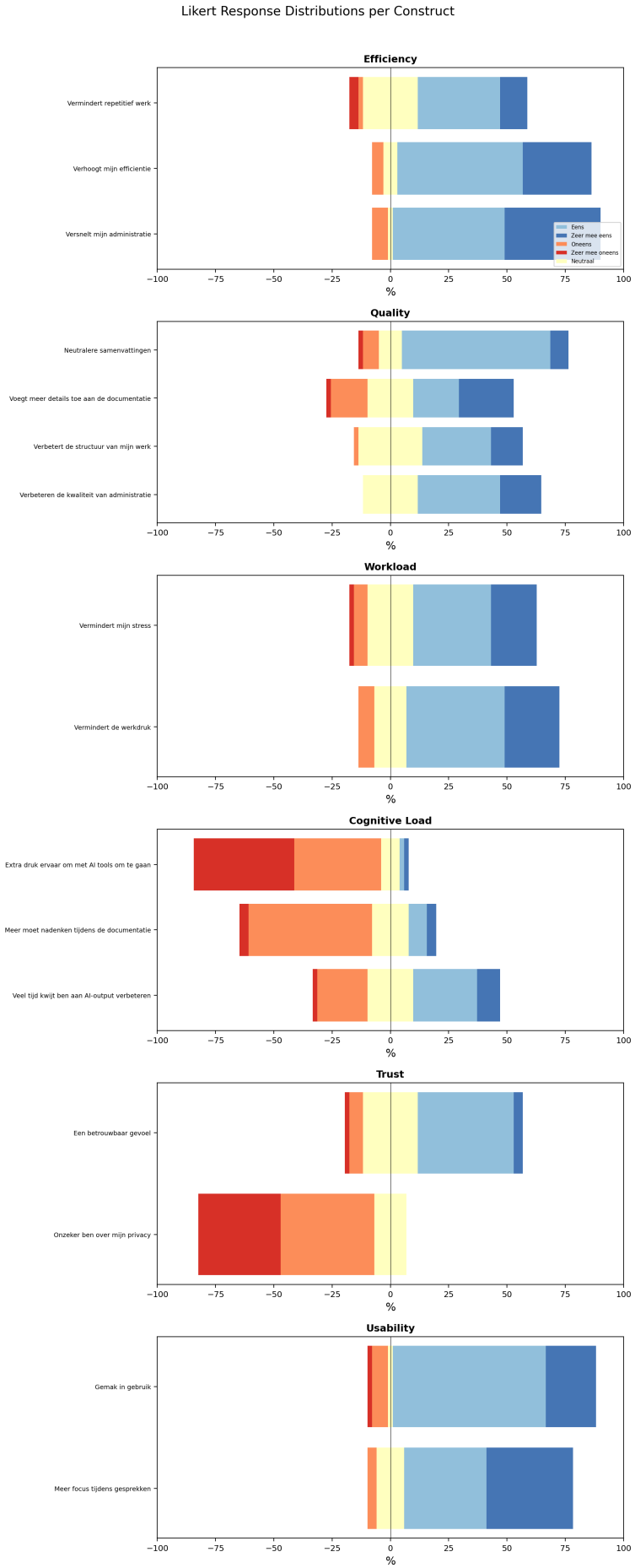

Figure 5.2: Distribution of responses across all Likert-scale survey items grouped by construct.

5.1.4 Reliability Analysis

Cronbach's alpha was calculated to assess the internal consistency of each construct.

Three constructs demonstrated acceptable reliability. Efficiency obtained a Cronbach's alpha of 0.777, Quality obtained 0.761, and Workload obtained 0.862, indicating satisfactory internal consistency.

The remaining constructs produced lower reliability coefficients. Cognitive Load obtained an alpha of 0.545, while Trust and Usability produced values of 0.165 and 0.294, respectively. These values indicate relatively low internal consistency within these constructs.

As described in the methodology, the reliability analysis was performed to evaluate the suitability of the measurement instrument and to identify constructs requiring refinement in future iterations of the survey.

5.1.5 Correlation Analysis

Pearson correlation coefficients were calculated between all construct scores.

The resulting correlation matrix is presented in Figure 5.3.

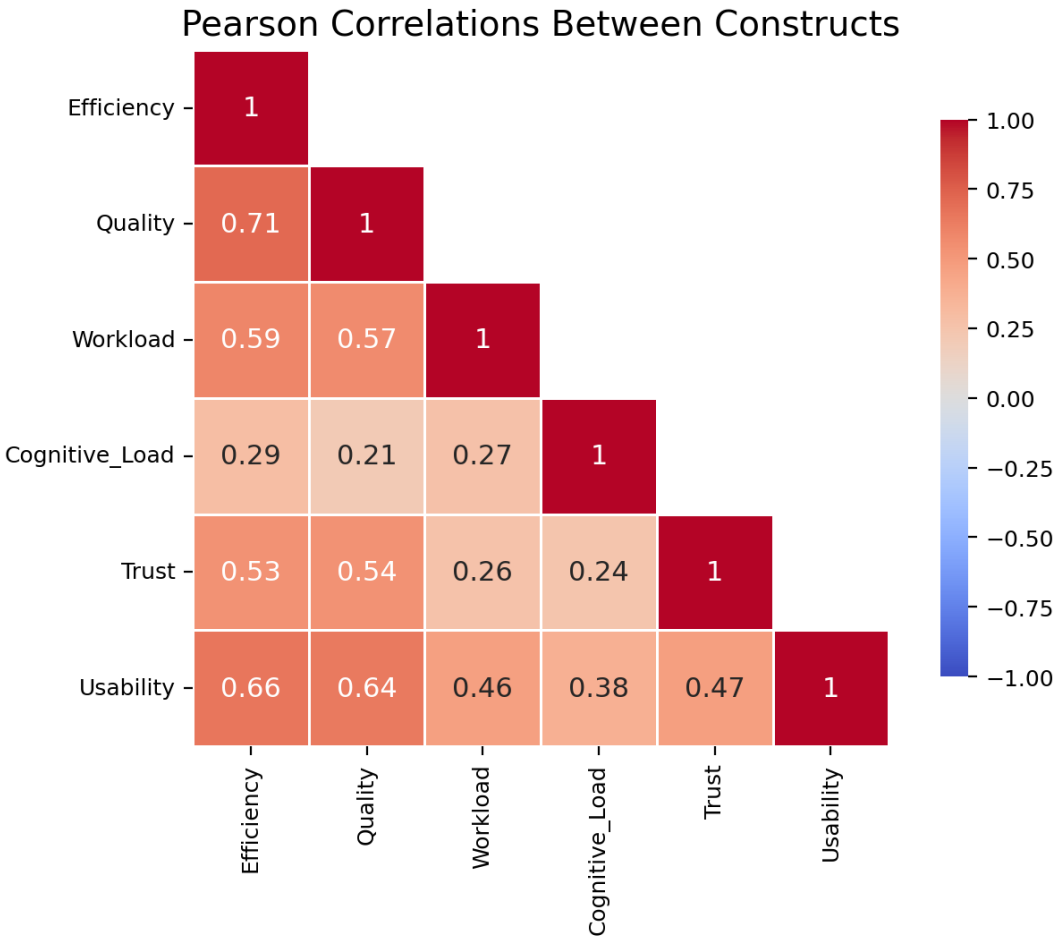

Figure 5.3: Pearson correlation matrix between the survey constructs. Darker colors indicate stronger positive relationships.

The strongest positive relationship was observed between Efficiency and Quality ($r = 0.714$). Strong positive correlations were also observed between Efficiency and Usability ($r = 0.657$), Quality and Usability ($r = 0.643$), and Efficiency and Workload ($r = 0.593$).

Trust demonstrated moderate positive correlations with Efficiency ($r = 0.532$) and Quality ($r = 0.537$).

The Cognitive Load construct exhibited comparatively weaker correlations with the remaining constructs, with coefficients ranging between $r = 0.205$ and $r = 0.378$.

Overall, the correlation matrix indicates positive associations between the majority of the measured constructs.

5.1.6 Exploratory Factor Analysis

An exploratory factor analysis was performed to examine the latent structure of the survey instrument.

Three factors were extracted, together explaining approximately 49.35% of the total variance.

The resulting factor loadings are presented in Figure 5.4.

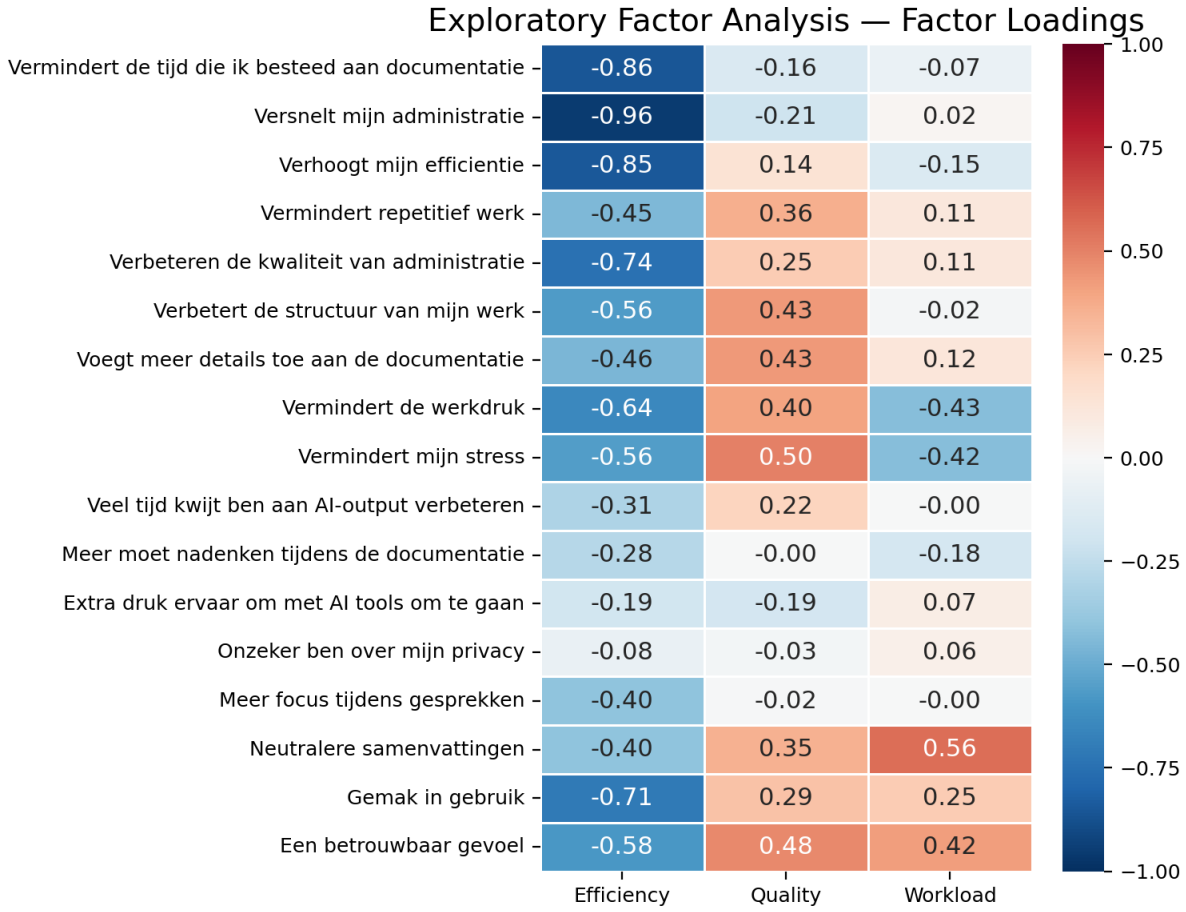

Figure 5.4: Factor loading matrix obtained from the exploratory factor analysis. Higher absolute values indicate stronger associations between survey items and latent factors.

Items related to administrative efficiency demonstrated strong loadings on the first factor, while documentation quality items generally loaded on the second factor. Workload-related items primarily contributed to the third factor, although several survey items exhibited cross-loadings across multiple factors.

Items measuring Cognitive Load and Trust generally produced weaker factor loadings and did not clearly cluster into separate latent dimensions.

Because of the exploratory nature of the analysis and the achieved sample size, these findings should be interpreted as an initial examination of the measurement structure rather than a confirmatory validation of the survey instrument.

5.1.7 Comparison Between User Groups

To investigate whether usage frequency influenced user perceptions, construct scores were compared between occasional users ($n = 30$) and frequent users ($n = 21$) using the Mann–Whitney U test.

Figure 5.5 presents the average construct scores for both user groups.

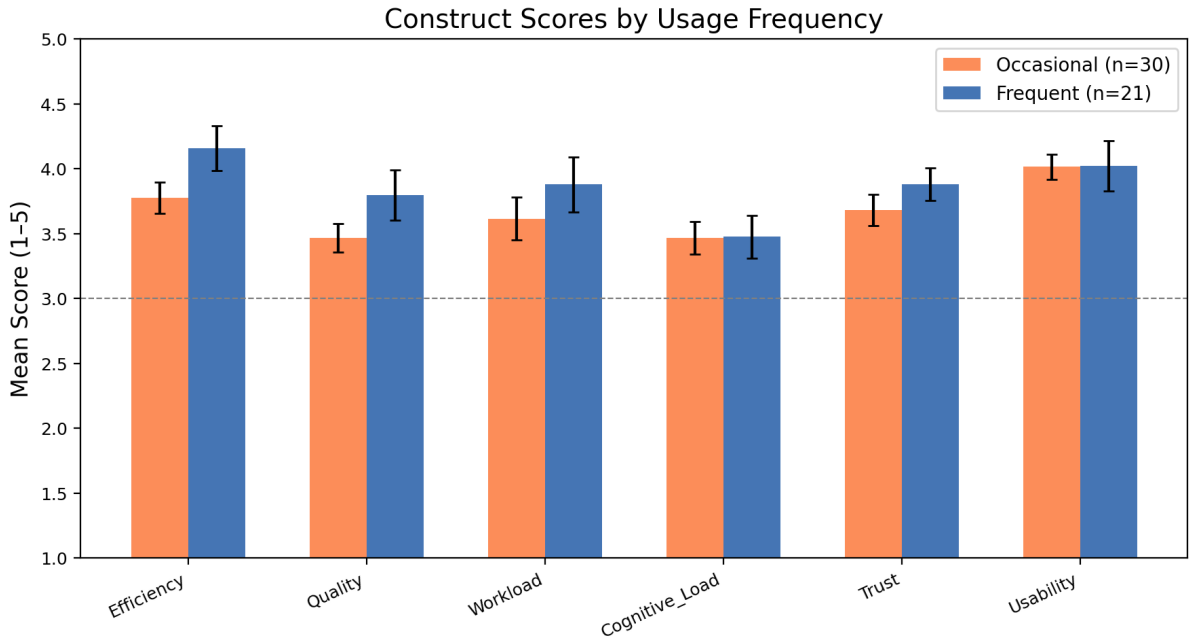

Figure 5.5: Comparison of construct scores between occasional and frequent users. Error bars indicate the standard error of the mean.

Frequent users reported higher average scores for all constructs except Cognitive Load, although the magnitude of these differences varied.

Efficiency demonstrated the largest difference between both groups. Frequent users obtained an average construct score of 4.16 compared to 3.78 for occasional users. The Mann–Whitney U test indicated that this difference was statistically significant ($U = 189.0,\ p = 0.014$).

Quality also demonstrated higher scores among frequent users (3.80 versus 3.47). Although this difference did not reach the conventional significance threshold of $p < 0.05$, the observed result ($p = 0.086$) suggests a possible trend that may warrant further investigation using a larger sample.

No statistically significant differences were observed for Workload, Cognitive Load, Trust, or Usability.

Table 5.2 summarizes the statistical comparison.

Table 5.2: Comparison between occasional and frequent users

Construct	Occasional	Frequent	U	p-value
Efficiency	3.78	4.16	189.0	0.014
Quality	3.47	3.80	225.5	0.086
Workload	3.62	3.88	269.5	0.374
Cognitive Load	3.47	3.48	314.0	0.992
Trust	3.68	3.88	256.0	0.250
Usability	4.02	4.02	261.5	0.290

5.1.8 Summary of Quantitative Results

The quantitative survey results demonstrate generally positive perceptions of LuisterLinie across all measured constructs. Respondents reported the highest scores for usability and efficiency, while quality and workload reduction also received favourable evaluations.

Reliability analysis indicated acceptable internal consistency for the Efficiency, Quality, and Workload constructs, whereas the Cognitive Load, Trust, and Usability constructs produced lower reliability coefficients.

The exploratory factor analysis extracted three latent factors explaining approximately 49% of the total variance, broadly corresponding with the conceptual dimensions of the survey instrument.

Comparison between occasional and frequent users indicated statistically significant differences only for perceived Efficiency, with frequent users reporting higher scores than occasional users.

The findings presented in this chapter provide the quantitative foundation for the discussion in Chapter 6, where the survey results are interpreted alongside the qualitative findings and application usage data.

5.2 Application Usage Data

In addition to the survey, log data from the LuisterLinie application was analyzed to obtain objective behavioral indicators of tool usage. This section presents six metrics extracted from Azure Application Insights.

5.2.1 Weekly Active Users

Figure 5.6 shows the growth in weekly active users since the public release of LuisterLinie. The user base grew from 125 users in early April to a peak of 465 users by mid-June, then stabilised around 430–450 active users through June and July. Table 5.3 summarises the monthly trends.

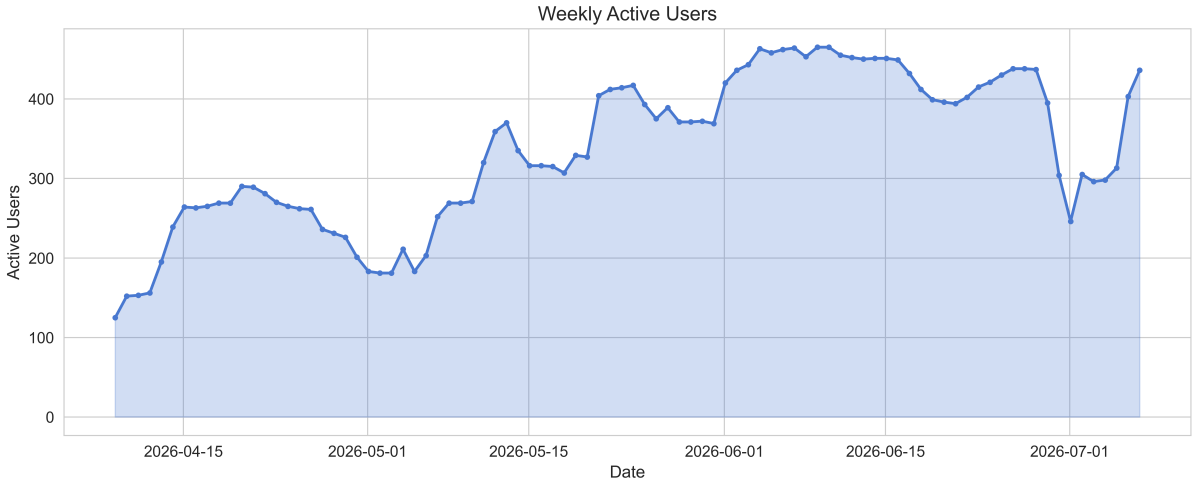

Figure 5.6: Weekly active users of LuisterLinie since public release.

Table 5.3: Weekly Active Users — Monthly Summary

Month	Min	Max	Avg	End-of-Month
Apr 2026	125	290	235	201
May 2026	181	417	316	369
Jun 2026	304	465	432	304
Jul 2026	246	436	328	436

5.2.2 Recording Length Distribution

Figure 5.7 presents the distribution of recording lengths. A total of 2,771 recordings were shorter than ten minutes, of which 1,180 were under two minutes (likely failed or abandoned attempts). The frequency declines gradually after ten minutes. A notable secondary spike occurs at approximately 240 minutes, corresponding to the automatic cutoff limit of the application. Table 5.4 provides the aggregated distribution.

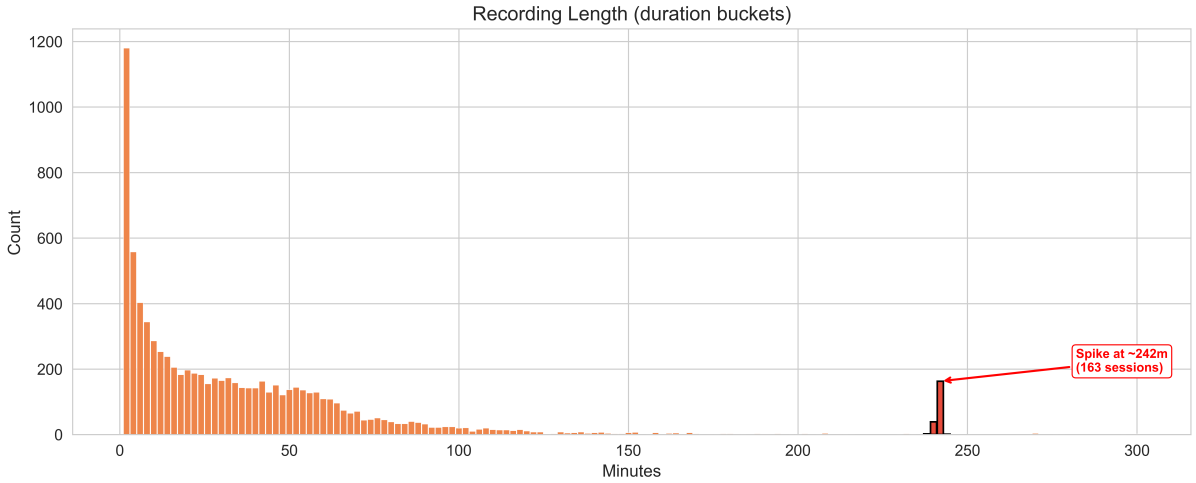

Figure 5.7: Distribution of recording lengths in minutes.

Table 5.4: Recording Length Distribution

Duration	Sessions
≤10 min	2,771
10–20 min	1,076
20–30 min	862
30–60 min	2,104
1–2 h	1,074
2–3 h	124
3–4 h	73
4–5 h	173
5+ h	0

5.2.3 Chat Turns per Conversation

Figure 5.8 displays the number of chat turns per conversation session. The majority of sessions consist of a single turn (1,207 sessions), with 724 sessions containing two turns. Beyond three turns, the counts follow a steady downward trend. Table 5.5 groups the data into meaningful ranges.

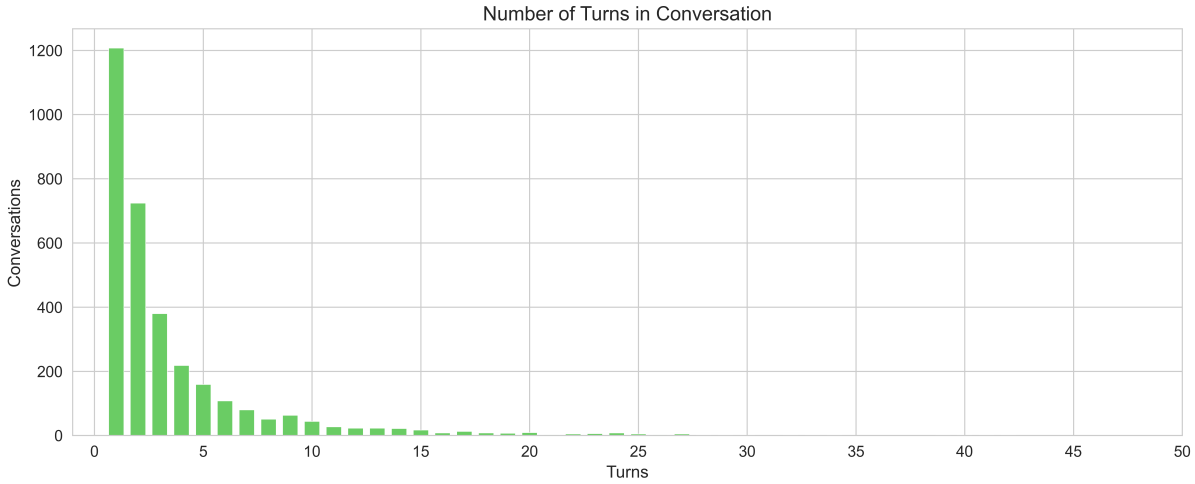

Figure 5.8: Distribution of chat turns per conversation session.

Table 5.5: Number of Turns in Conversation

Turns	Conversations
1 turn	1,207
2 turns	724
3–5 turns	757
6–10 turns	346
11–20 turns	157
21+ turns	54

5.2.4 Time from Recording End to Final Action

Figure 5.9 shows the time elapsed between the end of a recording and the user's final action within the same session. A pronounced peak is visible immediately after the recording ends (423 sessions within 30 minutes), after which the distribution remains relatively steady across the first six hours. The highest concentration occurs in the 6–8 hour window (2,217 sessions), suggesting that many users end their shift and return to finalise their notes later. Table 5.6 provides the full aggregated breakdown.

Table 5.6: Time Until Final Action After Recording

Time Window	Recordings
≤30 min	423
30–60 min	410
1–2 h	994
2–3 h	1,124
3–4 h	1,183
4–5 h	964
5–6 h	1,024
6–8 h	2,217
8–10 h	710
10+ h	48

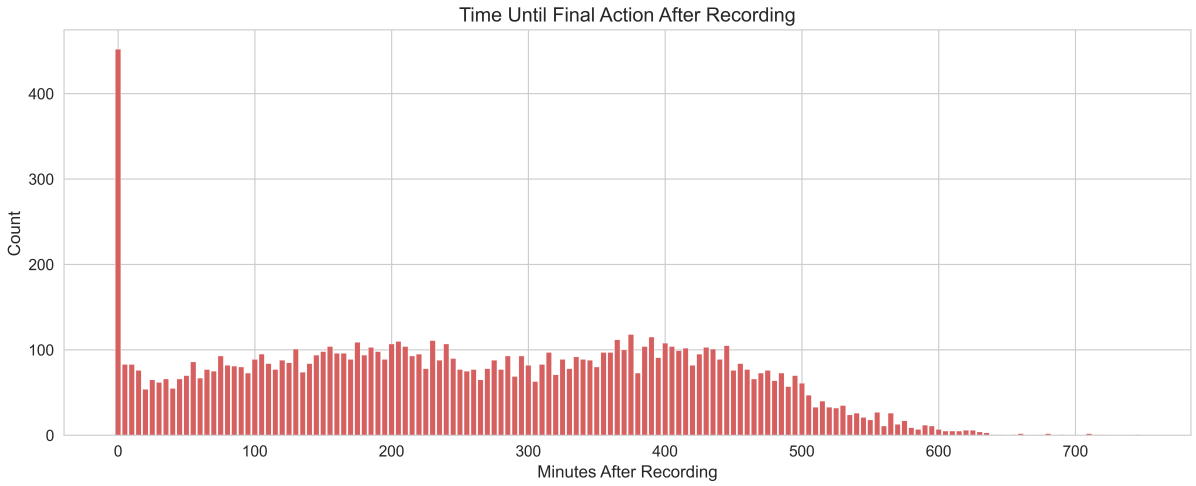

Figure 5.9: Time from recording end to final user action, in minutes.

5.2.5 Chat-to-Recording Ratio

A total of 11,988 recordings were made during the observation period, of which 3,245 sessions (27.1%) transitioned into chat-based follow-up. Table 5.7 summarises these figures. The rate indicates that a substantial minority of users engage with the chat feature after finishing a recording session.

Table 5.7: Chat Sessions Started

Metric	Value
Total Recordings	11,988
Chat Sessions	3,245
Chat Rate	27.1%

5.2.6 Summary Edit Rate

Of the 11,988 recorded conversations, only 179 transcripts or summaries were manually edited by users, representing an edit rate of 1.49%. This very low percentage suggests that the AI-generated output is generally accepted without modification, indicating satisfactory output quality from the users' perspective.

Table 5.8: Edited Summaries

Metric	Value
Total Recordings	11,988
Edited Summaries	179
Edit Rate	1.49%

Chapter 6

Discussion

6.1 Survey

The objective of this survey was to investigate how Generative Artificial Intelligence (GenAI), specifically the LuisterLinie application, influences perceived knowledge management, productivity, and workload within the Dutch Salvation Army. This chapter discusses the quantitative survey findings in relation to the research objectives, the existing literature, and the practical context in which the application is used.

6.1.1 Perceived Productivity Benefits

One of the clearest findings emerging from the survey is that users perceive LuisterLinie as improving their administrative efficiency. Respondents reported that the application reduces documentation time ($M = 4.24$), accelerates administrative work ($M = 4.24$), and increases overall efficiency ($M = 4.08$). These findings align with the developer interviews conducted during the exploratory interviews, in which reducing administrative burden and allowing professionals to spend more time with clients were identified as the primary objectives of the application.

The observed efficiency improvements are also consistent with previous research on Generative AI in knowledge-intensive work. Studies such as Noy and Zhang [16] and Brynjolfsson et al. [5] have shown that AI-assisted documentation can substantially reduce routine administrative activities, allowing professionals to allocate more time to higher-value tasks. The present findings suggest that these benefits are also experienced within a non-profit healthcare and social care environment, extending the evidence base beyond the commercial and experimental settings that dominate the current literature [10, 13].

Interestingly, respondents who used LuisterLinie more frequently reported significantly higher efficiency scores than occasional users ($M = 4.16$ vs. $M = 3.78$, $U = 189.0$, $p = 0.014$). Although causality cannot be established from the survey data, this finding suggests that familiarity with the application may influence users' ability to integrate AI into their daily workflow. As users become more experienced with the system, they may develop greater confidence in interpreting and utilizing AI-generated documentation, leading to greater perceived efficiency gains.

6.1.2 Knowledge Management and Documentation Quality

Knowledge management formed one of the central theoretical constructs of this research. The survey results indicate that respondents generally perceive improvements in documentation quality ($M = 3.71$), report structure ($M = 3.55$), and the completeness of administrative records ($M = 3.47$). Together, these items produced a Quality construct mean of $M = 3.60$ ($\alpha = 0.76$).

These findings support the notion that GenAI can contribute to organizational knowledge management by improving the consistency and accessibility of documentation. Rather than replacing professional judgement, the application appears to function as a supporting tool that assists employees in creating structured and comprehensive records.

The strong positive correlation observed between the Efficiency and Quality constructs ($r = 0.714$) further suggests that respondents who perceive greater productivity improvements also tend to perceive improvements in documentation quality. Although these relationships cannot be interpreted causally, they indicate that increased efficiency does not necessarily come at the expense of documentation quality. Instead, both constructs appear to improve simultaneously.

From a knowledge management perspective, this finding is particularly relevant, as higher-quality documentation contributes to improved organizational memory, more consistent reporting practices, and better information retrieval for future care processes.

6.1.3 Workload Reduction and Cognitive Workload

One of the research objectives was to distinguish between traditional workload reduction and cognitive workload introduced by AI-assisted work.

The survey demonstrates that respondents generally perceive reductions in administrative workload ($M = 3.82$) and work-related stress ($M = 3.63$) when using LuisterLinie (Workload construct: $M = 3.73$, $\alpha = 0.86$). These findings correspond with the intended purpose of the application and support previous literature suggesting that AI systems can reduce repetitive administrative activities.

At the same time, respondents reported moderate levels of cognitive workload associated with reviewing AI-generated output (Cognitive Load construct: $M = 3.47$, $\alpha = 0.55$). Users indicated that they still spend time correcting generated documentation ($M = 3.22$) and remain responsible for verifying its correctness before finalizing reports.

This distinction is important. Rather than eliminating work entirely, LuisterLinie appears to shift part of the workload from manual documentation toward quality assurance and verification. This observation supports previous research emphasizing that Generative AI changes the nature of professional work rather than fully automating it [12, 19]. Human oversight remains essential, particularly in care environments where documentation quality directly affects client care and organizational accountability. This aligns with the concept of human–AI symbiosis, where professional judgement and AI-generated suggestions complement one another rather than replacing human expertise [2].

Interestingly, cognitive workload demonstrated relatively weak correlations with the remaining constructs (maximum $r = 0.378$). This suggests that the additional mental effort required to review AI-generated output does not necessarily diminish users' overall positive perception of the application.

6.1.4 Trust and Usability

The survey indicates generally positive perceptions regarding usability. Respondents reported that the application is easy to use ($M = 3.98$) and enables greater focus during client conversations ($M = 4.06$). These findings correspond with one of the principal design objectives identified during the developer interviews: minimizing interaction with the technology during conversations while allowing documentation to be generated automatically. The Usability construct, despite being quite unreliable with a Cronbach Alpha score of 0.29, obtained the highest overall mean score of all constructs ($M = 4.02$), which indicates that the user perception is consistent with research emphasizing that perceived ease of use is a critical determinant of technology acceptance in organizational settings [2].

The Trust construct requires more careful interpretation. Although respondents generally expressed confidence in the application and reported relatively limited privacy concerns ($M = 1.86$), the internal consistency of the Trust construct was low ($\alpha = 0.17$). This suggests that privacy concerns and perceived trustworthiness may represent different underlying concepts rather than a single latent construct.

Similarly, the relatively low reliability observed for the Usability construct ($\alpha = 0.29$) indicates that ease of use and conversational focus may capture different dimensions of the user experience. Future versions of the survey should therefore include additional items to better operationalize these constructs.

6.1.5 Evaluation of the Survey Instrument

An important secondary objective of this research was the development of a quantitative measurement instrument suitable for evaluating GenAI applications within organizational settings.

The reliability analysis demonstrated acceptable internal consistency for the Efficiency ($\alpha = 0.78$), Quality ($\alpha = 0.76$), and Workload ($\alpha = 0.86$) constructs. These constructs therefore appear suitable for continued use in future evaluations.

The remaining constructs require refinement. In particular, the Trust construct ($\alpha = 0.17$) combined both positive and negative perceptions within only two items, resulting in low internal consistency. Similarly, Cognitive Load ($\alpha = 0.55$) may represent multiple dimensions of mental effort that cannot be adequately captured using the current survey design.

The exploratory factor analysis broadly reflected the theoretical structure proposed in the conceptual model, extracting three factors that together explained 49.35% of the total variance. This dimensionality is consistent with the Knowledge Management framework proposed by Alavi and Leidner [1] and the productivity–workload distinction discussed by Drucker [8]. However, several survey items exhibited cross-loadings. Given the achieved sample size of 51 respondents, these findings should be interpreted as exploratory, as recommended for studies with similar sample sizes in organisational research [23]. Nevertheless, the results provide useful guidance for improving the measurement instrument prior to future confirmatory analyses.

6.1.6 Implications for Practice

From an organizational perspective, the results indicate that LuisterLinie is generally perceived as a valuable support tool for administrative work within the Leger des Heils.

The reported improvements in efficiency and documentation quality suggest that the application contributes to one of its primary implementation goals: reducing administrative burden while supporting high-quality reporting.

However, the findings also emphasize that successful implementation extends beyond technical performance. Continued user training, clear organizational guidelines, and transparent communication regarding privacy and AI verification responsibilities remain important factors influencing long-term adoption.

Furthermore, the finding that frequent users reported higher perceived efficiency ($p = 0.014$) suggests that organizations may benefit from investing in onboarding and continuous training, enabling employees to become more proficient in integrating AI into their daily work processes.

6.1.7 Limitations

Several limitations should be considered when interpreting the findings.

First, the achieved sample size of 51 respondents was sufficient for descriptive statistics and exploratory analyses but remains below commonly recommended thresholds for confirmatory factor analysis and covariance-based Structural Equation Modeling. Consequently, the quantitative findings should be interpreted as exploratory.

Second, the survey measures user perceptions rather than objective productivity. Although respondents consistently reported reductions in documentation time and workload, these findings do not directly demonstrate actual improvements in organizational performance.

Third, the survey represents a single organization and evaluates one specific AI application. The findings may therefore not be directly generalizable to other healthcare organizations or different Generative AI systems.

Finally, although application log data were collected to complement the survey, these data primarily provide behavioral indicators rather than subjective user experiences. Combining both sources through methodological triangulation strengthens the overall validity of the research but does not eliminate the limitations associated with each individual data source.

6.1.8 Future Research

Future research should continue validating the proposed measurement instrument using a larger and more diverse sample. Achieving a larger number of respondents would enable confirmatory factor analysis and Structural Equation Modeling, allowing the conceptual relationships proposed in this thesis to be statistically tested.

In addition, longitudinal studies could investigate how user perceptions evolve over time as employees become more experienced with AI-assisted documentation. The observed differences between frequent and occasional users suggest that adoption maturity may play an important role in realizing the benefits of Generative AI.

Finally, future research could further integrate subjective survey findings with objective application usage data, including interaction logs and workflow analytics. Combining behavioral data with user perceptions may provide a more comprehensive understanding of how AI systems influence knowledge work in organizational settings.

6.2 Data

The behavioral metrics extracted from the application logs provide an objective complement to the survey-based perceptions.

6.2.1 User Adoption and Growth

The weekly active user data confirms that adoption has grown substantially since the public release, with the user base growing from 125 to a peak of 465 users between April and June, then stabilising around 430–450 active users in the weeks that followed. The monthly summary (Table 5.3) shows that average weekly active users rose from 235 in April to 432 in June before a slight decline to 328 in July. This growth, combined with sustained retention, suggests that the application is meeting a genuine organisational need and that users continue to find value in the tool over time.

6.2.2 Usage Patterns

The recording length distribution reveals two important patterns. A substantial proportion of recordings (2,771 sessions, or 23% of total) were shorter than ten minutes, and 1,180 of these were under two minutes, likely reflecting accidental activations or abandoned attempts, which is typical for newly deployed tools. The majority of recordings fall within the 30–60 minute window (2,104 sessions), representing typical client meeting durations. The secondary peak at the 240-minute automatic cutoff indicates that some conversations reach the maximum recording duration, which may be a limitation for longer client meetings or multi-session shifts.

The chat histogram demonstrates that most users engage briefly with the chat feature. Of the 3,245 chat sessions, 1,207 (37%) contain a single turn and 724 (22%) contain two turns, indicating that the chat is used primarily for quick follow-up questions or confirmation rather than extended dialogue. Only 54 sessions (1.7%) exceeded 21 turns.

The time-to-final-action metric provides insight into user workflow patterns. A bimodal distribution is visible: 423 sessions (3.5%) are closed within 30 minutes of the recording ending, while the largest concentration occurs in the 6–8 hour window (2,217 sessions, 18.5%). This latter group likely represents users who end their shift immediately after recording and return to finalise their notes later. This pattern suggests that LuisterLinie integrates well into existing workflows without requiring continuous attention.

6.2.3 Output Quality

The low edit rate of 1.49% is particularly noteworthy. Of 11,988 recorded conversations, only 179 transcripts or summaries were manually edited by users. This suggests that the AI-generated summaries and transcripts are of sufficient quality that users rarely feel the need to manually correct them. This finding strengthens the survey results, where users reported improvements in documentation quality ($M = 3.60$) and efficiency ($M = 3.94$).

6.3 Combining Survey and Data

The combination of survey data and application log data enables a more comprehensive evaluation of LuisterLinie's impact than either source alone would provide. The survey captures subjective user perceptions across six theoretical constructs, while the log data provides objective behavioral indicators of actual usage.

Several points of convergence between the two data sources strengthen the overall findings. First, the survey's Efficiency construct received the second-highest mean score ($M = 3.94$), and the application data confirms substantial adoption and sustained usage, as users continue to use the tool frequently, which would be unlikely if it did not improve their efficiency. Second, the low edit rate (1.49%) corroborates the positive Quality construct scores ($M = 3.60$), supporting the conclusion that users perceive AI-generated documentation as sufficiently accurate. Third, the survey's moderate Cognitive Load scores ($M = 3.47$) align with the behavioral observation that most users limit chat interaction to one to three turns (72% of sessions), suggesting that the additional mental effort of reviewing GenAI output is manageable in practice.

However, the integration also reveals areas where perceived and objective measures diverge. The large number of recordings under two minutes (1,180) suggests that a considerable proportion of recording attempts fail or are abandoned, a finding that is not captured by the survey, which asks only about completed recording sessions. This highlights the value of behavioral data in identifying usability issues that users may not explicitly report.

Overall, the methodological combination of survey and log data provides a more robust foundation for the conclusions of this study. Their combined analysis reveals a consistent picture: LuisterLinie is perceived as an effective tool that reduces administrative burden, produces acceptable documentation quality, and has achieved meaningful adoption within the organization.

Chapter 7

Conclusions

7.1 Answers to the research questions

7.1.1 RQ1: Perceived productivity and efficiency

The first research question examined how GenAI tools influence the perceived productivity and efficiency of administrative work within the Dutch Salvation Army. The survey results demonstrated that respondents evaluated LuisterLinie positively on efficiency, with a mean construct score of 3.94 out of 5. Item-level analysis showed particularly strong agreement that the application reduces documentation time ($M = 4.24$), accelerates administrative work ($M = 4.24$), and increases overall efficiency ($M = 4.08$). These perceptions were significantly stronger among frequent users compared to occasional users ($p = 0.014$), suggesting that familiarity with the tool amplifies perceived efficiency gains. The efficiency construct also demonstrated acceptable internal consistency ($\alpha = 0.78$), supporting the reliability of these findings.

7.1.2 RQ2: Knowledge management and documentation quality

The second research question addressed the impact of GenAI tools on knowledge management and documentation quality. Survey respondents reported moderate to positive perceptions of quality improvements, with a construct mean of 3.60. Users agreed that LuisterLinie improves documentation quality ($M = 3.71$), produces more neutral summaries ($M = 3.69$), and improves report structure ($M = 3.55$). The strong positive correlation between Efficiency and Quality ($r = 0.714$) indicates that perceived productivity gains do not come at the expense of documentation quality. The Quality construct demonstrated acceptable reliability ($\alpha = 0.76$), and the extremely low summary edit rate of 1.49% observed in the log data further corroborates that the AI-generated output is generally accepted without modification.

7.1.3 RQ3: Workload and cognitive load

The third research question investigated how GenAI tools affect perceived workload and cognitive load. The survey showed that users perceive reductions in both administrative workload ($M = 3.82$) and work-related stress ($M = 3.63$), with a Workload construct mean of 3.73 and high internal consistency ($\alpha = 0.86$). At the same time, respondents reported moderate levels of cognitive load associated with reviewing and verifying AI-generated output (Cognitive Load construct mean of 3.47). The low reliability of the Cognitive Load construct ($\alpha = 0.55$) suggests that this dimension may require refinement in future measurements. Overall, the findings indicate that LuisterLinie shifts the nature of work from manual documentation toward quality assurance, rather than eliminating the task entirely.

7.1.4 RQ4: Objective metrics and subjective perceptions

The fourth research question examined how objective behavioral metrics relate to subjective user perceptions. The log data analysis revealed that the user base grew from 125 to a peak of 465 weekly active users between April and July, that 27.1% of recordings transitioned to chat-based follow-up, and that only 1.49% of summaries required manual editing. These metrics converge with the positive survey findings: sustained adoption corroborates perceived efficiency gains, the low edit rate supports perceived quality, and the limited chat interaction depth aligns with moderate cognitive load scores. Divergences were also observed; the 1,180 recordings under two minutes suggest usability issues not captured by the survey, highlighting the complementary value of both data sources.

7.2 Contributions

This thesis makes several contributions to the understanding of Generative AI in non-profit organizational contexts.

First, it provides an empirical evaluation of a deployed GenAI tool within the Dutch Salvation Army, an organizational setting that has received limited attention in the existing literature. The study demonstrates that GenAI transcription and summarization tools can be effectively implemented in a mission-driven, volunteer-based environment where productivity is not measured in monetary value.

Second, the survey instrument developed for this study provides a quantitative framework for measuring perceived productivity, knowledge management, workload, and usability in AI-assisted documentation

work. The instrument demonstrated acceptable reliability for the Efficiency, Quality, and Workload constructs, offering a foundation for future evaluations in similar contexts.

Third, the study demonstrates methodological triangulation by combining subjective survey data with objective application log data. The log analysis revealed that the user base grew from 125 to a peak of 465 weekly active users between April and July, that the AI-generated summary edit rate was only 1.49%, and that 27.1% of all recordings transitioned to chat-based follow-up. These behavioral metrics corroborate the positive survey findings and provide converging evidence of the tool's effectiveness.

Finally, the study documents the process from the exploratory phase through evaluation, offering practical insights for organizations seeking to implement GenAI tools in knowledge-intensive work environments.

7.3 Future work

Future research could continue validating the proposed measurement instrument using a larger and more diverse sample to enable confirmatory factor analysis and structural equation modeling.

Longitudinal studies could investigate how user perceptions and behavioral metrics evolve over time as the organization gains more experience with GenAI tools. The observed growth from 125 to 465 users suggests that adoption is still in an early phase, and follow-up measurements could reveal whether the positive trends are sustained.

Further integration of survey and log data at the individual level, should anonymization constraints allow, would enable more powerful statistical analyses linking perceived effects to actual usage behavior. Additionally, deeper analysis of the LangFuse trace data could provide insights into how prompt design, response latency, and error rates relate to user satisfaction and output quality.

Finally, comparative studies across multiple non-profit organizations would help determine whether the findings reported here generalize to other mission-driven contexts with similar constraints and objectives.

Bibliography

[1] Maryam Alavi and Dorothy E. Leidner. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 25(1):107–136, 2001.

[2] Saleema Amershi, Daniel Weld, Mihaela Vorvoreanu, et al. Guidelines for human–ai interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2019.

[3] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

[4] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[5] Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. Generative ai at work. *The Quarterly Journal of Economics*, 2023. Working paper version prior to journal publication.

[6] Longbing Cao et al. A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. *arXiv preprint arXiv:2303.04226*, 2023.

[7] Thomas H. Davenport and Laurence Prusak. *Working Knowledge: How Organizations Manage What They Know*. Harvard Business School Press, 1998.

[8] Peter F. Drucker. *Management Challenges for the 21st Century*. HarperBusiness, 1999.

[9] Yogesh K. Dwivedi et al. So what if chatgpt wrote it? multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy. *International Journal of Information Management*, 71:102642, 2023.

[10] Xueyuan Gao and Hua Feng. Ai-driven productivity gains: Artificial intelligence and firm productivity. *Sustainability*, 15(11):8934, 2023.

[11] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in Psychology*, 52:139–183, 1988.

[12] Mohammad Hossein Jarrahi. Artificial intelligence and the future of work: Human-ai symbiosis in organizational decision making. *Business Horizons*, 61(4):577–586, 2018.

[13] Belayneh Yitayew Kassa and Eyob Ketema Worku. The impact of artificial intelligence on organizational performance: The mediating role of employee productivity. *Journal of open innovation: technology, market, and complexity*, page 100474, 2025.

[14] Reza Marvi, Pantea Foroudi, and Maria Teresa Cuomo. Past, present and future of ai in marketing and knowledge management. *Journal of Knowledge Management*, 29(11):1–31, 2025.

[15] Ikujiro Nonaka and Hirotaka Takeuchi. *The Knowledge-Creating Company*. Oxford University Press, 1995.

[16] Shakked Noy and Whitney Zhang. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, 381(6654):187–192, 2023.

[17] Daniel E O'Leary. Using ai in knowledge management: Knowledge bases and ontologies. *IEEE Intelligent Systems and Their Applications*, 13(3):34–39, 2002.

[18] OpenAI. Chatgpt. `https://chat.openai.com/chat`, March 2023. URL `https://chat.openai.com/chat`. Mar 14 version.

[19] Raja Parasuraman, Thomas B Sheridan, and Christopher D Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, 30(3):286–297, 2000.

[20] Sida Peng et al. The impact of ai on developer productivity: Evidence from github copilot. *arXiv preprint arXiv:2302.06590*, 2023.

[21] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, 4 edition, 2021.

[22] James P. Walsh and Gerardo R. Ungson. Organizational memory. *Academy of Management Review*, 16(1):57–91, 1991.

[23] Robert K Yin. *Case study research: Design and methods*. Sage Publications, 5th edition, 2014.

[24] Huan Zhang et al. Large language models in healthcare: Applications, opportunities, and challenges. *NPJ Digital Medicine*, 2024. Review article.

Appendix A

Exploratory Interview

A.1 Exploratory Interviews with Developers

The interviews focus on understanding the current state, developments, and priorities of the Generative AI tools within the organization. Special attention is given to differences between the more mature Luisterlinie tool and the newer ChatAI system, as well as identifying relevant focus points for subsequent end-user research.

Note: Interviews are conducted anonymously, recorded (with consent), and take approximately one hour. The focus is on the current situation and recent developments.

A.1.1 Context

- What is your role within the Leger des Heils (LdH)?
- What is your role in the development of Luisterlinie and/or ChatAI?
- How long have you been involved with these tools?

A.1.2 Development and Current Use

- What were the original problems or needs that led to the development and scaling of GenAI tools within LdH?
- How are Luisterlinie and ChatAI currently used in practice?
- What are the core functionalities of each tool?
- How would you ideally like employees to use these tools?

A.1.3 Changes Over Time

- What has changed in the use of these tools compared to last year?
- What improvements or developments have been made since earlier versions or pilots?
- Have new issues or challenges emerged after these changes?

A.1.4 Luisterlinie (Mature Tool – Scaling Phase)

- How has the use of Luisterlinie evolved since the previous pilot(s)?
- What lessons were learned from earlier implementations?
- What challenges arise when scaling the use of Luisterlinie across the organization?
- Are there differences in usage between teams or contexts?

A.1.5 ChatAI (New Tool – Early Stage)

- What is the main purpose of ChatAI within the organization?
- How is ChatAI currently being introduced and adopted?
- What early experiences or feedback have you received?
- What lessons learned from Luisterlinie are being applied to ChatAI?

A.1.6 Strengths

- What are the main strengths of the tools?
- In which areas do the tools provide the most value to users?
- What improvements in work processes or outcomes have you observed since implementation?

A.1.7 Limitations and Challenges

- What technical limitations currently exist (e.g., accuracy, integration, reliability)?
- What organizational barriers affect the development or effective use of the tools?
- What factors or conditions stimulate or enable successful use within the organization?

A.1.8 Data & Feedback

- What kind of feedback do you currently receive from users?
- What usage data or logs are collected?
- How do you currently evaluate the performance or success of the tools?
- Which metrics are most important to you (e.g., time savings, accuracy, user satisfaction)?

A.1.9 Future Improvements

- What are the main priorities for further development?
- What key problems do you still aim to solve with these tools?

A.1.10 End-User Research Focus

- What type of user feedback is most valuable for further development?
- Which aspects of use would you like to better understand (e.g., productivity, workload, usability, output quality)?
- What would define "success" of these tools from your perspective?
- Are there specific questions or hypotheses you would like this research to address?

A.1.11 Closing

- How can this research best support the development of the tools?
- What insights or outcomes do you expect from this study?
- Is there anything we have not discussed that you consider important?

A.2 Dutch Version used for interviews

NB: De interviews worden anoniem afgenomen, (met toestemming) opgenomen en duren ongeveer één uur. De focus ligt op de huidige situatie en recente ontwikkelingen.

A.2.1 Context

- Wat is uw rol binnen het Leger des Heils (LdH)?
- Wat is uw rol in de ontwikkeling van Luisterlinie en/of ChatAI?
- Hoe lang bent u al betrokken bij deze tools?

A.2.2 Ontstaan en huidig gebruik

- Wat waren de oorspronkelijke problemen of behoeften die hebben geleid tot de ontwikkeling en opschaling van GenAI-tools binnen LdH?
- Hoe worden Luisterlinie en ChatAI momenteel gebruikt in de praktijk?
- Wat zijn de belangrijkste functionaliteiten van beide tools?
- Hoe zou u idealiter willen dat medewerkers deze tools gebruiken?

A.2.3 Veranderingen in de tijd

- Wat is er veranderd in het gebruik van deze tools ten opzichte van vorig jaar?
- Welke verbeteringen of ontwikkelingen zijn er doorgevoerd sinds eerdere versies of pilots?
- Zijn er nieuwe problemen of uitdagingen ontstaan na deze veranderingen?

A.2.4 Luisterlinie (volwassen tool – opschalingsfase)

- Hoe heeft het gebruik van Luisterlinie zich ontwikkeld sinds eerdere pilots?
- Welke lessen zijn geleerd uit eerdere implementaties?
- Welke uitdagingen komen naar voren bij het opschalen van Luisterlinie binnen de organisatie?
- Ziet u verschillen in gebruik tussen teams of contexten?

A.2.5 ChatAI (nieuwe tool – vroege fase)

- Wat is het belangrijkste doel van ChatAI binnen de organisatie?
- Hoe wordt ChatAI momenteel geïntroduceerd en gebruikt?
- Welke eerste ervaringen of feedback zijn er ontvangen?
- Welke lessen uit Luisterlinie worden toegepast bij ChatAI?

A.2.6 Sterktes

- Wat zijn de belangrijkste sterke punten van de tools?
- Op welke gebieden leveren de tools de meeste waarde voor gebruikers?
- Welke verbeteringen in werkprocessen of resultaten heeft u waargenomen sinds de implementatie?

A.2.7 Beperkingen en uitdagingen

- Welke technische beperkingen bestaan er momenteel (bijv. nauwkeurigheid, integratie, betrouwbaarheid)?
- Welke organisatorische barrières beïnvloeden de ontwikkeling of het gebruik van de tools?
- Welke factoren of omstandigheden stimuleren juist het gebruik van de tools?

A.2.8 Data & feedback

- Welke feedback ontvangt u momenteel van gebruikers?
- Welke gebruiksdata of loggegevens worden verzameld?
- Hoe wordt momenteel het succes of de prestaties van de tools geëvalueerd?
- Welke metrics zijn voor u het belangrijkst (bijv. tijdsbesparing, nauwkeurigheid, gebruikerstevredenheid)?

A.2.9 Toekomstige ontwikkeling

- Wat zijn de belangrijkste prioriteiten voor verdere ontwikkeling?
- Welke problemen wilt u in de toekomst nog oplossen met deze tools?

A.2.10 Focus voor gebruikersonderzoek

- Welke gebruikersfeedback is voor u het meest waardevol?
- Over welke aspecten van gebruik wilt u meer inzicht krijgen (bijv. productiviteit, werkdruk, gebruiksgemak, kwaliteit van output)?
- Wat betekent "succes" van deze tools volgens u?
- Zijn er specifieke vragen of hypotheses die u graag beantwoord ziet in dit onderzoek?

A.2.11 Afsluiting

- Hoe kan dit onderzoek het beste bijdragen aan de verdere ontwikkeling van de tools?
- Welke inzichten of resultaten verwacht u van dit onderzoek?
- Is er nog iets dat niet aan bod is gekomen maar wel belangrijk is om te benoemen?

Appendix B

End-User Survey

B.1 Survey Instrument: End-User Evaluation of Luisterlinie

This survey aims to evaluate the impact of the GenAI tool *Luisterlinie* on administrative work, productivity, workload, and user experience among employees of the Salvation Army.

Unless stated otherwise, all statements are measured on a 5-point Likert scale:

1 = Strongly disagree
2 = Disagree
3 = Neutral / Don't know
4 = Agree
5 = Strongly agree

B.1.1 Background

1. What is your role within the Salvation Army?

---

2. How often do you record a conversation with Luisterlinie?
   (a) Less than once per week
   (b) 1–3 times per week
   (c) 3–5 times per week
   (d) More than once per day

B.1.2 Usage of Luisterlinie

Please indicate to what extent you agree with the following statements.

1. Luisterlinie reduces the time I spend on documentation.
2. Luisterlinie speeds up my administration.
3. Luisterlinie increases my efficiency.
4. Luisterlinie reduces repetitive work.
5. Luisterlinie improves the quality of documentation.
6. Luisterlinie improves the structure of my work.
7. Luisterlinie adds more detail to documentation.
8. Luisterlinie reduces my workload.
9. Luisterlinie reduces my stress.

B.1.3 Experienced Drawbacks

Please indicate to what extent you agree with the following statements.

1. Luisterlinie causes me to spend a lot of time improving AI output.
2. Luisterlinie causes me to think more during documentation.
3. Luisterlinie causes me to experience additional pressure when working with AI tools.
4. Luisterlinie causes me to be uncertain about my privacy.

B.1.4 User Experience

Please indicate to what extent you agree with the following statements.

1. Luisterlinie gives me more focus during conversations.
2. Luisterlinie gives me more neutral summaries.

3. Luisterlinie is easy to use.
4. Luisterlinie feels trustworthy.

B.1.5 Open Questions

1. What do you find most positive about Luisterlinie?

---

2. What do you find most negative about Luisterlinie?

---

3. What improvements would you suggest for Luisterlinie?

B.2 Vragenlijst: Evaluatie van Luisterlinie

Deze vragenlijst heeft als doel de impact van de GenAI-tool *Luisterlinie* op administratieve werkzaamheden, productiviteit, werkdruk en gebruikerservaring binnen het Leger des Heils te evalueren.

Tenzij anders vermeld, worden alle stellingen beoordeeld op een 5-punts Likertschaal:

1 = Zeer mee oneens
2 = Oneens
3 = Neutraal / Geen mening
4 = Eens
5 = Zeer mee eens

B.2.1 Achtergrondgegevens

1. Wat is jouw functie binnen het Leger des Heils?

---

2. Hoe vaak neem je een gesprek op met Luisterlinie?
   (a) Minder dan één keer per week
   (b) 1–3 keer per week
   (c) 3–5 keer per week
   (d) Meer dan één keer per dag

B.2.2 Het gebruik van Luisterlinie

Geef aan in hoeverre je het eens bent met de volgende stellingen.

1. Luisterlinie vermindert de tijd die ik besteed aan documentatie.
2. Luisterlinie versnelt mijn administratie.
3. Luisterlinie verhoogt mijn efficiëntie.
4. Luisterlinie vermindert repetitief werk.
5. Luisterlinie verbetert de kwaliteit van administratie.
6. Luisterlinie verbetert de structuur van mijn werk.
7. Luisterlinie voegt meer details toe aan de documentatie.
8. Luisterlinie vermindert de werkdruk.
9. Luisterlinie vermindert mijn stress.

B.2.3 Ervaren nadelen

Geef aan in hoeverre je het eens bent met de volgende stellingen.

1. Luisterlinie zorgt ervoor dat ik veel tijd kwijt ben aan het verbeteren van de AI-output.
2. Luisterlinie zorgt ervoor dat ik meer moet nadenken tijdens de documentatie.
3. Luisterlinie zorgt ervoor dat ik extra druk ervaar om met AI-tools om te gaan.
4. Luisterlinie zorgt ervoor dat ik onzeker ben over mijn privacy.

B.2.4 Gebruikerservaring

Geef aan in hoeverre je het eens bent met de volgende stellingen.

1. Luisterlinie geeft mij meer focus tijdens gesprekken.
2. Luisterlinie geeft mij neutralere samenvattingen.

3. Luisterlinie geeft mij gemak in gebruik.
4. Luisterlinie geeft mij een betrouwbaar gevoel.

B.2.5 Open vragen

1. Wat vind je het meest positief aan Luisterlinie?

---

2. Wat vind je het meest negatief aan Luisterlinie?

---

3. Wat voor verbeteringen zou je voorstellen voor Luisterlinie?
