Shape the Future of Swedish Healthcare with AI-Technology

How to Implement Large Language Models as a Tool to Streamline Clinicians' Administrative Tasks

Master's Thesis in Management and Economics of Innovation, MSc

AMANDA NACKOVSKA
ELIN BERTHAG

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF INNOVATION AND R&D MANAGEMENT

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024
www.chalmers.se

Shape the Future of Swedish Healthcare with AI-Technology

How to Implement Large Language Models as a Tool to Streamline Clinicians' Administrative Tasks

AMANDA NACKOVSKA
ELIN BERTHAG

Department of Technology Management and Economics
Division of Innovation and R&D Management
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2024

Shape the Future of Swedish Healthcare with AI-Technology
How to Implement Large Language Models as a Tool to Streamline Clinicians' Administrative Tasks
AMANDA NACKOVSKA
ELIN BERTHAG

Department of Technology Management and Economics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Gothenburg, Sweden 2024

Shape the Future of Swedish Healthcare with AI-Technology
How to Implement Large Language Models as a Tool to Streamline Clinicians' Administrative Tasks

AMANDA NACKOVSKA
ELIN BERTHAG

Department of Technology Management and Economics
Chalmers University of Technology

ABSTRACT
The demand for healthcare is continuously increasing in line with an ageing population. It is a recognized problem that clinicians have a significantly high administrative workload as a consequence of digitalization, which takes time away from valuable direct patient care. Clinicians perform multiple text-based administrative tasks, and it can be argued that large language models (LLMs) have potential to streamline these tasks. In recent years, LLMs have received great attention, led by the public introduction of ChatGPT by OpenAI in November 2022. Thus, the purpose of this study is to explore how LLMs can be used to relieve the administrative text-based workload for clinicians at Sahlgrenska University Hospital. The study is delimited to look at patient-related administrative text-based tasks performed by physicians and nurses at the neurology, ophthalmology and radiology department at Sahlgrenska University Hospital. This study is based on a strong empirical surface built upon a wide data collection of 46 semi-structured interviews where 37 have been conducted with healthcare professionals at the three mentioned departments and 9 have been conducted with 10 experts within the field of AI in healthcare. In addition, data have been collected by distributing self-completion forms at the hospital to measure the time clinicians spend on certain administrative text-based tasks, and through field notes from a number of observations at the hospital. The obtained data has been analyzed through thematic analysis.

The result of the study identifies that there is a vast potential to use LLMs to streamline patient-related administrative text-based tasks in healthcare. However, there are boundaries that need to be addressed. Technological concerns have been identified due to the novelty of the technology. Ethical concerns have been identified, mainly the risk that LLMs generate biased and incorrect information, and that the information can not be validated. The three practical cases of this study clearly show that there is a need to streamline the clinicians' patient-related administrative text-based tasks. While it can be concluded that there is potential to use LLMs for this, it should be noted that it has to be further researched in a practical setting. This research further concludes that there are clear differences in the clinicians' needs across the three different departments, which adds complexity to the process of prioritizing use cases to put into practice at the hospital.

Keywords: large language models, artificial intelligence, healthcare, administration, text-based tasks, patient-related.

Acknowledgements

This master's thesis was conducted during the spring semester of 2024 as part of our studies in Management and Economics of Innovation at Chalmers University of Technology. The research study was performed in collaboration with several key people to whom we would like to express our gratitude.

First, we would like to extend our heartfelt thanks to Andreas Hellström, our supervisor and examiner. Andreas has been an incredible source of guidance and support throughout this project. His positive attitude and willingness to share his expertise has been invaluable. We truly appreciate all his help and encouragement along the way.

Second, we would like to extend our gratitude to our clients at Sahlgrenska University Hospital: Erik Thurin, Maria Hansen and Robin Molander. Their leadership in initiating this project, coupled with their ongoing sharing of expertise, engagement, and passion for improving Swedish healthcare has been integral to the success of this project. Without you, this project would not have been possible.

Third, we would like to extend our gratitude to our collaborating master's thesis project partners, Albert Lund and Felix Nilsson, as well as the talented team at Kompetenscentrum AI at Sahlgrenska University Hospital, led by Isak Barbopoulos. We deeply appreciate the invaluable teamwork and knowledge sharing between our two groups, which has been incredibly beneficial to our project's progress and results.

Finally, we would like to sincerely thank the research participants, both all participating healthcare professionals from Sahlgrenska University Hospitals as well as all participating experts within the field of AI in healthcare. Without the contribution of each and every one of you, this project would not have been possible.

Amanda Nackovska

Elin Berthag

Gothenburg, June 2024

List of Acronyms

Below is the list of acronyms that have been used throughout this thesis listed in alphabetical order:

AI	Artificial Intelligence
EHR	Electronic Health Records
IT	Information Technology
LLMs	Large Language Models
SU	Sahlgrenska University Hospital

List of Concepts

Below is the list of concepts that have been used throughout this thesis listed in thematic order:

Medical fields	
Neurology	Focused on diagnosing and treating disorders of the nervous system, including the brain, spinal cord, and peripheral nerves.
Ophthalmology	Dedicated to the diagnosis, treatment, and prevention of eye and vision disorders.
Radiology	Specializing in imaging techniques to diagnose and treat diseases.

Healthcare task types	
Non-patient-related	Tasks that do not involve direct patient interaction, such as support services.
Patient-related	Tasks that involve direct interaction with and care for patients.

Patient care types	
In-patient care	Medical care provided to patients admitted to a hospital or clinic for an overnight stay or longer.
Out-patient care	Medical care provided to patients who visit a healthcare facility for diagnosis or treatment without staying overnight.

Professional healthcare titles	
Clinician	A general term for a healthcare professional involved in patient care.
Medical secretary	Supports medical staff by managing appointments, maintaining patient records, and performing other administrative duties.
Nurse	Provides patient care, administers medications, performs medical procedures, and educates patients and their families.
Nursing assistant	Support nurses by assisting with basic patient care tasks such as bathing, dressing, and feeding patients, and measuring vital signs.
Physician	Have completed medical school and residency training. Diagnoses and treats illnesses, prescribes medications, and may perform surgeries.
Physiotherapist	Helps patients improve their physical movement and manage pain through exercise, manual therapy, and other treatments.
Resident	A physician in training who has completed medical school and is undergoing specialized training in a particular field under the supervision of senior physicians.
Senior specialist	A physician who has completed advanced training and has extensive expertise in a specific medical field, often overseeing patient care and guiding junior physicians.
Surgeon	A physician who specializes in performing surgical procedures to treat diseases, injuries, or deformities, requiring extensive training in specific surgical techniques.

Different types of notes physicians write in the health records	
Out-patient care	
Clinicial note/visit note	Documentation of new or follow-up visits, including symptoms, examinations and treatments.
In-patient care	
Admission note	Summary of the patient's condition upon admission to the hospital, including medical history, examinations, and planned care.
Daily progress note	Daily documentation of the patient's condition, treatments, and progress during their hospital stay.
Discharge letter	Summary of a patient's care and treatment during their hospital stay, providing instructions for continued care and recovery following discharge.
Discharge summary	Summary of the patient's hospital stay, including diagnosis, treatments, and recommendations for continued care after discharge.
Ward round notes	Documentation of the patient's condition, treatment plan, and significant observations made each day of a patient's hospital stay.
Other notes	
Consultation note	Documentation of a specialist's assessment and recommendations after consulting another healthcare provider regarding a patient.
Operative report	Detailed description of a surgical procedure, including the intervention, any complications, and postoperative care.
Referral	Document sent from one healthcare provider to another for the transfer of a patient to a specialist or for further investigation and care.

Different types of notes nurses write in the health records	
Out-patient care	
Clinicial notes/visit notes	Documentation of new or follow-up visits, including symptoms, examinations and treatments.
In-patient care	
Admission note	Summary of the patient's condition upon admission to the hospital, including medical history, examinations, and planned care.
Discharge planning	Coordination of the patient's transition from the hospital to home or another care setting, including arrangements for follow-up care, medication management, and patient education.
Nursing assessment	Documentation of a patient's current health status, including vital signs, physical assessment findings, and any nursing interventions required.
Nursing discharge note	Documentation of the patient's condition at the time of discharge, including final assessments, instructions for ongoing care, and any referrals or follow-up appointments required.
Risk assessment note	Evaluation of potential risks to the patient's health and safety, including assessment of fall risk, pressure ulcer risk, and risk of developing complications.
Ward round notes	Documentation of the patient's assessments, interventions, and responses to treatment made each day of a patient's hospital stay.
Other notes	
Postoperative care note	Documentation of the patient's recovery and ongoing care following surgery, including assessments of wound healing, pain management, and monitoring for potential complications.
Referral for cancer rehabilitation	Request to another healthcare provider or specialist for a patient to receive specialized cancer rehabilitation services.
Referral for sutures	A request to another healthcare provider or specialist to have a wound closed with stitches (sutures).

Contents

List of Acronyms **ix**

List of Concepts **xi**

List of Figures **xix**

List of Tables **xxi**

1 Introduction **1**
1.1 Background . . . 2
1.2 Purpose . . . 3
1.3 Research Questions . . . 3
1.4 Delimitations . . . 3

2 Theoretical Framework **4**
2.1 The Development of Artificial Intelligence and its Role in the Healthcare Sector . . . 5
2.2 Large Language Models . . . 6
2.2.1 The Technology Behind LLMs . . . 7
2.2.2 Training LLMs . . . 8
2.2.2.1 Fine Tuning . . . 8
2.2.2.2 Retrieval-Augmented Generation . . . 9
2.2.2.3 Scaling Laws . . . 9
2.2.3 Challenges of LLMs . . . 10
2.2.4 The Future of LLMs . . . 11
2.3 Ambient AI Scribes . . . 12
2.4 LLMs in Healthcare . . . 12
2.5 Ethics and Legislation of LLMs in Healthcare . . . 13
2.5.1 Medical Device . . . 16
2.6 Healthcare in Sweden . . . 16
2.6.1 Definition of Administrative versus Clinical Work in Healthcare Settings . . . 17
2.6.2 Previous Studies on Time Utilization in Healthcare Settings . 19
2.6.3 Perspectives on the Administrative Burden . . . 20
2.6.4 Illegitimate Tasks: Conceptualization and Implications for Employees in the Healthcare Sector . . . 20

2.6.4.1 Illegitimate Tasks Impact on Well-Being 22

3 Methods 23
3.1 Research Process 24
3.2 Data Collection 25
3.2.1 Literature Review 25
3.2.2 Qualitative Interview 26
3.2.3 Sampling and Practicalities 26
3.2.4 List of Interviewees 28
3.2.5 Diary as a Self-Completion Questionnaire 29
3.2.5.1 Sampling for the Diary as a Self-Completion Questionnaire 30
3.2.6 Observations for Understanding and Learning 30
3.3 Data Analysis 30
3.4 Research Quality 31
3.4.1 Ethical Considerations 32

4 Results 34
4.1 Expert Perspectives Within the Area of AI in Healthcare 35
4.1.1 Potential LLM Applications in Healthcare 35
4.1.2 Exploring the Potential of LLMs at Region Halland 36
4.1.3 Risks, Benefits and Future Directions 37
4.1.4 The Challenges of Validating LLMs 40
4.1.5 Ethics and Legislation 41
4.1.5.1 Machines versus Clinicians 41
4.1.5.2 Who Takes Responsibility for AI Generated Outputs? 42
4.1.5.3 Regulations 44
4.1.6 Implementation 46
4.1.6.1 A Successful Case of Implementing AI Tools in Healthcare in Canada 46
4.1.6.2 Key Factors to Implement New Technology in Healthcare 48
4.2 The Neurology Department 50
4.2.1 The Needs of Physicians at the Neurology Department 51
4.2.1.1 Everyday Text Based Tasks Performed by Neurologists 51
4.2.1.2 The Perspective of the Neurology Physicians on Their Text Based Tasks 51
4.2.1.2.1 Time-Consuming and Troublesome Tasks . . 51
4.2.1.2.2 Tools and Characteristics 55
4.2.1.3 Time Estimates for a Number of Text Based Tasks . 59
4.2.2 Self-Completion Form at the Neurology Department 60
4.2.3 The Needs of Nurses at the Neurology Department 62
4.2.3.1 Everyday Text Based Tasks Performed by Neurology Nurses 62
4.2.3.2 The Perspective of the Neurology Nurses on Their Text Based Tasks 62
4.2.3.2.1 Time-Consuming and Troublesome Tasks . . 62

4.2.3.2.2 Tools and Characteristics 67
4.2.3.3 Time Estimates for a Number of Text Based Tasks . 68
4.3 The Ophthalmology Department 69
4.3.1 The Role of Physicians at the Ophthalmology Department . . 69
4.3.2 The Physicians' View on the Administrative Burden at the Ophthalmology Department 70
4.3.3 Task Shifting with Speech Recognition 72
4.3.4 The Ophthalmology Physicians' Perspectives on Patient-Related Administration 73
4.3.5 The Ophthalmology Physicians' Perspectives on Their Computer Systems 75
4.4 The Radiology Department 76
4.4.1 The Role of Nurses and Physicians at the Radiology Department 76
4.4.2 Administrative Tasks at the Radiology Department 77
4.4.3 Deep-Dive into Administration Related to Referrals 77
4.4.4 Suggestions to Make the Referral Management More Efficient 80
4.4.5 Other Administrative Tasks and Bottlenecks at the Radiology Department 80
4.4.6 Time Spent on Administrative Tasks 81

5 Discussion **83**
5.1 The Potential and Risks of LLMs in Healthcare 84
5.2 The Cases of the Three Departments 85
5.2.1 The Case of the Neurology Department 85
5.2.2 The Case of the Ophthalmology Department 89
5.2.3 The Case of the Radiology Department 92
5.2.4 Comparison Between the Three Departments 95
5.3 The Impact of Administrative Work on Clinicians' Health 96
5.4 The Ethical and Legal Perspective of LLMs in Healthcare 97
5.5 How to Implement LLMs in Healthcare 99
5.5.1 What to Learn from the Case of Unity Health Toronto 100
5.5.2 What to Think About When Implementing New Technology in Healthcare 100

6 Conclusion **102**

7 Recommendations **104**

8 Future Research **105**

Bibliography **106**

A Appendix 1 **I**
A.1 Interview Guide for Clinician Interviews I
A.2 Interview Guide for Expert Interviews II
A.3 Self-Completion Forms III

List of Figures

3.1 Data collection, sampling and data analysis methods. 25

5.1 Visualization of the different needs across departments. 96

A.1 Self-completion form for physicians at the neurology department, second update. IV
A.2 Self-completion form for physicians at the neurology department, first update. V
A.3 Self-completion form for nurses at the neurology department. VI
A.4 Self-completion form for physicians at the ophthalmology department. VII
A.5 Self-completion form for physicians at the radiology department. . . . VIII

List of Tables

2.1 Examples of direct patient work, patient-related administration, and non-patient-related administration. 19

3.1 List of clinician interviews. 28
3.2 List of expert interviews. 29

4.1 Estimated time physicians spend on administrative work in average on a normal day of clinical work at the neurology department. 52
4.2 Time estimates by neurology physicians for reading EHR and different types of writing tasks. 59
4.3 Data from the self-completion form at the neurology department, occasion 1. 60
4.4 Data from the self-completion form at the neurology department, occasion 2. 61
4.5 Mean values of self-completion form data. 61
4.6 Estimated time nurses spend on administrative work in average on a normal day of clinical work at the neurology department. 63
4.7 Time estimates by neurology nurses for reading EHR. 68
4.8 Time estimates by neurology nurses for different writing tasks. 69
4.9 Estimated time physicians spend on administrative work in average on a normal day of clinical work at the ophthalmology department. . 71
4.10 Estimated time physicians spend on administrative work in average on a normal day of clinical work at the radiology department. 82

5.1 A selection of administrative tasks performed by radiologist and radiology nurses. 93

1

Introduction

This chapter provides a background for the studied issue of this paper. It further presents the purpose of this study and two research questions, as well as the delimitations of this study.

1.1 Background

Numerous studies have shown that Swedish clinicians spend a significant portion of their time on administrative tasks (Anskär, 2019; Borelius, 2009; McKinsey & Company, 2019), which detracts from the time available for essential direct patient care (Barkman & Aasa, 2023). When it comes to the patient-related administration, such as reading electronic health records, pre-patient meetings, and writing medical notes such as referrals, discharge letters, and admission letters, a study from 2019 showed that physicians and nurses spend 32.3 % respectively 27.2 % on such tasks. Meanwhile 34.4 % respectively 42.2 % of the clinicians' time were spent on direct patient care (Anskär, 2019). In addition to taking time from critical clinical work, studies have shown that unnecessary and unreasonable administrative tasks are positively correlated with burnout and negatively impact job satisfaction, which can subsequently affect the quality of healthcare (Semmer et al., 2019; Werdecker & Esch, 2021). Furthermore, the Swedish healthcare system is projected to experience increased demands due to an aging population and a shrinking recruitment base (Ahlstedt et al., 2023; Barkman & Aasa, 2023; Grant Thornton, 2023).

Given these challenges, there is a critical need to explore solutions that can alleviate the administrative burden on clinicians. Implementing AI-technologies, particularly large language models, offers a promising avenue to streamline administrative tasks and enhance operational efficiency in healthcare settings (**Ilicki**; DiGiorgio & Ehrenfeld, 2023; Rosenberg et al., 2024; Zhang & Kamel Boulos, 2023; Zheng et al., 2023). However, the potential is also hindered by certain risks and barriers that need to be acknowledged and addressed, especially given the sensitive nature of healthcare (Haltaufderheide & Ranisch, 2023; Harrer, 2023). Moreover, successful deployment and implementation of AI-solutions in clinical settings require coordinated planning and collaboration among healthcare professionals, IT experts, researchers, data scientists, health service administrators, and policymakers (Dhalla et al., 2023).

Sahlgrenska University Hospital is Sweden's largest hospital, where over 17 000 coworkers take care of 350 000 patient every year (Sahlgrenska Universitetssjukhuset, 2023b). With the goal of becoming Europe's leading hospital by 2032, SU has created a strategy that partly emphasizes advancements in medical developments, new technologies, and digitization. Moreover, in 2021, SU launched Kompetenscentrum AI to support the development and integration of AI in clinical practices (Sahlgrenska Universitetssjukhuset, 2023a).

In collaboration with Sahlgrenska University Hospital, Kompetenscentrum AI and Chalmers University of Technology, this paper is part of a larger project that focuses on how LLMs can be used to make the patient-related administration at Sahlgrenska more efficient. This paper focuses on identifying and addressing challenges related to the administration of physicians and nurses at three different departments: neurology, ophthalmology, and radiology.

1.2 Purpose

The purpose of this study is to explore how LLMs can be used to relieve the administrative text-based workload for clinicians at Sahlgrenska University Hospital.

1.3 Research Questions

The purpose of this study is broken down into two research questions presented below:

1. How can LLMs be used as a tool for administrative tasks in healthcare?

2. What are the boundaries to use LLMs on the administrative side of healthcare?

1.4 Delimitations

This research examines the healthcare at Sahlgrenska University Hospital in Gothenburg, Sweden. Due to available resources, data collection is limited to three different departments: neurology, ophthalmology and radiology. The study is delimited to only examine the patient-related administrative text-based work of physicians and nurses at the three mentioned departments.

2 Theoretical Framework

This chapter delves into the research perspective of large language models in healthcare. The chapter is divided into multiple sections. First, it describes the development of artificial intelligence and its role in the healthcare sector. Second, it gives a comprehensive understanding of the LLM technology followed by a description of ambient scribes. Third, research on LLMs in healthcare is presented. Fourth, the ethical and legal research perspective on LLMs is described. Lastly, the topic of healthcare in Sweden and the definition of administrative work in a clinical setting is presented.

2.1 The Development of Artificial Intelligence and its Role in the Healthcare Sector

AI is a rapidly evolving domain within computer science (Alowais et al., 2023). It refers to the development of computer systems capable of performing tasks that typically require human intelligence like reasoning, perception and decision making (Dave & Patel, 2023). The field of AI incorporates techniques such as machine learning (ML) and deep learning (DL) (Alowais et al., 2023). While AI is a broad field that includes techniques to make machines intelligent, ML is a subset of AI that involves systems which can autonomously learn. DL is further nestled within ML, utilizing deep neural network models to discern patterns with minimal human intervention.

The domain of AI has undergone significant changes over the years (Alowais et al., 2023). The term “Artificial Intelligence” was first introduced during the Dartmouth Conference organized by John McCarthy in 1956. Following this event, AI research focused on expert and rule-based systems, an approach that was limited by the need for more data and computing power. Medicine was identified as one of the most promising areas to use AI in the mid-twentieth century (Yu et al., 2018). In the 1970s, researchers developed and proposed many clinical decision support systems and rule-based approaches were developed successfully. However, rule-based systems are costly to build and require explicit expressions of decision rules, comprehensiveness of prior medical knowledge and human-authored updates. These systems are also limited by their difficulty to encode higher-order interactions between different pieces of knowledge that have been authored by different experts.

It was not until the 1980s and 1990s that AI research pivoted towards machine learning (ML) and neural networks (Alowais et al., 2023). This enabled machines to learn from data and improve their performance gradually. The shift carved a path for new system developments, like IBM’s Deep Blue that defeated the world chess champion, Garry Kasparov, in 1997. Unlike the early rule-based AI system, new AI systems that leveraged machine learning methods made it possible for AI systems to discover patterns without the need to specify decision rules for each specific task or to account for complex interactions among input features (Yu et al., 2018). These achievements have been driven mainly by successful application of deep learning.

AI research continued to evolve in the 2000s (Alowais et al., 2023). During that time, the research was focusing on computer vision and natural language processing (NLP). NLP is a subset of AI dedicated to train machines to interpret, comprehend and generate natural human language. This focus resulted in the the development of virtual assistants capable of comprehending and reacting to natural language requests, like Apple’s Siri.

AI continues to evolve still, and today it is transforming large sectors in society like healthcare, transportation and finance, among other fields (Alowais et al., 2023).

Most recently, LLMs have made significant advancements in NLP (Kasneci et al., 2023). LLMs are a type of generative AI, which means that it can generate new content (Fruhlinger, 2023). LLMs are trained on extensive text datasets and can perform language-related tasks, such as generating human-like text and accurately answering questions. (Kasneci et al., 2023). Utilization of transformer architectures that use a self-attention mechanism to determine the relevance of different aspects of inputs when generating predictions and pre-training on large datasets combined with fine-tuning on specific tasks have been key to the most recent successes of LLMs. In 2022, the impact of generative AI exploded into the public consciousness (Bergmann, 2024). The LLM-chatbot ChatGPT by OpenAI was a major source of this breakthrough (Sætra, 2023). The technology was made openly and freely available, having reached a level of maturity that made it immediately accessible and useful for multiple users. In 2023 generative AI became gradually incorporated into the business landscape (Bergmann, 2024).

The accelerating pace of change in the domain of AI makes it challenging for policy makers to promote innovation while establishing safeguards (Anderson & Sutherland, 2024). This balance is particularly challenging to manage in healthcare, where there are significant opportunities to use AI, but also considerable risks. AI can transform healthcare by helping to discover cures for diseases such as cancer, enhancing diagnostic accuracy, speed, and access, and providing tools to support clinicians. However, AI in healthcare is faced with significant risks connected to ethical concerns, algorithmic biases and data privacy and security. Close collaboration between healthcare organizations, regulatory bodies and AI researchers are crucial to establish standards and guidelines for AI algorithms and their application in clinical decision making (Alowais et al., 2023).

According to Anderson and Sutherland (2024), there are cases where the risk of using AI in healthcare is low and the predicted benefits are high. One such area is automating tasks for clinicians with AI-applications to free more time to patient care. These represent potential areas of initial focus, and are estimated to reduce the administrative burden for clinicians by 10-30 %.

2.2 Large Language Models

As outlined above, LLMs represent a groundbreaking advancement in the field of AI (Hadi et al., 2023). LLMs are demonstrating remarkable capabilities in understanding and generating human-like language across a broad spectrum of tasks. Powered by deep learning algorithms, these models exhibit an extraordinary capacity to understand, generate, and interact with human language, based on the vast amount of data that is used to train them (Hadi et al., 2023; IBM, n.d.).

Since LLMs have the ability to generate text, they are frequently classified under the category of generative artificial intelligence (GenAI) (Raschka, 2023). Recent advancements in LLMs bolster confidence in their ability to address numerous challenges across real-world settings (Zhao et al., 2023). However, given the rapid advancements in this field, pinpointing unresolved challenges and areas of potential success remains a complex task (Kaddour et al., 2023). To gain a fundamental understanding of how LLMs work, this section provides an introductory background on LLMs, encompassing the technology behind LLMs, important concepts regarding how to train LLMs, a selection of recognized challenges and concludes with a future outlook.

2.2.1 The Technology Behind LLMs

The evolution process of language models has greatly expanded the range of tasks they can perform and markedly improved their performance (Zhao et al., 2023). The latest models, such as GPT-4, are engineered to tackle complex, real-world tasks.

The significant breakthrough in the field of LLMs was marked by the introduction of the Transformer architecture, detailed by Vaswani et al. (2017) in their paper titled "Attention is All You Need". Transformer is a type of neural network architecture, and has become the basis of LLMs. This architecture can contain trillions of parameters and be trained on several petabytes of textual data, forming the foundation of the model's knowledge base (Harrer, 2023; Liu et al., 2024; Raschka, 2023; Zhao et al., 2023). To illustrate the vastness of a single petabyte, consider that it can store over 200 million 5 MB photos.

The Transformer architecture is mainly composed of two parts: an encoder and a decoder, along with the attention mechanism that operates within these parts (Liu et al., 2024). The encoder-decoder structure involves an encoder that transforms input into a context vector, and a decoder that takes this context vector and generates an output (Ghojogh & Ghodsi, 2020). This process ensures that the output is connected to the input via the context vector within a concept called the hidden space. Furthermore, the attention mechanism concentrates on key details within a broad dataset, sidelining less significant information. A variant of the attention mechanism is self-attention mechanism, which is the core idea behind the transformer architecture (Wang et al., 2023).

The self-attention mechanism excels at processing input data across extensive ranges, enhancing local attention through parallel computing (Wang et al., 2023). It grasps diverse spatial information representations by considering the position and meaning of each word in the input. Moreover, the self-attention mechanism could be extended to multi-head self-attention, extracting hidden aspects of the data from every direction and angle (Liu et al., 2024).

Multi-head self-attention runs multiple self-attention processes at once, with each focusing on different parts of the input to uncover various patterns (Ghojogh & Ghodsi, 2020). This enables a deeper understanding of both close and distant elements in the input, enriching the model's grasp of complex word interactions (Liu et al., 2024).

In practice, this means that transformer-based models have the capability to analyze the full context of a sentence or whole documents and produce text that is both coherent and contextually appropriate, transforming fields such as chatbots, text summarization, and language translation (Hadi et al., 2023; IBM, n.d.). Thus, the multi-head self-attention mechanism, encoder and decoder all play crucial roles in how well the LLM performs, tackles downstream tasks as those previously mentioned as well as handles issues such as long-range dependencies and more (Ghojogh & Ghodsi, 2020; Liu et al., 2024; Wang et al., 2023).

2.2.2 Training LLMs

LLMs are incredibly flexible, meaning that one single model could perform a large variety of tasks related to text (Hadi et al., 2023). However, achieving superior performance and ensuring quality in specific application domains involve various complex and active areas of research regarding different techniques to enhance the models' efficiency.

As previously highlighted, the introduction of Chat-GPT represented a significant milestone of LLMs and revolutionized the AI landscape. However, ChatGPT is not available as an open-source tool and its primary usage is confined to interactions through OpenAI's website or its API (Liu et al., 2024). Consequently, there is a growing need for the development of alternative LLMs or those tailored to specific domains to provide alternatives to ChatGPT.

Choosing the right training approach is crucial for the success of a language model and training them necessitates extensive collections of textual data, with the data's quality playing a crucial role in determining the performance of LLMs (Raschka, 2023). The process of training LLMs unfolds in three primary stages (Liu et al., 2024). Initially, it starts with the collection and preparation of data. The next step is the pre-training process, containing selection of architecture and pre-training tasks, as well as applying appropriate algorithms for parallel training to accomplish the training process. The final step includes fine-tuning and alignment, a concept that will be further explained below.

2.2.2.1 Fine Tuning

In the rise of LLMs, many organizations are hopeful about grasping the potential to make their operations more efficient (Bowman, 2023; Nabwani, 2023; Shanahan, 2023). The potential of LLMs is significant, yet they may not always fulfill specific requirements directly after initial pre-training (Nabwani, 2023). This is due to several factors including that LLMs may lack tailored outputs necessary for certain applications. They might also miss crucial context, such as specific documents or instruction manuals. A LLM might not understand the context needed in clinical settings if it has never been trained on medical documents, for example (Singhal et al., 2023).

Additionally, LLMs often struggle with specialized vocabulary unique to certain industries or enterprises, making it difficult for them to effectively summarize or answer questions about complex topics like financial data or medical research (Nabwani, 2023). This is when fine-tuning comes in. Fine-tuning a LLM involves adjusting a pre-trained model to perform better on specific tasks or to adapt to particular types of data (Devlin et al., 2019; Raffel et al., 2019). This process is essential because while a general pre-trained model can handle a broad range of topics, fine-tuning can significantly improve its performance and accuracy in specialized areas. LLMs are statistical models that predict sentences of words based on probability distribution (Ramlochan, 2023). By fine tuning these models, they are exposed to new word relationships and structural conventions specific to a target task or domain.

2.2.2.2 Retrieval-Augmented Generation

As an alternative to full fine-tuning, Retrieval-Augmented Generation (RAG) serves as an innovative method to enhance LLMs by integrating prompt engineering with database querying, enabling the generation of context-rich answers (Gao et al., 2024; Nabwani, 2023). This approach not only offers a cost-effective solution to improve the relevance, and applicability of LLM outputs across diverse contexts (Gao et al., 2024; Nabwani, 2023) but also significantly reduces the production of factually incorrect content by leveraging external databases (Gao et al., 2024; Lewis et al., 2020; Nabwani, 2023). RAG is particularly effective for applications requiring tailored responses from extensive, context-specific document sets (Nabwani, 2023). Additionally, it proves invaluable when LLMs need to incorporate the latest updates, such as recent news or medical research findings not included in the initial training data, ensuring that responses remain current and factually accurate.

2.2.2.3 Scaling Laws

A key insight in pre-training is that scaling plays a vital role in the efficient development and deployment of LLMs, making the establishment of a quantitative method to measure the scaling effect invaluable (Zhao et al., 2023). The relationship between scale and model performance, also called scaling law, have the ability in most cases to predict a continued increase in certain capabilities as models get larger (Ganguli et al., 2022).

Scaling laws enable precise predictions of how future models' capabilities will evolve as they are scaled up in three dimensions: the volume of data they are trained on, their size in terms of parameters, and the computational effort (measured in FLOPs) utilized in their training (Bowman, 2023). This predictive capacity facilitates key design decisions, such as determining the optimal model size within a fixed resource budget, thereby avoiding the need for costly trial and error.

For transformer language models, the OpenAI team presented in 2020 (Kaplan et al., 2020) that performance of a language model improves as the parameters model size, data and compute appropriately is scaled up in tandem making the conclusion that larger models perform better. Moreover, the Google DeepMind team presents their investigation of the optimal model size and number of training tokens for a transformer language model in 2022 (Hoffman et al., 2022), reaching a slightly different conclusion. While both research groups concur on the choice of parameters, Kaplan et al. (2020) advocate for allocating a larger portion of the budget to model size over data, whereas Hoffman et al. (2022) contend that model size and data volume should be scaled up proportionately. Despite differing theories on the ideal balance, there's unanimous agreement among them and other researchers that the computational and energy costs for training LLMs are considerable. Consequently, precisely identifying the optimal model hyperparameters for a specific computational budget becomes crucial (Hoffman et al., 2022; Kaplan et al., 2020; Zhao et al., 2023).

While pre-trained LLMs deliver remarkable performance across a range of tasks, they are not universally applicable to every specific need. When encountering tasks beyond the capabilities, one solution is to fine-tune the model. As mentioned, fine-tuning involves adjusting the model's parameters to better suit new, specific objectives, aiming to tailor the model for particular tasks (Ganguli et al., 2022). This process typically uses more focused datasets to refine the model's abilities for specific uses (Hadi et al., 2023). However, just as training LLMs from scratch with extensive sequences presents computational hurdles, fine-tuning an already pre-trained LLM can also be a significant investment (Chen et al., 2024). However, various techniques for specifying LLM to specific application areas has been an attractive area of research.

2.2.3 Challenges of LLMs

Probably the most widespread challenges of LLMs today is the concept of hallucinations (Kaddour et al., 2023). Hallucinations of LLMs imply factual errors, incorrect or misleading outputs provided by the model (Zhao et al., 2023). This notable constraint frequently stems from the model's efforts to bridge gaps in knowledge or context by making assumptions grounded in the patterns it absorbed during its training phase (Hadi et al., 2023). As one of the most well-known implications it is also a topic of active research, with hypothesis that the problem is a result of the model's training process, dataset, and architectural design. Hallucinations substantially limits how LLMs can be responsibly used (Bowman, 2023), especially in certain high-risk fields (Li et al., 2023). However, recent research suggest that a solution to mitigate hallucinations could be achieved in a near future, by the LLMs own internally capability to track which statements are true with reasonably high precision (Bowman, 2023).

Another common implication of LLMs is that the models can demonstrate bias, when the training data the model is developed on is biased (Hadi et al., 2023). A statement well aligned with Baeza-Yates (2016) quote "the output quality of any algorithm is a function of the quality of the data that it uses". Viswanath and Zhang (2023) has listed a detailed quantitative analysis of various biases, including those related to race, gender, ethnicity, age, and more, which has been displayed by well-known LLMs as BERT and GPT-2. Well-known biases like these, related to minorities and disadvantage groups, are all examples of potential harmful content that need to be avoided (Navigli et al., 2023). To eliminate biases and other toxic content, it is important to filter low-quality text in the data preprocessing step (Liu et al., 2024). This could be done by adopting content moderation techniques, using methods like sentiment analysis, hate speech detection, and using specialized bias identification algorithms. The significant challenge posed by biases in LLMs is crucial to recognize, given the documented mistakes in critical sectors like law, medicine, and recruiting (Navigli et al., 2023).

Furthermore, as real-world information evolves, the knowledge embedded in LLMs can become obsolete or inaccurate (Zhao et al., 2023). According to Kaddour et al. (2023), with current methods, refreshing the outdated knowledge in models with new pre-training data proves to be expensive, and the task of removing old data while adding new information during the fine-tuning stage is complex. Thus, it is vital to explore cost-efficient approaches, such as various tuning techniques, to incorporate updated content into LLMs (Zhao et al., 2023). This challenge is crucial to manage due to it limiting the applicability of LLMs on real-world use cases, particularly where single outdated information can imply large implications for the end-user (Kaddour et al., 2023).

Another large implication of LLMs is the technology's lack of a human-like understanding of real-world problems (Ilicki, 2023a), where it is required strong reasoning skills and expert domain knowledge (Liévin et al., 2022).

2.2.4 The Future of LLMs

Future LLMs carry high expectations, with predictions pointing towards a swift enhancement in the technology's learning capabilities and overall performance (Liu et al., 2024). As one of the emerging trends, researchers believe the development will move towards fine-tuning LLMs against specific industries and real-world applications (Liu et al., 2024; Stringhi, 2023). Interestingly, there is a belief among some researchers that LLMs designed for specialized tasks or specific fields may surpass the modeling capabilities of general-purpose LLMs, such as ChatGPT (Raschka, 2023). However, it is then crucial that researchers in the field of AI will work in close collaboration with various sectors and professionals from diverse fields (Liu et al., 2024). By pooling expertise from various fields, joint endeavors will tackle challenges and explore potential solutions in a unified manner. Furthermore, it is crucial to clearly define the LLMs' application scope (Stringhi, 2023) and how to tackle it. For instance, there is a debate among researchers about the appropriateness of using LLMs in decision-making processes. Moreover, some researchers emphasize the necessity of human oversight for the activities of LLMs at all times, particularly in sensitive areas such as healthcare and law (Guo et al., 2023; Stringhi, 2023).

2.3 Ambient AI Scribes

Another promising AI tool that could be used in a clinical setting is ambient AI scribes. This technology utilize machine learning to analyze conversations, offering real-time, scribe-like capabilities (Tierney et al., 2024). It employs automatic speech recognition technology to extract and analyze relevant medical information and its context from patient-clinician dialogues. Employing a smartphone microphone, it transcribes encounters as they happen without retaining any audio recordings. This system then automatically generates electronic health record notes in near real-time (Crampton, 2019; Tierney et al., 2024). Thus, this technology has the potential to significantly reduce the documentation burden for healthcare staff and enhances physician-patient interactions.

However, to the best knowledge of the researchers conducting this study, there is a limited number of up-to-date scientific articles on this technology's performance, making it difficult to assess its effectiveness. In Tierney et al. (2024)'s study, the authors conclude that despite the early promises of this technology, careful and ongoing attention is necessary to ensure that it supports clinicians effectively. Additionally, optimizing ambient AI scribe outputs for accuracy, relevance, and alignment within the physician–patient relationship is crucial (Tierney et al., 2024).

2.4 LLMs in Healthcare

The recent focus on LLMs has caused unprecedented discussion of their potential application in various fields, including healthcare (Reddy, 2023). LLMs have revolutionized natural language processing (NLP), and state-of-the-art models like PaLM2 and GPT-4 have gained significant attention (Thirunavukarasu et al., 2023). LLMs have the potential to improve the effectiveness and efficiency of clinical, research and educational work in medicine. LLMs like ChatGPT use are described to have the potential to provide numerous benefits both for clinicians and patients (Zheng et al., 2023). Among these are to greatly improve the efficiency and accuracy of diagnosis and treatment plans, immediate 24/7 answers on healthcare questions, streamlining administrative tasks like billing and scheduling and support radiology departments with image analysis and interpretation.

In one recent study conducted in cooperation between actors is Sweden and Switzerland, orthopedic discharge summaries and letters were generated from ChatGPT-4 and orthopedic physicians which were evaluated blindly and then compared (Rosenberg et al., 2024). The study's result presented that both ChatGPT-4 and physician-generated notes were comparable in quality, but that ChatGPT-4 generated discharge documents ten times faster than the physicians did. This was a pilot study consisting of six cases, inviting further research and underlining that it exists technical and legal barriers before ChatGPT-4 could be used in clinical application using patient data, but the result indicates that significant gains can be realized in terms of time by utilizing LLMs in healthcare.

There are several issues and limitations preventing clinical deployment of LLMs like ChatGPT (Thirunavukarasu et al., 2023). Even though current applications have shown considerable potential in performing human-capable tasks, LLMs have also demonstrated significant drawbacks like falsifying data and generating misinformation (Reddy, 2023). These are concerning aspects in general, but in the context of healthcare, they can be even more severe. As LLMs are explored for utility in healthcare, including providing medical advice, interpreting health records and generating discharge summaries, it is necessary that safeguards are ensured around their use.

Even if AI-enabled tools can not diagnose and treat diseases today, DiGiorgio and Ehrenfeld (2023) argue that LLMs are ready to tackle many administrative tasks. It will require regulatory oversight, but less than for algorithms that recommend treatments, make diagnoses or otherwise impact clinical decision making. The researchers argue that the healthcare industry should reach for the low-hanging fruit in improving the efficiency for physicians, de-tethering clinicians from the computer. Being aware of the risks and limitations to implement LLMs in healthcare, (Ilicki, 2023c) agrees that LLMs should be implemented to free time for physicians through automating documentation and thus, de-tethering clinicians from the computer.

2.5 Ethics and Legislation of LLMs in Healthcare

AI's potential to revolutionize healthcare delivery is widely acknowledged and the development of pre-trained LLMs has opened new possibilities for generating medical content and support clinicians with decision making, diagnosis prediction and treatment options (Reddy, 2023). Since the launch of OpenAI's AI-based conversational LLM ChatGPT in November 2022, research has acknowledged that promising applications of ChatGPT can induce a paradigm shift in health practice Sallam (2023). However, the use of LLMs has raised multiple ethical concerns (Reddy, 2023).

According to Harrer (2023), there are three core limitations of LLM-generated data. First, models that have been trained on a large corpus of internet data with limited filtering, like ChatGPT, have ingested as much fair as biased content, as much harmless materials as harmful ones and as much facts as misinformation. This causes a risk of LLMs reproducing, disseminating or amplifying misinformation or problematic content. Second, the models do not know whether the material it produces contains misrepresentations, inappropriate content or falsehood or whether it tells the truth and they have no means to assess this by themselves, nor can they inform the user about this. Third, LLMs are probabilistic algorithms, meaning that when they are prompted multiple times with the same task or question, the model will return different responses, which can be different versions of previously problematic or wrong answers, or replacement of wrong answers with correct or improved ones and vice versa, or constitute different versions of previously correct replies or combinations on that. This behavior poses a reproducibility and reliability problem, requiring continuous human oversight of the model's output.

Zhang and Kamel Boulos (2023) describe how ChatGPT's responses have shown a wide and unpredictable fluctuation in quality and veracity, and the authors argue that this unpredictability is the main barrier for successful adoption of LLMs in healthcare. Similar to the argument by Harrer (2023) above, Zhang and Kamel Boulos (2023) explain that users can not know when the model is going to return a good answer and when its answers are going to be misleading or wrong, and thus, can not know when the model can be trusted or not, especially when the user is not qualified enough to assess the quality, the completeness and accuracy, of a given response. Haltaufderheide and Ranisch (2023) support above arguments and state that the most distinctive concern with LLMs in healthcare is the tendency to produce convincingly, but inaccurate content or harmful misinformation. Scientifically plausible, but factually inaccurate answers provided by LLMs are a phenomenon called hallucinations (Sallam, 2023). In healthcare, inaccurate algorithms output risks resulting in harm to health through for example misdiagnoses or inappropriate recommendations of treatment (Petročnik et al., 2023).

Another paramount concern regarding LLMs in healthcare is that LLMs risks perpetuating harmful racial, cultural and gender biases (Haltaufderheide & Ranisch, 2023). In healthcare, these models risk providing biased outputs inadvertently propagating social prejudices against already vulnerable groups (Petročnik et al., 2023). GenAI models are prone to various forms of bias depending on how they are trained (Zhang & Kamel Boulos, 2023). Biases are seen as a significant source of harm (Haltaufderheide & Ranisch, 2023).

Further, inputting patient data raises ethical questions regarding privacy, data security and confidentiality, especially in relation to commercial and publicly available models like ChatGPT (Haltaufderheide & Ranisch, 2023). The propensity of LLMs to disseminate patient data or other sensitive health information is a serious privacy concern. Personal health data is protected by the General Data Protection Regulation (GDPR) Article 9 (European Union, 2016).

Privacy concerns have further been raised in relation to ChatGPT's training data and code being kept secret, regarding its collection and storage of personal user data to train the model further (Zhang & Kamel Boulos, 2023). A need for transparency is evident, and the EU AI Act includes transparency requirements for GenAI, including a need that summaries of copyrighted data used for training is published. Transparency, together with explainability, is especially important in healthcare for operators to be able to validate LLMs performances (Harrer, 2023). The "black box" nature of generative AI models raises questions about the interpretability of the output they generate and calls for explainability and transparency (Zhang & Kamel Boulos, 2023).

In the EU AI Act by the European Union (2024), it is stated that AI systems that profiles individuals are always considered high-risk, that is, systems that process personal data to assess various aspects of a person's life automatically, such as health, economic situation, location or movement, etc. Researchers highlight that the use of ChatGPT in healthcare should be conducted with extreme caution considering its potential limitations (Sallam, 2023). Many further argue that LLM applications in healthcare should require close human oversight (Haltaufderheide & Ranisch, 2023).

World Health Organization (2021) has identified a number of ethical principles for the responsible design and application of AI technology: protect autonomy; promote human safety and well-being and the public interest; ensure transparency, explainability and intelligibility; foster accountability and responsibility, ensure equity and inclusiveness; and promote AI that is sustainable and responsive. Referring to these principles by World Health Organization (2021), Harrer (2023) argues that ethical and legal frameworks are needed for deployment and use of AI applications and for the selection and management of training data. There is also a need for measures to mitigate model biases. The author highlights six key factors that need to be regarded for responsible ethical design, use and governance of AI in healthcare.

1. Accountability: Users need to be informed about the capabilities and risks of the technology as well as the sensitivities and responsibilities involved in using it. The current legal state is blurry and there is a need for legal clarity regarding accountability, responsibilities and rights of developers and users.

2. Fairness: General LLMs like ChatGPT that have been trained on data from the internet are faced with the critical issue of bias and how to make an AI system unlearn problematic content is a complex research question. Ethic panels currently need to review current model implementations and audit the performance of already deployed models to identify and eliminate sources of misinformation and bias.

3. Data privacy and selection: Electronic health records range among the most sensitive and highly restrictive data sources and need to be treated as such. The healthcare sector is an evidence driven domain and the act of choosing and accessing suitable training data to develop generative AI applications has ethical, legal and model performance related implications.

4. Transparency: Prompts are bridged with responses by LLMs that lack inherent capability of showing the logic behind their work, leaving human operators with this task who will only be able to do a reliable job if the models provide insights into its data sources. Camouflaged generative AI content through automated chatbots could become a dangerous source of misinformation, leaving a toxic mark on knowledge bases, risking contamination of knowledge bases at scale. Labeled AI-generated content could be needed.

5. Explainability: For generative AI systems, explainability should be a key design feature as it provides an important datapoint to the human operator and user whose role it is to validate the soundness and correctness of AI generated

content before it is translated into action. Explainable and transparent AI systems could be assigned a truthfulness index, which could serve as a measurement for assessing the trustworthiness of an LLM when assisting clinicians and patients.

6. Value and purpose alignment: AI's so-called Alignment Problem describes the ethical and existential risk that emerge when AI machines violate or do not follow the the purpose and values of their human creators and users. To avoid this, a human value system needs to be in place as well as a clearly defined system of what the AI-system should do and why.

2.5.1 Medical Device

Some medical products in Sweden fall under the legal framework of medical device (Swedish Medical Products Agency, 2021). Below is a part of the definition of a medical device quoted from the Swedish Medical Products Agency (2021).

> *"A medical device can be an instrument, apparatus, appliance, software, implant, reagent, material or other article. The condition is that a medical device must be used on humans and have one or more of the following medical purposes:*
> - *diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease*
> - *diagnosis, monitoring, treatment, alleviation of, or compensation for, an injury or disability*
> - *investigation, replacement or modification of the anatomy or of a physiological or pathological process or state*
> - *providing information by means of in vitro examination of specimens derived from the human body, including organ, blood and tissue donations."*

2.6 Healthcare in Sweden

In a study by Anskär et al. (2019), the authors highlighted a significant concern about the allocation of healthcare resources in Sweden. At that time, Sweden had a relatively higher number of healthcare staff per capita compared to most other European countries. Despite these resources, the healthcare system in Sweden struggled with issues such as limited access and prolonged waiting times for diagnosis and treatment in both hospitals and primary care facilities (Anskär et al., 2019).

Fast forward to today, researchers forecasts a severe staff shortage in the Swedish healthcare sector, occurring simultaneously with an aging population (Barkman & Aasa, 2023). Given the aging population, the demand for healthcare will also increase (Ahlstedt et al., 2023). Additionally, Grant Thornton (2023) notes that care queues will continue to grow significantly each year. In 2023, the healthcare sector continued to face challenges in providing patients with their first consultation, surgery, or other scheduled treatments within the timelines specified by the national care guarantee (Grant Thornton, 2023; Janlöv et al., 2023). This scenario indicates that fewer healthcare workers will need to manage increasing demands, even as the system fails to resolve the previously mentioned critical issues, underlining the pressing need for significant improvements in efficiency.

Currently, there is an ongoing debate concerning the level of administration in the Swedish healthcare and its impact on the care provided to patients (Ilicki, 2023b). Then given the increasing challenges faced by the healthcare sector, a greater emphasis on planning and optimizing healthcare resources will be necessary (McKinsey & Company, 2019). However, to discuss the potential for reducing the administrative burden within healthcare effectively, it is essential to first define what is meant by 'administrative work' and what is meant by 'clinical work'.

2.6.1 Definition of Administrative versus Clinical Work in Healthcare Settings

Clinical work is related to the observation and diagnosis of patients (Strauss et al., 1985). Holten Møller and Vikkelsø (2012) describe clinical work as a circular process involving five activities: (a) assessment of the patient's health status and illness, (b) acquisition and analysis of clinical data, (c) making informed decisions regarding the patient's health and illness, (d) providing further medical treatment for the patient's condition and disease, and (e) observing the outcomes of the treatment. These work tasks are typically associated with direct patient care, which is generally regarded as more meaningful compared to other tasks performed by clinicians (Bringsén et al., 2012). Another way of phrasing clinical work is 'direct patient-related care', defined as work in direct interactions with patients as well as phone communications with patients or their close relatives (Anskär et al., 2018).

Broadly, administrative work is by Anskär et al. (2019) defined as communicating information between various professional roles, coordinating activities, and structuring systems, pointing out that information is one key word of administration. However, the word "administration" can have different meanings in different contexts, which might explain why there are so many views on its scope, importance, and even the emotional connotations it carries (Holberg & Bell, 2015).

In the Swedish healthcare, there is a lack of a sufficiently established definition for what is meant by 'administration' (Holberg & Bell, 2015). Occasionally, the term 'administration' covers activities indirectly related to patient care, such as managing health records and processing referrals. On the other hand, some reports specifically address tasks that are clearly administrative, like billing and scheduling. On this topic, Anskär et al. (2019) distinguishes between patient-related and organization related administrative work tasks.

The nature of patient-related work tasks typically encompasses activities such as documentation, dictation, scheduling appointments, managing health records and referrals, as well as inputting data into healthcare records and quality registries, to provide a few examples (Anskär et al., 2018). This "patient-related administration" could be defined as the administrative tasks that arise during a patient's care episode, where every stage of healthcare requires or necessitates some form of documentation and registration (Holberg & Bell, 2015).

Meanwhile, organization related administration tasks involve for example overseeing equipment and facilities, handling emails, organizing schedules, and procuring medical supplies, including items like laundry (Anskär et al., 2018, 2019). These tasks are part of an administrative process which is continuous, yet does not have a direct connection to the patient (Holberg & Bell, 2015).

Distinguishing between clinical and administrative work can be challenging, as these tasks often overlap and intertwine (Holberg & Bell, 2015). Also, given the substantial volume of administrative duties in Swedish healthcare, it might be beneficial to categorize these tasks into those related to patient care and those associated with organizational management (Anskär et al., 2018), for a clearer structure and overview of clinicians' everyday work.

The definitions used in this report are presented in the list below. A selection of examples of the subcategories are presented in table 2.1.

- Direct patient work: Tasks that involve direct physical interactions with patients as well as phone communications with patients or their close relatives.

- Patient-related administration: Administrative tasks that arise during a patient's care episode, which are directly linked to the patient's care.

- Non-patient-related administration: Administrative tasks that do not have a direct connection to patients.

Table 2.1: Examples of direct patient work, patient-related administration, and non-patient-related administration.

Direct patient work	Patient-related administration	Non-patient-related administration
Face-to-face contact with patients	Documentation of healthcare records	Meetings (non-patient-related)
Telephone contact with patients	Reading EHRs	Managing equipment and facilitation
Telephone contact with patients' next of kin	Signing journal entries	E-mail management (non-patient-related)
	Referral management	Scheduling
	Prescribing medical drugs	Managing computer problems
	Entering data into health records and quality registers	Ordering medical suppliers (as laundry)

2.6.2 Previous Studies on Time Utilization in Healthcare Settings

As in many other countries, patient care at Swedish hospitals is a collaborative effort, engaging various clinicians who rely on each other's expertise and thus must work together closely (Bardram, 1997). The collaborative efforts consists of various tasks with different character. Barkman and Aasa (2023) contend that there has been a significant rise in administrative duties in recent years. Numerous studies indicate that physicians spend one-third of their time on direct patient-related activities, whereas nurses allocate approximately 43 % of their time to such tasks (Barkman & Aasa, 2023; Grant Thornton, 2023).

Anskär (2019) investigated time utilization by healthcare staff in Swedish primary care. The study collected data using self-reporting forms, where participants both estimated and then documented in real-time the time dedicated to tasks categorized as direct patient work, patient-related administration, and other work tasks. According to the study, registered nurses reported spending 42.2 % of their time on direct patient care, 27.2 % on patient-related administration, and 30.6 % on other tasks. For physicians, the figures were 34.4 % for direct patient care, 32.3 % for patient-related administration, and 33.3 % for other tasks. Interestingly, both groups initially overestimated the time they would spend on direct patient care and underestimated the time for other tasks, as compared to their actual logged hours (Anskär, 2019).

For physicians, the the most time-consuming patient-related administration tasks were dictation (24.0 %), reading health records (16.8 %), signing documentation (13.0 %), and documentation of health records and order tests (11.9 %). For nurses, it were documentation of health records and order tests (51.6 %), reading health records (13.3 %), contact with other caregivers about patient cases (9.6 %) and administering appointments (6.8 %).

Another study conducted by McKinsey & Company and the Swedish Medical Association (Läkarförbundet) in 2018 found that Swedish physicians spent 7.5 hours per week on patient-related administrative tasks, which amounts to one full day each week dedicated to these activities (McKinsey & Company, 2019). Furthermore, a study in 2009 revealed that district nurses in Sweden allocated 30.47 % of their time to indirect patient care (Borelius, 2009).

2.6.3 Perspectives on the Administrative Burden

Views on the administrative burden vary widely, with some researchers asserting that all administration is necessary, while others strongly contend that unnecessary administration exists (*Effektiv vård*, 2016).

Ilicki (2023b) highlights a nuanced perspective on administrative tasks in healthcare, pointing out that while some administrative work has direct medical value, it is often too time-consuming. For instance, Ilicki (2023b) notes that tasks like writing discharge notes for patients who the physicians have never met, and reporting information to quality registers, are valuable for researchers and patients but not for the clinicians who perform these tasks. Additionally, Ilicki (2023b) emphasizes the inefficiency of the login processes to clinician data systems, which, while time-consuming, also play a role in ensuring computer security.

Regardless of opinions on whether the administration is unnecessary, it is clear that clinicians in Sweden face a significant administrative workload and that excessive administrative tasks divert time from essential medical activities (Barkman & Aasa, 2023). Additionally, there has been a gradual shift of administrative responsibilities from administrative personnel to medical staff, suggesting that the deployment of medical resources may not be optimized for efficiency (Anskär et al., 2019).

2.6.4 Illegitimate Tasks: Conceptualization and Implications for Employees in the Healthcare Sector

Illegitimate tasks encompass work assignments that are perceived by employees as falling outside their role responsibilities, thus violating normative expectations of their professional duties (Björk et al., 2013; Semmer et al., 2015). Such tasks increase the amount of administrative duties (Basinska & Dåderman, 2023). For instance, instead of delivering direct medical care, employees may be tasked with duplicative documentation, generating various reports and summaries. According to the Stress-as-Offense-to-Self (SOS) theory, developed by (Semmer et al., 2007), illegitimate tasks have been identified as a significant social stressor, endangering one's sense of self and infringing upon one's professional identity (Semmer et al., 2019). In licensed professions, such as in healthcare, there is a strong link between an individual's identity and their professional role (Aronsson et al., 2012). In these professions, stress often arises from conflicts between one's professional role, personal identity, and self-evaluation. The professional role defines and suggests what is expected, as well as what is not expected under typical conditions, including tasks that are deemed illegitimate.

Illegitimate tasks could be divided into two categories: unnecessary tasks and unreasonable tasks (Semmer et al., 2015). Unnecessary tasks are tasks that seem to lack a clear purpose or benefit and often stems from organizational inefficiencies, outdated ways of working or idiosyncratic demands (Kilponen et al., 2021). Employees may view tasks as unnecessary if they are fundamentally avoidable or if better organization could have eliminated the need for them altogether (Kilponen et al., 2021; Semmer et al., 2019). For instance, having to re-enter data due to incompatible computer systems is seen as unnecessary because it could have been prevented if the systems were compatible from the start (Semmer et al., 2015). In healthcare settings, these characteristics may be linked to the notorious need for double and triple documentation caused by incompatible data systems (Barkman & Aasa, 2023).

Unreasonable tasks on the other hand, are considered inappropriate for the person assigned to them, either because they do not align with their skills, are deemed a waste of time, or are tasks that should be performed by someone else (Björk et al., 2013; Semmer et al., 2015). This phenomenon is particularly prevalent in Sweden's healthcare sector. According to McKinsey & Company (2019), one significant factor contributing to low productivity within Swedish healthcare is that physicians devote a substantial portion of their time to tasks that should be performed by others.

Whether the task is classified as unnecessary or unreasonable, it signals a lack of respect for the person who is expected to do it (Semmer et al., 2015). However, it is crucial to recognize that what is considered illegitimate in the workplace can vary depending on the context. Understanding behaviors that might be viewed as illegitimate requires considering how individuals perceive their job responsibilities and the professional identity linked to their role (Semmer et al., 2009). Thus, could the same tasks be considered illegitimate for one person but not another (Semmer et al., 2015).

The valuation of tasks varies based on whether they are central to the profession, peripheral, or completely outside the role (Aronsson et al., 2012). Tasks aligning with the expectations of the role holder are seen as legitimate, while those that exceed these norms are viewed as illegitimate. However, peripheral tasks are not inherently viewed as illegitimate, although they are more prone to this perception compared to core tasks (Semmer et al., 2015). These tasks can be seen as legitimate if they support, rather than obstruct, primary activities. However, according to the article by Barkman and Aasa (2023) on healthcare administration, the primary source of frustration among Swedish clinicians is not the tasks themselves, but the manner in which they are executed. This suggests a connection to unnecessary illegitimate tasks, especially since the frustration often is related to data systems and organizational processes (Barkman & Aasa, 2023; Janlöv et al., 2023; McKinsey & Company, 2019).

The situation of how a task occur also influences legitimacy (Semmer et al., 2015). For instance, if a colleague is ill, it might necessitate another to temporarily take on duties beyond their usual role, which is typically not viewed as illegitimate. However, the administrative tasks that burden clinicians are not caused by unusual situations (Barkman & Aasa, 2023). The Official Report by the Swedish Government on healthcare in 2016 *Effektiv vård* (2016) highlights that over the past decade, legislative modifications have predominantly augmented procedural and administrative regulations within healthcare settings in Sweden. Moreover, this accumulation of requirements persists, with new regulations being introduced regularly without any removal of obsolete ones — a situation unchanged to this day (Barkman & Aasa, 2023).

2.6.4.1 Illegitimate Tasks Impact on Well-Being

Effectively assigning tasks can not only liberate clinicians to concentrate more on patient care but also boost job satisfaction, optimize skills, and cultivate a more efficient workforce (McKinsey & Company, 2019). Previous studies suggest that illegitimate tasks often cause stress for clinicians (Anskär et al., 2019; Thun et al., 2018) and have a detrimental effect on their well-being (Björk et al., 2013; Semmer et al., 2015; Thun et al., 2018). For example, illegitimate tasks has been discovered to correlate with adverse health and emotional consequences, such as diminished self-esteem (Eatough et al., 2016; Schulte-Brauck et al., 2019; Semmer et al., 2015; Sonnentag & Lischetz, 2018), decreased sleep quality (Semmer et al., 2015) and an elevated risk of burnout (Semmer et al., 2019; Werdecker & Esch, 2021). Furthermore, research indicates that illegitimate tasks can result in heightened cortisol release, greater occurrence of musculoskeletal pain (Kottwitz et al., 2013) and heightened feelings of anger and frustration (Eatough et al., 2016). Additionally, studies has found that higher degree of illegitimate tasks lead to decreased job satisfaction (Björk et al., 2013; Werdecker & Esch, 2021).

3
Methods

The objective of this chapter is to describe and discuss the research methods that have been applied for this study. The study is based on 37 interviews with healthcare professionals at the neurology, ophthalmology and radiology department at Sahlgrenska University Hospital as well as 9 interviews with 10 experts within the field of AI in healthcare from Sweden and Canada. Some data were collected via self-completion forms and field notes through observations at the hospital.

3.1 Research Process

Initially, data was gathered for constructing the theoretical framework and to facilitate learning. A literature review was conducted about LLMs in healthcare that encompasses theory about the development of AI in healthcare, the technology of LLMs, the ethics of LLMs in healthcare and the definition and perspective of administrative work in healthcare.

Furthermore, semi-structured interviews were conducted with clinicians from the neurology, ophthalmology and radiology department at Sahlgrenska University Hospital. These interviews were focused on identifying the clinicians' needs for their administrative text-based tasks. The semi-structured interviews were complemented with self-completion forms that aimed to collect numerical data on how much time clinicians spend on a selection of administrative tasks. The selection of tasks were derived from the identified needs in the interviews with clinicians'. Some data were collected as field notes from a number of observations the researchers did to gain understanding of the clinical setting at the hospital.

In addition, expert interviews were conducted with individuals that work in the field of AI in healthcare to gain deeper insights and learnings. The collection of expert interviews represents a small scale environmental scanning, which provides knowledge and inspiration from actors in Sweden and Canada on the topic of LLMs and AI in healthcare.

Lastly, the collected data were analyzed through thematic analysis in relation to the theoretical framework. The objective of the discussion was to evaluate partly how LLMs can be used in healthcare in general, partly how LLMs can be used to streamline clinicians' text-based tasks at the neurology, ophthalmology and radiology department at Sahlgrenska University Hospital. The potential of LLMs in healthcare was discussed, as well as the barriers.

Figure 3.1 below, visualizes the methods that have been used for data collection, sampling and data analysis in this study.

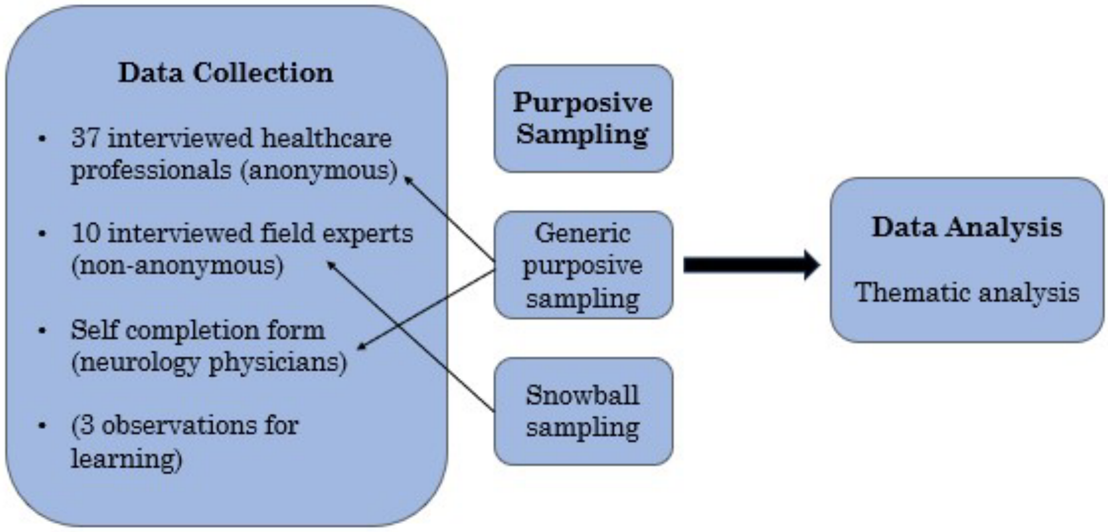

Figure 3.1: Data collection, sampling and data analysis methods.

3.2 Data Collection

This study relies on both primary and secondary data. Primary data were gathered through 37 semi-structured interviews with healthcare professionals at Sahlgrenska University Hospital (see table 3.1) and through 9 interviews with 10 experts in the field of AI in healthcare (see table 3.2). Some primary data were collected via self-completion forms and field notes from observation. Secondary data were collected as a complement to the primary data and were derived from an extensive literature review of reputable sources. Secondary data forms the base of the theoretical framework of this study. The initial data collection includes a broad spectrum of perspectives, and were followed by a more focused data collection in line with evolvement and refinement based on emerging results.

3.2.1 Literature Review

A literature review was conducted in order to collect relevant theory to this research. According to Easterby-Smith et al. (2015) a literature review describes, clarifies and evaluates what is already known about the topic. It is also a valuable tool for the authors to achieve in-depth learning about the topic early on in the process. This research requires a multifaceted knowledge base, including technical understanding of LLMs as well as knowledge of the complexities in healthcare where ethical, legal and organizational aspects are intertwined. Relevant literature were retrieved from public domains related to healthcare and medical technology, and also through the use of keywords in databases such as Google Scholar and ResearchGate.

The challenge of doing a literature review is determining what information is trustworthy and relevant to the research (Eriksson & Wiedersheim-Paul, 2014). Bell et al. (2019) supports this argument and suggests four criteria for quality assessment: authenticity, credibility, representativeness and meaning. These four criteria are fundamental for this research when retrieving data from existing literature.

3.2.2 Qualitative Interview

Interviews have been a central part of this research to collect qualitative in-depth data. According to Bell et al. (2019), there are two main types of interviews: the unstructured interview and the semi-structured interview. The term "qualitative interview" is sometimes used to encapsulate these two types of interviews. In an unstructured interview the interviewee is asked very few questions and is allowed to respond freely, with the interviewer responding in turn only to points that seem worth following up on (Bell et al., 2019). Unstructured interviews tend to be similar to conversations. In a semi-structured interview the interviewee still has a great deal of leeway in how to respond, but the interview is somewhat structured to keep within fairly specific topics that the interviewer are prepared to cover with the help of a list of questions (an interview guide) (Bell et al., 2019).

In this study, all 46 interviews have been semi-structured. Two interview guides have been used, one for interviews with clinicians and one for interviews with experts. All interviews showed large variations, not least between clinical departments, making each interview unique. The two interview guides are presented in Appendix 1 (see section A.1 and A.2). The length of time for the clinician interviews was approximately 30 minutes and the expert interviews were one hour, with the exception of two expert interviews that were 30 minutes long.

3.2.3 Sampling and Practicalities

All interview objects were collected through purposive sampling. In qualitative research, most sampling entails purposive sampling of some kind (Bell et al., 2019). Purposive sampling implies that the sampling is conducted with reference to the goals of the research. This means that the selection of units of analysis are in terms of criteria that will allow the research questions to be answered.

The clinicians' from Sahlgrenska University Hospital that were interviewed were sampled through generic purposive sampling. In generic purposive sampling, the researcher defines criteria, identifies suitable cases, and samples from them to address the research questions (Bell et al., 2019). The defined criterion in this study was to a beginning with healthcare professionals. It was later narrowed down to physicians and nurses. Participating healthcare professionals were collected by the help of the three clients for this study, who themselves work as residents at SU at the three departments that this study focuses on.

The interviewed field experts were sampled through snowball sampling. Snowball sampling involves the researcher initially engaging with a small group of individuals pertinent to the research topic and subsequently leveraging these connections to establish contacts with others (Bell et al., 2019)). Initially, experts were sampled by the help of the supervisor for this study. Experts that participated in interviews in turn referred the researchers to other experts that were interviewed.

In the beginning of each interview, the researchers asked for approval to record. All interviews were recorded and later transcribed with the help of Turboscribe and Microsoft Office Word. Both researchers participated in all interviews with a few exceptions due to practical reasons. The majority of the interviews were conducted via Microsoft Teams, only a few were conducted on site at SU. The two researchers had clear roles throughout the data collection of this study. One of the researchers had the responsibility to conduct the interviews, and the other researcher was responsible to take notes during the interviews.

All clinicians who participated in interviews for this study are anonymously presented in the report and all experts are mentioned by name. All experts who participated in interviews got the opportunity to be anonymous. The ten experts received a draft of the final result-text with yellow markings of what they had said and got the opportunity to delete or change the context of the text. In each draft the other experts were anonymously marked with X in the researchers' wait of approval. All experts approved to be mentioned by names in the report. The report's results from the clinician interviews where looked over and approved by the clients for this study for each department.

3.2.4 List of Interviewees

The clinician and expert interviewees are listed in table 3.1 and 3.2 below.

As mentioned above, the scope of this study was narrowed down from healthcare professionals to physicians and nurses. However, as table 3.1 presents, a few interviews were conducted with other healthcare professionals at SU as well. The material is not included in the result of this study because of the study's deliminations that were made because of time and resource constraints. However, the interviews in question were of high value for the researchers' learning and understanding of the clinicians work at SU. The material from the physiotherapist and the assisting nurses showed similarities mainly with the administrative nurses' text-based tasks, indicating that the results of this study could be generalized to other clinicians than physicians and nurses.

Table 3.1: List of clinician interviews.

Department	**Professional title**	**Amount**
Neurology	Nurse	8
	Nursing assistant	1
	Physiotherapist	1
	Resident	4
	Senior specialist	4
Ophthalmology	Resident	9
	Senior specialist	1
Radiology	Medical secretary	1
	Radiology nurse	2
	Radiology nursing assistant	1
	Resident radiologist	1
	Senior radiologist	4

Table 3.2: List of expert interviews.

Name	Company	Professional title
Amrit Krishnan	Vector Institute	Technical Team Lead
Carolyn Chong	Vector Institute	Senior Product Manager
Derek Beaton	Unity Health Toronto	Director, Advanced Analytics
Hans van den Brink	Västra Götalandsregionen	Digital Strategist
Isak Barbopoulos	Sahlgrenska University Hospital	Machine Learning Engineer
Jonathan Ilicki	Industrifonden	Principal
Magnus Kjellberg	Sahlgrenska University Hospital	Head of Kompetenscentrum AI & Chief Data Scientist
Markus Lingman	Region Halland	Chief Strategy Officer & Senior Consulting Cardiologist & AI Swede of the Year 2020
Michael Page	Unity Health Toronto	Head of AI Commercialization
Rosita Ahlstedt	Region Halland	Department Manager Decision Support and AI

3.2.5 Diary as a Self-Completion Questionnaire

An effective alternative to traditional self-completion questionnaires is a diary, particularly useful for precise estimates of various behaviors. Bell et al. (2019) identified two types of diaries: structured and free-text, where the structured is resembling a questionnaire with closed questions, commonly known as a time-use diary (Bell et al., 2019). Thus, this kind of diary as a self completion questionnaire is an effective method for collecting data when fairly precise estimations of time spent on a certain behavior are required (Bell et al., 2019). Often, for a certain objective where sequencing of a behavior is required, it offers more accurate results than interviews or estimation questionnaires.

In this study, five structured diaries as self completion questionnaires were developed, one for radiologists, one for physicians at the ophthalmology department, one for nurses at the neurology department and one for the physicians at the neurology department. Additionally, a new diary was developed for the neurology physicians after receiving feedback on the first. However, due to time and resource constraints, only the diary for neurology physicians received valid responses. Although the number of responses was too small to be representative of the entire population, they were still considered useful for this study's results.

The diary as a self completion questionnaire that has been distributed among clinicians at the hospital to measure the time they spend on certain administrative text-based tasks is called 'self-completion form' throughout this study.

3.2.5.1 Sampling for the Diary as a Self-Completion Questionnaire

For the diary as a self-completion questionnaire, generic purposive sampling was conducted. This method involves defining criteria and selecting samples based on them to address the research questions (Bell et al., 2019). Similar to the clinician interviews, the criteria for this sample included physicians and nurses from the neurology department, radiologists from the radiology department, and physicians from the ophthalmology department.

3.2.6 Observations for Understanding and Learning

In order for the researchers to gain a more comprehensive understanding of the clinical setting a number of observations were conducted at Sahlgrenska University Hospital. One day was spent at the neurosurgery department observing two neurosurgical procedures, one day was spent observing a senior specialist at the stroke department and one half day was spent observing radiologists at the radiology department. Bell et al. (2019) explain how shadowing is a form of observation that affinities with the notion of passive participant observation. The researchers follow a member of an organization throughout his or her working day, and can ask the observed participant questions about what he or she is doing. The researchers may write field notes, recording the times, conversation subjects, mood and body language. During the observations that were conducted in this study, the researcher's asked questions to learn of things clinicians do. Some field notes were collected during these occasions.

3.3 Data Analysis

A thematic analysis was conducted to analyze the data obtained from the semi-structured interviews. According to Bell et al. (2019), a thematic analysis imply focusing on identifying concluding themes. The three cases from neurology, ophthalmology and radiology, and the expert interviews was all handled as separate cases, where the data categories were generated by applying a general framework for qualitative analysis proposed by Grodal et al. (2021). The proposed framework follows three steps: 1. generating initial categories, 2. refining tentative categories and, 3. stabilizing categories (Grodal et al., 2021).

During the first stage, generating initial categories, two methods were used: asking questions and focusing on puzzles. The questions asked, were predefined related to this study's research questions. The interview questions for clinicians centered around three main themes, patient-related administrative text-based tasks, primary bottlenecks and suggested efficiency improvements. For the interviewed experts, the interview questions also centered around three themes, LLMs' potential in healthcare, risk evaluation, and implementation strategies. By asking questions, initial (possible) categories were crafted during the collection and analysis of data, allowing for proactivity, which is recommended by (Grodal et al., 2021). Moreover, in formulating the initial data categories from the responses to the interview questions, emphasis was placed on discerning patterns and isolating the most pertinent data for addressing the research questions of the study. This approach involved selectively prioritizing relevant data to navigate the complexity and volume of information.

Secondly, the initial data categories underwent analysis and restructuring based on connections between them. Certain categories were subdivided to enhance clarity and differentiation. For instance, risk evaluation was divided into technological risks and ethics and legalisation, while patient-related tasks were divided into patient-related tasks where information is retrieved from EHRs and patient-related tasks that is not. Categories were then organized sequentially, acknowledging dynamic relationships. This method facilitated the discovery of novel interconnections between concepts (Grodal et al., 2021).

Third, in the concluding phase of the analysis, the findings underwent a reevaluation, incorporating insights from the theoretical framework and the other cases. This involved a comparison of the results with relevant theories to discern supporting and conflicting perspectives, facilitating the formulation of conclusions. This stabilizing stage often assists researchers in offering answers to or clarifying their initial questions or puzzles (Grodal et al., 2021).

3.4 Research Quality

To review the quality of a qualitative research study Bell et al. (2019) points out credibility, transferability, dependability, and confirmability as important criteras that need to be considered.

Credibility refers to the confidence in the truth of the data and the interpretations of them (Bell et al., 2019). It involves ensuring that the research findings accurately represent the participants' perspectives. Techniques to enhance credibility include for instance prolonged engagement with participants, persistent observation, and respondent validation. Hence, after compiling of the interviews and results for this study, expert interviewees were notified that they could either confirm the information, change it, or reject it. The results from the clinicians' interviews and observation were sent to the three clients from Sahlgrenska University Hospital for approval of the included information.

Transferability addresses the extent to which the study's findings can be applied to other contexts (Bell et al., 2019). Qualitative research does not aim for broad generalization but seeks to provide rich, detailed insights that others can relate to their own situations. Providing thorough descriptions of the research context, participants, and methodology allows others to judge the applicability of the findings to different settings. Therefore, by utilizing the interview guides provided in Appendix 1, A.1 and A.2, as part of the data collection tools, other researchers can replicate the study's concepts, thereby enhancing the transferability of the findings.

Dependability is concerned with the consistency and stability of the research process if it was repeated (Bell et al., 2019). To ensure dependability, researchers maintain an audit trail documenting all phases of the study, from data collection to decision-making processes. Furthermore, problem formulation, selection of research participants, fieldwork notes, interview transcripts, data analysis decisions should be documented to ensure dependability. Throughout this study, the researchers aimed to ensure transparency in problem formulation, participant selection, data collection, and data analysis methods. All interviews were transcribed, and all field notes were thoroughly documented.

Confirmability ensures that the findings are shaped by the participants' responses rather than researcher bias (Bell et al., 2019). Researchers can maintain objectivity by keeping a reflexive journal to document their reflections, biases, and decision-making processes. To mitigate potential biases in this study, some measures were implemented. The inclusion of a large number of interviewees from Sahlgrenska University Hospital potentially reduced bias and increased the confirmability of this study. However, the selection of interview participants could have been more strategically planned to include individuals with varying levels of experience and diverse backgrounds. This approach would help limit the risk of role-dependent biases, particularly in the ophthalmology department, where all participants except one were residents. Also, during the compilation of the results and conclusions, researchers worked independently and then presented their ideas to one another before finalizing the results. This collaborative process helps to reduce the risk of personal bias.

3.4.1 Ethical Considerations

Ethical considerations is a largely important part of research, where four areas is underlined by Bell et al. (2019). First, avoidance of harm, where harm could take various forms, physical injury, damage to participants' development or self-esteem, stress, or negative impact on career prospects, for instance (Bell et al., 2019). Second, informed consent, implying to provide prospective research participants with comprehensive information about a study. This ensures they can make an informed decision about whether or not to participate. Third, privacy, relates to the necessity of protecting the privacy of research participants. Last, preventing deception, where deception occurs when researchers misrepresent the true nature of their research.

In this study, participant involvement was entirely voluntary. At the start of each interview, participants were informed about the study's purpose and the reasons for their participation, to ensure informed consent. Permission to record the interviews was requested, with participants assured that the recordings would only be used for this study. Clinician participants were informed that their participation would remain anonymous, while experts were asked if their contributions could be attributed to them by name, after their approval of the included content.

As explained above, upon completion of the expert interviews, the results were distributed to each expert. The sections involving their contributions were highlighted, and other names were anonymized. Each expert was given the opportunity to choose whether to participate, to remain anonymous, or to have their name attributed to their contributions. Additionally, they were allowed to modify their input to avoid any invasion of privacy. Moreover, once the results from the three departments, neurology, ophthalmology, and radiology, at Sahlgrenska University Hospital were finalized, it were sent to the clients of this study at SU for approval.

4 Results

In this chapter, the results of this study are presented. The chapter first presents the results from the conducted expert interviews. Further, it presents the results from the conducted interviews with healthcare professionals at the neurology, ophthalmology and radiology department at Sahlgrenska University Hospital as well as the results from the self-completion forms that were distributed among physicians at the neurology department.

4.1 Expert Perspectives Within the Area of AI in Healthcare

> *"The greatest challenge we see, one that has been present for ages but that is only growing, is that the demand for healthcare exceeds our capacity to deliver it. So, there is a need for more care than we can provide. And it is exacerbated by an ongoing demographic shift, with an increasing number of elderly individuals in society who statistically require more care.*
>
> *Meanwhile, our ability to hire as many healthcare professionals as we truly need to meet future healthcare demands is not increasing at the same pace. So, we have three options.*
> 1. *Discontinue certain forms of care.*
> 2. *Reduce the need for care by ensuring that patients can do more for themselves.*
> 3. *Streamline our operations. We consider administration as something that does not add value."*
>
> Hans van den Brink

This study takes a foothold in the third option.

4.1.1 Potential LLM Applications in Healthcare

It is recognized that LLMs hold substantial potential to streamline various tasks in healthcare settings. However, this potential is not without its barriers and risks. Jonathan Ilicki elaborates on a range of tasks that LLMs could potentially address. He identifies two main categories: general administrative processes, which involve non-patient-close interactions, and clinical-specific processes. Ilicki believes that while the use of LLMs in administrative tasks shows potential in a closer future, their application could eventually extend to clinical processes, which, although promising, involve greater risks and uncertainties.

> *"And I think we will begin in the very general administrative part. Because there is a lot of administrative healthcare that can be automated. And then over time it will move further towards clinical tasks. Clinical tasks can have greater value in terms of increasing efficiency for clinical staff. But that will come later as the risks are larger and there is a greater need for controls, which take time to create, implement and trust."* Jonathan Ilicki

Hans van den Brink discusses the pressing need to enhance the efficiency of the Swedish healthcare system in response to both current and emerging challenges. He highlights a critical issue: the demand for healthcare is consistently outstripping the available capacity to provide it. This problem is anticipated to intensify due to demographic shifts, notably an aging population that requires increasingly more healthcare services.

Given that the growth of the healthcare workforce is not keeping pace with this escalating demand, van den Brink proposes three potential strategies to address these challenges effectively. Firstly, reduction of certain healthcare services, which he emphasizes, could sound drastic but he explains carefully that this approach has always existed, giving an example of pregnancy tests that were done in hospitals in the past. Secondly, reduction of healthcare demand on-site by adding functionality that empowers patients and enables self-management of diseases. Thirdly, focus on streamlining operations within healthcare services.

Related to the third approach, van den Brink highlights the inefficiencies in administrative tasks, which are often seen as non-value adding. He says that the goal should be to reduce secondary work that healthcare staff engage in, which does not directly involve patient care. By minimizing these tasks, clinicians can focus more on care related to direct interaction between patients and themselves. Consequently, van den Brink mentions five administrative tasks that could benefit from automation through LLMs: creating various journal notes, writing different types of referrals, issuing certificates to insurance funds, producing informational documents for patients post-visit, and preparing summaries for clinicians pre-visit about the reasons for patient consultations.

4.1.2 Exploring the Potential of LLMs at Region Halland

At Region Halland, Markus Lingman is at the forefront of analyzing the potential impact of LLMs in healthcare. Lingman says that his team has experimented with functionalities like summarizing medical journals and automatically generating epicrisis. The initial results have been promising, with the technology performing well, although it still requires oversight by medical professionals for technical accuracy and legal compliance. This validation process ensures that outputs meet the stringent requirements of medical practice, though Lingman views this as a realistic path forward in integrating LLMs into daily healthcare operations.

Further, Lingman mentions a commercial company in Sweden that has developed a product closely aligned with the potential use cases of LLMs identified by van den Brink. These products include ambient scribes using GPT-4, which actively listen and transcribe medical discussions in real-time as they occur within patient rooms. The transcriptions are not only recorded but also converted into formal medical journal entries in Swedish. Lingman says that this system can automatically generate documents for sick leave and referrals, such as to radiology, based on the content of the conversation, streamlining these administrative tasks significantly. He says that they have tried the product in Region Halland's acute care, however, he emphasizes that it is a product from a commercial company.

Furthermore, Lingman notes that Region Halland have begun to explore the potential of providing clinicians with a summarized patient reference pre-visit, especially useful when encountering less familiar patients. The aim of the summaries is to offer quick access to guidelines and recommendations tailored to individual patient needs, although Lingman says that this application is still in its early stages and does not work well enough yet.

In discussing the prioritization of such technological initiatives, Lingman emphasizes the importance of addressing the most pressing pain points in healthcare. He says that the principal challenge today is the scarcity of time from highly qualified medical professionals. By automating routine tasks, such as reporting to quality registers and summarizing patient interactions, LLMs can free up valuable time for these professionals, allowing them to focus more on direct patient care rather than administrative duties.

> *"One should focus on the pain points in healthcare. The biggest pain point in healthcare today is the availability of qualified workforce time. Doctors and nurses with special skills that are currently insufficient in Sweden or the world for that matter. Replacing work for them is the first important step. Then, this can be gradually expanded to many applications. For example, automatic reporting to quality registers, which is done manually today. Thousands of work hours from highly qualified employees. If a solution to this were available, it would be quickly integrated into healthcare. Summarizing information both before and after a patient contact, is again time-consuming work for highly qualified employees. That would be something to reduce. Those are the biggest issues. One could say that the greatest measure of efficiency is time. Wanting to save time for these highly skilled doctors. Employee time for qualified labor is a priority."* Markus Lingman

4.1.3 Risks, Benefits and Future Directions

Despite a potential bright future for successfully implementing LLMs in healthcare settings, the experts also deliver contrasting views highlighting barriers. van den Brink identifies three major concerns:

1. Transparency: there is insufficient clarity about how LLMs make certain decisions and the question of who is accountable when AI makes an untraceable decision.
2. Experience accumulation: the impact on clinicians' ability to gain experience is questionable if tasks are partially or fully automated by LLMs.
3. Decision understanding: there are risks involved if LLMs make decisions and clinicians can not comprehend how those decisions were reached or why certain actions were taken with AI's assistance.

Another expert highlighting risks such as insufficient clarity in decisions and potential impact on clinicians' experience curve is Rosita Ahlstedt. She put emphasis on the importance for clinicians to feel secure about what they are bringing to the patient, which can be damaged by a tool where certain decisions can not be explained. She worries that using AI in healthcare might cause physicians to lose their decision-making abilities and not develop the skills that today's physicians acquire through personal expertise and judgment.

> *"Let me give you an analogy. If you drive a car using GPS navigation, think about when you arrive at your destination. You might find yourself saying, 'I am already here. How did I get here?' The active memory of the specific driving route is not as strong when you are the one behind the wheel. Now, consider this in a healthcare context. If there is a safety issue or something goes terribly wrong during a patient's treatment, we might ask the clinician, 'What were you thinking?' The clinician might respond, 'I actually do not know. It was the decision support system that suggested that course of action.' However, just like with driving, the clinician is ultimately the one 'behind the wheel,' not the decision support system. While such systems can provide recommendations, the clinician is still responsible for actively monitoring and making crucial decisions throughout the treatment process. They can not simply defer responsibility to the technology."* Hans van den Brink

Despite the acknowledged risks, van den Brink and Ahlstedt maintain a positive outlook, confident that the benefits of developing LLMs far surpass the potential downsides. van den Brink advocates for a methodical approach to addressing these risks, tackling them one by one and Ahlstedt says that Region Halland would never continue the research regarding these tools if they were not under the impression that the gains will outweigh the risks.

Additionally, Ilicki points out challenges specific to implementing LLMs within the Swedish healthcare system. Beyond regulatory and technical barriers, there are also commercial barriers. Many LLM use cases will require integrations with health data in electronic health records (EHRs) in order to generate large time savings. However, EHRs are often procured for long periods of time, and tenders can't include future technologies which aren't known at the time. When Sweden's EHRs were procured, several years ago, no provider could have foreseen or demanded that an EHR be ready to integrate with LLMs (which didn't exist then). This is why integrating LLMs to EHRs will be a new request from providers, and something that will need to be prioritized among other requests.

> *"For many healthtech solutions it is difficult to gain access and integrate with EHRs, simply due to limited incentives for the EHR. However, if and when customers see significant efficiency gains due to LLMs, and start prioritizing it highest among other requests to their EHR provider, then we'll start seeing integrations… I believe we'll see significant time savings, similar to what has been seen in the US when LLMs for admin have been implemented. And then we'll start seeing integrations, partnerships and acquisition."* Jonathan Ilicki

Furthermore, Ilicki highlights regulatory challenges in developing medical software, which also affect LLMs for healthcare. For example, issues such as predictability and reproducibility are tangible obstacles that must be overcome through effective validation methods and ensuring that the LLMs stay within specified limits or datasets. Ilicki underscores that tackling these hurdles is crucial for the broader adoption of such technologies in healthcare.

Michael Page, Derek Beaton, and Amrit Krishnan all highlight the issue that the technology behind LLMs is still novel, but is regardless rapidly entering the industry.

> *"To be honest even in the most top tier hospitals that we work with, we have not yet seen LLMs being deployed yet. They are still in research and they are still being assessed for these risks. People are still trying to understand how do the benefits compare with the associated risks."* Amrit Krishnan

In Beatons discussion on the integration of LLMs into clinical practice, he identifies three primary concerns. First, he stresses the importance of trust in the technology among clinicians. Without trust, the adoption of any new technological tool is unlikely. Second, he highlights the necessity of time efficiency; the technology must save time for clinicians, or it will not be utilized. Third, he points out the challenges related to accuracy, noting the current difficulty in verifying the correctness of LLM outputs.

However, Beaton recognizes the significant administrative burden faced by family clinicians in Canada and suggests that LLMs could alleviate some of this. *"Maybe some aspects of LLMs could help here, but probably a mixture of a person and an LLM working together. So that family physicians can put more time into practicing instead of filling out forms."* Derek Beaton says. Nevertheless, he advocates for a cautious approach to using LLMs in clinical settings, suggesting that other technologies such as classification tools and optimizing algorithms might also hold potential. Discussing the use of a generative component, Beaton says:

> *"In many healthcare cases that have come up as potential ideas make us incur more time on staff. Because if the point is to summarize a bunch of notes, and you also need to go back to these notes, you add another step. So there is a lot of potential, but I am cautious, especially due to what we have seen out in the wild."* Derek Beaton.

Lingman holds an optimistic view regarding clinicians' acceptance of new technology, noting that clinicians already use advanced technological tools daily. However, he also emphasizes the necessity of proper validation but believes that once LLMs have undergone these procedures, clinicians will readily adopt them.

> *"But getting healthcare staff to accept new technology is not as problematic as some perceive, because we already use a lot of technology in*

> *healthcare. It is more about how to build trust in the new technology."* Markus Lingman

Furthermore, both Page and Krishnan raise concerns about the occurrence of hallucinations as a significant issue when considering the deployment of LLMs in healthcare. Page explains that during the testing of LLMs within healthcare environments, there have been instances where the models incorrectly mix up various symptoms and conditions.

> *"And in healthcare, we have seen hallucinations where the LLM will confuse something like hypoglycemia and hypothyroidism. No clinician would ever make that mistake. It is not that it is just like, oh, it made a mistake. A clinician would never make that mistake, right? So that is part of what we get concerned about."* Michael Page

However, Page also put emphasis on the increasing demand on clinicians in Canada to do administrative tasks. Like other experts previously mentioned, he sees potential in the use of ambient scribes to alleviate these demands, noting that this technology could significantly reduce the hours physicians spend on such tasks each day. However, Page stresses that clinicians must review and ensure the accuracy of the outputs produced by these technologies.

4.1.4 The Challenges of Validating LLMs

A challenge highlighted by several experts is the potential difficulties with validating the performance of the LLM technology. Beaton said that today, there is no good way of evaluating and verifying that the information a LLM provides is accurate. Further, he emphasizes that before you put anything into practice, there has to be a way of verifying that the information a product provides is correct, which he says could be difficult without involving clinicians. Beaton notes the impracticality of extensively involving physicians in the initial testing phase due to time constraints and emphasizes the need to automate part of the validation process. Moreover, van den Brink discusses the current issue of validating LLMs and suggests, by referring to a previous research, that academia and university hospitals should collaborate in finding a solid solution for addressing this issue.

> *"So, for an example: Me or someone in my team is developing a LLM for some task in the medical setting, like summarization aspects, on typical medical conditions, I have no idea to see if the answers is correct. And I am not sure if I can get physicians to sit down and use it for a long time before it will be rolled out, it is a lot of time that needs to be required. So we need to find a way to automate some things, but we don't have a sense of how we automate things before it goes out. And then, how do you keep track that things are still okay?"* Derek Beaton

In a discussion about model accuracy and validation, Page highlights the Retriever-

Augmented Generation (RAG) as a notable tool because it provides physicians with sources of its information, enhancing transparency. However, he questions whether it fully addresses the foundational issues, such as the time required for tasks like generating discharge letters.

Furthermore, Krishnan notes that LLMs have not yet been fully deployed in practice and remain under research to evaluate potential risks and benefits. He emphasizes that methods of evaluation are in their nascent stages but are improving, especially with new datasets emerging in the medical field for better benchmarking and assessment. However despite the advancements, challenges persist if LLMs are implemented into different hospitals that use different terminologies and codes. A way of going around this is to fine-tune the model into these different settings, however, that will complicate the benchmarking process further, according to Krishnan.

4.1.5 Ethics and Legislation

In this section, the experts' perspective on LLMs in relation to ethics and legislation is presented.

4.1.5.1 Machines versus Clinicians

> *"I think we always will, and one could argue that we should, place higher demands on machines than humans when it comes to these tasks. I say "will", as we are not prepared to accept technology that is as bad as we are, and that is because we underestimate how bad humans are at applying our best knowledge to individual cases. And I say "should" as we should place higher demands on machines that have a greater scale and potential impact than any individual human. That is why I expect higher demands."* Jonathan Ilicki

All experts agree that we should have high demands on machines that are used in patient care and that machines performances should exceed that of humans. Some experts further raise that more and more research that compares clinicians with LLMs indicate that LLMs perform better than clinicians for some healthcare tasks, and still better alone than combined with clinicians. If this were to become a fact, it would cause another ethical problem.

> *"How long should the human being, the average physician, be allowed to influence the results from the optimized language model, which is then trained on the golden standard at best. No one knows the answer to this yet."* Markus Lingman

In the current, early state where the technology of LLMs is now, there are uncertainties about what the demands are on the technology to be implemented in healthcare. Kjellberg describes that it sometimes feels like the requirements are one hundred percent on LLMs performance. Ilicki believes that it is not unreasonable that LLMs will reach in principle 100 % accuracy over time, it is just a question of how far away it is in time. However, the expert emphasizes that it will make mistakes along the road to get there. Ilicki compares the transition towards auto generated medical notes by LLMs with the transition to generating medical notes, first with medical secretaries and then with speech recognition.

> *"I would argue that it is comparable with the transition to secretaries. There was a time when you as a physician had to write down everything yourself. Then you knew what you wrote was right. Then, someone came up with the idea that it would be more efficient if someone else did the writing, so the physicians could prioritize seeing patients. So we got medical secretaries, who listened a little to what the physician said, wrote it down, and then the physicians still had to double check if it was right. It does not always turn out right when you dictate in that way. Then came speech recognition. Similarly, the physicians dictates and then the system automates text. These systems also make mistakes, and the physicians still need to double check the result. LLMs is the next step in this evolution, but clinicians will still maintain control and need to sign off on the final output in order to be responsible for it."* Jonathan Ilicki

Ilicki believes that clinicians will have to validate AI outputs for legal and practical reasons, but does not think it will be a big obstacle. Ahlsted thinks that machines and clinicians will have to cooperate, and describes this situation as a "one plus one equals three-situation".

4.1.5.2 Who Takes Responsibility for AI Generated Outputs?

> *"I remember once in the Netherlands, there was a truck that drove into a river. The police asked the driver, 'What were you doing?' It was nighttime, I must add. The driver replied, 'The GPS instructed me.' So he drove into a river.*
>
> *But that is a very poor defense, is it not? You can not use that as a defense.*
>
> *And in healthcare, it is the same. What is the organization's actual responsibility regarding this type of tool?"* Hans Van Den Brink

The topic of responsibility is brought up by many of the experts. Who will take responsibility for AI output? There is no clear answer to this today, not in research or in practice. However, if LLMs are to be introduced in the healthcare sector, answering this question is of top priority.

> *"The legislation we have today is based on the premise that there should be a person to blame when things go wrong. If a physician makes an obvious mistake, they have to face the consequences. But if a model makes a mistake, who is responsible then?"* Markus Lingman

There is a consensus among the interviewees that the healthcare professionals will bear some responsibility for AI generated outputs they use. That clinicians need to be responsible to review, correct and approve AI generated text. Ilicki argues that the technology or software providers are responsible for products' performances, but that clinicians are still ultimately responsible for the usage of approved products in patient care.

> *"Healthcare is to some extent about bearing risk. Technology and software suppliers can not bear certain risks, instead they are borne by the clinicians. If you have a syringe that is used as it should, but something strange happens, the physician is still ultimately responsible given that the syringe is built as the syringe should be built."* Jonathan Ilicki

Lingman agrees, but adds complexity to this discussion by involving the aspect that clinicians often work under time pressure, making it easier for inaccurate data to be approved.

> *"You always have to have a physician's signoff so that the physician is still responsible for the content, diagnosis, or summary, or whatever it may be. The challenge here is that the healthcare is so stressed. Physicians and nurses are under time pressure all the time. It is so easy to buy into what you got from the language model if you think it seems good. You are nudged into a certain behavior, which is a bit problematic."* Markus Lingman

Beaton provides two real life examples, outside the healthcare sector, that puts light on the question of responsibility for AI generated outputs, and underlines the need for direction.

> *"Air Canada airline had an LLM chatbot on their website and one customer had ordered plane tickets under the bereavement policy. A family member had died. So the customer had used the chatbot to ask about the bereavement policy and the chatbot gave it all sorts of wrong answers. And then the customer sued Air Canada and Air Canada was found at fault for the LLM giving wrong answers. And then the person was compensated.*
>
> *So it was interesting to see that at first, these things are put in place without really any oversight or checking. And second, we have formally crossed that line where someone now has to take responsibility for this."* Derek Beaton

> *"The second version of this, which is unfolding now, and we do not know how it is going to pan out, say, in the legal sense, but New York City just put a chatbot up on their website so that people can ask questions about laws, bylaws, regulations. And it is giving not just verifiably wrong*

> *information, but in some cases, illegal information. So for example, if you were to ask it, as a landlord, I have a tenant that I do not want living here anymore, can I just kick them out? It will say yes. And you can not actually do that. There are rights tenants have and there are very strict legal steps you have to take in order to evict someone. [...]*
>
> *And then someone's going to have to take the blame and we will find out if it is New York City or if it is the landlord or something."* Derek Beaton

4.1.5.3 Regulations

> *"We already have quite a lot of regulations to deal with and there will be even more. But I think that in healthcare we live in such a complex environment already, with many different regulations and ethical balances and stuff."* Magnus Kjellberg

Due to the novelty of the LLM technology, there is a lack of legal guidance of its usage in society, not least in healthcare. At the same time, older laws affect the area as well, like GDPR, laws on Medical Devices, etc. There is a significant need for legal guidance fast as the technology is evolving rapidly.

> *"One problem I would say is that we have so many different laws that must interact. We have laws at both national and EU level. There will also be more and we have not really adapted the national ones."* Magnus Kjellberg

There is a clear need to sort out if certain LLM applications will be defined as medical devices. For LLM applications to fall under this legal framework they have to affect patient treatment. While LLMs that are used as support for medical decisions or predictions clearly do, it is not as clear if LLMs that are used to streamline administrative tasks and involve patient data do.

> *"And what determines whether you end up there is if you do something that affects the patient's treatment, then it is basically a medical device. There is a bit of a gray area there. But basically, that is where the limit goes. If it can affect a patient's treatment, then it needs to be tested super carefully.*
>
> *Our problem is that, if what we do does not affect the patient, then it might not even make sense to do it. So I think we probably need to expect that we will end up there."* Isak Barbopoulos

Some of the experts highlight the need for guidance regarding cloudbase processing. Cloudbase processing is not allowed for Swedish healthcare settings today due to the risk of losing patient data to commercial actors, which contravenes GDPR. However, it is currently the best option for implementing LLMs in relation to performance.

Until anything else regarding cloudbase processing is decided, Swedish healthcare actors need to rely on open source models for LLM applications that can be used on internal servers, which currently performs worse than commercial models, or more specifically, than Azure OpenAIs GPT-4.

> *"There are some difficulties. One of them is purely legal, then the question is, can we? For example, are we allowed to use these Azure OpenAI models? For they are in the cloud. It is not certain that we are allowed to have patient data there. If we are not allowed to do that, we will use Open Source models like Llama, because they can be on our own servers. So we kind of have a back-up plan. That is how we start now."* Isak Barbopoulos

> *"We are not allowed to use cloud processing in the regions right now because we handle sensitive information. Legally, it is a bit uncertain. We do checks and are denied to use such public services. But we also know that technically speaking, that is where development is fastest. So there is something that is hindering the regions today. That we can not use the technical possibilities that are always available. We are forced to look at local tools, local models. Then we need to have the computing power internally which costs a lot."* Rosita Ahlstedt

LLMs require enormous computing power during their training. Local implementations would require huge investments for Swedish healthcare actors. However, the training phase is limited and could be done in a couple of months, leaving the vast processing power unused when the model is fully trained and has been put into production. If cloud-based services were allowed, it would have been possible to only pay the high costs for increased processing power during the training phase, and then no longer when the model is fully trained.

> *"We believe that the cloud would have been a facilitating factor. Because then we can set it up for three months on a cloud service, and we might incur a high cost, but then we have no cost for the rest of the year. If we invest in computing power to train it, we will have a cost for several years to come. And then it is crucial that we use it a hundred percent of the time, otherwise, we have unused capacity that we are paying for. So it is a bit of a hurdle for the regions, actually."* Rosita Ahlstedt

In the United States, which has a different set of regulations to deal with, these barriers are not as evident resulting in faster application of LLMs tools in the United States. This risks the LLM market moving from Europe to the United States. However, two experts highlight that Sweden has a different health care system compared to the United States, which means it does not translate completely. The United States has a more business-like way of running healthcare and the technology development they do in hospitals. They also see an opportunity to sell their AI algorithms, which Swedish actors do not.

4.1.6 Implementation

This section presents the case of Unity Health Toronto which has been successful in implementing LLM in healthcare. Furthermore, the section describes key factors for implementing new technology in healthcare, factors that can be applied to for LLMs.

4.1.6.1 A Successful Case of Implementing AI Tools in Healthcare in Canada

Unity Health Toronto has been recognized around the world for its successful implementation of different AI tools in healthcare. It has been acknowledged that the organization has some unique characteristics compared to traditional healthcare organizations when it comes to implementation of software technology, which experts from Unity Health Toronto regard as fundamental for its success.

Unlike other hospitals, Unity Health Toronto decided not to focus on traditional research, instead they decided to be on the opposite end of the spectrum and focus on the use of innovation, technology and data science in an applied way. From day one, their mission statement has been: how can we use data to help patients and help clinicians?

> *"So we were never going to be a group that was contributing significantly to fundamental or basic research. But we wanted to be the group that figured out the really messy side of the problem and that is how do you build a bespoke AI tool and integrate that into a clinical setting and ultimately get clinician feedback from nurses and physicians and the interprofessional team. So I think that part of this is our mission statement, part of this is the culture of focusing on applied work rather than research work."* Michael Page

One of the first decisions the group did, was to establish a contemporary data governance process and policy to build out their data infrastructure. Page underlines that this was really, really important, because they are dealing with the hospital's most sensitive data. They need to be aware of regulation. It is a massive privacy and security concern. They hired independent third party consultants in a wide variety of disciplines, everything from legal advice, advising on regulation of what you can and can not do with patient data, to cybersecurity and privacy experts, advising on what are the best policies, how to train whether any additional insurance policies or checks and balances were required. It was a significant financial investment. They thus started with a strong foundation in governance, and they continue to have a commitment to that and regard it as ongoing with continuous improvement, not just a one-time-thing. Besides this, they spent time talking to patients asking how they felt about the use of their data and about AI, and they have continually involved dedicated ethics boards.

> *"We are building the plane as we are flying it because there is a massive*

> *absence around other hospitals that are doing this work and a lot of kind of vague direction in any regulation. Organizations are required to build their own framework, and I think that ours is quite robust, but we have had to do it independently."* Michael Page

Unity Health Toronto eventually formed four dedicated teams that work closely together to apply AI in healthcare. A data team that will get the data from the different systems, making sure it is usable for the advanced analytics team, who takes over and builds the model. When the advanced analytics team is done with the model, the product development/software team puts it into production. Meanwhile, the project management team is keeping track of everything.

> *"I do not know of any other hospital that does this. When I have talked to a lot of other places, they might have a data team. They are getting data, and they will just make data available to other people. But they are not working together in one place. The data team might be part of research, or they might be part of IT. And when they get a request for data, they just go give people data."* Derek Beaton

In addition, the organization has a hands-on way to identify what hospital needs they should focus on solving. They work closely with the hospital employees throughout their projects, putting emphasis that the need should come from them, as they are the end users. To identify what needs they should focus on, Unity Health Toronto's The Data Science and Advanced Analytics (DSAA) team sends out an intake form that is filled out primarily by someone who is coming to them with a problem. The form collects what the problem is, what the future state would look like, where data might be, etc. The team further compares the information that has been collected in the forms with a background of existing literature and knowledge as well as their own data, to validate the size of the problem and of the potential gains from solving it, to then be able to prioritize fairly what will give the highest value.

> *"It starts with this intake form and it starts with really candid conversations like, what is the problem that you are experiencing right now as a clinician? We have proposals predominantly from clinician scientists who are physicians. But we get them from nurses. We get them from some people within inter professional teams. We also get administrative examples, we have gotten some from human resources and other internal groups as well.*
>
> *And they fill out the form, they describe in plain English what is the problem that they are experiencing. And, you know, we talk about it and we are like, you know, AI might not be the solution. Or we have a really detailed conversation where it is like, oh, this is really interesting. And you have begun to imagine what a future state solution could be. Let us have some conversations and dig into feasibility a bit more. But the actual problem, the place where we start, it is very important that it comes*

from clinicians and that it comes from the end users." Michael Page

However, the Canadian team is currently not implementing LLMs in healthcare in the same ways others currently intend to. Page explains that Unity Health Toronto uses LLMs, but not related to the clinicians/patient interface where most people currently are fixated. Beaton confirms this, saying that Unity Health Toronto uses some LLM architectures, but that they do not use it for anyone else's use, but for their own use, and that they are not using LLMs as a generative tool or a summary tool, but as a classifier.

Beaton adds that they are experimenting with generative LLMs a little with a very focused task that is non-clinical in order to 1. try to bring something into production and, 2. find out what the challenges and barriers are there. It is something that is helping them go through a small set of documents that are available to everyone in the hospital. He underlines that they are seeing results that seem right, but also that it sometimes makes up an answer that does not seem right, and they do not currently know how to evaluate whether it is actually giving a correct or incorrect response unless they are reading it.

> *"I think there is a lot of potential, but I think the barriers are much, much larger than most are expecting. These tools have grabbed a lot of attention. They are very powerful in part because of just how much data have gone into them. And for us so far, what we have seen is that when you switch out the generative component for a classification component, they work exceptionally well on classification tasks. So there is still, in the giant complex architecture of that neural network, critical information that could be used for something besides a chatbot."* Derek Beaton

4.1.6.2 Key Factors to Implement New Technology in Healthcare

All experts agree that the needs for new technological solutions should come from the hospital employees, the end users. Carolyn Chong explains how they, at Vector Institute, have a tight relationship with the hospital network in Toronto. Vector Institute faces two types of users whom they interview in relation to their projects. First and mainly the hospitals' data scientist teams that are deploying the machine learning models and who are the ones that are actually using Vector Institute's tooling. Second, the clinicians who are using the outputs of the models to make decisions. As mentioned above, Page and Beaton describe how Unity Health Toronto has a close relationship within their hospital network and identifies the needs formally with the intake form described above.

> *"And what is really, really important when we are presented with those things is that we are really hearing from the individuals that are going to use this. And so we will go and conduct user testing or user feedback sessions of, will you use this if the tool does X? Because we might have a more theoretical, optimal way of solving something, but if no one is going to use it, who cares? So we are very much, I would say, more than 80 %*

of the time leaning towards making decisions based on our user group." Michael Page

Furthermore, Ilicki argues that it is important, especially with new technologies in healthcare, to show benefits for the individuals who use it. Lingman agrees with this statement and describes access to time of qualified labor as the most critical issue.

> *"But these products will be loved as soon as they have gone through validation processes and you have understood how much time they can save in their everyday life, like being able to pick up from daycare on time for once, or that you had time to eat lunch today. It is things like that that count. It is amazing how the day you make the connection between a full stomach and that you actually had time to say goodbye to the patient in a dignified way, then you have found something valuable and then the health care will pay."* Markus Lingman

> *"Saving time for qualified labor is priority 1A."* Markus Lingman

Chong and Beaton further support this view. Chong mentions that time saved is one of the main KPIs so far when evaluating the use of LLMs in healthcare, together with patient outcomes and patient mortality. Beaton expresses that if LLMs are not saving time, people are not going to use them.

Furthermore, Ilicki argues that change management generally speaking in healthcare requires that you have resources set aside for it and have the right anchoring based on existing needs across different levels in the organization. Besides this, the expert believes that it is important that the change leaders have a clinical background and a clinical anchoring to be able to speak the right language. He further believes that it is important that, in order to speed up the implementation, you do clear evaluations and make early gains and make sure that you can show it like, "look at this clinic, they have started using it, and it saves a lot of time this way".

Ahlstedt highlights the importance of being aware of and planning for future maintenance for a software product that has been put into production. Software products, like LLMs, require ongoing work. Ahlstedt explains how maintaining these types of products is clearly embedded in their typical workflow.

> *"Once you have done something and put it into production, the business does not always think about managing it over time. It is not certain that the AI service will be as good in a year, or two years, or six months. We do not know that. How soon until we might even need to retrain that AI algorithm? Because the input has changed, the patient structure has changed, new medications have been added that mean we can no longer use the version we have. It is not as reliable anymore."* Rosita Ahlstedt

Ilicki sees a clear barrier when it comes to integrating new technology with existing systems. He believes that this is partly due to lack of incentives for the existing system vendors until the customer clamor for it. Partly due to that large and complicated systems, like the health record system, are procured in large procurements that are certain for many years to come and it is often difficult to set requirements in advance for technology that does not exist, and it tend to be a lot of bugs that need to be fixed, which are prioritized. Therefore, new functions are often lagging behind. To overcome this barrier, Ilicki argues that the healthcare actors, the customers, need to put clear demands on the health record system suppliers, and thus create incentives for them to integrate LLMs. He believes the customers will do this if the use of LLMs in healthcare shows large time savings.

Ilicki put emphasis that innovative suppliers that have the outside-in perspective need to find the right champions within the organization, that is, employees who are passionate about the issue and want to improve healthcare and sees the value of the technology. Page and Krishnan also express the value of identifying champions within the organization for implementing new technology.

It has further been acknowledged in some interviews that a few private actors have already started to implement LLMs and seen early gains. It is argued that the regions are more rigorous in their demands on LLMs compared to some private actors. It will arguably be difficult over time to retain this difference in the healthcare system, because those who implement LLMs will get a significant head start. It becomes an ethical issue rather than a legal issue in the end.

Lastly, Page believes that the best way to move forward when it comes to improving healthcare through innovation and technology is to collaborate and work together.

> *"My advice or my hope is we find new ways to collaborate and work together. There is a period of time where using technology or being an innovator feels like being in a competitive race, and I hope what we see during this next phase of AI usage and development is greater collaboration, and collaboration that could not exist before. Because I do not think that it is going to be one hospital, I do not think it is going to be one clinician, that is going to solve this. I think that it is going to take all of us working together in new ways to reimagine what the health systems globally could be."* Michael Page

4.2 The Neurology Department

In this section the results of the case at the neurology department is presented. The section is divided into two parts, the needs of physicians at the neurology department and the needs of the nurses at the neurology department.

4.2.1 The Needs of Physicians at the Neurology Department

The results below are based on data from eight interviews with physicians at the neurology department at Sahlgrenska University Hospital, four residents and four senior specialists. All eight work with out-patient care and in-patient care. Two of the senior specialists are surgeons and thus work with surgical care. The results below are also based on data from a full day observation of a senior specialist at the stroke department.

4.2.1.1 Everyday Text Based Tasks Performed by Neurologists

At the neurology department, the physicians perform a number of text-based tasks. They write different types of health record notes. When they work with out-patient care they write clinical notes and when they work with in-patient care they write admission notes, daily progress notes, discharge summaries and discharge letters. The neurologists further receive and write referrals and consultation notes. The surgeons write operative reports after every surgical procedure.

Besides writing notes in the health record system, the physicians at the neurology department write certificates to various external bodies, fill in registers for quality and research, and study each patient's health record before encountering the patient. Some physicians have further responsibilities to write additional documentation like routine documents and deviation reports. All physicians are further obliged to read through, correct and sign all documents that are published in their name, both documents that are conducted by themselves or by someone else, like medical secretaries.

4.2.1.2 The Perspective of the Neurology Physicians on Their Text Based Tasks

> *"It is always a mystery how much time disappears. At the end of the days, you realize that you have met the patients very little. You have been sitting at the computer all the time and it is difficult to understand exactly what you are doing. But a lot of time is spent waiting for various programs to open, clicking, getting there and logging in, opening tabs and so on. Then you produce text and summaries."* Resident, neurology

4.2.1.2.1 Time-Consuming and Troublesome Tasks There is a consensus among the physicians that the time they have for administrative tasks in front of the computer is too short, but that the amount of time that these tasks require is unreasonably much. All physicians agree that there are clear bottlenecks that should be eliminated to minimize the time they have to spend in front of the computer with administrative tasks and free up time they can allocate to patients instead. They further agree that there is clear potential to do this. Table 4.1 below presents the amount of time the eight physicians estimate that they put on administrative text-based tasks on a normal day of clinical work at the neurology department.

Table 4.1: Estimated time physicians spend on administrative work in average on a normal day of clinical work at the neurology department.

Professional title	**The estimated time spent on administrative work on one day**	**Is the amount of time reasonable or should it be more or less?**
Resident	60 %	*"It should be less."*
Resident	30 %	*"It should be able to be less."*
Resident	Did not get the question.	Did not get the question.
Resident	90 %	*"I think clinical decision-making and examination of the patient should make up a much larger part."*
Senior specialist	10-20 %	*"The problem is finding the time to do it. So really, you would need more time set aside for such tasks. I am a scheduler too. So I have those types of tasks as well that I do not have allocated time for. So that is the downside. But otherwise it is relatively reasonable, I think."*
Senior specialist	30-40 %	*"Obviously I think it should be less."*
Senior specialist	*"It is a difficult question to answer. A very large part of working life is spent reading journals, yes. And to work in the records system."*	*"Everything is definitely not reasonable."*
Senior specialist	At least 50 %, probably more.	*"In a way, it should be less, I think. But if I have to carry out the tasks imposed on me, it is not too much time. The way we work today, it is hard to see how I could do it that much faster. But I can also find it unreasonable. A lot of time goes into such things."*

Furthermore, there is a consensus that the time required for text-based tasks highly depends on the particular patient and the complexity of the case. For many tasks, what takes time is finding the correct and relevant information. One of the residents estimates that he typically dictates a discharge summary in 10-15 minutes, but underlines that there can be great variations when completing those tasks. A simple case, like a patient who has done an MRI scan that showed good results and was allowed to go home, can take only two minutes to dictate. However, getting an overview, reading and summarizing a case where the patient stayed at the hospital for five months and moved around in different departments can take an hour to dictate.

One senior specialist adds to this and explains how she, the same day, studied one patient's health record for 20 minutes, some other patients' health records for 5 minutes and someone for only 2 minutes, only glancing through the health record realizing it was a patient she was familiar with from before. Another senior specialist underlines that it clearly varies depending on how well they know the patient. Writing a discharge summary for a patient they have been working with themselves and thus know well, is easier than writing a discharge summary for a patient that they have not met. In those cases they have to read through the health records to understand the case and to know what their colleagues have done and assessed for the patient, which can be hard.

A second resident further describes how admission notes can be time-consuming, not because it takes a long time to create the note itself, but because it can take time to find the right information in the different, complex systems about the patient's current state and especially which medicines the patient is currently taking. The physicians can find information in three different systems to update the medicine list: Pascal (Apodos), a national system via the pharmacy; NPÖ, the national health record system that contains information from a number of health centers and regions, but not all; and Melior, Sahlgrenska's health records system that contains information documented from patients' visits at Sahlgrenska.

A third senior specialist provides a similar picture as his colleagues above of how text-based administrative tasks vary greatly in relation to the time it takes to complete them. He explains how an average referral takes 10 minutes to handle, but that it sometimes takes twice as long and other times only 2-3 minutes. Sometimes it can take 10-15 minutes to just find relevant information. Information can be missing or be incomplete and sometimes the physician needs to ask the remittance to complement the referral, which prolongs the entire referral process, etc. He however, underlines that the greatest bottleneck is their digital systems. Physicians have to go in and out of different IT-systems to retrieve the information they need. The different systems require multiple steps, can only be opened one at the time and are often lagging, making the physicians need to wait for them to respond.

All eight physicians agree that the complex and inefficient IT-environment they currently work in, together with traditional processes with faxes and letters, is the largest bottleneck when it comes to time spent on administrative tasks.

> *"The biggest problem is that there are so many inflows. I have to actively seek out all the information myself. So even if I have not received anything, I will have to spend a few minutes checking, on multiple different*

> *locations, so that nothing new has come."* Resident, neurology

> *"Getting to the information is the biggest bottleneck. The programs we have today are slow and the paths are long. You have to click forward. It may not always be that something loads, but many times that you have to click back and forth and can not have it open at the same time."* Resident, neurology

> *"What I wish for is a health records system and interconnected systems that talk to each other in a functional way."* Senior specialist, neurology

They underline that having to receive and send faxes and letters for some types of documents is time-consuming and causes unnecessary complexity. One of the senior specialists describes how these traditional ways to communicate information can be bad for the patients' safety due to long waitings for the information to reach its recipient.

Furthermore, the physician's experience is also regarded as a factor that affects the time it takes to complete administrative text-based tasks. The four senior specialists all express that they complete their text-based tasks faster than they did earlier in their career when they had less experience, and that they have learned to better prioritize what to do among the huge amounts of tasks they have. All four believe that physicians with less experience in general need more time to complete their text-based administrative tasks, partly because they need to learn by doing the job, partly because they have a harder time to prioritize and therefore do everything, which takes longer time than what they often have.

A senior specialist explains how the senior specialists delegate discharge summaries to residents so the more experienced physician can use that time for other tasks, and for the residents to learn. He notes that it takes longer for the residents than it does for the senior specialists to write discharge summaries. Further, it was observed at the stroke department that senior specialists act as mentors for junior physicians. Additionally, it was observed that the physicians at the department cooperate to a large extent, often asking each other questions about different patient cases they are handling.

It should be noted that all eight physicians from the neurology department dictate almost all their notes, with the exception of some short notes like daily progress notes, shorter referrals, like to the radiology department, and certificates and forms. All agree that it goes fast to dictate notes, and that time-consuming tasks depend more on the complexity of the case as explained above. However, some physicians describe that they can not dictate discharge letters, because they have to give that directly to the patient when he or she leaves the in-patient care. One of the residents explains that writing discharge letters is time-consuming and stressful due to this characteristic.

However, dictating notes to secretaries is not done without problems. First, the physicians explain that it can take quite a long time to get the notes back from the secretaries, sometimes even weeks. Secondly, the physicians describe how the secretaries always make multiple mistakes. The secretaries can write "..." where they have not heard clearly what has been dictated, they sometimes write words wrong and they can miss negations or other important words which induce incorrectness in the text. It can be hard for the physicians to recall and correct the text if they created it weeks ago, forcing them to study the patient's health record once again to confirm that the right information is published. These problems are well known, therefore the physicians' are obliged to read through, correct mistakes and sign under all notes that can be published in their name.

> *"And I never have time to sit down and sign my notes. I have a mountain of signatures I have to complete piled up in my signature inbox."* Resident, neurology

> *"It could go wrong. But what is especially troublesome is that, if I now get it printed and I see that there is something wrong, then I write by hand which word is missing. Because sometimes they write dot, dot, dot, I do not hear this. Or a word is missing. Then it goes back to the secretary. Sometimes I go into the journal myself and change the writing to make it go faster. But sometimes I just write on the paper that I got printed that there is a mistake in this sentence, this is how it should be written. But then it ends up back to the medical secretary, and then the corrected discharge summary comes back to me again."* Senior specialist, neurology

However, dictating to secretaries is being phased out and replaced by speech recognition. Some of the physicians have started to try out TIK ("taligenkänning", speech recognition in Swedish), but are not optimistic. One resident explains that she uses it for daily progress notes, but that it is making many mistakes. A senior specialist has not started to use TIK yet, and believes that the idea behind it is positive, but that it does not work in practice and fears that the physicians' administrative workload will increase as secretaries are replaced by TIK. Another senior specialist has tried it out a little and agrees that it is making many mistakes, but believes that if it works well, it could be a smooth tool.

4.2.1.2.2 Tools and Characteristics Some physicians' believe that receiving a summary of a patient's health record could make their administrative text-based tasks easier, but it is of high importance that the summary contains correct information and does not leave any important information out. One senior specialist explains that a tool like this would be especially valuable for patient health records that are long and complex. It is clear that the neurology physicians sometimes need to get an overview of a patient's health record, and sometimes they are looking for specific things which some use Melior's search function for. Another of the senior specialists underlines that if information could be retrieved from the health records by the use of keywords, it is of great importance that there are references of where the information comes from.

> *"It is important that there is a clear reference to where the information can be found in the text, so that it is not just retrieved loosely. Because it is very easy for something that enters the health record and is repeated many times to suddenly become a truth, but which may not really be so. For example, a care provider has thought about whether a patient is an alcoholic, and then it is presented in the text that the patient is an alcoholic. So you should be able to go back and see where it comes from."* Senior specialist, neurology

Many of the physicians are optimistic about receiving a draft for discharge summaries and referrals, if the information can be trusted. It is however important that the draft can be changed and corrected if needed. The observed senior specialist thinks it would be good to receive a draft even if she has to double check the information.

> *"If I do not have to write, it would still be better."* Senior specialist, neurology

She explains that when the physicians reads a text that has been dictated, you many times read what you have dictated and thus, some mistakes can be missed. A draft from an external party could be easier to criticize because then you can look at the text with fresh eyes. Further, a resident and a senior specialist do not believe that receiving AI generated drafts of for example discharge summaries will inhibit the learning of residents who are learning the job by doing it.

> *"No I do not think so. I think the opposite actually. That would mean that you read through the entire process and you probably get more data points than you would have come up with yourself and then you are reminded that those parts come into play. It also depends on how well an auto-generated discharge summary turns out. Because if it were to systematically forget an essential part, it would affect the perception of what is important to include. I think it would align one's own understanding of the importance of the information that is important to bring. A better auto-generation would provide a better understanding and not prevent the learning."* Resident, neurology

The physicians explain that they do a lot of "double work" (performing tasks that have already been completed previously). In many of the medical notes they need to add the patient's background, which include information like known current and previous diagnoses, previous medical procedures, current medicines, etc. For a patient that has had multiple hospital stays, this type of information is repeated every time the patient comes to the hospital. Adding this in note after note, clearly becomes a "double work" for the physicians.

Two of the residents think that a way to make their text-based administrative tasks more effective would be to have a dedicated patient backgrounds as a health record feature, which some other hospitals in Sweden have (that have another EHR-system supplier).

> *"An obvious thing that you could do, which we actually had when I worked in [another Swedish city], is that the patient background is a separate template at the top of the health record where it says what diseases the patient has. So that you do not have to go back in the health record every time and read how it was now, then it becomes a game of whispers, like "on the emergency note they wrote that she had a migraine", but where did that come from?"* Resident, neurology

> *"For example, using a patient background that is elsewhere would streamline our text tasks. Often, a large part of referrals and discharge summaries is made up of things that have happened before. An effective patient background, using that would be helpful. A patient's background is something that is repeated extremely often in the journal and which makes it more difficult to find information in it."* Resident, neurology

The first of the two residents describes that this type of dedicated patient background should be updated regularly with the most current information about the patient. The second believes that if the information would be updated automatically, a physician should approve the update.

Furthermore, the physicians describe that the discharge summary is in large parts based on information that has already been added to the health records. It is normal to copy information about the patient's background and current state from the admission note for example. The plan ahead does however not build on existing information. The discharge letter is similarly built on existing health record data, but it is written to the patient, in an everyday, non-medical language in contrast to the discharge summary that uses medical language.

Daily progress notes generally add new information to the health records. Clinical notes/visit notes mainly contain new information, but could contain some existing, previous information as well if it is a revisit.

Referral tends to consist of existing information, except the question at issue. The neurology physicians usually writes a clinical note, followed by one or more referrals which partly contain the same information as the clinical note. A senior specialist explains that he sometimes writes radiology referrals first, copying what he wrote there and adds it to his clinical note, because he can copy the text from the radiology system, Agfa, and put into the EHR-system, Melior, but not the other way around.

> *"But it only underlines that there is similar information in two places.*

> *Usually written twice a day."* Senior specialist, neurology

It is clear that medical notes are highly individual for patients. However, the physicians describe that sometimes they write fairly standardized formulations for certain medical conditions or procedures. A senior specialist describes that it happens at the stroke department that the same things can be written for different patients in discharge summaries. Another senior specialist explains how some referrals always follow certain surgical procedures and are written in a standardized way. The same senior specialist further expresses that discharge summaries about shunts and biopsies contain especially standard formulations.

> *"Perhaps to have standardized formulations for discharge summaries for certain surgical procedures. It feels like you are doing the same thing many times unnecessarily. Then there are certain interventions where it is very individualized. But a large part, the large volume of our interventions are so similar that it could be more standardized, that you skim through and add something if necessary. It would go faster than having to dictate and get it written."* Senior specialist, neurology

4.2.1.3 Time Estimates for a Number of Text Based Tasks

Table 4.2 below presents time estimates that have been collected from the eight interviews that were conducted with physicians at the neurology department. Each physician that made one or more time estimates are presented in the table (note that not all interviewed physicians made time estimates).

Table 4.2: Time estimates by neurology physicians for reading EHR and different types of writing tasks.

Professional title	Resident	Resident	Resident	Senior specialist	Senior specialist	Senior specialist
Read EHR	20 min				Examples: 2 min 5 min 20 min	3-10 min
Clinical note	15 min					
Admission note			20-30 min			
Daily progress note		Short: 5 min Long: 10 min	10-15 min			
Discharge summary		Short: 2 min Long: 1h Typical: 10-15 min	30 min			10 min
Discharge letter			20 min			
Referral	5 min		20 min	Short: 2-3 min Long: 20 min Typical: 10 min		
Certificates			5-10 min			

4.2.2 Self-Completion Form at the Neurology Department

Table 4.3 and 4.4 below present the result of the self-completion forms that were distributed at the neurology department at two different occasions. Ten physicians filled in parts of the forms. The time presented in the tables are measured in real time by the physicians themselves.

Note:
- P.1 stands for physician nr 1, P.2 for physician nr 2, etc.
- S stands for student.
- D stands for dictate.
- ND stands for no dictate.
- - stands for dictate unknown.

Table 4.3: Data from the self-completion form at the neurology department, occasion 1.

Physician	**P.1**	**P.2**	**P.2**	**P.3**	**P.4**	**P.4**	**P.5**	**P.5**	**P.6**
Year of experience	20	6-7	6-7	12	S	S	<1	<1	1
Read EHR					25 min	20 min			
Create a referral	10 min ND	8 min ND		3 min ND	30 min ND		8 min ND	12 min ND	4 min ND
Create a discharge summary		6 min D	6 min D		45 min ND				

Table 4.4: Data from the self-completion form at the neurology department, occasion 2.

Physician	P.7	P.8	P.8	P.9	P.9	P.9	P.10	P.10
Year of experience	5	4	4	5	5	5	5	5
Read EHR				7 min	6 min		9 min	
Create a referral	1 min D	2 min D						
Create a discharge summary	7 min D	4 min D	6 min D	12 min D	13 min D	10 min D	7 min D	15 min -
Create an admission note				6 min -				

In table 4.5 below, the calculated mean values of above data in table 4.3 and table 4.4 is presented. Note that the student is excluded from the mean value calculations below.

Table 4.5: Mean values of self-completion form data.

Text-based task	Mean value	Nr of data points
Read EHR	7.3 min	3
Create a referral D	1.5 min	2
Create a referral ND	8 min	6
Create a discharge summary D	8.6 min	10
Create an admission note	6 min	1

4.2.3 The Needs of Nurses at the Neurology Department

The results below are based on data from eight interviews with nurses at the neurology department at Sahlgrenska University Hospital, of which two also work as coordinators. The group includes nurses that work only in in-patient care, only in out-patient care, and nurses that work in both.

4.2.3.1 Everyday Text Based Tasks Performed by Neurology Nurses

At the neurology department, the nurses perform a number of text-based tasks. They write different types of health record notes. In out-patient care they write nursing assessment notes, and in in-patient care they write admission assessment notes, nursing progress notes, ward round notes, discharge planning notes and nursing discharge notes. The nurses also perform risk assessments when they meet with patients at admission and can write post operative notes if they work with post-op. Besides writing notes in the health records system, the nurses at the neurology department fill in quality registers, they prepare care planning documents and they submit documentation to the municipality and aid assessors.

4.2.3.2 The Perspective of the Neurology Nurses on Their Text Based Tasks

> *"We work a lot with documentation. For example, we update care plans, we make risk assessments, write down what has happened during the day, fill in changed checks, we always have an admission assessment when the patient arrives. Almost everything we do boils down to documenting it."* Nurse, neurology

4.2.3.2.1 Time-Consuming and Troublesome Tasks All but one of seven nurses agree that they spend too much time on administrative tasks (the eighth nurse did not get the question), see table 4.6 below. The nurse who thinks they spend a reasonable amount of time on administrative tasks comments *"It is part of my job, so I do not know if I think it takes too much time."* Four of seven nurses underline that they should reduce the time they spend on administrative tasks to get more time with the patients and five believe that there is potential to do so.

> *"I get to sit almost every day after work and document. That is not what is prioritized. If a patient needs help, then we take it first. It is a balancing act, but it simply does not work. If they want us to do more administrative work, they need to get us a little more away from patient care. But if we are to be more involved in patient care, they need to reduce our administrative workload. And I think the latter is better because I am not a nurse to write notes, it is to take care of my patients."* Nurse, neurology

Table 4.6: Estimated time nurses spend on administrative work in average on a normal day of clinical work at the neurology department.

Professional Title	**The estimated time spent on administrative work on one day**	**Is the amount of time reasonable or should it be more or less**
Nurse/ coordinator	40 %	*"What we document must be documented, and if it can be done faster, that is only good. But if we say it is just going to go down without anything else, I feel like it is at the patient's expense or at the business' expense or at the national expense. Because then it is to ignore those assignments."*
Nurse	75 %	*"Yes I think it is reasonable."*
Nurse	60-70 %	Q: *"And do you think this is a reasonable amount? Or should it be..."* A: *"No."* Q: *"So it should be less?"* A: *"Yes, I am here to care for patients."*
Nurse	At least 50 %	*"It should be less, of course. So that you have more time for the patients. But I do not know anything else. I have worked in several places before. So it is reasonable. If you compare it to how much time I spent on administration in previous places, it is no more here than what I experienced in previous places. A lot of time is spent on administrative work."*
Nurse	20-40 %	*"It should be less."*
Nurse	20 %	*"Of course, I think that you should spend more time with the patients. So I think it is too much time."*
Nurse/ coordinator	70-75 %	Did not get the question.
Nurse	50-60 %	*"I think it should be much less."*

Some nurses describe that the nursing discharge note can be time-consuming. It depends on the patient. If the patient has stayed at the hospital for weeks and has moved around between different departments it can take a long time to write the nursing discharge note. However, if a patient has stayed only for a few days at the department, the nursing discharge note can be done in just a few minutes. The same goes for when the nurses are studying a patient's health record before meeting with the patient. However, the nurses usually help each other out to write nursing discharge notes, and write small pieces continuously during a patient's hospital stay. A few nurses further experience that the admission note can be time-consuming, because they have to find and add all relevant information for the patient in question. Some nurses explain that what is time-consuming for these tasks is to find and collect all information that is needed from the health records. The nurses clearly express that the amount of administrative text-based tasks they have are too many and that their administrative work as a whole takes too much time.

A majority of nurses describe that several of their administrative text-based tasks are "double work" (performing tasks that have already been completed previously), which is time-consuming, feels unnecessary and like something that could be solved and free up time they can spend on patients instead. They have to write the same information in several different systems, multiple times a day for every patient.

> *"When a patient arrives you read the admission assessment note. The day after an operation, I often receive the patient at NUVA (neuro-awakening department). Then I write a post-op report. In relation to that, I usually write a nursing discharge note. All three notes are similar. You answer the same questions all the time. [...] It is quite a lot of double work or triple work almost, when writing the same things."*
> Nurse, neurology

Some nurses explain that they write certain types of referrals. Referrals that have been mentioned in the interviews are about suture removal or about cancer rehabilitation. One nurse describes that she sometimes sends referrals about patients' problems with blood sugar or urinary function to the health centers. The nurses who write referrals confirm that large parts of the texts are identical for all referrals of the same type and that you can copy and paste information. Some information varies between patients, and that information can be found in the health records. However, sometimes the physicians miss to enter some information that the nurses need to write the referrals, such as after how many days sutures will be removed. When that happens the nurses need to find the responsible physician, which can be difficult. One of the nurses has put a request to add a sentence into the existing template for suture removal referrals. Currently the nurses need to manually write the same sentence in every suture removal referral about who the patients should contact if they suspect infection after suture removal.

Furthermore, a majority of the nurses experience that SAMSA, the program they use to submit documentation to the municipality and aid assessors, is a bottleneck. The nurses explain that they write the same information into SAMSA as they have already entered into Melior, and that they can not copy the information directly from Melior to SAMSA, but need to write the same information manually twice in both systems. The nurses underline that this clearly is “double work” for them.

> *“When writing in SAMSA, it really feels like everything you write is double. It is information that... We take it from one system and write it into another. So it would have been easier if you had somehow been able to link it.”* Nurse, neurology

> *“What is written in this SAMSA system, I can copy that into Melior. It is very smooth. It is super nice. But what we have entered in Melior, we can not copy over into SAMSA. No. And we usually document in Melior, it is our main work tool after all. If we had been able to copy the text to SAMSA, it would have taken much less time. [...] They are exactly the same sentences. And, unfortunately, you only sit with one screen. And you have to jump between these two systems, if you are not very alert.”* Nurse, neurology

One of the nurses further explains that they have the same problem with their system for booking transport/ambulance. There they also need to submit the same information that they have already entered in Melior, but as mentioned, they can not copy and paste text from Melior to another system. For a third system they use, the national health record system (NPÖ), they even have to close Melior completely down every time they enter NPÖ. They thus need to write down the information they want to enter from Melior to NPÖ on a piece of paper, close Melior down, open NPÖ and then enter the information. If they miss any information they have to go back again and log in and log out again. One nurse explains the same procedure as for NPÖ, for when they write suture removal referrals. Many of the nurses express that it would be more efficient if they could copy information from Melior into the other systems that they work in and need to enter the same information into.

All nurses feel that their IT-environment is difficult to work in. They describe their systems as a major bottleneck that causes great inefficiency when it comes to their administrative work. Everyone expresses different problems that they are experiencing when moving around between all the different systems that they have for adding and retrieving information.

> *“Partly the systems are in the way, and partly it is also the fact that we are nurses. We are not technicians, so we would really like to see more work done on how to do what we want to convey, or what we want to do. So I was very happy when you came and asked, because it is somehow something that you miss.*
>
> *It is a bit like this, here you get a system, yes, but what are we going to use it for? Or, you very rarely look the other way, and I think we all feel a little bit like that. It will be like share point, or whatever it is, we do not understand how to use it, because that is not what we actually*

> *work with. But we have been thrown into it, here you have a system, and here you have a system, but how do we use them in a smart way? So that knowledge is needed, and I do not really know if it should be left to us to find out. It does not really feel that way sometimes. Maybe we can be nurses and help the patients and get help with the technology instead."* Nurse/coordinator, neurology

Furthermore, one nurse explains how she spends much time working with "Min vårdplan", a national care plan that applies to all patients who develop a tumor in the central nervous system. Min vårdplan is part of the 1177-system, a national system for advice on healthcare. The nurse explains how Min vårdplan consists of text documents which a group, her included, revise every few years. One large problem she experiences with these documents is however that all text is in Swedish, which becomes problematic for patients that are in need of the information these documents include, but who do not understand Swedish. The nurse explains that she would like a system that can translate these documents into other languages without getting the information wrong.

Moreover, one nurse describes how the nurses often need to look for information about different drugs in yet another system, FASS (Pharmaceutical specialties in Sweden). She however describes that FASS is not updated, and that they then have to look further. If they can not find the information they need, they have to ask the physicians. However, the physicians too, do not know more than what is documented in FASS.

Three of the nurses explain that the nurses at the neurology department have started to use TIK (speech recognition). One of the nurses likes it and explains that they can use it to create health record notes and word documents. She describes that it sometimes makes mistakes, but that they just will have to learn to use it properly. The second nurse agrees with the first that the TIK can make mistakes and that they just have to learn how to use it and further expresses that she does not really know it yet. The third nurse has not tried it, but is skeptical about it.

Lastly, some nurses feel like they are being asked to do work that is not their work to do. Three of the nurses express that there are situations when the nurses have to do the physician's job. Four nurses for example describe that they fax referrals the physicians write. One of them explains how referral recipients, like radiologists, contact them instead of the physician directly, because the physicians do not answer the phone. The nurse then becomes an intermediary who needs to pass the information on to the referrer.

> *"No, actually I think this is for the physicians. Because actually, we nurses have nothing to do with referrals. It is the doctors who write them and it is they who have to fax them. Sometimes they ask us to do things, but that is their job. It is."* Nurse, neurology

> *"I usually get faxes and referrals for some reason. I usually get the doctors' referrals faxed. Which is a bit strange because I have previously worked at reception and so on, then I wrote my own referrals to various agencies and I always faxed them myself because I liked to know that they got sent away."* Nurse, neurology

4.2.3.2.2 Tools and Characteristics It is clear that the nurses do much "double work" between systems and have to add the same information at several different places. In addition, the nurses write similar information multiple times in different notes in their health records system, Melior. In other words, the nurses often write notes into the health records that contain already existing health record information. For example, one nurse explains how she uses the patient background from the admission assessment note for other nursing notes as well.

A majority of the nurses express that it would help them to get a summary of a patient to study a patient's health record or to write for example a nursing discharge note. One of the nurses comments that there has to be trust that the summary is correct, otherwise you have to double check if the information in the summary is correct. Another nurse thinks that there is a risk that newly graduated nurses will not learn how to study a patient and know what information is relevant to look for, if they get used to getting a summary and the program stops working at some point.

Five of the nurses explain that they generally know what they are looking for when looking for information in the health records, rather than needing to read through all the notes in the health records in order not to miss something they had not thought of. One nurse explains that it can vary, sometimes you know exactly what you look for, and other times you look through everything to understand the case. Many of the nurses explain how they have keywords like "nutrition" and "elimination" in their notes and that they have templates for some of their notes, for example the admission assessment note, the nursing progress note and the nursing discharge note.

The majority of the nurses explain that they often write the same things for a number of nursing categories, like nutrition and elimination mentioned above. It is not just yes and no answers, but sentences. However, how they write the same type of information differs from nurse to nurse. Five of the nurses believe it would be good to have standardized sentences that they could click on for some of their text-based tasks, for example for the nursing progress notes, nursing discharge note and post-op notes, categorized by the keywords they already use. One of the five nurses already uses standardized sentences that she has collected in her head from long nursing experience. One nurse mentions that they have certain categories with standardized choices in the nursing discharge note, like for fall risk, and that you can also enter more text if needed, which she thinks is good.

One nurse puts great emphasis on the need of a coordination tool, something that gives the nurses notifications. Currently the nurses are keeping track of a large number of patients, but they do not have a system that tells and reminds them of what they have to do and when. The information exists, but is spread out in their complex IT-environment with several different systems. The nurse further believes that getting notifications and reminders would be great for the physicians as well, for things like updating prescriptions or expiring sick leave, etc.

Lastly, one nurse has worked with another health records system in another Swedish city called Cosmic, which she thinks is much better than Melior, the health record system they have in Gothenburg and at Sahlgrenska. She explains how Cosmic is easier to use and that there is a nursing status that can be updated with nursing progress notes, and that Cosmic stores the old information. In contrast to the nursing status in Melior where statuses from previous days get overwritten and are not stored anywhere. She explains how this becomes a problem, because the nurses sometimes need to know the status of the patient from previous days at the hospital. Therefore, many of the nurses at Sahlgrenska collect “old” information in today’s nursing status box in Melior, making today’s status long and complicated. If the information gets deleted, it is lost. The nurse finds Cosmic better in every way. She explains how it is easier to find information in Cosmic and that it is less time-consuming to use Cosmic compared to Melior. She explains how she spoke to a physician who also had experience with Cosmic who shared the nurse’s opinion.

4.2.3.3 Time Estimates for a Number of Text Based Tasks

Table 4.7 and 4.8 below present time estimates that have been collected from the eight interviews that were conducted with nurses at the neurology department. Each nurse that made one or more time estimates are presented in the table (note that not all interviewed nurses made time estimates).

Table 4.7: Time estimates by neurology nurses for reading EHR.

Professional title	Read EHR time estimates
Nurse	1 hour for a patient you do not know and who will undergo an investigation for example (out-patient care).
Nurse	1 hour in the morning, about 7-8 patients (in-patient care).
Nurse	A few minutes, but then I go and meet them too (in-patient care).
Nurse	5-10 min
Nurse/coordinator	10 min

Table 4.8: Time estimates by neurology nurses for different writing tasks.

Professional title	**Nurse/ coordinator**	**Nurse**	**Nurse**	**Nurse**	**Nurse**
Nursing admission note				30 min	
Nursing assessment note				20 min	
Nursing discharge note		Short: A few minutes Long: 20 min	10-15 min	15 min	Short: 5 min Long: 30-60 min
Post-op note			5 min		
Suture referral			10-15 min		
Referral (cancer rehabilitation request)	10 min				

4.3 The Ophthalmology Department

This section presents the case that examines the practices and experiences at the ophthalmology department. The ophthalmology department specializes in the treatment of eye-related conditions. Data were collected through semi-structured interviews with nine residents and one senior specialist from the department.

4.3.1 The Role of Physicians at the Ophthalmology Department

At the ophthalmology department, physicians work with both out-patient care and in-patient care. However, the out-patient care provide a larger part of the work tasks of the physicians interviewed. The department sees a high volume of patients daily, with one resident noting that each physician meets between 10 and 30 patients, depending on the specific department they are working at that day.

The ophthalmology department has medical and surgical teams that take care of various eye diseases. As a physician in ophthalmology, you meet patients daily in different settings. Documentation is required for all patient visits, taking various forms. Examples of patient-related administration at the ophthalmology department is identified as writing medical notes including admission notes, discharge summaries and referrals, checking lab results as well as reading medical notes as preparation before a patient's visit. Recently, there has been a shift in responsibilities at this department. Most documentation tasks have transitioned from physicians dictating notes to a secretary, to physicians themselves using speech recognition software to write their own medical notes.

4.3.2 The Physicians' View on the Administrative Burden at the Ophthalmology Department

> *"I think we have far too much administration. It is a major frustration in the work that detracts from the joy of meeting patients. I never feel that I have time to do what I am good at. I spend a lot of time on peripheral tasks instead of seeing more patients. I definitely think it is too much. I find it unnecessarily complicated, which creates frustration."* resident, ophthalmology

In the ophthalmology department, 8 out of 10 of physicians report an excessive administrative burden, while 2 consider it manageable. Notably, one resident that addresses the administrative workload as manageable attributes their perspective to being new and seeing fewer patients, which reduces their administrative duties. Several physicians have identified certain aspects of this workload as unreasonable, arguing that their expertise could be better utilized in clinical work. Additionally, the adoption of speech recognition technology has unexpectedly increased their administrative responsibilities, as tasks traditionally performed by administrative staff and secretaries are now being delegated to them.

> *"Practically speaking, we could see more patients if there were less administrative work. And what we are best at is diagnosing, investigating, and treating patients. Just because I can write a good note does not mean I should spend time writing it. So, absolutely, it should become better, easier, and more efficient to perform the administrative tasks."* resident, ophthalmology

Furthermore, some of the more experienced physicians believe that the administrative burden has increased over time. They feel that more documentation is being added to their daily tasks and that certain organizational processes have made administrative work less efficient than before. Two physicians describe this transition as a move in the wrong direction.

Table 4.9 below summarize the physicians estimation of how much time they spend on administrative tasks, followed with an opinion on how reasonable the administrative burden is.

Table 4.9: Estimated time physicians spend on administrative work in average on a normal day of clinical work at the ophthalmology department.

Professional title	The estimated time spent on administrative work on one day	Is the amount of time reasonable or should it be more or less?
Resident	50 %	"*Just because I can write a good note does not mean I should spend time writing it. So, definitely, it should become better, easier, and more efficient.*"
Resident	25 %, of which half is patient-related administration.	"*I may think that it is reasonably to too much. But I do not appreciate the content of the administrative time. There are many other shortcomings.*"
Resident	60 %	"*I think it is completely unreasonable. I have now worked as a doctor for more than 10 years, and things are going in the wrong direction. There is less and less time for the patient and more and more for administration.*"
Resident	12.5 %	"*I believe it is reasonable, at least a large part of it.*"
Resident	Could not estimate.	"*I think it is way too much. It is a major frustration in the work that detracts from the joy of meeting patients. I never feel like I have enough time. I spend a lot of time on peripheral tasks instead of seeing more patients.*"
Resident	Could not estimate.	"*It is a bit difficult to estimate how much time we actually spend on administration. But my administrative tasks take a lot of time. I feel like I work more administratively than clinically.*"
Resident	Between 12.5-25 %	"*Reasonable? No! We have a lot of administration.*"

Resident	Could not estimate.	"*I do not know how much, but it is a significant portion and it should be less, definitely less. I also believe there is great potential to streamline due to today's technology*"
Senior specialist	Did not get the question.	"*Depending on the type of patient, somewhere between 6 to 15 patients do not get an appointment because we doctors have to sit and work with data systems. We are over-administering healthcare, it is frightening.*"
Resident	10 %	"*I believe the time for administration is reasonable, however, I am also in the beginning of my specialist career and do not have as many patients as the rests of my colleagues*"

4.3.3 Task Shifting with Speech Recognition

Opinions on the effectiveness of the speech recognition technology are mixed. One physician express extreme dissatisfaction, suggesting that some might resign if required to use this technology regularly. Conversely, another physician praise the system, highlighting its efficiency, particularly the ability to program short commands into the tool. Others take a more diplomatic approach, acknowledging that while the tool does not currently perform well, it shows potential for improvement once physicians have been properly trained to use it. Also, the majority of physicians point out that the speech recognition tool needs to be trained to recognize their specific voices in order to maximize its effectiveness. This might increase the potential for streamlining documentation, according to some physicians. However, some still believe the speech recognition will not work well after the training either. Moreover, one interviewee pointed out that clinicians have not been adequately introduced to the tool, nor have their needs been sufficiently considered during its development.

> *"They have surely done it properly but they have never worked with patients before, since it is the IT people who are handling it. And then it has been tested on administrative staff, not on clinicians. There is a lot of that. Unfortunately, we are always last in line when perhaps we should be first in line for changes."* resident, ophthalmology

Although many physicians acknowledge that speech recognition technology has increased their administrative workload, some see a clear advantage in no longer needing to dictate notes to a secretary. Previously, when dictating medical journal entries, it could take weeks or months for the physicians to receive their notes for approval. This delay often led to inefficiencies and required physicians to perform tasks twice. One resident highlight the inefficiency of the old system, noting that the lengthy turnaround time forced them to double-check the transcribed information, as they can not reliably recall the accuracy of the details after such a long interval. This issue was compounded when secretaries used question marks in place of words they could not hear clearly, adding another layer of verification for the physicians.

> *"It can take several weeks before you see the secretary's response - but with the speech recognition technology, it takes longer for me but in the long run, it saves time."* resident, ophthalmology

Moreover, one resident highlights that with speech recognition, as medical notes are promptly documented and made available, it allows both healthcare providers and patients to review the recorded details without delay.

4.3.4 The Ophthalmology Physicians' Perspectives on Patient-Related Administration

Opinions differ regarding which patient-related administrative tasks are the most burdensome at the ophthalmology department. Among the physicians interviewed, 4 out of 10 identifies pre-visit review of medical journals as particularly time-consuming, with a couple noting the necessity of dedicating additional time outside of regular hours to complete this task effectively. One resident specifically mentions spending evenings before workdays catching up on reading. The time spent on this task can vary significantly, ranging from 2 to 20 minutes, depending on the patient's medical history. However, some physicians argue that reviewing previous medical journals is not overly time-consuming, emphasizing that the current examination tends to be more crucial than previous ones at the ophthalmology department.

Meanwhile, 3 out of 10 of physicians point out that writing referrals is another time-consuming activity, often involving the redundant task of repeating information already recorded in health records. One resident elaborated,

> *"Writing referrals can take a long time - but above all, it feels unnecessary. When we write referrals, it is a summary of the note plus the diagnosis we believe it to be and the examination we are requesting. It would be most efficient to just read the medical note that was just written, or automatically summarize it into a referral if possible."* resident, ophthalmology

Discharge summary are another administrative task mentioned, though most physicians do not view them as particularly time-consuming. One resident notes that writing a discharge summary typically takes about two minutes, while another emphasizes that they do not write many such letters throughout the year. Moreover, the discharge summary is partly described as double documentation as the physicians use information that already exists in the medical journal.

> *"It is not like writing discharge summaries as at many other clinics that have very complicated care processes, with various assessments during the course of treatment and different treatment regimes. But I think it could certainly be made much more efficient and standardized. However, on the whole, it does not take much time. And I do not think there are that many either. That is, as an resident in ophthalmology, you do not write many discharge summaries per year."* resident, ophthalmology

Despite varying opinions on the specific tasks that consume the most time, there is a general consensus among physicians that the overall administrative load is excessively heavy. This sentiment is further encapsulated by another resident who emphasize,

> *"That is really the problem, it is not just one specific thing that takes time but rather all of the administrative tasks that are individually a bit inefficient, which ultimately makes everything very inefficient."* resident, ophthalmology

Regarding the handling of referrals, half of the physicians report that the process of questioning or denying incoming referrals is excessively time-consuming. As a result, they tend to accept nearly all referrals, even those patients that could potentially be managed by primary care or treated at home. Several of the physicians also report that there is a large share of the total referrals that are inadequate, which makes this problem even larger.

> *"I believe that perhaps 30-40 percent of the visits would have been completely eliminated if the referrals were better or if it were easier to reach the referrer. Then we could have addressed it on the spot, determining what needs to be done, without needing to see the patient."* Resident hysician, ophthalmology

However, some physicians do not see this as a large problem, due to them preferring to examine the patients as ophthalmology is such a specialized field.

> *"Often, when a primary care physician do not have much knowledge about eye conditions. So, we still need to look at it regardless of what is stated in the referral."* resident, ophthalmology

The reasons for the inadequacy of referrals are viewed differently, but two primary causes have been identified. The first is that ophthalmology is a highly specialized field, which can make it challenging for non-specialized physicians. The second reason is that the remitters may not know what specific information is required in the referrals. Some physicians believe this issue could be addressed by providing templates to guide the remitters in compiling more complete and informative referrals.

4.3.5 The Ophthalmology Physicians' Perspectives on Their Computer Systems

All of the physicians at the ophthalmology department express significant concerns regarding the complexity and inefficiency of the computer systems they are required to use. These systems are described as not only cumbersome and time-consuming but also as potentially compromising patient safety due to their disjointed and unreliable nature. The department operates with approximately seven to eight different IT systems to manage various aspects of patient care. Physicians report that this multiplicity is highly time-consuming, as they need to continually switch between systems to access different types of patient information. This fragmentation of data across multiple platforms often leads to inefficiencies that not only delay the process but also increase the risk of overlooking critical patient information.

> *"All the IT systems are horrible and take a lot of time, and they lag, and you have to restart them, and they do not work, and they are in different programs. All this makes it unsafe for patients. Because you can lose information, you can misinterpret information, sometimes you can not access certain information, and it takes far too much time away from patient care."* resident, ophthalmology

Physicians highlight several issues related to system performance, including frequent lags and the necessity to reboot systems that fail to function properly. Such technical problems not only interrupt the workflow but also lead to delays in patient care. In some instances, these delays can jeopardize patient safety, as timely access to health records and test results is crucial for making informed treatment decisions. Another example from one physician is that accessing information as medical images can be an issue. This is due to locating particular imaging that might require logging into multiple systems, complicating what should be a straightforward task. This cumbersome process adds unnecessary steps and time to the physicians' workload.

Another large issue highlighted by the physicians is the absence of system notifications, which is a particular point of frustration among the departments staff. Physicians say that they must actively log into each system to check for updates rather than receiving automated alerts. This not only demands constant vigilance to ensure no crucial information is missed but also adds to the already considerable administrative load carried by the physicians.

It is further described by the physicians that the combination of these system-related issues significantly detracts from the time physicians can dedicate to patient care. The administrative burden imposed by the inefficient and often unreliable IT environment is extensive, suggesting an urgent need for system integration and optimization to enhance both efficiency and patient safety.

> *"A commercial company with this kind of IT environment and structural*

organization could never exist. It would go bankrupt immediately if its management considered this situation acceptable, which I firmly believe can never be the case." resident, ophthalmology

4.4 The Radiology Department

This section presents the case that examines the practices and experiences of radiology physicians and nurses, specifically within the musculoskeletal disorders and thorax radiology department. Data were collected through semi-structured interviews with staff from the musculoskeletal department and observations at the thorax radiology department. The interviewees included four senior radiologists, one resident radiologist, and two radiology nurses.

4.4.1 The Role of Nurses and Physicians at the Radiology Department

Physicians in the radiology department have less direct patient interaction compared to those in specialties like neurology and ophthalmology. The primary role of a radiology physician (radiologist) is to analyze and interpret medical images, such as ultrasounds, MRIs, and CT scans, to assess patient conditions. As part of their duties, radiologists also perform tasks such as collaborate closely with referring physicians, offer consultations on treatment options, and provide guidance on selecting the most appropriate imaging methods for specific medical cases. During examinations, such as ultrasounds, radiologists may provide consultations to patients. However, these consultations are informal and are not documented in the same manner as other patient meetings. Moreover, radiologists are responsible for certain interventions guided by ultrasound or x-ray.

Radiology nurses play a crucial role in medical imaging examination. They are responsible for meeting patients, preparing them for the examination by setting up equipment and providing detailed instructions. Their involvement varies depending on the complexity of the procedure. For more intricate examinations, nurses have increased interaction with the patient to ensure everything runs smoothly. This is small selection of radiology nurses responsibilities.

One common main task for both radiologists and radiology nurses is that they handle referrals from other specialties. At Sahlgrenska, this implies referrals both from within Sahlgrenska and from other external care clinics. A referral to the radiology department includes information regarding a patient's medical background, the patient's current condition, reason for radiological examination and suggested examination, e.g., CT scanning. This referral is the basis of which type of examination a patient will undergo.

Radiologists describe two main referrals coming into their department, emergency referrals and elective referrals. The difference between them is how quickly the evaluation must happen. Emergency referrals may require immediate action. For example, if pulmonary embolism, heart attack, stroke, or appendicitis is suspected, the evaluation might need to happen the same day. Some elective referrals can be urgent but not emergencies, they can be delayed without risk for patient's health.

4.4.2 Administrative Tasks at the Radiology Department

In relation to the referrals received by the radiology department at Sahlgrenska, both prioritization and interpretation of these referrals are two interconnected patient-related administrative tasks performed by radiologists. Also, radiologists document the results of an examination, adjust image quality, radiation levels and determine settings on the medical imaging devices, which could all be defined as patient-related administration tasks. Moreover, preparation for medical evaluation rounds, which is a meeting where healthcare professions from various specialities gather to go through specific patient cases, is another patient-related administrative task. Also, radiology nurses' patient-related administration contains referral management.

For non-patient-related administration, some radiologists and radiology nurses at Sahlgrenska have specific areas of responsibilities which they write protocols and job routines descriptions for. Protocols could for example imply descriptions of guidelines of how a CT scanning should be performed. Moreover, some healthcare staff are responsible for preparing materials for internal training sessions, and most highly experienced radiologists need to educate the more junior physicians. More non-patient-related administration includes healthcare staff handling e-mail and the responsible physician at the radiology department handles the internal complaint system called MedControl.

4.4.3 Deep-Dive into Administration Related to Referrals

All radiologists interviewed at the musculoskeletal department expressed frustration about handling the incoming referrals. The interpretation and prioritization of the referrals is explained as being delayed and creating extra work many times a day.

> *"We have issues with referrals almost every day, several times a day. Not only the referrals that come from other specialties within Sahlgrenska but also with external referrals, from primary care centers."* Senior radiologist, radiology

The expressed challenges with handling these referrals is that the information provided by the referring physician is often inadequate. The lack of information hinders the radiologist from correctly prioritizing the referral. All the radiologists agree that there is no clear sort of referral that is more or less adequate, whether the referral is very detailed and long or contains a few sentences, it could lack crucial information. One radiologist at the musculoskeletal department testifies that 50% of the emergency referrals and 20% of the elective referrals that come into the department require unnecessary long handling.

When a referral is inadequate, radiologists often need to take extra actions before they can sufficiently handle the referral. These actions imply that radiologists either need to discuss the patient's case with a colleague, contact the referring physician, or search for details in the health records. All radiologists describe the process as very time-consuming and frustrating. The approach a radiologist takes to solve problems can differ individually. However, one radiologist noted that while the missing information is often available in the health records, retrieving it tends to be too time-consuming.

> *"Referral management is obviously an important part. But clearly, every time you pick up the phone to look for a referring physician and have to wait in line and they do not answer, and then you need to call a department instead, I wish it would take less time. I just wish this path to clarity was shorter."* Senior radiologist, radiology

A recurring theme among radiologists discussing the frequent inadequacies in referrals is the belief that these issues often arise when less experienced physicians write the referrals. Additionally, one radiologist suggests that this inexperience could explain the significant disparity in the quality of emergency versus elective referrals, noting a hypothesis that emergency referrals tend to be written by less experienced physicians. A concrete fault in a referral could be that the referring physician writes more than one reason for the radiology evaluation, which could require more than one evaluation. This in turn leads to the patient being exposed to more radiation than necessary.

> *"They can write, for example, several different questions that the remitter wants the answer to, and to answer these questions, we often need to perform various examinations on the patient. This means more radiation and more contrast, sometimes the examination could even have been done with ultrasound or MRI, which are examinations where the patient is not exposed to any dangerous radiation at all."* Senior radiologist, radiology

At the thorax department, while the issue of inadequate referrals is less common, there is also noticeable frustration regarding the task of prioritizing these incoming referrals. The work is seen as highly repetitive and mundane. The department handles approximately 100 referrals daily, and the list to be prioritized is described by the radiologists as endless. Similarly, radiologists at the musculoskeletal department find the prioritization of referrals to be a standardized yet frustrating task.

Additionally, radiologists at the thorax department explain that when they receive referrals, they often need to review previous evaluations and medical images. This process can be particularly time-consuming when the patient's history is extensive and complex.

> *"If you look at lung cancer, it may have developed a little every year. When you compare a new X-ray with the one done the time before, you might hardly see any change, but when you compare it with the first one done maybe five years ago, you can see a huge difference. Then it is important to go through the X-ray history, but it is tedious and time-consuming to go through the X-ray history to find the first X-ray for the case."* Senior radiologist, radiology

Also the radiology nurses have expressed difficulties in interpreting referrals. In line with the complaints from radiologists, they explain the problematic that you often need to take extra action before you can correctly interpret a referral.

> *“There can be a lot of extra work with the incoming referrals, as we may need to ask a colleague, a radiologist, or try to contact the referrer and ask them.”* Radiology nurse, radiology

Like the radiologists, the nurses also stress the importance of managing referrals, as the information they contain forms the basis for determining the appropriate radiological evaluation for the patient. This should put both responsibility on that the referring physician writes the correct information, but also that the staff at the radiology investigates if the information is inadequate.

> *"So just managing and interpreting the referrals that come in is a huge part of the administrative work and also extremely important for me as a radiology nurse, because the information I have for this patient can greatly influence which images I should take or how the examination should be conducted overall."* Radiology nurse, radiology

Additionally, the nurses explain that occasionally, the administrative complexities lead to unnecessary radiation of patients. One nurse says that in some instances, referrers issue multiple referrals for procedures that could be consolidated into a single referral. Further, the nurse explains that by using one referral instead of three, they could not only streamline the processes but importantly, minimize harmful radiation exposure. This issue is also recognized by radiologists, who share similar concerns about the inefficiencies and potential overuse of harmful radiation, explained above.

> *"It is extremely important what information the person writing the referral provides to me, because it naturally plays a role in which examination is performed. For my part, it means a great deal since I irradiate patients every day. We operate under a rule where we must give as little radiation as possible, and therefore the referral is extremely important to us and a major part of everything we do."* Radiology nurse, radiology

Additionally, a concern shared by both radiologists and radiology nurses is the communication gap between physicians and nurses. While communication among physicians is generally effective, interactions between physicians and nurses can sometimes face challenges. This is often because the two groups use different terminology and abbreviations. As a result, nurses may struggle to interpret referrals written by physicians.

4.4.4 Suggestions to Make the Referral Management More Efficient

As the need to streamline referral management has been widely recognized, it has been prompting discussions among radiologists and nurses about potential improvements. Several interviewees suggested that assisting the referrer in sending better-quality referrals could be beneficial. As a suggestion, this could be done by templates that the referring physician can use. For example, it is suggested that standardized referral templates be created for common patient conditions. These templates would enable the referrer to modify only certain parts of the referral, thereby simplifying the process and promoting uniformity in the information submitted. However, opinions on the effectiveness of this approach are mixed. Some clinicians argue that patient conditions frequently vary too much for standardization, while others believe that many conditions can indeed be standardized effectively.

However, some radiologists and nurses say that some initiatives in this direction are already underway. For instance, the musculoskeletal and orthopedic departments have collaborated on addressing these issues. Additionally, radiologists at the thorax department highlighted the medical evaluation rounds at Sahlgrenska, where specialists from various departments convene. Except for these meetings' main objectives, these can serve as an excellent platform for cross-disciplinary feedback on referrals, fostering better communication and coordination among clinicians.

4.4.5 Other Administrative Tasks and Bottlenecks at the Radiology Department

In addition to the administrative challenges related to managing referrals, radiologists also identify mail management as a time-consuming task. Moreover, a recurring theme among all interviewees is the inadequacy of the computer systems they use. These systems are described as a major bottleneck, not only due to their individual inefficiencies but also because of their lack of interoperability.

> *"As with everything else, you get used to a shockingly bad system. Eventually, you invent your own patterns to manage it."* Senior radiologist, radiology

For instance, entering the computer systems is mentioned as the largest bottleneck by one radiology nurse, while another nurse suggested that consolidating these systems into a single platform would greatly enhance daily efficiency.

Moreover, both radiology nurses and radiologists at the musculoskeletal department mention the lack of highly specialized professionals as a issue, which is described to make the overall work more inefficient.

> *"We have lack of a certain level of competence, as there are many more residents than senior specialists at the Radiology. Sahlgrenska relies on resident doctors but per specialist, there are very many residents. The work becomes inefficient because we can not work as quickly as specialists, we do not know as much, and we can not make the same decisions that a specialist can."* Resident radiologist, radiology

An hypothesis regarding the lack of specialists is described as a combination of unfavorable working conditions and a challenging work environment. A radiologist notes that many private radiology providers offer significantly higher salaries, which attract specialists away from Sahlgrenska, even though the work at Sahlgrenska can be considered more exciting.

Moreover, both the radiologists and nurses report occasionally receiving referrals by post, a process they describe as both time-consuming and potentially unsafe for patients.

> *"People are still sending paper referrals within the hospital, where someone dictates that a referral needs to be sent, someone else prints it, another person posts it, someone opens it, and another person reads it, and then scans it in. It feels like the Stone Age to me."* Resident radiologist, radiology

4.4.6 Time Spent on Administrative Tasks

All interviewees concur that they invest significant time and effort in performing administrative tasks, however also emphasizing the importance of executing these tasks accurately. Nonetheless, there is a consensus that these tasks could be streamlined for greater efficiency. The senior radiologists report spending the most time on administrative duties compared to their less experienced colleagues and radiology nurses. This disparity is attributed to the senior's greater responsibility in prioritizing the diverse array of referrals received by the department.

> *"It is very important to have this planning phase, to have time to plan for an examination and handle the administrative matters before the patient arrives. But I would like to make it more efficient in a way that would make it easier for me to understand the referrals."* Radiology nurse, radiology

The interviewees also agree that the department could have time to handle more patients if the administrative workload was lower alternatively made more efficient.

Table 4.10 below summarize the clinicians estimation of how much time they spend on administrative tasks, followed with an opinion on how reasonable the administrative burden is.

Table 4.10: Estimated time physicians spend on administrative work in average on a normal day of clinical work at the radiology department.

Professional title	**The estimated time spent on administrative work on one day**	**Is the amount of time reasonable or should it be more or less?**
Senior radiologist	30 %	"*If I get designated time in the schedule for administrative tasks, it suffices.*"
Senior radiologist	20 %	"*Of course, it is an important part to do, but I wish it could be done much faster.*"
Senior radiologist	30 %	No answer.
Senior radiologist	30 %	"*It would be best if it were only 10%.*"
Resident radiologist	Did not get the question.	Did not get the question.
Radiology nurse	30 %	"*That is how it looks. I can imagine that we work even more with administration. It depends on the system.*"
Radiology nurse	40-50 %	"*I think my administrative tasks take a lot of time. I believe we could have seen more patients if the administrative work were reduced.*"

5 Discussion

In this chapter, the results of this study are discussed in relation to the theoretical framework. The chapter is divided into multiple sections. First, the potential and risks of LLMs in healthcare are discussed. Second, the three cases of the neurology, ophthalmology and radiology department at Sahlgrenska University Hospital are evaluated in relation to the opportunity to streamline the clinicians patient-related administrative text-based tasks by implementing LLMs. Third, the impact on clinicians of the administrative workload is discussed. Fourth, the ethical and legal perspective of LLMs is discussed. Lastly, the case of Unity Health Toronto and nine key factors to implement new technology in healthcare are outlined under the topic of implementation.

5.1 The Potential and Risks of LLMs in Healthcare

The exploration of LLMs in healthcare presents a multifaceted landscape where the potential benefits are substantial but accompanied by significant risks and challenges. The results from expert interviews present a nuanced perspective of the potential for LLMs to be used in various healthcare settings and use cases, a view the theoretical framework also shares. As explained by (Hadi et al., 2023; IBM, n.d.), LLMs are capable of handling enormous datasets and analyzing the full context of documents as well as providing coherent and contextually appropriate text from these documents. Thus, it can be argued that the technology is a great fit for many of the patient-related administrative tasks highlighted as challenging in the results. Given these attributes of LLMs, the technology has potential to significantly reduce the burden of double and triple documentation in healthcare.

Specifically, LLMs show great promise for documentation tasks where information can be retrieved from existing health records (Ganguli et al., 2022; Wang et al., 2023). However, ambient scribes, could streamline the creation of documentation that includes new information not previously found in health record notes (Tierney et al., 2024). As described by (Crampton, 2019; Tierney et al., 2024), this technology excels at listening to conversations and without recording, transcribing them in real time, providing a basis for generating health care notes from new patient meetings. This can thus be applicable to both patient meetings in out-patient care and admission in in-patient care, which both require documentation in various kinds post-meetings. However, as far as the researchers conducting this study are aware, there is a scarcity of current scientific research demonstrating the performance of ambient scribes. Tierney et al. (2024) conclude that early promises exist, a sentiment echoed by expert interviews. Nonetheless, Tierney et al. (2024) also express reservations about the current state of the technology. Although, it is worth noting that this type of technology advances rapidly, potentially outpacing the speed at which studies can be conducted.

More precisely, both text summarizing and chatbots, areas LLMs excel at (Ganguli et al., 2022; Hadi et al., 2023; Liu et al., 2024; Zheng et al., 2023), hold significant potential to enhance the efficiency of administrative work for physicians and nurses. Chatbots can efficiently retrieve specific information requested by clinicians, while text summarizing can streamline tasks such as producing discharge summaries and reviewing healthcare records.

However, for LLMs to reach their full potential in medical settings, they must be fine-tuned on health record data to produce the tailored outputs necessary for healthcare (Devlin et al., 2019; Liu et al., 2024; Raffel et al., 2019; Stringhi, 2023). Suggested approaches include using an open source pre-trained model and fine-tuning it with appropriate health record documents or developing a specialized LLM from scratch for healthcare tasks (Chen et al., 2024). However, these approaches face significant barriers due to the sensitive nature of health record data (Guo et al., 2023; Haltaufderheide & Ranisch, 2023; Harrer, 2023), concerns about patient privacy, data security, and associated costs. These concerns are discussed further in Chapter 5.6 below.

Despite the large potential in using LLMs for patient-related administration, it also provides great risks that must be acknowledged and addressed. The most relevant concern described by (Bowman, 2023; Kaddour et al., 2023; Sallam, 2023) and in the results, is the issue of hallucinations and how to validate that the model provides the correct information to the physicians and nurses using LLMs. However, studies as (Rosenberg et al., 2024) has shown that LLMs can perform equally to physicians, but the sensitive nature of healthcare settings should require the technology to be next to faultless. Therefore, it is argued that the technology must be properly validated before being put into production.

One way to increase the safety is by letting the physicians and nurses validate the generated outputs by LLMs, which is highlighted in the results from the expert interviews and by (Haltaufderheide & Ranisch, 2023). This has larger potential in patient-related administration happening right after a patient meeting or in relation to the creation of a related document, since the memory of the clinician should be more clear. Also, the development of LLMs-technology moves incredibly fast, and the issues of hallucination is being successfully mitigated (Bowman, 2023). However, it remains crucial to address this issue, in particular due to the sensitive nature of healthcare.

Moreover, addressing the issue of outdated knowledge explained by (Kaddour et al., 2023; Zhao et al., 2023), necessitates a structured approach to maintaining LLMs once they are deployed. Ensuring that these models remain current is particularly crucial if there are changes in the hospital's documentation practices. This maintenance involves regular updates and continuous monitoring to align with evolving healthcare standards and protocols.

Furthermore, the issue of experience accumulation must be acknowledged, as raised by the expert interviews. This implies the risk to impact the clinicians' ability to gain experience if tasks are partially or fully automated by LLMs.

5.2 The Cases of the Three Departments

In this section, the results from the neurology, ophthalmology, and radiology department are discussed in relation to the theoretical framework and insights gained from the expert interviews. Moreover, the section outlines a brief comparison of the findings across the three departments.

5.2.1 The Case of the Neurology Department

In the case at the neurology department it is clear that:

1. There is consensus among the neurology clinicians that they spend too much time on administrative work.

2. For neurology physicians and nurses, the time required for text-based administrative tasks varies, correlating with the complexity of individual patient records and the clinicians' information systems.

3. What is time-consuming for long and complicated patient cases is to find all relevant information and get an overview of the case.

4. Neurology physicians and nurses need to do a lot of administrative double work.

5. Physicians' and nurses' agree that their information systems are regarded as the largest bottleneck when it comes to their administrative work.

At the neurology department, both physicians and nurses agree that there are large variations in time required to complete text-based administrative tasks. This clearly correlates with the length and complexity of the patient's health records and the complex structure of Melior (Sahlgrenska's EHR-system) and all other portals where physicians and nurses at Sahlgrenska store and receive information about their patients. For physicians, it is noted that factors like length of experience, having worked directly with the patient and being able to dictate notes affect the time spent on administrative text-based tasks.

Physicians' and nurses underline that what takes time in complicated cases is to find all relevant information, get an overview, and summarize the case. Relevant information is stored in many different places, between systems and in systems. Two physicians note that it can take up to 20 minutes to read the health record of a complicated patient. The mean for reading health record notes, calculated from data collected via the self-completion form, is 7.3 minutes (3 data points).

LLMs show promising potential for summarization tasks based on existing data (Hadi et al., 2023). Both physicians' and nurses believe that receiving a summary of a patient's health record could make their administrative text-based tasks easier, especially for long and complex patient cases. It is however of high importance that it contains correct information and does not leave any important information out. One of the main concerns when it comes to LLMs is that these models sometimes miss to include relevant information, and that they risk generating biased or incorrect information (Harrer, 2023). Some interviewed experts argue that LLM generated medical notes will have to be read, corrected and approved by the clinicians who use them, as a step to avoid incorrect and inappropriate information to be incorporated in the medical systems. Neurology clinicians' as well underline the importance of having references of where the information comes from to avoid incorrect information, that for example could follow from human misunderstandings among the clinicians, to become truth. RAG makes it possible to give references to where pieces of information are retrieved from (Gao et al., 2024). It can thus be argued that by using RAG when implementing an LLM summarization tool, the clinicians would get the opportunity to validate the LLMs generated material.

Furthermore, LLMs could be used as a tool that generates drafts of different notes that neurology clinicians' write. Rosenberg et al. (2024) study shows that LLMs can generate discharge summaries ten times faster than human orthopedists, indicating that time gains can be realized for these text-based tasks by using LLMs. One neurology physician explains that a discharge summary typically takes 10-15 minutes, but that it can also take only two minutes, and sometimes up to 1 hour. The mean for a physician to write a discharge summary, calculated from the data collected via the self-completion form, is 8.6 minutes (10 data points). In relation to the results of Rosenberg et al. (2024), these time examples propose that significant time savings can be achieved. Especially if one includes the additional time it takes for the secretaries to complete notes physicians' have dictated.

The neurology nurses paint a similar picture for their nursing discharge notes, underlining that there are clear variations from case to case and that it sometimes can be very time-consuming. This suggest that time saving can be achieved for nurses as well. However, it should be noted that the nurses at the neurology department usually help each other out to write nursing discharge notes, and write small pieces continuously during a patient's hospital stay, which reduces the time it takes for the individual nurse.

Moreover, LLMs research describes that LLMs are statistical models and that they are especially suitable for completing tasks that follow a certain structure (Ramlochan, 2023). Arguably, it would be suitable to use LLMs for medical notes that use a template and that are fairly standardized in their structures. Many physicians are optimistic about receiving a draft for discharge summaries and referrals, if the information can be trusted and the draft can be changed and corrected if needed. A senior specialist thinks it would be good to receive a draft even if they have to double check the information and that it could be easier to criticize a text that an external program has produced than one you have produced yourself. The nurses too, could be helped by LLMs generated drafts of for example nursing discharge notes. However, a few interviewed experts highlight that there is a risk that automatically generating medical notes will inhibit the learning of junior clinicians. One nurse expresses the same concern. However, two physicians do not believe that AI generated drafts will inhibit the learning of residents, if the drafts are of high quality and do not miss important information. A resident argues that it on the contrary could help the residents' learning, if the drafts are of high quality.

Moreover, medical notes are highly individual for patients, but some physicians' believe that some medical notes could include choices of standardized formulations. The same goes for nursing notes, which often include standardized sentences, and typically follow certain keywords. Transformer based LLMs have the capability to analyze the full context of a sentence or of a whole document, and can produce text that is both coherent and contextually appropriate (Hadi et al., 2023). Arguably, it is possible to train these models to generate standardized formulations, which might be helpful in some cases. However, especially for standardized nursing notes, pre-formulated sentences can be argued to be an easier solution.

Besides this, the characteristics of LLMs make the models suitable for language translation tasks (IBM, n.d.). One nurse describe how she works with documents for patients who develop a tumor in the central nervous system found in 'Min vårdplan', and that all documents are in Swedish. This becomes problematic for patients who do not understand Swedish. The nurse explains how there is a need for a system that can translate these documents into other languages without getting the information wrong. It can be argued that LLMs could be used for this, but that tools for validation is necessary as discussed above.

Furthermore, both neurology physicians and nurses do a lot of double work. They have to write the same information in several different systems, multiple times a day for every patient. Semmer et al. (2015) define duplicative documentation as an illegitimate task. Illegitimate tasks have been identified as a significant social stressor, endangering one's sense of self and infringing upon one's professional identity (Semmer et al., 2007). It can thus be argued that the large amount of double work the physicians and nurses have to do have a negative impact on the clinician individuals.

The problem of double work at the neurology department is largely a consequence of the inability to copy and paste from Melior to external systems. The majority of the nurses experience this problem with SAMSA, where they need to add the same information as they have added into Melior, but can not copy from Melior and thus need to write it manually twice. The easiest solution to this would be to allow copy-and-paste to Melior. A senior specialist further explains how physicians' often write clinical notes followed by one or more referrals that partly contain the same information as the clinical note, but that Melior hinders the information to be copied to the external system. LLMs can infer from context, generate contextually relevant and coherent responses (IBM, n.d.). LLMs could help generate information from clinical notes to referrals, and be trained to formulate the same information differently in relation to the receiver, which would be suitable in this case. It can further be argued that it is easier for the physician to validate referrals that are generated right after he or she has just written a clinical note.

Another type of double work both physicians and nurses do is adding patient backgrounds multiple times in different medical and nursing notes. For patients' that have had multiple hospital stays, parts of this information is repeated every time the patient comes to the hospital. Two of the residents believe that their administrative tasks would become more effective if the health record system had dedicated patient backgrounds that are updated continuously and that store old information. One nurse agrees, and explains that she had experience with another health record system with that function. According to her, it made the nurses administrative work much more efficient. This is in other words something that other EHR-system suppliers provide. Thus, LLMs are not necessarily needed, but the need should be noted. LLMs could be used to update the patient background when new information enters the system. However, one resident argues that in that case, a clinician should approve those types of updates.

Lastly, it should be noted that both physicians and nurses agree that the biggest bottleneck in their administrative work is the complex and inefficient IT-environment they currently work in, together with traditional processes with faxes and letters. This highly affects the time they spend on administrative text-based tasks, and causes great frustration among the physicians. Solving these problems of functionality and interaction between systems is out of the scope of this study. It arguably requires other tools than LLMs to be solved, as well as cooperation with the health record system suppliers.

5.2.2 The Case of the Ophthalmology Department

Based on the results from the ophthalmology department, four key takeaways can be identified:

1. Physicians experience an excessive administrative burden.

2. High volume of documentation tasks.

3. Inefficient computer systems hinder the efficiency of work processes.

4. Mixed opinions on speech recognition technology.

Despite varying estimates of the time physicians spend on administrative tasks, the majority report an excessive administrative burden. The results clearly indicate a consensus among most physicians that the administrative workload is overwhelming and detracts valuable time from patient care. However, there is a varied view when identifying specific tasks that are the biggest time thieves. Some physicians consider reading EHRs before patient meetings to be very time-consuming, while others do not. Similarly, opinions differ on the time consumption of referral management and producing discharge summaries.

It is evident that physicians are burdened with a substantial volume of documentation tasks, including writing medical notes as discharge summary and referrals, reading EHRs, and keeping track of lab-results. Collectively, these tasks contribute to a significant administrative workload, even though it is challenging for the physicians to pinpoint specific tasks that consume the most time. Moreover, physicians agree that the time required to complete text-based administrative tasks varies significantly. This variation clearly correlates with the length and complexity of the patient's health records, as well as the intricate structure of their current computer systems.

All physicians agree that the inefficient computer system consumes a significant amount of time and negatively impacts their work environment. Issues include the lack of system notifications and the need to log into multiple systems for patient information. It can be argued that the inefficient computer system is the biggest bottleneck in the ophthalmology department. However, finding a solution for this issue is beyond the scope of this research.

Moreover, while some physicians see potential in the speech recognition tool, others find it increases their workload. The tool is recognized for its potential because the current process of dictating, sending to a secretary, and then returning to the physician for signing is seen as too long and complicated. This process often results in double work for the physician during the signing stage, as it can take several weeks for the document to be returned from the secretary, by which time the physician may forget the details of the patient case. Although the speech recognition tool shortens this process, it requires more administrative time from physicians.

Hence, some physicians believe that the speech recognition tool increases their administrative burden. This perception stems from the task shifting explained above, where the responsibility for creating the entire medical document has shifted from the secretary to the physician. Additionally, many physicians report that the speech recognition tool is perceived as inadequate, making the process more time-consuming. This issue parallels the concept of illegitimate tasks. Most physicians believe that having the full responsibility for writing medical notes is not part of their job, thus viewing it as unreasonable (Björk et al., 2013; Semmer et al., 2019). The perceived inadequacy of the tool contributes to the feeling that these tasks are unnecessary, often resulting from organizational inefficiencies (Kilponen et al., 2021). These tasks can lead to serious health risks, such as burnout, diminished self-esteem, and increased stress (Anskär et al., 2019; Eatough et al., 2016; Thun et al., 2018; Werdecker & Esch, 2021), underlining the importance of addressing physicians' concerns to ensure their well-being and a satisfying work environment.

The implementation of the speech recognition tool offers valuable lessons. It is crucial to address the actual problems faced by clinicians and involve them in developing new solutions. Recognizing that new technology may not function perfectly initially, a robust implementation strategy is essential for physician acceptance. Early gains should be demonstrated to physicians, and change leaders should have a clinical background to communicate effectively with end-users. All factors presented by the expert interview results.

Three main use cases for LLM implementation in the ophthalmology department have been identified, each with associated risks that must be acknowledged and addressed. Also, the LLM needs to be properly trained and customized to understand ophthalmology-specific terminology and workflows to produce relevant and precise documentation, as described in relation to fine-tuning by (Devlin et al., 2019; Nabwani, 2023; Raffel et al., 2019; Ramlochan, 2023).

1. Ambient scribes for efficient documentation.

2. Summarizing EHRs.

3. Referral management.

Given the high volume of patients visits, using ambient scribes in the out-patient care section of the ophthalmology department has significant potential. By automating the documentation process, ambient scribes can reduce the time physicians spend on administrative tasks (Tierney et al., 2024), allowing them to focus more on patient care. This technology can greatly improve workflow efficiency by extracting and analyzing relevant medical information directly from patient-clinician dialogues and generating EHR notes in near real-time. Implementing ambient scribes could address challenges such as the large documentation workload, streamlining the process and making it quicker and more efficient. This is particularly beneficial in high-volume departments like ophthalmology.

However, there is a risk that ambient scribes may produce incorrect information, as the technology is still novel (Tierney et al., 2024). This risk can be mitigated by having physicians validate the documentation immediately after patient meetings, while the details are still fresh in their memory. Validation is crucial, as highlighted in the expert results and by (Guo et al., 2023; Haltaufderheide & Ranisch, 2023; Stringhi, 2023), who emphasize the need for human oversight in LLM applications in healthcare. Furthermore, it is important to evaluate how much time this solution actually saves, particularly if physicians need to review the documentation to verify its accuracy. This consideration, as highlighted in the expert results, is crucial for the practical implementation of any solution.

The second use case involves using LLMs to efficiently summarize patient histories and previous EHRs (Hadi et al., 2023; IBM, n.d.), saving time during pre-visit reviews. This solution is beneficial as many physicians find reading EHRs time-consuming. However, potential barriers include the need for physicians to still read the entire health record for a comprehensive overview, and the time required closely relates to the complexity of the patient's medical history. Hence, variations in patient complexity should be investigated to assess the solution's value accurately. Additionally, LLMs carry the risk of generating biased or incorrect information (Harrer, 2023), potentially harming patients (Petročnik et al., 2023) in this type of solution.

The third use case is referral management, implying that LLMs can automatically generate referrals by summarizing patient visits and relevant medical history (Hadi et al., 2023; IBM, n.d.), ensuring all necessary information is included without redundant data entry. This solution can both directly help the physicians by the mitigating the double documentation that today is required when writing referrals, according to the results. It can also help the physician or opticians referring to the ophthalmology department to write more complete referrals, hopefully minimizing the amount of incoming inadequate referrals.

However, several risks must be addressed for effective implementation of this use case. The LLM must generate accurate and comprehensive referrals, as inaccurate or incomplete referrals can lead to patient harm, misdiagnoses, or delays in treatment. Zhang and Kamel Boulos (2023) describe how LLMs can vary in quality and veracity, arguing that this unpredictability is a primary barrier to successful adoption in healthcare. Users cannot predict when the model will provide accurate answers or when it will be misleading, making it difficult to trust, especially if they lack the qualifications to assess the quality, completeness, and accuracy of the response (Haltaufderheide & Ranisch, 2023; Harrer, 2023; Zhang & Kamel Boulos, 2023). Furthermore, hallucinations by LLMs pose significant risks in healthcare, as the model's output can lead to misdiagnoses and inappropriate treatment recommendations (Petročnik et al., 2023; Sallam, 2023; Zhao et al., 2023).

Validation of automatically generated referrals by physicians immediately after completion can mitigate these risks. This process should be streamlined to save time while maintaining high standards of care, as underlined in the expert interviews. Additionally, the complexity of the patient's health record must be considered to determine if the solution can effectively handle complex cases. Cost-effective alternatives, such as developing templates collaboratively with the ophthalmology department, primary care centers, and optometrist clinics, should also be considered before implementing LLM technology.

In conclusion, there is significant potential to streamline patient-related administration in the ophthalmology department using LLMs. Also, it is evident that there is a significant need to make these tasks more efficient, given their impact on physicians' work and well-being. However, this potential comes with ethical and technological risks, uncertainties about time savings, and concerns about addressing the department's actual problems. For instance, the performance of ambient scribes remains uncertain, as, to the best of the researchers' knowledge, there are limited scientific resources demonstrating the technology's potential. However, both the first solution based on ambient scribes and the third solution involving referral management have significant potential, particularly with human oversight and opportunities for successful real-time validation of the LLMs' output. This approach can help mitigate the highlighted technological and ethical risks. Additionally, the second use-case, summarizing patients' previous EHRs, also has significant potential, especially since LLM technology excels at summarizing existing information (Ganguli et al., 2022; Wang et al., 2023). However, this solution lacks an efficient validation process, which can increase the risk to patients.

5.2.3 The Case of the Radiology Department

Based on the results, a selection of administrative tasks identified that is performed by radiologists and radiology nurses are listed in table 5.1 below.

Table 5.1: A selection of administrative tasks performed by radiologist and radiology nurses.

Patient-related administration	Non-patient-related administration
Interpreting referrals	Preparing training materials
Prioritizing referrals	Writing protocols
Documenting examination results	Educating junior staff
Adjusting image quality and settings	Managing emails
Preparing for evaluation rounds	Handling internal complaints

The primary bottleneck identified by radiologists and radiology nurses is the administration related to referrals. From the results, four primary challenges related to referral administration have been identified. Firstly, there are frequent inadequacies in referral information, leading to delays and extra work, particularly in the musculoskeletal department. Secondly, there is significant frustration among radiology clinicians due to the variability in referral quality, especially from less experienced physicians, which directly exacerbates the first challenge. Thirdly, both departments spend considerable time interpreting and prioritizing referrals. Lastly, radiology clinicians emphasize the risk of potential overexposure to harmful radiation for patients due to poorly written referrals.

Thus, the largest challenges is particularly related to the interpretation and prioritization of referrals. Although these two tasks are interconnected since a referral can not be prioritized without being interpreted, it can be argued that these tasks can be streamlined in separate ways.

Based on the results from the radiology department, three main use cases for LLMs with the potential to significantly enhance the efficiency of managing referrals were identified. However, each use case also carries associated risks, with the highest risk linked to the first suggestion and diminishing with the subsequent methods.

1. Prioritization support of incoming referrals.

2. Standardization and automation of incoming referrals.

3. Quality check and enhancement of incoming referrals.

The first use case suggested is prioritization support of incoming referrals. This involves utilizing an LLM to assist in prioritizing referrals by analyzing the information in incoming referrals and categorizing the urgency and evaluation method based on predefined criteria. If implemented, it is recommended that the LLM automate the prioritization of adequate referrals while sorting out inadequate referrals for manual management by a radiologist. This solution has potentially high value, as it would save significant time for senior radiologists, who bear the greatest responsibility in prioritizing incoming referrals. Also, this approach aligns with experts' recommendations to prioritize use-cases that save time for highly specialized staff.

However, this solution also carries significant risk. Due to the risk for LLMs to provide inaccurate outputs (Haltaufderheide & Ranisch, 2023; Harrer, 2023) automating referral prioritization could have serious consequences for patients. Both inaccuracies in determining the appropriate evaluation method and the urgency of the evaluation can lead to substantial negative impacts on patient care, as highlighted in the results. This solution is also related to the idea of an LLM making clinical decisions, which can have potential in the longer term, but involves significant risks and uncertainties, as described in the results from expert interviews. Additionally, utilizing LLMs to interpret and prioritize referrals places considerable pressure on the referring physician to accurately formulate the referral. Thus, this must be considered if developing this suggested solution.

The second use case suggested is standardization and automation of incoming referrals. This approach involves implementing LLMs to develop and use standardized templates for common conditions, ensuring more consistent and comprehensive referral information. Addressing the significant challenge of inadequate referrals, this solution will assist the referring physician in writing better referrals. LLMs can autofill parts of these templates based on initial inputs made by the referring physician, ensuring that crucial information is not missed. This can significantly reduce the number of inadequate referrals, saving time for radiology clinicians. As senior radiologists are the professionals prioritizing the most referrals, this solution aims to save time for highly specialized staff.

However, this solution does have barriers. There remains a risk that automating referrals with LLMs could result in faulty referrals being sent to the radiology department, which could have significant consequences. This risk is somewhat mitigated by having the referring physician validate the LLM output before sending the referral, aligning with the view that LLM applications in healthcare should require human oversight (Haltaufderheide & Ranisch, 2023). Additionally, this approach might have a simpler alternative. As reported in the results, other solutions besides LLMs should be considered before prioritizing a potential use-case. One potential solution could be to implement standardized templates through collaboration between the radiology and referring departments, a strategy already being explored according to the interview results.

The third use case is quality check and enhancement of incoming referrals. This approach involves using LLMs to analyze incoming referrals, identify missing or inadequate information, and suggest necessary additions or modifications before the referral reaches the radiologist. This method directly addresses the issue of inadequate referrals. The process could either involve sending inadequate referrals back to the referring physician for validation or having the LLM correct the information and automatically forward it to the radiology department. The first option carries less risk since the LLMs' output is validated by the referring physician.

Moreover, LLMs can assist with non-patient administration tasks such as creating content for training materials and protocols. However, these tasks are performed relatively infrequently, with one radiologist noting that writing protocols occurs only once a year or every two years.

Furthermore, clinicians at the radiology department identify the computer systems as a major bottleneck. However, finding solutions for better performance, improved user experience, and enhanced communication between different systems falls outside the scope of this research. Therefore, while it is important to highlight this as a major issue, this research will not propose any solutions.

It can be concluded that radiologists and radiology nurses, particularly senior clinicians, spend considerable time on referral management, making a solution to this problem essential. However, the appropriateness of using LLMs for this purpose remains debatable. Integrating LLMs into the radiology department at Sahlgrenska University Hospital has the potential to significantly streamline referral management and other administrative tasks. Despite inherent risks, careful implementation, validation, and continuous improvement can mitigate these risks and enhance overall efficiency, allowing radiologists and nurses to focus more on patient care rather than administrative burdens.

Additionally, it can be argued that prioritization support has the greatest potential value capturing but also the highest risk. This solution is closely tied to clinical decisions, adding to its complexity. Also, the other two suggested solutions, standardization and automation and quality check and enhancement of incoming referrals, might be achievable through means other than LLM technology. Thus, further investigation is necessary before determining the best approach to addressing these challenges.

5.2.4 Comparison Between the Three Departments

It is evident that there is a clear need to streamline the patient-related administration in all three departments. Most clinicians strongly express that their current text-based administrative tasks are time-consuming and frustrating, detracting from more valuable direct patient care (Bringsén et al., 2012). It is further clear that the needs differ between the three departments. In addition, the clinicians' needs vary depending on whether they work in in-patient care or out-patient care. This thus adds a layer of complexity to prioritizing which use cases to move forward with. Therefore, it is important to evaluate which use cases will capture the most value, either across hospital departments or within a specific department. Figure 5.1 below illustrates these relationships.

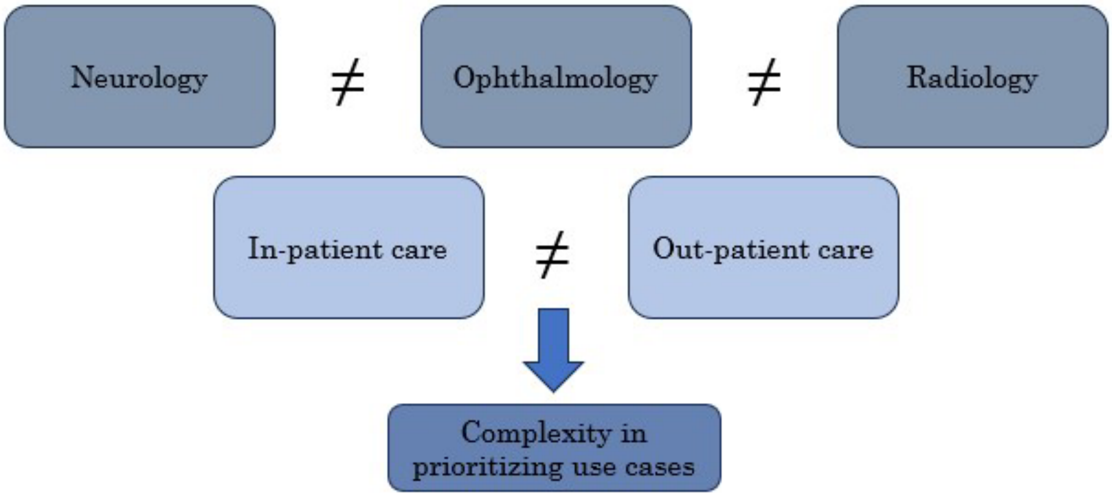

Figure 5.1: Visualization of the different needs across departments.

5.3 The Impact of Administrative Work on Clinicians' Health

As seen in the results, a significant majority of the physicians and nurses in all three departments find their administrative workload burdensome. This burden has a clear negative impact on the time available for direct patient care as described by (Anskär, 2019; Borelius, 2009; McKinsey & Company, 2019), but it also significantly affects the mental and physical health of clinicians as described by (Werdecker & Esch, 2021). The issue is particularly pronounced with administrative tasks that are perceived as illegitimate tasks, both unreasonable or unnecessary, as described by (Basinska & Dåderman, 2023; Björk et al., 2013; Semmer et al., 2015, 2019)

One notable example of an unnecessary task is managing the slow and disconnected computer systems that the clinicians use. This issue was highlighted as a major problem by all physicians at the ophthalmology clinic and neurology department, all nurses at the neurology department, and the majority of the clinicians from the radiology department. Another example of an unnecessary task is double documentation, which was predominantly reported as a significant problem by the neurology department clinicians.

Additionally, as explained by (Janlöv et al., 2023; Kilponen et al., 2021; Semmer et al., 2019), inefficiencies in organizational processes or outdated ways of working are also regarded as contributing to unnecessary tasks. At SU, the organizational processes can require clinicians to redo work, such as duplicating documentation tasks or performing extra documentation to comply with governmental regulations, as explained in the results. Further, a specific example of outdated ways of working is the continued use of physical mail for sending documents instead of digital transmission. Another example of organizational inefficiencies is the need to double check information on health record notes before signing them because a significant amount of time has passed since the notes were initially written. These issues are prevalent across all three departments according to the results.

Regarding unreasonable tasks, explained as tasks outside one's occupational role, position or competence (Kilponen et al., 2021; Semmer et al., 2019), several physicians at the ophthalmology department believe that they are performing administrative tasks that should not be part of their responsibilities. Specifically, physicians regard writing health record notes as unreasonable. They spend considerable time writing these notes, which detracts time from their primary duty of direct patient care. This sensation of encountering unreasonable tasks might be linked to the task shifting related to speech recognition, as outlined in the results. Furthermore, the physicians at the ophthalmology department have observed an overall increase in their administrative workload, which may lead to a heightened perception of performing unreasonable tasks. Nurses in the neurology department also encounter unreasonable tasks, specifically being asked to perform duties typically assigned to physicians. This not only diverts them from their own responsibilities but also impacts their efficiency and focus on nursing-specific tasks.

It can be concluded that clinicians from all three departments experience parts of their administrative tasks as illegitimate, and tasks that are considered illegitimate have a serious impact on the mental and physical health of clinicians (Anskär et al., 2019; Björk et al., 2013; Eatough et al., 2016; Kottwitz et al., 2013; Thun et al., 2018). Specifically, they can lead to lower self-esteem (Eatough et al., 2016; Schulte-Brauck et al., 2019; Semmer et al., 2015; Sonnentag & Lischetz, 2018), decreased sleep quality (Semmer et al., 2015), increased cortisol release, heightened feelings of anger and frustration, and a greater incidence of musculoskeletal pain (Eatough et al., 2016; Kottwitz et al., 2013). Additionally, illegitimate tasks have been found to correlate positively with burnout and negatively with job satisfaction (Björk et al., 2013; Werdecker & Esch, 2021). Thus, reducing unnecessary and unreasonable administrative tasks at Sahlgrenska University Hospital is crucial not only for improving efficiency but also for enhancing the well-being of clinicians. Streamlining these duties can lead to improved job satisfaction, lower health risks related to occupational stress, and higher staff retention rates. Ultimately, such measures help maintain a supportive and effective healthcare environment, ensuring that skilled professionals remain engaged and able to deliver high-quality patient care.

5.4 The Ethical and Legal Perspective of LLMs in Healthcare

It is clear that the potential to use LLMs is vast in various fields, including in healthcare (Reddy, 2023). Researchers and interviewed experts however agree that there are ethical concerns about this as well. In the ethical research literature about LLMs there is an agreement that the core limitations of the technology is the risk of the models generating information that is biased and propagate social prejudice against already vulnerable groups, that is incorrect, but gives the appearance of being correct, and that the user can not validate the generated information. Zhang and Kamel Boulos (2023) argues that the user can not know when the model is going to return a good or a misleading or wrong answer, and thus can not know when the model can be trusted or not, especially when the user is not qualified enough to assess the quality, the completeness and the accuracy of a given response.

In healthcare, this would mean that the clinicians, which need to be assessed as qualified enough, need to validate the information that LLMs generate. The interviewed experts agree that clinicians will bear some responsibility for AI generated outputs, but that there is no clear answer to the question of responsibility today. It is however of top priority to answer this question if LLMs are to be introduced in the healthcare sector. There is consensus that clinicians will need to double check and approve what LLMs generate. However, it should be taken into account that doctors often work under time pressure, which can create situations with less accurate approvals of LLM generated documents.

Due to the fact that the healthcare sector is particularly sensitive to mistakes, given that its outcome affects people's health directly, it is inevitable that the demands on technology in healthcare are especially high, more or less towards perfection. However, to realize the vast potential of implementing LLMs in healthcare, it is difficult to require a performance that is 100 %, since all solutions have some flaws, especially in the early phases of their development. One interviewed expert believes that the models can come to achieve in practice 100 % performance with time, but that mistakes will be made along the road to get there. The expert compares this to the evolution of delegating the writing of medical notes to medical secretaries, and most recently to be produced with speech recognition, which has not worked without fault.

All interviewed experts agree that the demands on machines in healthcare should be higher than that on humans. Some experts however raise the question of what is ethically right if machines will come to perform much better than humans. Some new literature, for example the study by Rosenberg et al. (2024), shows results of LLMs performing better than humans at some healthcare tasks, in this case on generating discharge summaries and discharge letters.

Moreover, the interviewed experts as well as the ethical research on LLMs highlight that there is a need for legal guidance when it comes to the usage of LLMs in society, not least in healthcare. The technology is novel, but is evolving quickly. In Sweden there are old laws that affect the area, like GDPR and Medical Devices, and new legal frameworks are currently entering the field.

To streamline clinicians administrative text-based tasks LLMs require the input of patient data. It is noted in the ethical research that inputting patient data into LLMs raises ethical questions regarding privacy, data security and confidentiality (Haltaufderheide & Ranisch, 2023). Personal health data is protected by the General Data Protection Regulation (GDPR) Article 9 (European Union, 2016). The propensity of LLMs to disseminate patient data is regarded as a serious privacy concern (Haltaufderheide & Ranisch, 2023). In some expert interviews, it is highlighted that guidance is needed when it comes to cloudbase processing. Today, using the cloud to store patient data contravenes GDPR and is thus not allowed in Sweden. It risks losing sensitive patient data to commercial actors. However, the best performing LLMs are currently cloud based. One interviewed expert further argues that it is cheaper to train local LLMs in the cloud to avoid investing in vast amounts of processing power, which is needed if the healthcare actors are to train their own LLMs on local servers. The vast processing power is only needed for the limited time for when the model is trained.

Furthermore, some interviewed experts emphasize the need to define if LLMs will be regarded as medical devices. The LLMs have to affect patient treatment to fall under this legal framework. It is clear that LLMs that are used as support for predictions and medical decisions do (Swedish Medical Products Agency, 2021). It is however not clear today if LLMs that are used to streamline tasks and involve patient data do, but if that would be the case, the demands on quality and testing of the LLMs will be significantly higher and it will take much longer time before they can be deployed in practice.

Zhang and Kamel Boulos (2023) underlines that there is an evident need for transparency with these models, and note that the EU AI Act includes transparency requirements for generative AI, including a need to publish summaries of copyrighted model-training data. OpenAI's models that currently perform best are kept secret to the public. Harrer (2023) agrees with (Zhang & Kamel Boulos, 2023), and adds that transparency, together with explainability, is especially important in healthcare for operators to be able to validate LLMs performances.

World Health Organization (2021) outlines a number of ethical principles for the responsible design and application of AI technology: protect autonomy; promote human safety and well-being and the public interest; ensure transparency, explainability and intelligibility; foster accountability and responsibility, ensure equity and inclusiveness; and promote AI that is sustainable and responsive. Harrer (2023) refers to WHO's principles and argues that there is a need for ethical and legal frameworks that can guide how to deploy AI applications and select and manage training data, and there is a need for measures to mitigate model biases. He highlights six key factors for responsible ethical design, use and governance of AI in healthcare: accountability, fairness, data privacy and selection, transparency, explainability, value and purpose alignment.

5.5 How to Implement LLMs in Healthcare

Below sections outlines for what can be learned from the case of Unity Health Toronto, as well as nine key factors to think about when implementing new technology in healthcare identified through the expert interviews.

5.5.1 What to Learn from the Case of Unity Health Toronto

Unity Health Toronto has been successful in implementing different AI tools in healthcare and has been recognized around the world for it. Actors can learn from their example. From day one, Unity Health Toronto's mission statement has been: how can we use data to help patients and help clinicians? The organization made a decision to focus on innovation, technology and data science in an applied way, instead of focusing on traditional research. Their mission statement and applied approach is described as the key to their success.

Unity Health Toronto began with making a decision to establish a contemporary data governance process and policy to build out their data infrastructure, something that was of great importance when they were dealing with the hospital's most sensitive data. They hired independent third party consultants in a wide variety of disciplines, such as legal advice, advising on regulation of what you can and can not do with patient data, cybersecurity and privacy experts and advising on policies. It was a significant financial investment, which they regard as ongoing with continuous improvement, not just a one-time-thing. In addition, they involved dedicated ethics boards and asked patients how they felt about the use of their data and about AI. Eventually, Unity Health Toronto formed four dedicated teams that work closely together to apply AI in healthcare: a data team, an advanced analytics team, a product development/software team and a project management team. This way of organization is regarded as unique and another key factor to Unity Health Toronto's success with implementing AI in an hospital environment.

Furthermore, the organization put emphasis on the importance that the need should come from the hospital employees, the end users. The Data Science and Advanced Analytics (DSAA) team identifies needs through intake forms they send out, that are filled out primarily by someone who is coming to them with a problem. The team compares what they receive via the forms with existing literature and knowledge as well as their own data to validate the size of the problem and the potential gains from solving it. They prioritize what gives the highest value.

However, Unity Health Toronto is currently not implementing LLMs in healthcare in relation to the clinicians or the patient interface. They use some LLM architectures for their own use, but not as a generative tool, but as a classifier.

5.5.2 What to Think About When Implementing New Technology in Healthcare

From the interviews with experts nine key implementation factors were identified.

First, the needs for new technological solutions should come from the hospital employees, the end users. If no one is going to use the solution, it makes no difference.

Second, benefits should be shown to the individuals who use it, the end users. Saving time for qualified labor is of top priority in the healthcare sector. Saving time is naturally followed by better allocation of resources which can result in higher quality of care.

Third, change management in healthcare requires resources set aside for it and the right anchoring based on existing needs across different levels in the organization. One expert argues that the change leaders need to have clinical backgrounds and clinical anchoring to be able to speak the right language.

Fourth, in order to speed up implementation, do clear evaluations, make early gains and make sure to show those early gains to gain attraction.

Fifth, be aware of and plan for future maintenance for software products that are set into production. These types of products require ongoing work and that should be embedded in the workflow for a dedicated group of employees.

Sixth, healthcare actors need to put clear demands on the health record suppliers to overcome the barrier of integrating new technology with existing systems and like that create incentives for these suppliers to integrate the new technology, in this case the LLMs.

Seventh, innovative suppliers that have the outside-in perspective need to find the right champions within the organization, that is, employees who are passionate about the issue and want to improve healthcare and sees the value of the technology.

Eighth, when a new technology comes with large potential, those who implement it will get a significant head start. If the healthcare actors do not follow the technological development evenly, it will cause inequality between different healthcare actors and thus, inequality among patients.

Ninth, the best way to move forward is to collaborate and work together when it comes to improving healthcare through innovation and technology.

6
Conclusion

This study has investigated how large language models can be used to streamline patient-related administrative text-based tasks for physicians and nurses at the neurology, ophthalmology and radiology department at Sahlgrenska University Hospital. It is clear that the demand for healthcare is continuously increasing in line with an ageing population. It is further a noted problem that clinicians have a significantly high administrative workload as a consequence of the digitalization, which take time away from valuable direct patient care. Clinicians' perform multiple text-based administrative tasks and it can be argued that LLMs have potential to streamline these tasks.

In recent years, LLMs have received great attention, led by the public introduction of ChatGPT in november 2022. Many agree that there is a vast potential to use LLMs, not least in the healthcare sector. However, there are ethical concerns as well, mainly the risk that LLMs generate biased and incorrect information, and that the information can not be validated. Due to the risk of LLMs generating biased and incorrect outputs, it is of great importance that healthcare actors that aim to implement these models have a strategy on how to validate the LLMs outputs. In this study, it is clear that implementing LLMs today, as a tool to streamline clinicians administrative text-based tasks, would require clinicians approval of the LLM generated texts.

It is clear that the clinicians at the neurology, ophthalmology and radiology department at Sahlgrenska University Hospital experience their administrative workload as too high. There is consensus among the clinicians that there are large variations in time required to complete text-based administrative tasks, and that this clearly correlates with the length and the complexity of the patient's health records and the current complex structure of their health records system, Melior, and accompanying systems.

Across the three departments, this study has found distinct use cases where LLMs could be used. While there is some similarities across the departments, the differences are even more pronounced. One clear difference that has been identified is between in-patient care and out-patient care. In in-patient care, the clinicians could benefit from auto generated drafts of notes that builds on existing data, for example discharge summaries. In out-patient care clinicians could benefit from using ambient scribes that generated notes based on a transcribed dialogue between a visiting patient and the clinician.

Furthermore, some notable similarities observed are the potential benefits for both the neurology and ophthalmology departments in receiving summaries before meeting with patients. Additionally, all departments, but especially the radiology and ophthalmology departments would gain from improved referral management processes. However, each use case within every department is constrained by varying risks and barriers, with the extent of limitation depending on the specific scenario. It should however be noted that some solutions that can be streamlined by using LLMs need to be investigated to see if they are more effectively addressed by alternative technologies or methods.

The pronounced differences are thus clear between the departments in this study. This complexity adds a factor to how Sahlgrenska University Hospital's management group should prioritize use-cases to pursue.

In conclusion, strategic advice for Sahlgrenska University Hospital and their potential future work with LLMs has been provided on the basis of this research's results. The research is argued to not only contribute to Sahlgrenska University Hospital, but also to other hospitals and regions in Sweden with a focus on strategic decisions related to LLMs in healthcare.

7
Recommendations

We recommend that Sahlgrenska University Hospital should:

1. Allocate resources to continue researching how LLMs can be used to streamline the clinicians' administrative text-based tasks.

2. Restructure the organization to enable AI projects.
   - Follow Unity Health Toronto's example and invest in competencies that can help organize how to successfully implement AI tools in a healthcare environment.
   - Organize dedicated teams for all process steps to work with deploying AI tools, like Unity Health Toronto.

3. Identify use cases that make a clear difference.
   - Identify the needs of the end users. In the case of this study, the clinicians'.
   - Identify and first-hand target use-cases that will save time for highly specialized clinicians.
   - Identify solutions that create the highest value, either by finding a solution that can be utilized for multiple departments or by targeting a solution for a specific department.

4. Actively take part in the development of LLMs in healthcare.
   - Integrate the end users in the development of LLM technology.
   - Identify and give champions within the organization the opportunity to take ownership of LLMs projects.
   - Create and communicate incentives for the health record suppliers to integrate LLMs.
   - Strategic options must be considered before moving forward with LLMs solutions.

8
Future Research

Throughout this study a few topics have been identified that would require future research. A selection of topics are presented here. First, it is recommended to conduct further research on how to manage the boundaries related to implementing LLMs in a clinical setting. Second, it is recommended to investigate the size of potential time gains that can be achieved by implementing LLMs to streamline clinicians' time-consuming text-based tasks. Third, it is recommended to research on how LLMs could help streamline the tasks for other healthcare professionals than physicians and nurses as well as for other departments than those included in this study.

Bibliography

Ahlstedt, C., Moberg, L., Brulin, E., Nyberg, A., Anell, A., & Merkur, S. (2023). Do illegitimate tasks matter for registered nurses' work motivation? a cross-sectional study based on a nationally representative sample of swedish nurses. *International Journal of Nursing Studies Advances*, *5*. Retrieved April 10, 2024, from https://doi.org/10.1016/j.ijnsa.2023.100159

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC medical education*, *23*(1). https://doi.org/https://doi.org/10.1186/s12909-023-04698-z

Anderson, B., & Sutherland, E. (2024). Collective action for responsible ai in health. (10). https://doi.org/https://doi.org/https://doi.org/10.1787/f2050177-en

Anskär, E. (2019). Time flies in primary care: A study on time utilisation and perceived psychosocial work environment. Retrieved April 10, 2024, from https://doi.org/10.3384/lic.diva-156119

Anskär, E., Lindberg, M., Falk, M., & Andersson, A. (2018). Time utilization and perceived psychosocial work environment among staff in swedish primary care settings. *BMC Health Serv Res*, *18*(166). Retrieved April 10, 2024, from https://doi.org/10.1186/s12913-018-2948-6

Anskär, E., Lindberg, M., Falk, M., & Andersson, A. (2019). Legitimacy of work tasks, psychosocial work environment, and time utilization among primary care staff in sweden. *SCANDINAVIAN JOURNAL OF PRIMARY HEALTH CARE*. Retrieved April 4, 2024, from https://doi.org/10.1080/02813432.2019.1684014

Aronsson, G., Bejerot, E., & Härenstam, A. (2012). Onödiga och oskäliga arbetsuppgifter bland läkare [unnecessary and unreasonable tasks among physicians]. *Läkartidningen*, *48*, 2216–2219. Retrieved April 2, 2024, from https://lakartidningen.se/klinik-och-vetenskap-1/2012/11/onodiga-och-oskaliga-arbetsuppgifter-bland-lakare/

Baeza-Yates, R. (2016). Data and algorithmic bias in the web. Retrieved March 29, 2024, from https://doi.org/10.1145/2908131.2908135

Bardram, J. (1997). "i love the system—i just don't use it!". Retrieved April 4, 2024, from https://doi.org/10.1145/266838.266922

Barkman, C., & Aasa, L. (2023). Onödig administration i sjukvården. Retrieved January 20, 2024, from https://healthpolicy.se/wp-content/uploads/2023/06/Rapport_Onodig-administration-sjukvard-_juni-2023.pdf

Basinska, B., & Dåderman, A. (2023). Psychometric properties of the bern illegitimate tasks scale using classical test and item response theories. *Scientifc Reports*, *13*(7211). Retrieved April 10, 2024, from https://doi.org/10.1038/s41598-023-34006-0

Bell, A., Bryman, E., & Harley, B. (2019). *Business research methods* (Vol. 5). Oxford University Press.

Bergmann, D. (2024, February). The most important ai trends in 2024. *IBM*. Retrieved March 24, 2024, from https://www.ibm.com/blog/artificial-intelligence-trends/

Björk, L., Bejerot, E., Jacobshagen, N., & Ha renstam, A. (2013). I shouldn't have to do this: Illegitimate tasks as a stressor in relation to organizational control and resource deficits. *Work Stress*, *27*(3), 262–277. Retrieved April 2, 2024, from https://doi.org/10.1080/02678373.2013.818291

Borelius, C. (2009). The district nurses work time in focus. Retrieved April 10, 2024, from diva2:518735

Bowman, R. (2023). Eight things to know about llm. Retrieved March 15, 2024, from https://doi.org/10.48550/arXiv.2304.00612

Bringsén, Å., Andersson, H. I., Ejlertsson, G., & Troein, M. (2012). Exploring workplace related health resources from a salutogenic perspective. results from a focus group study among healthcare workers in sweden. *Work (Reading, Mass.)*, *42*, 403–414. Retrieved April 8, 2024, from https://doi.org/10.3233/WOR-2012-1356

Chen, Y., Qian, S., Tang, H., Lai, X., Liu, Z., Han, S., & Jia, J. (2024). Longlora: Efficient fine-tuning of longcontext large language models. Retrieved March 15, 2024, from https://doi.org/10.48550/arXiv.2309.12307

Crampton, N. (2019). Ambient virtual scribes: Mutuo health's autoscribe as a case study of artificial intelligence-based technology. *Healthcare Management Forum*, *33*, 34–38. Retrieved May 20, 2024, from https://doi.org/10.1177/0840470419872775

Dave, M., & Patel, N. (2023). Artificial intelligence in healthcare and education. *British Dental Journal*, *234*(10), 761–764. https://doi.org/10.1038/s41415-023-5845-2

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. Retrieved April 2, 2024, from https://doi.org/10.48550/arXiv.1810.04805

Dhalla, A., Akinli Kocak, S., Wan, F., Da Silva, J., & Jain, T. (2023). Health ai implementation toolkit. *V.10.* Retrieved April 30, 2024, from https://vectorinstitute.ai/health-ai-implementation-toolkit/

DiGiorgio, A., & Ehrenfeld, J. (2023). Artificial intelligence in medicine chatgpt: De-tether the physician. *Journal of Medical Systems*, *47*(1). https://doi.org/https://doi.org/10.1007/s10916-023-01926-3

Easterby-Smith, M., Thorpe, R., & Jackson, P. R. (2015). *Management and business research* (Vol. 5). Sage Publications Ltd.

Eatough, E., Meier, L., Igic, I., Elfering, A., Spector, P., & Semmer, N. (2016). You want me to do what? two daily diary studies of illegitimate tasks and employee well-being. *Journal of Organizational Behavior*, *37*, 108–127. Retrieved April 8, 2024, from https://doi.org/10.1002/job.2032

Effektiv vård. (2016). Swedish Government Official Reports. Retrieved April 9, 2024, from https://www.regeringen.se/contentassets/42b0aef4431c4ebf9410b8ee771830eb/effektiv-vard---slutbetankande-av-en-nationell-samordnare-for-effektivare-resursutnyttjande-inom-halso--och-sjukvarden_sou-2016-2.pdf

Eriksson, L. T., & Wiedersheim-Paul, F. (2014). *Att utreda, forska och rapportera* (Vol. 10). Liber.

European Union. (2016). Art. 9 gdpr – processing of special categories of personal data. Retrieved April 14, 2024, from https://gdpr-info.eu/art-9-gdpr/

European Union. (2024). High-level summary of the ai act. Retrieved April 14, 2024, from https://artificialintelligenceact.eu/high-level-summary/

Fruhlinger, J. (2023). What is generative ai? artificial intelligence that creates. Retrieved March 24, 2024, from https://www.infoworld.com/article/3689973/what-is-generative-ai-artificial-intelligence-that-creates.html

Ganguli, D., Hernandez, D., Lovitt, L., Dassarma, N., Henighan, T., Jones, A., Joseph, N., Kernion, J., Mann, B., Askell, A., Bai, Y., Chen, A., Conerly, T., Drain, D., Elhage, N., El Showk, S., Fort, S., Hatfield-Dodds, Z., Johnston, S., . . . Clark, J. (2022). Predictability and surprise in large generative models. Retrieved March 20, 2024, from https://doi.org/10.48550/arXiv.2202.07785

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). Retrieval-augmented generation for large language models:

A survey. Retrieved May 20, 2024, from https://doi.org/10.48550/arXiv.2312.10997

Ghojogh, B., & Ghodsi, B. (2020). Attention mechanism, transformers, bert, and gpt: Tutorial and survey. Retrieved March 26, 2024, from https://doi.org/10.31219/osf.io/m6gcn

Grant Thornton. (2023). *Vårdrapporten 2023: Så mår den privata vård- och omsorgssektorn i Sverige*.

Grodal, S., Anteby, M., & Holm, A. (2021). Achieving rigor in qualitative analysis: The role of active categorization in theory building. *Academy of Management Review*, *46*. Retrieved May 23, 2024, from http://doi.org/10.5465/amr.2018.0482

Guo, Z., Renren, J., Liu, C., Huang, Y., Shi, D., Linhao Yu, S., Liu, Y., Li, J., Xiong, B., & Xiong, D. (2023). Evaluating large language models: A comprehensive survey. Retrieved April 20, 2024, from https://doi.org/10.48550/arXiv.2310.19736

Hadi, M., Tashi, Q., Qureshi, R., Shah, A., Muneer, A., Irfan, M., Zafar, A., Shaikh, M., Akhtar, N., J., W., & Mirjalili, S. (2023). Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects. Retrieved March 20, 2024, from https://doi.org/10.36227/techrxiv.23589741.v3

Haltaufderheide, J., & Ranisch, R. (2023). The ethics of chatgpt in medicine and healthcare: A systematic review on large language models (llms). https://doi.org/https://doi.org/10.48550arXiv.2403.14473

Harrer, S. (2023). Attention is not all you need: The complicated case of ethically using large language models in healthcare and medicine. *eBioMedicine*, *90*. https://doi.org/https://doi.org/10.1016/j.ebiom.2023.104512

Hoffman, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., de Las Casas, D., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., van den Driessche, G., Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., . . . Sifre, L. (2022). Training compute-optimal large language models. Retrieved March 20, 2024, from https://doi.org/10.48550/arXiv.2203.15556

Holberg, P., & Bell, L. (2015). Vårdnära administration. Retrieved April 10, 2024, from 978-91-7585-312-3

Holten Møller, N., & Vikkelsø, S. (2012). The clinical work of secretaries: Exploring the intersection of administrative and clinical work in the diagnosing process. Retrieved April 8, 2024, from https://doi.org/10.1007/978-1-4471-4093-1_3

IBM. (n.d.). What are large language models (llms)? Retrieved May 22, 2024, from https://www.ibm.com/topics/large-language-models

Ilicki, J. (2023a). A framework for critically assessing chatgpt and other large language artificial intelligence model applications in health care. Retrieved March 29, 2024, from https://doi.org/10.1016/j.mcpdig.2023.03.006

Ilicki, J. (2023b). Why does healthcare have so much administration? Retrieved April 23, 2024, from https://emmetropia.substack.com/p/why-healthcare-has-so-much-administration

Ilicki, J. (2023c). Ett stort steg stora språkmodeller i sjukvården. *Forum för Health Policy*. Retrieved March 24, 2024, from https://healthpolicy.se/rapport-ett-stort-steg-stora-sprakmodeller-i-sjukvarden/

Janlöv, N., Blume, S., Glenngård, A., Hanspers, K., Anell, A., & Merkur, S. (2023). Sweden: Health system review. *Health Systems in Transition*, *25*(4). Retrieved April 10, 2024, from ISBN%20978%209%202890%205947%203

Kaddour, J., Harris, J., Mozes, M., Bradley, H., Raileanu, R., & McHardy, R. (2023). Challenges and applications of large language models. Retrieved February 20, 2024, from https://doi.org/10.48550/arXiv.2307.10169

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. Retrieved March 10, 2024, from https://doi.org/10.48550/arXiv.2001.08361

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). Chatgpt for good? on opportunities and challenges of large language models for education. *ResearchGate*. Retrieved March 24, 2024, from https://www.researchgate.net/publication/367541637_ChatGPT_for_Good_On_Opportunities_and_Challenges_of_Large_Language_Models_for_Education

Kilponen, K., Huhtala, M., Kinnunen, U., Mauno, S., & Feldt, T. (2021). Illegitimate tasks in health care: Illegitimate task types and associations with occupational well-being. *Journal of Clinical Nursing*, *30*, 2093–2106. Retrieved April 8, 2024, from https://doi.org/10.1111/jocn.15767

Kottwitz, M., Meier, L., Jacobshagen, N., Kälin, W., Elfering, A., Hennig, J., & Semmer, N. (2013). Illegitimate tasks associated with higher cortisol levels amongmale employees when subjective health is relatively low: Anintra-individual analysis. *Scandinavian Journal of Work environment health*, *34*. Retrieved April 30, 2024, from doi:10.5271/sjweh.3334

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Retrieved May 20, 2024, from https://doi.org/10.48550/arXiv.2005.11401

Li, Y., Li, Z., Zhang, K., Dan, R., Jiang, S., & Zhang, Y. (2023). Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge. Retrieved March 15, 2024, from https://doi.org/10.7759/cureus.40895

Liévin, V., Egeberg Hother, C., Geert Motzfeldt, A., & Winther, O. (2022). Can large language models reason about medical questions? Retrieved March 29, 2024, from https://doi.org/10.48550/arXiv.2207.08143

Liu, L., He, H., Han, T., Zhang, X., Liu, M., Tian, J., Zhang, Y., Wang, J., Zhang, J., Gao, X., Zhong, T., Pan, Y., Xu, S., Wu, Z., Liu, Z., Zhang, X., Zhang, S., Hu, X., Zhang, T., . . . Ge, B. (2024). Understanding llms: A comprehensive overview from training to inference. Retrieved March 26, 2024, from https://doi.org/10.48550/arXiv.2401.02038

McKinsey & Company. (2019). Tid till vård ger vård i tid. Retrieved March 26, 2024, from https://www.mckinsey.com/~/media/mckinsey/featured%20insights/europe/time%20for%20care%20means%20care%20on%20time/tid-till-vard-ger-vard-i-tid.pdf

Nabwani, N. (2023). Full fine-tuning, peft, prompt engineering, and rag: Which one is right for you? Retrieved April 20, 2024, from https://deci.ai/blog/fine-tuning-peft-prompt-engineering-and-rag-which-one-is-right-for-you/

Navigli, R., Conia, S., & Ross, B. (2023). Biases in large language models: Origins, inventory, and discussion. Retrieved March 26, 2024, from https://doi.org/10.1145/3597307

Petročnik, T., Palmieri, S., & Marot, J.-A. (2023, May). The ai act and european health data space proposal: Seeing 'ai to ai' with each other [Blogpost 26/2023 at European Law Blog, News and Comments on EU Law]. Retrieved April 14, 2024, from https://europeanlawblog.eu/2023/05/30/the-ai-act-and-european-health-data-space-proposal-seeing-ai-to-ai-with-each-other/

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, *21*, 1–67. Retrieved April 2, 2024, from https://doi.org/10.48550/arXiv.1910.10683

Ramlochan, S. (2023). Fundamental abilities of production ready large language models. Retrieved May 22, 2024, from https://promptengineering.org/fundamental-abilities-of-large-language-models/

Raschka. (2023). *Build a large language model (from scratch).*

Reddy, S. (2023). Evaluating large language models for use in healthcare: A framework for translational value assessment. *Informatics in Medicine Unlocked*, *41*, 101304. https://doi.org/https://doi.org/10.1016/j.imu.2023.101304

Rosenberg, G., Magnéli, M., Barle, N., Kontakis, M., Müller, A., Wittauer, M., Gordon, M., & Broden, C. (2024). Chatgpt-4 generates orthopedic discharge documents faster than humans maintaining comparable quality: A pilot study of 6 cases. *Acta Orthopaedica*, *95*, 152–156. https://doi.org/10.2340/17453674.2024.40182

Sætra, H. S. (2023). Generative ai: Here to stay, but for good? *Technology in Society*, *75*, 102372. https://doi.org/https://doi.org/10.1016/j.techsoc.2023.102372

Sahlgrenska Universitetssjukhuset. (2023a). Kompetenscentrum ai – datadrivet arbete i vården - sahlgrenska universitetssjukhuset. Retrieved January 20, 2024, from https://www.sahlgrenska.se/forskning-utbildning-innovation/samverkan/kompetenscentrum-ai/

Sahlgrenska Universitetssjukhuset. (2023b). Om sjukhuset - sahlgrenska universitetssjukhuset. Retrieved January 20, 2024, from https://www.sahlgrenska.se/om-sjukhuset/

Sallam, M. (2023). Chatgpt utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare 2023*, *11*(6). https://doi.org/https://doi.org/10.3390/healthcare11060887

Schulte-Brauck, J., Dormann, C., & Vahle-Hinz, T. (2019). Get even and feel good?moderating effects of justice sensitivity and counterproductive work behavior on the relation-ship between illegitimate tasks and self-esteem. *Journal of Occupational Health Psychology*, *24*, 241–255. Retrieved April 8, 2024, from https://doi.org/10.1037/ocp0000112

Semmer, N., Jacobshagen, N., Elfering, A., & Meier, L. (2007). Occupational stress research: The "stress-as-offense-to-self" perspectiv. *Occupational health psychology: European perspectives on research, education and practice*, *2*, 43–60.

Semmer, N., Jacobshagen, N., Meier, L., Elfering, A., Beehr, T., Kälin, W., & Tschan, F. (2015). Illegitimate tasks as a source of work stress. *Work Stress*, *29*(1), 32–56. Retrieved April 2, 2024, from https://doi.org/10.1080/02678373.2014.1003996

Semmer, N., Jacobshagen, N., Tschan, F., Elfering, A., Beehr, T., Kälin, W., & Tschan, F. (2019). Stress as offense to self: A promising approach comes of age. *Occupational Health Science*, *3*, 205–238. Retrieved April 10, 2024, from https://doi.org/10.1007/s41542-019-00041-5

Semmer, N., Tschan, F., Meier, L., Facchin, S., & Jacobshagen, N. (2009). Illegitimate tasks and counterproductive work behavior. *Applied Psychology*, *59*(1), 70–96. Retrieved April 10, 2024, from https://doi.org/10.1111/j.1464-0597.2009.00416.x

Shanahan, M. (2023). Talking about large language models. Retrieved March 10, 2024, from https://doi.org/10.48550/arxiv.2212.03551

Singhal, K., Tu, T., Gottweis, J., Sayres, R., Wulczyn, E., Hou, L., Clark, K., Pfohl, S., Cole-Lewis, H., Neal, D., Mansfield, P., Prakash, S., Green, B., Dominowska, E., Agüera y Arcas, B., Tomasev, N., Liu, Y., Wong, R., Semturs, C., ... Natarajan, V. (2023). Towards expert-level medical question answering with large language models. Retrieved April 20, 2024, from https://doi.org/10.48550/arXiv.2305.09617

Sonnentag, S., & Lischetz, T. (2018). Illegitimate tasks reach into afterwork hours: A multilevelstudy. *Journal of Occupational Health Psychology*, *23*(2), 248–261. Retrieved April 2, 2024, from https://doi.org/10.1037/ocp0000077

Strauss, A., Fagerhaugh, S., Suczek, B., & Wiener, C. (1985). *Social organization of medical work.* (Vol. 10). The University of Chicago Press.

Stringhi, E. (2023). Hallucinating (or poorly fed) llms? the problem of data accuracy. Retrieved March 26, 2024, from https://doi.org/10.6092/issn.1825-1927/18877

Swedish Medical Products Agency. (2021). Is it a medical device? [Published: January 25, 2021. Last updated: January 25, 2021.]. Retrieved May 27, 2024, from https://www.lakemedelsverket.se/en/medical-devices/manufacture/is-it-a-medical-device#hmainbody1

Thirunavukarasu, A., Ting, D., Elangovan, K., Gutierrez, L., Tan, T., & Ting, D. (2023). Large language models in medicine. *Nature Medicine*, *29*(8), 1930–1940. https://doi.org/https://doi.org/10.1038/s41591-023-02448-8

Thun, S., Halsteinli, V., & Løvseth, L. (2018). A study of unreasonable illegitimate tasks, administrative tasks, and sickness presenteeism amongst norwegian physicians: An everyday struggle? *BMC Health Serv Res*, *18*(407), 248–261. Retrieved April 2, 2024, from https://doi.org/10.1186/s12913-018-3229-0

Tierney, A., Gayre, G., Hoberman, B., Mattern, B., Ballesca, M., Kipnis, P., Liu, V., & Lee, K. (2024). Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. *NEJM Catal Innov Care Deliv*, *5*. Retrieved May 20, 2024, from DOI:10.1056/CAT.23.0404

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, N. A., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Retrieved March 2, 2024, from https://doi.org/10.48550/arXiv.1706.03762

Viswanath, H., & Zhang, T. (2023). Fairpy: A toolkit for evaluation of social biases and their mitigation in large language models. Retrieved March 23, 2024, from https://doi.org/10.48550/arXiv.2302.05508

Wang, Y., Yang, G., Li, S., Li, Y., He, L., & Liu, D. (2023). Arrhythmia classification algorithm based on multi-head self-attention mechanism. Retrieved March 26, 2024, from https://doi.org/10.1016/j.bspc.2022.104206

Werdecker, L., & Esch, T. (2021). Burnout, satisfaction and happiness among german general practitioners (gps): A crosssectional survey on health resources and stressors. *PloS one*, *16*(6). Retrieved April 10, 2024, from https://doi.org/10.1371/journal.pone.0253447

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health*. Retrieved April 13, 2024, from https://www.who.int/publications/i/item/9789240029200

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, *2*(10). https://doi.org/https://doi.org/10.1038/s41551-018-0305-z

Zhang, P., & Kamel Boulos, M. N. (2023). Generative ai in medicine and healthcare: Promises, opportunities and challenges. *Future Internet*, *15*(9). https://doi.org/10.3390/fi15090286

Zhao, X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., . . . Wen, J.-R. (2023). A survey of large language models. Retrieved March 8, 2024, from https://doi.org/10.48550/arXiv.2303.18223

Zheng, Y., Wang, L., Feng, B., Zhao, A., & Wu, Y. (2023). Innovating healthcare: The role of chatgpt in streamlining hospital workflow in the future. *Annals of biomedical engineering*, *52*(4), 750–753. https://doi.org/https://doi.org/10.1007/s10439-023-03323-w

A
Appendix 1

The interview guides and self-completion forms that are used in this study are presented below.

A.1 Interview Guide for Clinician Interviews

Name, Role

Questions:

- Would you like to tell us briefly what you do/what your role is?

- What kind of administrative work do you do?
  - Do your administrative tasks take up a lot of time? Give examples of time-consuming tasks.
  - Are there any bottlenecks that make it take a long time that you thought of?
  - External factors?
  - Do you have any idea how it can be made more efficient?

- Do you write and/or read/interpret referrals?

- How much time do you spend on administrative work on average per day?
  - Do you feel it should be more, less or is it reasonable?

- Is there anything you would like to add?

A.2 Interview Guide for Expert Interviews

Name, Role

Questions:

- Tell us briefly about your work/your role.

- For what tasks do you see potential to implement LLMs in healthcare? (How do [organization] work with LLMs today?)

- What would you define as the low hanging fruit when it comes to AI in healthcare, and could it be approached with LLMs?

- Many researchers highlight the vast potential of integrating AI in healthcare, but they are also aware that there are large barriers to overcome in order to reach this potential. What would you say are the most central limitations/barriers to implement LLMs in healthcare (or text generating AI-models?).

- When implementing AI in healthcare, how do you conduct risk assessments beforehand?

- In our project we are asking for the clinicians' needs first hand and are trying to find solutions based on that.

  - How do/can you integrate with the clinicians to identify their needs?

  - How do/can you prioritize which needs to solve?

  - How do/can you measure the performance of the AI solutions? (Do [your organization] have any specific KPIs)?

- What do you think is important when implementing digital solutions in clinical setting? From idea to implementation of a solution.

  - What are the largest obstacles, and how do/can you tackle them?

- Do you know of any hospital in the world that is very prominent in LLMs/AI solutions, that we can take inspiration from?

- How do you think Sahlgrenska/VGR/Sweden should tackle the increased challenges in healthcare (aging population, limited resources, etc.)? Is AI part of the solution?

- What is the most important thing that VGR/Sahlgrenska can learn from your work with LLMs?

A.3 Self-Completion Forms

In the figures below, the self-completion forms are presented.

Figure A.1 presents the second, updated form that was updated after the feedback that was received from the first time the form was distributed among physicians at the neurology department, differentiating between TIK, dictate and manual writing.

Figure A.2 presents the first form that was distributed among the physicians at the neurology department.

Figure A.3 presents the form that was prepared for the neurology nurses, which did not receive any answers.

Figure A.4 presents the form that was prepared for the ophthalmology department, which did not get the opportunity to be distributed. Note that this form would have been update similarly as the updated neurology form, differentiating between TIK, dictate and manual writing, if an opportunity to distribute it would have been realized.

Figure A.5 presents the form that was distributed at the radiology department, which did not receive any valid answers.

Yrkesroll + mottagning eller avdelning

Antal år i tjänst (månader om <1år)

Neurologi/Neurokirurgi, Läkare

Syftet med undersökningen är att ta reda på hur mycket tid ni lägger på nedan administrativa uppgifter. Vänligen fyll i hur lång tid hela processen tar för remisser och slutanteckningar, dvs. inkludera förberedelser som inläsning av patientbakgrund, läsa vårdförlopp, skriv-/dikteringstiden osv.

Läsa på om en patient inför ett besök		
Exempel	*Start: 13.03*	*Slut: 13.28*
1	Start:	Slut:
2	Start:	Slut:
3	Start:	Slut:
4	Start:	Slut:
5	Start:	Slut:
Totalt antal idag (behöver inte motsvara ovan): …		

Skapa remiss						
Exempel	**Antal: 2*	*Start: 13.03*	*Slut: 13.28*	*TIK*	*Dikterat***	*Själv****
1	Antal:	Start:	Slut:	TIK	Dikterat	Själv
2	Antal:	Start:	Slut:	TIK	Dikterat	Själv
3	Antal:	Start:	Slut:	TIK	Dikterat	Själv
4	Antal:	Start:	Slut:	TIK	Dikterat	Själv
5	Antal:	Start:	Slut:	TIK	Dikterat	Själv
Totalt antal idag (behöver inte motsvara ovan): …						

* Skriver du fler remisser vid ett tillfälle, vänligen skriv antalet remisser du skrivit inom tiden du angivit.
** Diktera innebär att diktera och skicka till sekreterare.
*** Själv innebär att du har skapat remissen genom att skriva direkt på datorn.

Skapa slutanteckning					
Exempel	*Start: 13.03*	*Slut: 13.28*	*TIK*	*Dikterat*	*Själv*
1	Start:	Slut:	*TIK*	*Dikterat*	*Själv*
2	Start:	Slut:	TIK	Dikterat	Själv
3	Start:	Slut:	TIK	Dikterat	Själv
4	Start:	Slut:	TIK	Dikterat	Själv
5	Start:	Slut:	TIK	Dikterat	Själv
Totalt antal idag (behöver inte motsvara ovan): …					

Figure A.1: Self-completion form for physicians at the neurology department, second update.

Yrkesroll + mottagning eller avdelning

Antal år i tjänst (månader om <1år)

Neurologi, Läkare

Syftet med undersökningen är att ta reda på hur mycket tid ni lägger på nedan administrativa uppgifter. Vänligen fyll i hur lång tid hela processen tar för remisser och slutanteckningar, dvs. inkludera förberedelser som inläsning av patientbakgrund, läsa vårdförlopp, skriv-/dikteringstiden osv.

Läsa på om en patient inför ett besök		
Exempel	*Start: 13.03*	*Slut: 13.28*
1	Start:	Slut:
2	Start:	Slut:
3	Start:	Slut:
4	Start:	Slut:
5	Start:	Slut:
Totalt antal idag (behöver inte motsvara ovan): …		

Skapa remiss						
Exempel	**Antal: 2*	*Start: 13.03*	*Slut: 13.28*	*Dikterat:*	*JA*	*NEJ*
1	Antal:	Start:	Slut:	Dikterat:	JA	NEJ
2	Antal:	Start:	Slut:	Dikterat:	JA	NEJ
3	Antal:	Start:	Slut:	Dikterat:	JA	NEJ
4	Antal:	Start:	Slut:	Dikterat:	JA	NEJ
5	Antal:	Start:	Slut:	Dikterat:	JA	NEJ
Totalt antal idag (behöver inte motsvara ovan): …						

*Skriver du fler remisser vid ett tillfälle, vänligen skriv antalet remisser du skrivit inom tiden du angivit.

Skapa slutanteckning					
Exempel	*Start: 13.03*	*Slut: 13.28*	*Dikterat:*	*JA*	*NEJ*
1	Start:	Slut:	Dikterat:	JA	NEJ
2	Start:	Slut:	Dikterat:	JA	NEJ
3	Start:	Slut:	Dikterat:	JA	NEJ
4	Start:	Slut:	Dikterat:	JA	NEJ
5	Start:	Slut:	Dikterat:	JA	NEJ
Totalt antal idag (behöver inte motsvara ovan): …					

Figure A.2: Self-completion form for physicians at the neurology department, first update.

Yrkesroll + mottagning eller avdelning

Antal år i tjänst (månader om <1år)

Neurologi, Sjuksköterska

Syftet med undersökningen är att ta reda på hur mycket tid ni lägger på nedan administrativa uppgifter. Vänligen fyll i hur lång tid hela processen tar för slutanteckningen, dvs. inkludera förberedelser som inläsning av patientbakgrund, läsa vårdförlopp, skrivtiden osv.

Läsa på om en patient inför ett besök		
Exempel	*Start: 13.03*	*Slut: 13.28*
1	Start:	Slut:
2	Start:	Slut:
3	Start:	Slut:
4	Start:	Slut:
5	Start:	Slut:
Totalt antal idag (behöver inte motsvara ovan): …		

Skapa en slutanteckning			
Exempel	*Start: 13.03*	*Slut: 13.28*	**Vilken del: Hela*
1	Start:	Slut:	Vilken del:
2	Start:	Slut:	Vilken del:
3	Start:	Slut:	Vilken del:
4	Start:	Slut:	Vilken del:
5	Start:	Slut:	Vilken del:
Totalt antal idag (behöver inte motsvara ovan): …			

*Skriv vilken del av slutanteckningen som skapats inom tiden du angivit (hela/bakgrund/etc.)

Figure A.3: Self-completion form for nurses at the neurology department.

Yrkesroll ______________________ Antal år i tjänst (månader om <1år) ______________________

Ögon, Läkare

Syftet med undersökningen är att ta reda på hur mycket tid ni lägger på nedan administrativa uppgifter samt hur stor andel bristfälliga remisser ni hanterar.

Hantera remiss: *Exempel:*

Antal remisser som du hanterar idag	*Bristfälliga remisser som kräver extra tid idag*

***Hantera remiss:** Fyll i:

Antal remisser som du hanterar idag	Bristfälliga remisser som kräver extra tid idag

*Vänligen fyll i antalet mottagna remisser du hanterar idag samt antalet remisser som bedöms som bristfälliga och kräver mer tid att hantera än om remissen varit av förväntad/normal kvalitet. Remisser som är bristfälliga kan sakna information, ha motsägelsefull information, efterfråga en undersökning som ifrågasätts av er, etc.

Läsa på om en patient inför ett besök		
Exempel	*Start: 13.03*	*Slut: 13.28*
1	Start:	Slut:
2	Start:	Slut:
3	Start:	Slut:
4	Start:	Slut:
5	Start:	Slut:
Totalt antal idag (behöver inte motsvara ovan): …		

***Skapa slutanteckning**			
Exempel	*Start: 13.03*	*Slut: 13.28*	*Dikterat: JA NEJ*
1	Start:	Slut:	Dikterat: JA NEJ
2	Start:	Slut:	Dikterat: JA NEJ
3	Start:	Slut:	Dikterat: JA NEJ
4	Start:	Slut:	Dikterat: JA NEJ
5	Start:	Slut:	Dikterat: JA NEJ
Totalt antal idag (behöver inte motsvara ovan): …			

*Vänligen fyll i hur lång tid hela processen tar för slutanteckningen, dvs. inkludera förberedelser som inläsning av patientbakgrund, läsa vårdförlopp, skriv-/dikteringstiden osv.

Figure A.4: Self-completion form for physicians at the ophthalmology department.

Yrkesroll ______________________ Antal år i tjänst (månader om <1år) ______________________

Radiologi

Syftet med undersökningen är att ta reda på hur många remisser som kräver extraarbete och därmed extra administrativ tid.

Vänligen fyll i antalet remisser du hanterar på en dag samt antalet remisser som bedöms som bristfälliga och kräver mer tid att hantera än om remissen varit av förväntad/normal kvalitet. Remisser som är bristfälliga kan sakna information, ha motsägelsefull information, efterfråga en undersökning som ifrågasätts av er, etc.

Exempel:

Antal remisser som du hanterar idag	*Bristfälliga remisser som kräver extra tid idag*

Fyll i:

Antal remisser som du hanterar idag	Bristfälliga remisser som kräver extra tid idag

Vänligen fyll i tidsåtgången för ett urval av remisserna du hanterat idag. Markera om remissen varit bristfällig och krävt mer tid att hantera än om remissen varit av förväntad/normal kvalitet.

Tid att hantera remiss			
Exempel	*Bristfällig: JA NEJ*	*Start: 13.03*	*Slut: 13.28*
1	Bristfällig: JA NEJ	Start:	Slut:
2	Bristfällig: JA NEJ	Start:	Slut:
3	Bristfällig: JA NEJ	Start:	Slut:
4	Bristfällig: JA NEJ	Start:	Slut:
5	Bristfällig: JA NEJ	Start:	Slut:

Figure A.5: Self-completion form for physicians at the radiology department.

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF INNOVATION AND R&D MANAGEMENT
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY