SCHOOL OF ECONOMICS AND MANAGEMENT
LUND UNIVERSITY

Managing Responsibility & AI in Healthcare

Facilitating Managerial Learning on Responsibility when Implementing Artificial Intelligence Decision Support Systems in Healthcare

By

Sindhu Charan Pisipati & Sandra Lofs Midelf

June 2022

Master Programme in Management

Supervisor: Ola Mattison
Examiner: Rikard Larsson

Abstract

The increased use of artificial intelligence (AI) in a multitude of fields comes with many benefits but also managerial and ethical challenges. One such challenge relates to how we attribute and distribute responsibility when artificial intelligence decision support systems (AI-DSS) are part of human decision-making, also known as the "responsibility gap". This paper explores what factors can facilitate managerial learning regarding responsibility when AI-DSS is introduced in clinical healthcare, and thus aim to contribute to the research gap on this matter. An overview of the current literature is provided relating to AI, responsibility, and organizational learning, including legal and ethical frameworks that have been established or suggested to guide behavior. Based on semi-structured interviews with experts on AI, ethics, regulations, and healthcare professionals, the researchers present findings concluded by five factors that can facilitate managerial learning on the issue of responsibility when introducing AI-DSS in healthcare: diverse and cross-functional teams, critical and up-to-date assessments of legal and ethical frameworks, comprehensive cost-benefit analyses, education, and meaningful human control.

Keywords: artificial intelligence, responsibility gap, clinical ethics, managerial learning

Acknowledgments

Before all else, we would like to thank our amazing supervisor Ola Mattison from Lund University School of Economics and Management for his guidance and support throughout the entire process of the research. Ola has shown great interest and belief in our project and has given us plenty of constructive feedback. We are truly impressed by and thankful for Ola's ability to juggle the many responsibilities that come with teaching, supervising, and directing the program, while at the same time being a very friendly and positive human being. Having written this thesis, we know that responsibilities are not an easy matter.

We would also like to express our appreciation to the other teachers, our fellow students, and everyone else that has been a part of the Master's Program in Management 2021/2022, making it an incredible learning experience full of hard work and fun moments. Furthermore, we want to express our gratitude to all the experts in AI, ethics, regulations, and healthcare that gave us their time and provided us with valuable insights. The thesis would not have been possible without them. Lastly, the researchers would like to acknowledge those that have supported them on their life journey:

I, Sandra L. Midelf, would like to express my deepest gratitude to my mom who has always pushed me to be my best independent self and supported me in whatever interests I have had. To my dad, I'm thankful that he always reminds me not to worry too much. To the love of my life, Søren Mai, whom I can always be myself around and who always encourages me to think critically. To my mentor Hanna Beneker, who has shared her experience and supported me. I would also like to thank my tutor, Thekla Schneede, for the incredibly devoted support she has given me and my team during this year. Lastly, I must thank my dog Fjönt, my emotional support who is always by my side.

I, Pisipati Sindhu Charan, would like to express my gratitude to my parents who always gave me the best they could provide and always encouraged me to chart a path of my own and carve a place for me in this world, their undeterred support and unshakable trust made me what I am today. I would also like to thank my sister who always instills in me a competitive spirit and motivation. Lastly, I thank my dog Ricky who I am sure supports my endeavors.

Table of Contents

List of Tables & Figures

Abbreviations

AI	Artificial Intelligence
AI-DSS	Artificial Intelligence Decision Support System
CE	Conformité Européenne
GCP	Good Clinical Practice
GDPR	General Data Protection Regulation
GMAIH	Governance Model for AI in Healthcare
EU	European Union
IBM	International Business Machines Corporation
MHC	Meaningful Human Control
RQ	Research Question
R&D	Research and Development
SMACTR	Scoping, Mapping, Artifact Collection, Testing, and Reflection
TRIPOD	The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis
WHO	World Health Organization
WIPO	World Intellectual Property Organization

Terminology

Accountability	Responsibility used retrospectively, i.e. holding something or somebody accountable for something they have done.
Agent	Something or somebody performing an action.
Algorithm	An algorithm is a procedure used for solving a problem or performing a computation.
Epistemic	Relating to knowledge or knowing.
Expert	A person that has a lot of knowledge or skill in a particular field.
Patient	1) An individual receiving medical treatment, or 2) something or somebody that is affected by an agent's action.
Researchers	Those conducting this thesis, i.e. Sandra Lofs Midelf and Sindhu Charan Pisipati.
Regulators	A person or an organization within a national or international authoritative body controlling an area of business or industry to ensure that it operates fairly and legally.
Physician	Individual qualified to practice medicine.
Healthcare Manager	Individuals responsible for general management and administration of hospitals, healthcare units, and public healthcare systems.

1. Introduction

Imagine the following scenario: Dr. Smith has a patient, James 65 years old, who is admitted because of lung cancer and seems to have severe bleeding because of it. The artificial intelligence decision support system (AI-DSS) that Dr. Smith uses in his work recommends that James should be treated through chemotherapy and be prescribed a drug called Bevacizumab, often used to treat certain types of cancer, including lung cancer. After Dr. Smith has gone over the recommendation, he finds himself doubting whether he should follow through with the recommendation or trust his intuition that advises against using Bevacizumab. However, Dr. Smith decides to administer the drug because he has been told that the accuracy of this AI-DSS is very high. However, it turns out that there is a warning on the drug that says it significantly increases the risk of severe or fatal blood loss, especially in combination with chemotherapy. As a result of Dr. Smith's decision to follow through with the AI-DSS's recommendation, James suffers from more bleeding and his health becomes critical. Dr. Smith gets in a meeting with management in his healthcare organization about the incident. You are a part of the management - how do you deal with the situation?

1.1 Background

The example above is fictional and simplified but based on a case from the real world, where the patient was part of a system testing of the AI-decision support system "Watson for Oncology" developed by International Business Machines Corporation (IBM). It serves well as an illustration of what can potentially go wrong when we let AI-DSS inform our judgments and decisions within a field where a wrong decision can have undesired and unintended consequences, i.e. consequences that are such that they lead to harm or injustice.

In recent years there has been an explosion in the usage of artificial intelligence (AI), and there is virtually no sector that remains untouched by AI and its potential applications (EU Commission, 2021). AI is primarily used for making decisions when human cognition is limited and when we want to make much more informed decisions. The healthcare industry has also seen a significant uptick in the usage of AI (Bohr & Memarzadeh, 2020). The applications range from assessments and decisions within diagnostics, prognostics, and prevention to robotics-assisted surgery, administration, and discovery of new drugs (Brynjolfsson & Mitchell, 2017). This thesis focuses on the use of AI as a decision support system, but the term AI and AI-DSS will be used interchangeably since the problems that relate to AI-DSS often are applicable to AI in general as well.

The example above is specifically intended to encapsulate the question of who is to be held responsible for the damages done to the patient, in this case, James. Since the development of AI involves many stakeholders, for example, AI developers, physicians, regulatory bodies, ethicists, patients, etc. (WHO, 2021), it is a challenge to attribute and distribute responsibility properly, both in terms of holding somebody accountable for something that they have done (retrospective responsibility, also referred to as accountability) and clarifying what obligations and duties somebody have when looking into the future (prospective responsibility) (Noorman, 2020). In the literature, this challenge has been referred to as the "responsibility gap" (Matthias, 2004).

To meet this challenge along with many others, legal and ethical frameworks have been developed to guide the management of AI in general (e.g. EU Commission, 2021; Santoni de Sio & van den Hoven, 2018) and in healthcare in particular (e.g. Reddy et.al., 2020). There is also extensive literature on organizational learning (Mazmanian, Davis & Galbraith, 2009) and on how to manage stakeholder diversity (e.g. Curseu & Schruijer, 2017).

1.2 Problem statement

AI technology is rapidly being developed and used in many fields, and how to best govern and regulate it is globally contested (Cath, 2018). Due to the high-stake interests involved in the field of healthcare, responsibility is an important concept and even more so when introducing technology that assists human decision-making (Reed,2018). In the introductory example, responsibility relates to how Dr. Smith and other stakeholders should be held at fault after James has suffered from the consequences of the recommended treatment. Prospectively, it should be determined what duties and obligations Dr. Smith and other stakeholders should have so that the incident does not happen again. In the chain of process, it is important to understand that AI, often, is a product of many actors and interactions between them, which includes developers, companies, AI itself, and its users (Fosso & Wamba, 2022).

There is extensive philosophical literature on how responsibility should be understood (Talbert, 2019), and ethical and regulatory frameworks do consider issues relating to AI and responsibility (e.g. Bleher and Braun, 2022; EU Commission, 2021; GAO, 2021). However, there is relatively little said in the management literature about what may facilitate managers in learning how to address the issue of responsibility when introducing AI in healthcare (Baker-Brunnbaur, 2021). Yet it is important that the issue does not get left unresolved, since it risks making everybody's problem nobody's responsibility (Floridi, 2016) in an industry that has great importance to a great number of people, and the likelihood of the development and usage of AI becoming common practice in healthcare is high.

1.3 Research Purpose & Question

The purpose of this study is to identify factors that enable managerial and organizational learning by exploring the concept of responsibility when using AI-DSS within the public health sector. Hence, an overview of the literature on AI and healthcare is provided, as well as a discussion on how responsibility is understood, how legal and ethical frameworks address the issue, and how organizational learning is central to the implementation of AI-DSS in public healthcare. In order to evaluate the research purpose, the following question is of interest for the study:

> R.Q. What factors can facilitate managerial learning when introducing AI-DSS in public clinical healthcare in terms of responsibility?

1.4 Delimitations

The primary focus and objective of the thesis is to explore the managerial aspects of responsibility when AI-DSS is used in healthcare. The scope of the thesis excludes core technical and design aspects, as well as any broader socio-ethical or legal analysis of AI. Methodological limitations are elaborated further in section 3.6.

1.5 Outline

The entire study comprises six chapters. Chapter one introduces the reader to the background of the research and presents the research questions and problem statement. Chapter two provides the reader with a more in-depth understanding of AI in healthcare, responsibility, healthcare management, and organizational learning. Chapter three presents the choice of methodology, including how empirical data was collected and analyzed. In chapter four empirical data is outlined, followed by chapter five which consists of discussion, analysis, and answer to the research question. Chapter six summarizes and concludes the thesis with gained insights and points to potential future research.

2. Theoretical Framework

This chapter focuses on presenting existing research that contextualizes the current study, covering the field of AI, healthcare, ethics, and managerial learning, including ethical and legal frameworks. Doing so showcases that there is a gap that justifies the research question, and thereby, along with empirical data presented in chapter four, lays the groundwork for performing an analysis and subsequently presenting the findings.

2.1 AI in Public Healthcare

2.1.1 Description of AI

Artificial intelligence or AI, in general, is defined as the science and engineering of making intelligent machines, through algorithms or a set of rules, which the machine follows to mimic human cognitive functions, such as learning and problem solving (McCarthy, 1955). It has wide-reaching usage along with increasing shaping of daily life and some of its applications include advanced web search engines like Google, self-driving cars, automatic language translation, etc.

AI can be used to mimic any intellectual task (Russel and Norvig, 2009), and hence it offers scope to be used in multiple fields potentially impacting different areas of human activity (WIPO, 2019). Research on AI was confined to academia for a long time after its initial discovery and was revived after the 1980s with increased funding from the public and private sectors after realizing viability for commercial success (Crevier,1993). Currently, AI is emerging as the most focused area for research and this phenomenon is being described as the “AI wave” and is said to have the potential to fundamentally alter the course of the human way of life (WIPO, 2019).

2.1.2 AI in Public Setting

Implementation of digital and electronic services across public sector organizations has been shown to improve efficiency and productivity (Bertot, Estevez, & Janowski, 2016) and improve satisfaction amongst the public (Sangki,2018). However, legacy systems that exist within public organizations have tied down the progress of advanced technologies (Mehr, Ash, & Fellow, 2017). In this context, the private sector is well placed in adopting such technologies and the public sector needs to catch up to stay on par with improved customer satisfaction and consumer confidence (van Deursen, van Dijk, & Ebbers, 2006). To take advantage of using AI in public services, managers from public organizations have encouraged and are considering the adoption of such intelligent decision-making systems (Mehr, 2017). It is also imperative for public organizations they consider ingraining different ethical and moral aspects to ensure there is trust, transparency, and fairness when embedding AI into their day-to-day functions (Leslie,2019).

2.1.3 Adoption of AI in Healthcare

Healthcare has always been a forerunner in adopting the latest technologies. The need for taking such initiatives in healthcare was to ensure that doctors make informed decisions (Lysaght, Lim & Ngiam, 2019) leading to patients receiving better medical care and potentially saving lives (Karunanithi, 2007). AI in healthcare is an enveloping term used to illustrate the application of different machine-learning algorithms and software, or artificial intelligence (AI), to ape human cognition in the analysis, presentation, and synthesis of complex medical and healthcare-related data in any form varying from structured to unstructured.

One of the key healthcare-related applications is to analyze relationships between clinical techniques and patient outcomes (Coiera E, 1997). With the explosion of availability of data (Hilbert & Lopez, 2011) and increased computing power (Ahrens et. al., 2011), the scope of usage of AI in the healthcare sector has increased manifold. Usage of AI has also proved to be cost-effective in healthcare (Higgins & Madai, 2020). Robotics along with AI are also emerging as key technologies in patient care and in the European Union, where new compliance and safety rules are in effect (Tsang et al., 2017) and they are now being complemented by significant investment by governments and technology companies (Milliard, 2018). The deployment of AI-DSS has huge ethical and moral implications (Ludwin & Murry, 2017) that have to be considered before using it.

Currently, where AI serves a supporting/augmenting role in diagnosis and/or treatment/operation processes, some may assume that the role of physicians/doctors will be rendered obsolete. However, it is important for us to evaluate the role of AI in different opportunities and challenges in its applications in the healthcare industry. Some of the important challenges or risks associated with widespread usage of AI-DSS in healthcare are data privacy concerns, cybersecurity, data integrity issues, data ownership, the problem of cross data-sharing by different organizational silos, clinical ethics issues, responsibility for medical errors, and risks of system failures (Yoon & Lee, 2019). Currently, governance policies and ethical guidelines for clinical healthcare services that deploy AI and its assorted applications fall behind the pace of advancements in AI (Rigby, 2019).

2.1.4 Explainability of AI in healthcare

In a highly regulated and life-critical field like healthcare, any adoption of new technology in diagnosis, treatments, or drug discovery is important to ensure that it is adequately tested and analyzed before its deployment (Wienert, 2019). In this context of the adoption of AI in Healthcare, the concept of explainability has emerged as a counterweight to the famous black-box (Pasquale, 2015; Rudin, 2019) interpretation of AI. A black box is generally defined as a complex or complicated system whose internal workings are hidden or not readily understood (Wienert, 2019). Explainability in AI has many definitions and interpretations, one of the definitions is the idea of understanding how algorithms in AI function in a way that is accountable, transparent, and fair (Köhl et. al., 2019). Investigations into the function of explainability in AI in clinical settings concluded that to mitigate challenges to ethical principles, explainability is of utmost importance (Amann et. al., 2020). This can help different stakeholders involved in offering clinical healthcare services to ensure that their patient's interests are at the core of the treatment they receive and also the patients can make informed and independent decisions about their health with the assistance of medical professionals.

Explainability in AI is generally divided and understood in terms of transparency, interpretability, and explainability. Transparency can be defined as the processes that extract model parameters from training data and generate labels from testing data that can be comprehended, described, and configured by the designer (Roscher et. al.,2020). Interpretability can be defined as the ability to understand the ML model and present the underlying basis for decision-making in a way that is comprehensible to everyone (Arrietaet. al.,2020). Lastly, explainability in AI can be considered as a collection of features of the interpretable domain that have contributed to a given example to produce a decision (Montavon et. al., 2018).

2.1.5 Limitations of AI in healthcare

There is a myriad of limitations that public healthcare units might encounter when using AI in clinical settings (Coiera, 1996), some of the limitations they may encounter are lack of availability of data (Gijsberts, 2015), the inability of the system to process different data points, integration and transferability issues like a mismatch between theoretical simulation and practical results(Nolan, McNair & Brender, 1991) and other ethical and privacy concerns (Ludwin and Murry, 2017) which are covered in detail in the subsequent sections. It is important for a manager to understand the limitations posed by automated decision systems like AI (Janssen & Kuk, 2016) and ensure that said system is augmenting existing capabilities (Miller, 2018), and also ensure that it is equitable, stakeholder inclusive, and delivers quality service for the end-user (Kuziemski, & Misuraca,2020).

2.1.6 Conclusion

This part of the theoretical framework focuses on introducing AI and different aspects of AI that are associated with the scope of usage of AI in clinical healthcare. Together they carve out a dimension of AI-DSS that managers need to acclimatize in order to introduce AI-DSS in clinical healthcare. This begins with providing a standard definition of AI and proceeds with understanding its context in the public sector and then proceeds to the adoption of AI-DSS in healthcare, putting together these topics permits a manager in getting a better understanding of the applicability of AI-DSS in a public healthcare setting. It next focuses on the concept of explainability in AI and points out different limitations of AI in healthcare, these topics point to critical aspects of AI-DSS that need to be observed before implementing AI-DSS in healthcare. In summary, this part helps provide a foundational knowledge of AI-DSS in the setting of public healthcare that helps managers understand the dimension of responsibility.

2.2 Responsibility & AI

As mentioned previously, AI is created and used in healthcare with the purpose to advance human health and solve the wicked problem of increased costs and troublesome outcomes faced by healthcare agencies (Morley et.al. 2020; Topol, 2019). However, the use of AI also brings ethical issues with it and is a relatively new field within applied ethics (Müller, 2021). This part focuses on one central ethical dilemma, *the responsibility gap(s)*, which is about how to distribute responsibility when clinical AI is involved in decision-making. First, the definition of responsibility and its distribution is discussed, and second, the dilemma of responsibility when introducing AI is described more in-depth.

2.2.1 Defining Responsibility

Holding people accountable for what they have done and asking what their responsibilities are is central to human interaction and moral practice (Eshleman, 2016), but there is a widespread philosophical debate on how these concepts should be understood. For example, it is not completely clear what makes moral responsibility different from other types of responsibility, and how it relates to notions such as liability and blame (Talbert, 2019; Noorman, 2020). From a managerial perspective, the concept of responsibility is relevant from both a causal, moral, and legal point of view (Bleher & Braun, 2022).

Generally, to ascribe responsibility is to say that there is a link between a person or group often referred to as an *agent*, and whatever or whoever is affected by their actions often referred to as a *patient* (Noorman, 2020). In this context, “patient” is not to be confused with a patient in the sense of an individual receiving medical treatment, but more generally as somebody who is affected by another one’s actions. In our introductory example, James is a patient in both these senses. Another example would be a person A that is drowning in a pond and a person B that is considering whether to jump in and save A or not. In this case, person A is a patient that will be affected by B’s choice to jump in or not, and B is an agent that will, depending on how B acts, be blamed or praised.

Most accounts on (moral) responsibility hold that the following three conditions must be met for someone to be held responsible: 1) there must be a level of control, in the sense of causality, between the agent’s action and the outcome, 2) the agent must know of and be able to consider potential consequences of the actions, and 3) the agent must be free to choose the course of action and not be forced (Noorman, 2020). The epistemic condition (2) relates to explainability and transparency (Coeckelbergh, 2020) which were discussed in section 2.2.4.

The link between agent and patient that makes up the responsibility can be established both retrospectively and prospectively (Noorman, 2020). When using the term retrospectively, it concerns who is accountable for the action and consequences in question and rightfully subject to blame and punishment. This is also what is called accountability. When talking about responsibility prospectively, it concerns what duties and obligations that person has. This distinction matters since it shed light on related but divergent considerations for managerial practice. On the one hand, a manager must know what responsibilities there are to take the right precautions through routines and procedures and involve the right individuals in this. On the other hand, a manager must know how to deal with the failure of living up to these responsibilities. Henceforth, the term responsibility will refer to both retrospective and prospective responsibility, unless specified otherwise.

Distributed Responsibility

There are various positions on what conditions are required to attribute moral (and/or legal) responsibility, and whether it can be attributed to for example children, humans with dementia, and non-humans (Talbert, 2019). This debate is of relevance to whether or not machines can or should be seen as moral agents, which is briefly discussed under section *2.3.2 The AI Accountability Gap(s)*. Furthermore, it is not always easy to establish the link between an agent and a patient, especially when considering the complexity of human interaction and the use of technology. This complexity introduces the notion of *collective responsibility* where groups as a whole are seen as an agent, for example, organizations or states, and *shared responsibility* where responsibility is distributed between individuals that have acted together in a group (Smiley, 2017).

The distribution of responsibility is often referred to in the context of AI as a support system in human decision-making since there are many actors involved in its development, implementation, and use (Sullivan & Fosso, 2022; WHO, 2021; Taddeo & Floridi, 2018), but the notion of collective and shared responsibility is theoretically controversial (Smiley, 2017) and pragmatically difficult (Floridi, 2016). It is difficult to distribute responsibility because each of the humans' or artificial agents' individual actions might not be morally troublesome in themselves, it is only the product of many agents' actions that might lead to that which is morally wrong (Floridi, 2016). In our introductory example, where Dr. Smith is prescribing a treatment to James by using a recommendation from an AI-DSS, it was not only Dr. Smith's action that led to James's suffering. Dr. Smith was, arguably, not the only agent responsible for the outcome, but also for example the developers and managers behind the creation and implementation of the AI-DSS.

Whatever our theoretical position is on the matter of distributed responsibility, it comes with practical managerial challenges that need to be addressed (Baker-Brunnbaur, 2021; Gualdi & Cordella, 2021). If the issue is left unresolved the risk is that "everybody's problem becomes nobody's responsibility" (p. 11, Floridi, 2016).

2.2.2 The AI Responsibility Gap(s)

Defining the Gap

Matthias (2004) introduced what he called the *responsibility gap*. The gap occurs when humans give up some of their control and knowledge about their decisions to AI since it is generally thought that people are to be held responsible only for actions they have control over (Santoni de Sio & Mecacci, 2021; Noorman, 2020). Responsibility may, in the context of discussing the gap, refer to either moral or legal responsibility, and henceforth the term "responsibility" will refer to both unless specified otherwise. As touched upon in the previous section, the notion of distributed responsibility sheds light on the difficulty of identifying individual agents when the outcome is a result of interactions between several agents. It may lead to what has been called the "diffusion of responsibility" in social psychology and legal contexts (Bleher & Braun, 2022).

The responsibility gap has received extensive attention in recent years (Braun et.al., 2021; Coeckelbergh, 2012; Nyholm, 2018). Some have argued that the gap does not exist (e.g. Tigard, 2021), and it has also been argued that the gap should be seen as a set of at least four problems with different sources, and not only one (Santoni de Sio & Mecacci, 2021). The four problems that Santoni de Sio and Mecacci (2021) are pointing at bring interesting nuance to different agents and what responsibilities they have in both public and private contexts. For example, the authors argue that the gap in moral accountability highlights that the developers of AI, and not only physicians or users, have limited knowledge about the logic of the AI. Particularly when AI is based on deep learning techniques (Castelvecchi, 2016) which ultimately raises questions about the issue of explainability and when we can trust AI and those developing it to be reliable enough (von Eschenbach, 2021).

Moral Agency

Matthias (2004) argues that the responsibility gap will grow as AI becomes more autonomous since that implies that humans have less control and knowledge about the decisions being made. This begs the question of whether a machine itself can ever be held morally responsible, i.e. be seen as a moral agent and not only a tool, and what conditions must be in place for that to be the case. There is extensive literature that discusses what the scope of moral agency is, and it is often thought to include both natural and legal persons but has gained further attention because of the increased autonomy of AI (Nyholm, 2018; Floridi, 2013).

Sullivan’s and Fosso’s (2022) study, for example, shows that mind perceptions and how we attribute agentic characteristics to AI play a role in how we evaluate AI behavior and that outcomes of AI are a matter of distributed agency, i.e. outcomes that come about because of decisions and actions made by many agents (Taddeo & Floridi, 2018). Since the responsibility gap may exist even if AI is not fully autonomous, and AI may be understood as “quasi-moral” agents (Coeckelbergh, 2010), the debate on how to understand moral agency will not be further elaborated here. However, it is a debate of relevance to the issue of responsibility in a potential future with artificial consciousness (Müller, 2021).

2.2.3 Conclusion

In order to understand how to manage the introduction of AI-DSS in healthcare in terms of responsibility, an account of what is meant by “responsibility” was necessary. Three central aspects were identified in order to attribute responsibility to someone or something, namely that the agent has sufficient control, knowledge, and freedom of choice. Responsibility has also been presented as something which can be applied prospectively or retrospectively, which makes a difference to managerial practice in the sense that responsibilities can cover both preventive actions and actions of reparation.

Since many stakeholders are involved when introducing AI-DSS in healthcare, responsibility as something which can be distributed among these was important to explore. However, the distribution of responsibility is not an easy endeavor, particularly when AI-DSS is involved, which has been described as the “AI responsibility gap”. This issue is at the heart of what managers need to learn when engaging with the introduction of AI-DSS in healthcare.

2.3 Managerial Guidance & Organizational Learning in Healthcare

In the field of healthcare, organizational learning is associated with significant improvement in clinical outcomes (Richter, 2013), improved staff morale and satisfaction (Sadeghifar, 2014), and subsequently psychological safety inside an organization (Edmondson, 1999). Organizational learning can be defined as “a process of inducing positive change in an organization's collective, cognition, knowledge, and actions, which subsequently enhances its ability to achieve its desired outcomes” (Lyman, Hammond, & Cox, 2018). Subsequent sections help understand the phenomenon of organizational learning in the context of healthcare and act as a foundational step towards fostering and developing a theoretical groundwork needed to guide research and accelerate, stimulate, and sustain improvement in healthcare. The first two sections offer an overview of the theoretical and regulatory frameworks that are in place to guide managerial decision-making when engaging with AI and dealing with the issue of distributing responsibility. The third section concerns management in healthcare and organizational learning, including the management of stakeholder diversity.

2.3.1 Scholarly Frameworks for Managing AI & Responsibility

There is a global discussion among academics, industry representatives, regulators, and society at large on how AI should be governed from an ethical, legal, and technological perspective (Cath, 2018). Although the discussion about how to practically deal with the issue of responsibility when using AI in healthcare has been sparse (Reddy et.al. 2020), some frameworks can be applied to the governance of AI in general, and healthcare in particular. Some of these frameworks have been developed by scholars and others by national and international authorities. In this section, some of the frameworks presented by scholars will be discussed and compared in regard to how they address the issue of distributing responsibility.

Closing the Gap

Santoni de Sio and Mecacci (2021) argue that the offered solutions to the problem of distributing responsibility so far are flawed because they only offer a partial understanding of the problem and its different sources. For example, they argue that increasing transparency and explainability of AI is not sufficient to close the responsibility gap since the problem of distributing responsibility remains similar to what has more generally been called “the problem of many hands” (Coeckelbergh, 2020; Thompson, 1980).

In for example debates on lethal autonomous weapons, automated driving systems, as well as in medical automation, an approach has been suggested to minimize the responsibility gaps through technical, organizational, and legal design, referred to as the idea of “meaningful human control” (MHC). The bottom line of the approach is to keep humans in the loop and in meaningful control, which consists of *tracking* and *tracing* (Santoni de Sio & van den Hoven, 2018). This means mapping out the relevant agents and their reasons, intentions, and relations to the AI, as well as realizing what technical, motivational, and moral capacities the agents can be expected to have. The idea of MHC has been criticized for not being defined enough to derive legal regulations and guidelines (see e.g.

Roff & Moyes, 2016), but points to some very important aspects of building a framework on responsibility and AI.

Bleher and Braun (2022) offer a preventative resource-based approach to the issue, which includes managing responsibility on three levels: the design of the machine, the human-machine interaction, and the decision stage. The approach overlaps with the MHC approach in many ways, as it empathizes for example 1) the importance of human control and the role of causality, 2) the need for including all agents in decision-making and ethical reflection from the beginning 3) identifying what expectations there are of all the agents, 4) that the distribution of responsibility is highly contextual, and 5) the need for managing risks.

Reddy et.al. (2020) suggests what they call the Governance Model for AI in Healthcare (GMAIH) which consists of four components of which the fourth one is accountability. Similar to the approaches discussed above, GMAIH also emphasizes the inclusion of all agents throughout the whole process of development and implementation of AI. The GMAIH needs to be used in conjunction with other models, for example benchmarking systems (Salathé, Wiegand & Wenzel, 2018), to help individual health care institutions make the right choices on the market.

Other models of relevance to the GMAIH are the TRIPOD model (the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) (Collins et. al., 2015), as well as the more general notion of clinical governance that is there to ensure quality care (Halligan & Donaldson, 2001). In regard to the auditing of AI, which is a part of the GMAIH, there are many accounts to be found in the AI governance literature, for example, the auditing framework SMACTR (Scoping, Mapping, Artifact Collection, Testing, and Reflection) developed by Partnership on AI and Google (Raji et. al., 2020). As a part of integrating the GMAIH into the clinical workflow (Bowens, Frye, & Jones, 2010), Reddy et.al. suggest that clinical governance committees should be arranged and include “clinicians, managers, patient group representatives, and technical and ethics experts” (p. 495, 2020).

2.3.2 Legal Frameworks for Managing AI & Responsibility

It is not always clear how to identify responsible parties when something goes wrong in medical settings even when there are medical regulations in place, and it further complicates the matter when

AI is used as a support. In the previous section, a brief snapshot of the scholarly discussion on how to manage responsibility was considered. This section will instead give a very brief overview of how some international and national agencies address the issue.

International Frameworks: EU & WHO

The AI Act is a proposed law by the European Commission, developed through "extensive consultation with all major stakeholders" (p. 7, EU Commission, 2021), intended to become a global standard much like the EU's General Data Protection Regulation (GDPR). The Act is still a work in process, "[t]aking into account the rapid process in AI technologies and the tardiness of law-making" (Kiseleva, 2020), and has been praised as a good initiative but also criticized for lacking definition and clarity on several points (see e.g. Ministry of Infrastructure, 2021; Kiseleva, 2021).

The Act does not have a particular section about the issue of responsibility, but the recognition of several stakeholders and their obligations shows concern for this issue. The Act describes what obligations there are on providers and users, in terms of for example training, traceability, transparency, accuracy, and human oversight throughout the AI's life cycle. Those people that are assigned to oversee the AI must "have the necessary competence, training, and authority to carry out that role" (p. 30, EU Commission, 2021), which is in line with the conditions for holding somebody responsible, as discussed in section 2.3.1. The recognition of the need to have human oversight over AI is following the earlier mentioned MHC approach. High-risk AI, i.e. for example medical devices (Sioli, 2021), should according to the AI Act be CE marked to indicate that they are conforming with the regulation.

While the AI Act is a suggested law on AI in general, the World Health Organization (WHO) has provided ethical guidance on AI in healthcare in particular and covers a lot of issues on the topic, including that of distributing responsibility (WHO, 2021). The WHO guidance encourages what they call "human warranty" and human supervision of AI to assure responsibility (p. 28, WHO, 2021). Furthermore, the WHO guidance also recognizes that distribution of responsibility among several agents (including the public) is necessary when using AI technologies. To avoid diffusion of responsibility it is argued that "a faultless responsibility model" should be deployed to "encourage all actors to act with integrity and minimize harm" (p. 28, WHO, 2021). However, despite the multi-stakeholder approach, the WHO states that in certain cases of complex AI where the outcome is not foreseeable, assigning

> “responsibility to the developer might provide an incentive to take all possible steps to minimize harm to the patient. Such expectations are already well established for the producers of other commonly used medical technologies, including drug and vaccine manufacturers, medical device companies, and medical equipment makers.” (p. 42, WHO, 2021).

The WHO also provides practical guidance, based on the principles and ideas in the report, for developers and designers of AI, ministries of health, and healthcare providers and institutions.

National Frameworks

The World Health Organization (2021) states that the challenge of distributing responsibility from a legal perspective remains unsettled in many countries. However, there are several national guidelines on AI and ethics, many of which at least consider the issue of responsibility. The U.S. was “one of the first countries where AI-based helathcare applications were authorized to be placed on the market as medical devices” (Kiseleva, 2020). The U.S. Government Accountability Office (GAO, 2021) in order to “help managers ensure accountability and responsible use of artificial intelligence (AI) in government programs and processes” (GAO Fast Facts, 2021). The framework suggests that managers can promote accountability through several key practices, questions, and audit procedures relating to each of four complementary principles: governance, data, performance, and monitoring.

Other examples of national frameworks are Canada’s government’s “Directive on Automated Decision-Making” (Canada Directive, 2021), Singapore’s government’s “Model AI Governance Framework” (2nd edition) (Info-comm Media, 2020), In the United Kingdom’s “Ethics, Transparency and Accountability Framework for Automated Decision-Making” (UK Guidance, 2021), and Sweden's “National approach to artificial intelligence” (Ministry of Enterprise and Innovation, n.d) and “A handbook for information-driven healthcare - Insights from the inside” (authors’ translation, original title in Swedish: “En handbok för information-driven vård”; Lingman et. al., 2021).

Most of the frameworks mentioned in this section are not specifically aimed at AI technology in healthcare, and they are more or less detailed in terms of recommended practices. However, to conclude, they at least share an emphasis on the following features that relate to responsibility: 1) the importance of including all the numerous stakeholders from a multitude of disciplines throughout the whole AI life cycle, 2) the importance of risk management in order to appropriately govern the AI, 3) the importance of clarifying tasks and responsibilities to the relevant actors, 3) the importance of providing training and necessary knowledge for employees to fulfill their assigned responsibilities, 4) the importance of human supervision to ensure accountability, and 5) the importance of ensuring compliance with relevant laws and regulations.

2.3.3 Healthcare Management & Managerial Learning

Healthcare management refers to a specific area of healthcare that intertwines social and technical functions and activities within a healthcare organization with the aim of accomplishing various predetermined objectives with help from humans and other resources (Rakich et. al., 1993). The need for managerial decision-making in healthcare and public healthcare was identified by WHO in the early 2000s and was subsequently incorporated into their framework of governance (WHO, 2014). There is demonstrable evidence of such decision makings effectiveness in acting as a critical element in providing sustainable clinical outcomes (Duffy & Lemieux, 1995).

In this context of an environment of increased awareness of managerial decision-making in healthcare, facilitating organizational learning and understanding the resultant change it brings about in an organization is important for a manager to ensure that such change disseminates throughout the organization. Organizational learning can be described as collective learning inside an organization in such a way that it has an impact on its performance (Goh, Chan, & Kuziemsky, 2013). Organizational learning provides a framework for collective safe patient care in inter-connected and complex organizations (Mazmanian, Davis & Galbraith, 2009). Hence managers in healthcare need to accommodate these changes to their system iteratively and continuously to become a learning healthcare organization.

As described in section 2.2.1 on distributing responsibility, introducing AI-DSS in healthcare involves many stakeholders from different fields of expertise. The general literature on knowledge management usually describes exploiting existing knowledge within an organization as part of the endeavor to becoming a learning organization (Mårtensson, 2000) and diversity as a competitive advantage if done right (Bolman & Deal, 2017), this includes organizations in the public sector where "[e]ffective government and stakeholder partnerships" are important for creating good public policy (p. 36, Riege & Lindsay, 2006). However, stakeholder diversity risks inducing conflict because of differences in terms of "interests, power, perspectives, and identities, etc." (p. 115, Curseu &

Schruijer, 2017). In order to deal with this challenge and foster collaboration, it has for example been argued that personal abilities like "being able to listen, to communicate openly and fairly, and to understand and summarize complex issue[s]" (p. 247, Roloff, 2008) are important, as well as each partys' awareness in regard to their responsibility and impact on the collaborative success (Gray & Schruijer, 2010; Reypens et.al., 2021).

2.3.4 Conclusion

Section 2.3 has introduced scholarly and legal frameworks that are intended to guide the behavior of, among others, managers that are dealing with AI. The overview puts into context what is currently being done, relating to the research question in this thesis. The frameworks mentioned, some specific to the field of healthcare, all touch upon some important aspects regarding responsibility, such as the importance of ensuring meaningful human control, including all stakeholders, education and training, and the management of risk. Furthermore, the topic of managerial and organizational learning in a healthcare environment was discussed both generally and more specifically about stakeholder diversity. Given the many actors involved in the introduction of AI-DSS, and the importance of including all of them in the process, managing stakeholder diversity is particularly central to managers that engage with AI-DSS in healthcare.

2.4 Chapter Summary

The theoretical framework laid out in this chapter began with an introduction to AI, its use, and limitations in healthcare, as well as the more specific issue of explainability which is central to how responsibility should be understood and managed. The second part of the chapter defined the concept of responsibility and how the distribution of responsibility becomes a complex issue when working with AI. The final part of the chapter addressed the managerial and organizational aspects of AI in healthcare. In particular, the theoretical and regulatory frameworks that are established to guide managerial decision-making when engaging with AI, the importance of organizational learning and accommodating to change, and how to manage stakeholder diversity.

3. Methodology

This chapter outlines the chosen research methodology. It begins by describing the setting of the research design, followed by an illustration of data collection methods that consisted of semi-structured interviews, sampling techniques, and recording practices. It then proceeds to explain the method of analysis and then narrates what has been considered to meet the scientific rigor of the research. The chapter ends with a summary.

3.1 Research Design

This study attempts to understand the aspect of responsibility in the management of AI-DSS in healthcare. The study does not attempt at being scientific in a positivist sense, since the nature of ethical issues such as responsibility is hard to objectively measure and replicate. The approach is rather conceptual in nature since it to a large extent deals with theoretical assumptions and builds on philosophical definitions, but also pragmatic in the sense that the researchers want to inform (managerial) practice from theory (Bougie & Sekaran, 2016). Quality assurance and rigor of the research are discussed in part 3.4.

The researchers have followed a qualitative approach because it enabled bridging the managerial research gap that exists with reference to the concept of responsibility when using AI-DSS in healthcare vis a vis the research questions and contextualizing them with informed opinions from experts (Bell, Bryman & Harley, 2019). Moreover, in the setting of understanding the managerial implications of AI, qualitative research enables a broader perspective and motivates the need for weaving different social, cultural, and technological factors (Mona & Emmanuel, 2019).

The researchers combined a deductive (Saunders, Lewis & Thornhill, 2009) and inductive approach (Bryman & Bell, 2015), by first letting the theoretical framework inform the direction of the empirical study, and then letting the findings from the data to develop general conclusions. The researchers conducted their research by using these methods which helped them to explore different existing relevant theoretical concepts that cover topics such as AI, ethics, healthcare, and organizational learning and juxtaposing them with the gathered data to conclude their findings.

The time available for conducting the study played an important role since it was a practical and inherent limitation to the research. The total time available was eight weeks and the scope was adjusted accordingly by using a cross-sectional method, which means the study was undertaken to gather data at one point in time and is used in many studies because of the time, impact, and cost factors considered (Bougie & Sekaran, 2016).

3.2 Screening, Selection, & Choice of References

This section provides a brief overview of how the researchers approached the theoretical framework and it also gives a general idea of the literature flow and discusses the choice of references used for the purpose of this study.

3.2.1 Screening & Selection of Literature

To select the relevant theory, the researchers have made extensive use of available resources at their disposal ranging from visiting the university library and using its internal catalog like LUBCAT and also external sources like Google Scholar for gathering information that assisted with gaining deeper insights into the topic of study. Periodic meetings with the supervisor ensured that the researchers received expert quality support that was needed to forge a concrete and coherent understanding of the underlying theory that enabled a high-quality study (Anderson, 2006).

3.2.2 Choice of References

References plays a key role in persuading the strength of the theoretical framework (Nigel Gilbert, 1977). The researchers tried to incorporate well-cited, published, and peer-reviewed material from notable journals and made use of different frameworks and studies commissioned by governmental bodies like the EU, Swedish authorities, etc. The aforementioned sources are usually considered reliable; however, the researchers cannot validate their authenticity independently nor do they believe they are bias-free. The field of AI is rapidly evolving rendering older knowledge less reliable (Tegmark, 2017), considering this the researchers have tried to use the latest sources of information for this study.

3.3 Empirical Data Collection Method

3.3.1 Sampling

The subjects chosen for the interviews were a matter of purposive sampling, more specifically judgment sampling, which means that the researchers targeted specific types of people in terms of expertise to obtain the information necessary for the study. Although this method does not allow for a generalization of the findings, judgment sampling is a rich data source (Bougie & Sekaran, 2016) which the researchers found most suitable given the information needed and the scope of the study. The researchers chose to gather data from subject matter experts rather than for example the public or politicians to ensure that the data was qualitative and well-informed (Saunders, Lewis & Thornhill, 2009).

The demographics of the interviewees are important to consider when relevant to the study in terms of appropriateness (Connelly, 2013). The researchers had the following roles in mind when looking for respondents: AI developers, ethicists, physicians, and people with knowledge of legal/regulatory aspects. These groups of roles were chosen since they were all identified as key stakeholders involved in the process of the introduction of AI in healthcare. In total, the researchers interviewed ten people. The researchers intended to use quota sampling initially and get four respondents from each group but were unable to find that number of people in time. However, it was ensured that all the roles were represented and that there was a mix of both females and males. The interviewees were of different nationalities, though the majority were from Sweden. The sample is consequently not representative of any specific healthcare system, though slightly biased towards the Swedish public healthcare system. The distribution between the different roles can be seen in Table 1 below.

Interviewee	**Roles**	**Nationality**
1	Physician	Sweden
2	Physician	Sweden
3	Physician	Sweden
4	Physician/ AI Developer	U.S./India
5	AI Developer	Norway/India
6	Ethicist	France

7	Ethicist	Sweden/Poland
8	Ethicist	Sweden
9	Legal/Regulatory	Sweden
10	Legal/Regulatory	Sweden/Hungary

Table 1: Overview of the interviewees that participated in the research, their role, and nationality.

3.3.2 Semi-structured Interviews

For the collection of data, the researchers conducted semi-structured interviews. The researchers found interviewing individuals to be the most efficient way to gather data since this made it easier to schedule meetings along the way during a short period of time, as opposed to for example coordinating a focus group or expert panel which would have required all interviewees to attend at the same time.

The researchers had a prepared list of questions (see Appendix B) based on the themes in Table 2 below, but gave the interviewees the opportunity to speak freely, and the researchers could respond with follow-up questions that were not decided beforehand. According to Bryman (2012), semi-structured interviews allow the interviewees themselves to interpret and make sense of the topic at hand, something which the researchers found desirable given the explorative nature of the study and the fast-changing nature of technological development. The prepared list of questions allowed the researchers to ensure that the interviews did not go beyond the scope of the study and made it possible to compare the data.

Identified Theme	**Description**
Role of AI-DSS in Healthcare	This theme aims to describe the scope of use of AI in clinical healthcare.
Benefits of AI-DSS in Healthcare	This theme lists various advantages of introducing AI-DSS in clinical healthcare.
Challenges with AI-DSS in Healthcare	This theme highlights potential concerns and challenges that might be associated with using AI-DSS in clinical healthcare.
Relevant stakeholders	This theme is aimed at identifying stakeholders that can be responsible for the process of implementing and using AI-DSS in clinical healthcare.

Distributing responsibility	This theme relates to how responsibility should be distributed among the identified stakeholders.
Regulations and AI-DSS	This theme addresses the role of regulation when implementing AI-DSS in healthcare and how it can be applied to different stakeholders.
Lack of Moral Awareness, Capability, & Motivation	The theme addresses the issue of moral awareness, capability, and motivation in regards to the design, control, and use of AI-DSS.
Managerial learning	This theme relates to which inputs are helpful to a manager to ensure that AI is working in the best interest of the patient.

Table 2: Overview and description of the identified themes that served as a guide for formulating the interview questions.

Both researchers were present at every interview and had an assigned role to either ask the questions or make notes and make sure that all the prepared topics and questions were covered. The duration of the interviews ranged between 30 to 60 minutes. The interviews were either held in person or via video conferencing tools, despite one interview that was conducted over the phone. The researchers could not limit themselves to in-person interviews because of geographical limitations and the need to conduct the interviews in a short time.

The researchers have tried to minimize the risk of introducing bias by giving the respondents information about consent and how the information they provide will be used since this establishes trust. Furthermore, the researchers asked unbiased and open-ended questions and tried not to influence the responses given (Bougie & Sekaran, 2016).

3.3.3 Audio-Recording & Transcription

All the interviews were audio-recorded on a hand-held device which was important in order to let the researchers focus on what is being said and later on make a detailed, qualitative analysis (Bryman, 2012). The researchers were aware that recording might be perceived as more obtrusive than simply writing notes by hand, especially for the interviews held in person where the recording device was visible to the interviewee. However, the advantages of the recordings were considered prominent.

Prior to the interviews, terms, and agreements of confidentiality were discussed. The interviewees were also informed that they would be sent a summary of the interview before the thesis was to be finalized, to make sure that the information had been represented correctly. When transcribing the recordings, the researchers used an automated transcription service called Otter. The service provider states under their Privacy policy that the user (i.e. the researchers) has a right to request any personal information be erased. Although there is always a risk with data transfers, for which the researchers are fully responsible, all the data was deleted after finalized version of the thesis. The automated transcriptions do not come out perfect and therefore the researchers have also listened through the audio while writing the summaries of the interviews. Full transcripts can be given upon request.

3.4 Data Analysis

When analyzing the data, the researchers have followed the three steps approach of qualitative data analysis as presented by Bougie and Sakaran (2016), which is based on Miles' and Huberman's book *Qualitative data analysis: An expanded sourcebook* from 1994. The three steps consist of data reduction, data display, and drawing conclusions.

The first step, data reduction, consists of reducing and categorizing the gathered data through coding. The researchers identified themes from the theoretical framework, as presented in Table 2 above, to use as a guide when formulating the research questions and after the interviews to filter out what was not relevant to the themes. A theme is a larger coding unit that can contain any size of text as long as it represents a sole expression of an issue or idea. Once the data was gathered, these themes were adjusted inductively to any other patterns found in the data (see next step). As argued by Bougie and Sakaran (2016), this exemplifies that the three-step process is not linear but rather iterative.

The second step, data display, concerns presenting the data in an "organized, condensed manner" (p. 347, Bougie & Sakaran, 2016). For each of the themes, the researchers put together similar points from the interviews but made sure to highlight ideas that went against the general opinion or point out ideas that were mentioned by only a few of the interviewees. This was supplemented by quotes from the interviewees when found appropriate.

The last step consists of drawing conclusions from the collected and organized data and answering the research questions by "determining what identified themes stand for, by thinking about explanations for observed patterns and relationships, or by making contrasts and comparisons." (p. 347, Bougie & Sakaran, 2016). This step is represented in chapter five (discussion) and six (conclusion) of the thesis, where the researchers compare and conclude the findings from both the literature and the interview data in five main points.

3.5 Authenticity & Trustworthiness

There are different approaches to evaluating qualitative research. Some argue in favor of adopting similar concepts as in quantitative research (e.g. Mason, 2002; LeCompte & Goetz, 1982), while many others suggest that a qualitative study must be evaluated to a different set of criteria (e.g. Finlay, 2006; Shenton, 2004; Guba & Lincoln, 1994). Bryman (2012) suggests that qualitative research should be assessed for its *authenticity* and *trustworthiness*. The authenticity of the study refers to the "degree to which the research is transformative and emancipatory for the people studied and society at large" (p. 204, Bryman, 2012). Authenticity has been considered in this study through the inclusion of several different perspectives in the selection of interviewees and collection of data, as well as through a fair analysis of these perspectives.

Trustworthiness consists of four components (Bryman, 2012):

1) *credibility* is about making sure that the interpretation of the information given by the interviewees corresponds with the interviewees' reality.
2) *transferability* is about ensuring that enough information about the context in which the study was made is provided by the researchers to be transferable to other contexts.
3) *dependability* concerns making sure that proper research procedures have been followed throughout the research process.
4) *confirmability* is about reducing bias and not letting the researchers' personal values overly affect the research or the results.

The researchers fulfilled criteria (1) through respondent validation (see section 3.2.3), criteria (2) has been fulfilled through what has been presented in the introductory chapter and the methodology, and criteria (3) have been fulfilled by communication with the research supervisor throughout the whole process as well as through peer-review sessions, and criteria (4) has been fulfilled through the researchers' transparency in terms of research method and preferred theories, as well as through the scrutiny by peers.

3.6 Limitations

Researchers had to deal with constraints largely beyond their control but that may have had an impact on the study. It could be some inherent limitations of qualitative studies such as problems of validity and reliability because they do not occur in a natural setting (Wiersma,2000) or it could also be that the framing of semi-structured interviews by researchers might exhibit some bias (Sekaran & Bougie, 2016). Regarding both the researchers and the interviewees, a lack of sufficient knowledge (Strauss & Corbin,1998) in the field of AI, healthcare, and ethics and an understanding of its socio-cultural settings and inter-disciplinary implications (Papadopoulos & Lees, 2002) could hinder the formation of theoretical knowledge and philosophical underpinnings. Subjectivity and personal bias are unavoidable, but the researchers made the strength of their diverse academic and cultural backgrounds when integrating their perspectives during this study, doing this helped minimize conformity (N.K Denzin, 2008) and rendered additional credibility to the study.

3.7 Chapter Summary

This chapter covered the methodology used for the research in this thesis. The research has been done through a qualitative and both inductive and deductive approach within the given time frame. Because of limited time, a cross-sectional study was considered appropriate. The data were collected and managed by purposive sampling, semi-structured interviews, and audio-recording. The analysis of the data followed a three-step approach that included data reduction, data display, and drawing conclusions. Finally, quality assurance was discussed in terms of authenticity and trustworthiness, as well as any limitations with the chosen methodology.

4. Research Results

In this chapter, the data collected from the semi-structured interviews are summarized and displayed. The data is organized by the themes of the research questions: role of AI, benefits of AI, problems with AI, relevant stakeholders, distributing responsibility, regulations and AI, lack of moral awareness, capability and motivation, and managerial learning.

Role of AI-DSS in Healthcare

Every interviewee unanimously recognized AI in healthcare as currently having an assistive, supplementary role (as opposed to autonomous) that speeds up the clinical process and helps make informed decisions by using huge amounts of data. Interviewee #8 stresses that we must define where on the timeline we are to understand what is meant by AI:

> "One thing that I think it's important when you're talking about AI is that you always need to kind of specify where in the timeline you are, right? Because AI today, and AI tomorrow and next week can be very different things."

Interviewees with clinical experience identified radiology as the area which extensively uses AI, especially to identify malignant cells, but most interviewees consider it something that offers enormous potential (and challenges) in healthcare. Interviewee #8 speculates that AI might be something that patients are more eager to try, while physicians may see more problems.

Benefits of AI-DSS in Healthcare

Interviewee #8 said that the benefits of AI are dependent on what goals are intended to be achieved by using it and how we want its development to unfold. However, some of the benefits that most interviewees saw with AI-DSS, were that it works like a "good peer" (interviewee #2) if trained properly, reducing the risk of human failure by suggesting more accurate recommendations, uncovering findings where human cognition is limited, putting multiple findings into a single diagnosis, and increase success rates when planning surgeries. All in all, the interviewees think that AI-DSS increases efficiency and saves time and resources, for example by decreasing the need for input of multiple physicians. Interviewee #10 considers it as a frontier for improved healthcare.

Furthermore, interviewee #4 argued that AI-DSS adds a layer of transparency:

> "There's a lot of transparency if we use AI models, because right now, you know, [...] do we really know why a doctor is recommending a particular thing? You know, in like a 15 minutes appointment, they may probably give you a 10 seconds explanation [...] recommending this drug, but that's that. Like we don't really know how a decision is made and the clinicians' (?). Whereas an AI model is data points going in, this is how the model is working, this is the output. So I just feel it improves transparency."

A similar point was made by interviewee #8 arguing that the issue of transparency is an issue not only relating to AI since physicians need to deconstruct what led them to a certain decision when explaining why they did what they did. Interviewee #9 argued that AI-DSS makes healthcare more equal in some ways, and interviewee #8 said that it has the potential of benefiting developing countries with poor access to healthcare.

Challenges with AI-DSS in Healthcare

Regarding challenges with implementing AI-DSS in healthcare, the interviewees mention several, one of which is related to that distributing responsibility, accountability, and liability.

Some challenges related to the development of AI, such as

- Risk of bad and/or biased input data when training the AI.
- AI can't be better than those who built it.
- AI requires a lot of resources, and there is a lack of availability and accessibility of quality data to train the AI.
- It takes time to build customized, user-friendly AI.
- How to evaluate whether a system works well and is safe for patients, and no defined limits for risk tolerance and objectivity.
- Unmanaged development of AI could lead to irrelevant or unusable output, and new knowledge means new ethical problems that we might not be able to think about in advance.

Some challenges related to the use of AI and its effect on human interactions, such as

- Risk of over-reliance on the AI systems.
- The training of physicians or lack thereof.
- The risk of missing out on relevant aspects, and dilemmas arising out of weighing a physician’s judgment over the AI system’s.
- AI makes healthcare practice less personal and might have a detrimental impact on human interactions since AI is lacking the human aspect itself. AI adds to the complexity of the relationship between patient and physician.
- It can be hard to understand the output of a black-box model, in terms of transparency (this might also be an issue without AI though), interpretability, and explainability.
- Not everyone trusts the AI and may not feel comfortable with being yielded, treated, or addressed by an AI system. This raises the issue of informed consent and GDPR.

Other challenges that are identified by the interviewees are more general, such as

- There is a lack of an inter-disciplinary environment that has a common knowledge base, and it is difficult for experts to communicate with non-experts.
- There is too much focus on forecasting the development of AI instead of thinking about how we can develop it in the desired direction.
- Higher-level issues about how to govern and regulate AI.
- Increased use of AI might lead to high unemployment.
- Potentially only benefiting those who are already well off, while the risks are distributed to everyone or to a certain group. Relates to the issue of putting a monetary value on people’s health.

Relevant Stakeholders

The interviewees list many different stakeholders as relevant. Interviewee #10 says it is an issue that is currently being researched and that it is dependent on who the user is, their knowledge, what shortcomings there are, etc. Interviewee #3 says that the stakeholders are like those involved in drug approvals and that in Sweden, the Medical Products Agency can help define this issue. Interviewee #8 explains that there are two abilities that can help identify some stakeholders: 1) knowledge, and 2) influence. Both abilities then relate to how we distribute responsibility among the identified stakeholders (see next section). However, interviewee #8 stresses that there are other stakeholders, who might lack both expertise and influence yet need to be included, for example patients.

The stakeholders mentioned are engineers, designers, developers (of different kinds), AI providers, company leaders, regulatory bodies (including local municipalities and state-level health agencies), marketers, psychologists, physicians, surgeons, nurses, patients and patient representatives, hospital administration dealing with procurement of AI, lawyers, philosophers and ethicists, researchers (both within academia and companies), and politicians. Out of these, a few of the interviewees highlights the physicians as key stakeholders.

Distributing Responsibility

There seems to be a consensus that the issue of distributing responsibility is complex, and interviewee #6 says it is highly context-dependent to individual cases, something which interviewee # 4 also points to regarding distributing responsibility when something goes wrong:

> "I would assign it so if the AI model is wrong, not the patient decision, it is the AI team's responsibility. If the patient outcome is wrong, it is still the clinician's responsibility because the computer didn't hold the clinician's hand and made them prescribe that thing, the computer just recommended it. [...] I am responsible for what my model produces, not what the clinician prescribes."

Interviewee #3 brings attention to the complexity of risk in healthcare in general and that almost everything comes with risks. It can be hard to know what made things go wrong, and AI does not make this less complicated.

> "if 1 % of patients die from a specific surgery, or surgical procedure and the rest 99 % goes good, and they are healthier than they were before. The one person in a 100, who dies, is that statistics or has anyone done something wrong? [...] from the patient's perspective, it's complicated. Because if you have just a 1 % chance of dying, you shouldn't die. If you die, it has to be someone's fault. And if you are going to judge this afterward, you have to see, is these statistics? Or has someone really made the wrong judgments? And that will be complicated with AI because we don't know if it's a good judgment because we don't know how the judgment is made."

AI itself cannot be held accountable; the interviewees agree. Interviewee #8 says that long term if AI becomes conscious, we may be able to hold it accountable, yet we should ask ourselves if we want to do that (since that might take away the human perspective) or keep the human in the loop (which currently is important for the sake of responsibility). As reflected in the previous section, most of the interviewees recognize several relevant stakeholders, and most interviewees claim that responsibility must be shared among these. Interviewee #8 argues that this should be done as equally as possible, but that it is a challenge since our abilities, in terms of knowledge and influence, vary and thus also our responsibilities.

> “if you have lots of influence in this area, you have a responsibility. You could also say, and look at it the other way around, and say that it’s important to assign responsibility to this underrepresented group, because by doing that you will also give them influence.” (Interviewee #8)

Furthermore, several interviewees point out the importance of best practices (e.g. double-checking and standard check-lists with the patients’ interest in mind) and regulations to follow, in order for it to be possible to distribute responsibility. From a developer’s point of view, interviewee #4 says:

> “[E]very step of the way, all of the stakeholders need to review everything that we’re [developers] doing, But like, it shouldn’t be that, okay, everything’s done and you’re taking it for review. It should be more of an agile methodology. So it should be like every step, every increment has to be reviewed.”

Several interviewees brought up the physicians’ responsibility in particular and said that physicians are medically responsible for their patients (interviewee #2), should be involved in the feature selection process (interviewee #5), and have a high-level view of the algorithms used (interviewee #4). Interviewees #7 argue that although physicians have the final say, their burden should be taken from them to a large extent on a systemic level. Interviewee #10 argues that a large part of the responsibility is on the physician, the healthcare organization, and those representing it. Interviewee #10 also points out that stakeholders in the hospital must be understood in terms of complex organizational layers and hierarchies of management.

Ethical & Legal Frameworks

Several interviewees mentioned that any regulations need to be backed by adequate pre-assessments, clinical trials, ethical frameworks of which all stakeholders should be a part, and have detailed requirements on performance and safety. Interviewee #8 said that it is important to realize that law and ethics are two different things, even if laws have a basis in ethics by setting certain limitations and promoting or discouraging certain ways of thinking. It is not sufficient to just follow the law, ethical considerations must be taken into account too.

Some of the interviewees mention organizations that are important on this topic, and interviewee #10 says that regulations are geo-specific and differ depending on the socio-cultural setting. Interviewee #7 explains that AI is seen as an augmenting system rather than an autonomous one from a regulatory perspective. The EU is one of the organizations mentioned, and interviewee #7 thinks that the proposed EU AI Act touches upon many ethical issues. Interviewee #2 refers to the Swedish Public Procurement Act as important for Sweden in particular, though not much thought has been given to it yet.

Interviewee #6 is critical of the current regulations and argues, with several examples, that they are very problematic because they are skewed towards the Western perspective on fundamental, universal values (which is philosophically disputable) and indifferent towards other cultures and values. In the second quote below, interviewee #6 compares France to China but later says that the differences are equally varied within Europe.

> "[...] even if you set the limits of what is acceptable, and not acceptable, that might change depending on the situation. Ethics and artificial intelligence are really contextual."

> "Asserting the existence of fundamental values such as, let's say privacy, is problematic when you go to for example China, in Confucius's thinking, privacy does not make sense because it's a collectivist society where what belongs to you belongs to your family. [...] So privacy does not mean the same thing for me as a French guy, as it will mean for a Chinese person." (Interviewee #6)

Finally, interviewee #7 says we must watch out for over-regulation to not impede innovation, and interviewee #10 argues that going forward we need to future proof the process and create flexible regulations that can adapt to our future needs. Interviewee #10 also points out that there is a discussion on whether AI should be legally seen as a product or a service, but the regulations currently have a product-based approach. A service approach would give greater leeway to think about responsibility and accountability.

Lack of Moral Awareness, Capability, & Motivation

Regarding dealing with a lack of moral awareness, capability, and motivation, participant #7 stated that even if we have stricter laws and education in ethics, abuse will still occur, and interviewee #5 said it is a highly contextual issue. Nevertheless, several interviewees (including #7 and #5) that education is a key aspect of reducing this abuse, and that there's a need for increased awareness of ethical and moral obligations among non-clinical stakeholders. System developers must be trained to see ethical ramifications that result from the design of their systems. For example, interviewee #10 said they must be aware of the lack of diverse datasets since it has a huge impact on algorithmic decision-making.

Interviewee #6 argued that there is a big problem of lack of literacy in ethics and philosophy in general, and it is hard to teach ethics. Interviewee #8 says that there needs to be more work on applied ethics and real cases, as opposed to hypothetical problems. Education is really important, and the research community needs to develop guidelines with high standards, but motivating people to act in a certain way is a huge problem within academia. Researchers (philosophers) should publish outside traditional philosophy. Also, AI ethics courses directed at professionals that work with AI in their normal line of work, e.g. managers and medical staff, might be helpful. Interviewee #8 mentions the course "Artificial Intelligence: Ethics & Societal Challenges" that is upon Coursera.

> "What we are seeing with this epidemic today is that there is a growing interest in education, for ethics, real ethics, outside the box ethics, not the ethics that have been set by organizations such as the European Union. So this is great, but it's not enough. Lots of people are aware of the fact that ethics applies to artificial intelligence is important. But when it comes to for example pay for education, pay for training, pay for consultancy, they do not place AI ethics on the top of their priority." (Interviewee #6)

Interviewees #3 and #9 compared the issue to the balancing of commercial and moral interests in the pharmaceutical industry, two aspects that do not always go hand in hand. Interviewee #10 argued that it is important to align perspectives when designing such systems to avoid risking divergent interests in the whole process. Interviewees #5 and #3 underline the importance of including the patients' perspectives. Explainability is essential to this, and people involved in these systems usually must verify and streamline the workflow and prioritize resources in the best interests of the patient.

> "I guess everyone in this chain if you would ask them, they will say 'I work for the patients'. But many have also other interests to make a profit and make a name for themselves and get famous and make a career in the company. And there are a lot of incentives for people. And I guess for many people, the other incentives are greater than the incentive to help [the patients]." (Interviewee #3)

Lastly, several interviewees mentioned that laws and regulations play a huge role in this matter, and interviewee #7 said that transparency and access to information are important regarding legislation. Interviewee #2 says that there are extremely rigorous clinical studies done before it is introduced on the market and mentions the importance of keeping in line with for example Good Clinical Practice (GCP) and the Declaration of Helsinki.

Managerial Learning

In regard to managerial learning and how to manage AI in healthcare, several interviewees highlighted a multi-stakeholder approach. Interviewee #2 said that there is a need to create diverse teams that include for example economists, technicians, innovation and R&D specialists, ethicists, and someone representing the patients. Interviewee #4 also suggested cross-functional or multidisciplinary teams to get a holistic perspective.

> "I think what's important [here] is to, again, keep [a] pretty open mind about what you need to know here because it is very easy to fall back on the technology and say that we need a lot of knowledge, a lot of input about how this technology works. And that's obviously true, but it doesn't cover everything. And you need, again, you need input from different stakeholders and groups, because we know that information is always biased to some extent, and that means if you only get your information from a certain group, then you can be pretty sure that to a certain extent it is biased in favor of that group." (Interviewee #10)

Two interviewees (#5 and #10) underlined the importance of making the physician a part of the process from the beginning. Interviewee #7 pointed out that physicians need to be trained on a systemic level in a public healthcare system. Interviewee #5 highlighted the need for everyone to have a basic AI literacy.

Another aspect that was brought up by several interviewees was the importance of managers keeping updated on procurement laws, regulatory recommendations, ethical guidelines, and international best practices. However, participant #6 argued that managers need to go beyond compliance and look at ethical responsibilities as well, they need to be aware of their own responsibilities and make sure that others know what their responsibilities are. Ethics must be introduced early on, continuously, in people's careers and education, by providing tools for them to think for themselves rather than tell them what is right or wrong, and how they can apply it to their field of interest. Interviewee #7 pointed out that in addition to following regulations, evidence for performance and safety must be carefully considered, and managers must not blindly follow political pressures.

> "You have to avoid just doing what you've been told to do because that's just conformism. You have to take your own responsibility, you have to think by yourself: what are the consequences, am I comfortable with what I'm doing?" (Interviewee #6)

Furthermore, interviewee #10 pointed out the importance of enriching the legal perspective by receiving feedback and learning from those that work with the systems and their potential and shortcomings. A similar point was made by interviewee #4:

> "I think I need that healthcare manager to always know that the AI model is a learning child. And anytime the child has something that it is not supposed to it needs to be highlighted ASAP."

Relating to the performance and safety of the systems, interviewee #1 said that implementation should start gradually in low-risk areas of clinical healthcare. Several interviewees underlined the importance of making a cost-benefit analysis of such implementation of the systems, and interviewee #8 pointed out that there is an ethical difference between taking a risk and being exposed to a risk, that should be considered. Interviewee #10 pointed out that as a manager, or chief physician, you have a responsibility to keep your teams informed on how to use the system and what to watch out for.

5. Discussion

5.1 Answering the Research Question

The discussion is divided into five parts that serve the purpose of answering the research question below.

> "What factors can facilitate managerial learning when introducing AI-DSS in public clinical healthcare in terms of responsibility?"

The parts are the researchers' way of analyzing the data presented in chapter four by integrating the literature from chapter two and drawing conclusions from this analysis. The topic for each part was decided by filtering out what was most relevant in terms of answering the research question, thus each topic can be considered a factor that facilitates managerial learning in the process of introducing AI-DSS in healthcare in terms of responsibility.

1. **Need for diverse and cross-functional teams**

A large majority of the participants have stressed the need for incorporating diverse and cross-functional perspectives throughout the whole process of introducing AI-DSS. This multi-stakeholder approach was also a central aspect of the literature on AI, responsibility, and legal and ethical frameworks related to the issue (Santoni de Sio & van den Hoven, 2018; EU Commission, 2021; Bleher and Braun, 2022; Sullivan & Fosso, 2022). Furthermore, the presented literature expresses the importance of diversity to establish a learning organization (Mårtensson, 2000; Riege & Lindsay, 2006). The data presented suggest that the inclusion of all stakeholders in the process of implementing AI-DSS in healthcare is necessary for at least three reasons. First, it is necessary in order to deal with biased interests and the potential conflict between for example commercial and moral interests. Second, it is necessary to distribute responsibility as equally as possible while also being sensitive to for example varying abilities in terms of knowledge and influence, as pointed out by interviewee #8 in particular. Third, it helps to build trust in AI, especially when considering the inclusion of the patient as a stakeholder in the process.

Despite the expressed need for including all stakeholders, some of the interviewees expressed that there is currently a lack of interdisciplinary environments. This might in part relate to the challenges associated with collaborations between diverse stakeholders (Curseu & Schruijer, 2017). One of these challenges was brought up by interviewee #8, who expressed that experts and non-experts (or experts from different fields) might find it difficult to communicate and collaborate. To address this challenge, managers might turn to the extensive literature on how to manage diversity and leverage it (e.g. Roloff, 2008; Gray & Schruijer, 2010; Reypens et.al., 2021). Another aspect of addressing the issue that can be found in the data is the importance of education and increasing ethical literacy. This is elaborated in point four below.

Another issue relates to how the relevant stakeholders should be identified. Clarifying who the relevant stakeholders are is important when we want to distribute responsibility among them since we need to establish a link between the agent and whoever is affected by the agent's actions (Noorman, 2020). Identifying relevant stakeholders is still a matter of research when it comes to AI in healthcare, as described by interviewee #10. The data also suggests that identifying relevant stakeholders is complicated because of context-dependency and the layers or hierarchy of healthcare management. As pointed out by interviewee #3, the needed multi-stakeholder approach is comparable to the process of drug approval, which also involves many different stakeholders. Relatedly, several interviewees suggested that the adoption of AI can be compared to the adoption of other medical devices. This may suggest that in the ongoing development of frameworks intended to guide the management of AI-DSS in healthcare, much can be learned from already established procedures and identified stakeholders in other areas of the industry.

2. **Critically assess and stay up to date with legal and ethical frameworks and best practices**

The literature and the data presented in this thesis suggest that legal and ethical frameworks play an important role when guiding the management of AI-DSS in healthcare. Several interviewees made the point that managers need to stay updated on these frameworks to identify responsibilities, be able to distribute them, and hold people accountable. However, the data also suggests that there are at least three challenges to managers that wish to comply with ethical and legal frameworks. The first two are related and concern legal frameworks in particular. The third challenge concerns the relation between ethical and legal frameworks.

First, legal regulations are currently not settled and are a matter of ongoing development, both on a national and international level, as suggested by the theoretical framework presented in this thesis. The proposed European AI Act (EU Commission, 2021) is an attempt at regulating AI large-scale, and is indeed touching upon many ethical issues as argued by interviewee #7 but is yet to be settled.

With the rapid development and use of AI technology (WIPO, 2019), it is not unlikely that it will take time before the legal frameworks catch up and cover any ethical dilemmas that might get actualized. As pointed out by interviewee #8, the frameworks need to be updated and adapt as AI is redefined with new possibilities and challenges.

Second, legal frameworks need to be flexible enough to not impede innovation and adapt to future needs, as argued by interviewees #7 and #10. This is a challenge when formulating the frameworks, but arguably also a challenge for practicing managers since the flexibility might lead to vagueness and uncertainty when applying it, something which the proposed AI Act by the European Commission has been criticized for (Ministry of Infrastructure, 2021).

Third, as argued by interviewee #6, it can be a challenge to apply general principles as they are formulated in legal and ethical frameworks to specific cases. Several of the interviewees expressed how context-dependent the issue of responsibility in healthcare is, both with and without AI-DSS involved. If the frameworks are too loose and flexible it might end up in ambiguity and uncertainty, as expressed in the previous paragraph, but if the frameworks are too rigid they don’t give room for context. Furthermore, interviewee #6 argued that the general principles put forward by regulatory bodies might not align with the principles and values accepted in particular cultures. Interviewee #10 mentions that regulations are geo-specific and differ depending on the socio-cultural setting, but if the proposed AI Act from the EU will be enforced in all of Europe and have global influence, then the concern expressed by interviewee #6 is a serious one.

Although much of the underlying issues to the challenges described above cannot be resolved by individual managers, it highlights the need for managers to think about responsibilities as something beyond compliance, as suggested by the data. It also points to the importance of all stakeholders’ responsibility (including managers) to provide continuous feedback on both the potentials and shortcomings of the implemented AI-DSS to regulators to inform and improve the developing legislation and ethical frameworks on AI, as was pointed out by interviewee #10.

3. Need for a comprehensive cost-benefit analysis

One of the primary motivations behind using technology is to provide efficiency and cost savings for its users and related stakeholders (Bertot, Estevez, & Janowski, 2016) and this holds true for AI-DSS too. In the context of public healthcare interviewee #3 believes, that it is healthcare managers’ responsibility that each and every resource allocated is vital to the healthcare unit and judicious use of it needs to be made, it is also important for healthcare managers to deliberate on potential benefits that AI may bring to the table before actually implementing AI-DSS across the system (Ludwin & Murry, 2017). In addition to this interviewee #3 points out that, it is also of utmost importance to understand that the initial cost and efforts to use AI-DSS are high because they include training staff to be AI-literate enough to make use of it, and it might offer limited to no benefits at the inception of its use. Given this, it is also important for a manager to minimize the collateral impacts of prioritizing AI in a clinical setting which could come at the cost of hiring new staff or investing in other resources that a healthcare unit might require.

When discussing potential benefits interviewee #6 highlighted that it is important to define objectives or goals often long term because there is no linear relationship between cohorts who are paying the price for the development of AI and cohorts who are reaping the benefits it offers, it is a moral imperative for a manager to ensure that the development of AI is an equitable process that is accessible fo everyone and remains free of marginalizing anyone or any group.

Another dimension that is central to evaluating AI-DSS in healthcare from a cost-benefit perspective is its explainability or lack thereof according to interviewee #4, there is a consensus among experts in AI about the trade-off between the performance of an AI algorithm and how better it can be understood. The lack of tangible metrics in the field of AI that can be communicated to the receiver of AI-assisted decisions compounds this problem, it impedes a manager’s ability to make objective decisions (Yoon & Lee, 2019). It can be concluded that healthcare managers should strive to balance this sensitive inflection point and make reasonably informed decisions at the behest of the patient.

4. Educating and motivating everyone about their responsibilities

To create awareness, capability, and motivation to act on one’s responsibilities, the data presented suggest that education, as well as regulations, play an essential part. In terms of education, this includes both ethical literacy and, as pointed out by interviewee #5, basic AI literacy for all stakeholders. Within healthcare, the data suggests that special concerns must be taken to include patients’ perspectives in terms of both safety and trust. Interviewees #3 and #8 speculated that patients might be more eager to try new technology, and desperate for solutions, while physicians might be more hesitant due to seeing more problems. This is an argument for creating awareness of AI and the challenges it brings with it not only among experts but also to the public, which is considered a stakeholder in the proposed EU AI Act (p. 7, EU Commission, 2021).

The need for education is partly a societal and institutional matter, as several interviewees stressed that there is currently a lack of literacy in ethics, especially among non-clinical stakeholders. To solve this issue, interviewees #6 and #8 argued that ethics needs to be incorporated throughout the whole educational process. The data suggest that given that ethics is a difficult subject to teach, it is important that the curriculum focus on real cases that are applicable to the field of study, for students to see the relevance ethics will have in their work.

Even if ethical literacy is increased, the issue of balancing commercial and moral interests remains, much the same way as in the pharmaceutical industry, a comparison made by interviewee #9. Looking at the issue from a managerial point of view, the data suggests that making AI ethics a priority might be done most effectively through regulations, for example by requiring CE marking, which means the machine complies with certain standards (EU Commission, 2021). Individual managers can make sure that there are people in place within the organization that is appointed to consider ethical issues, much like there are people responsible for example economy and human resources. Furthermore, if managers make AI ethics a priority, there are opportunities to train and educate their personnel via consultancies and universities, as interviewee #6 pointed out. The literature and the data presented in this thesis hint that managers have a responsibility to keep their teams educated on how to use the AI systems and what to watch out for.

The importance of education in regards to managing the introduction of AI-DSS reconciles with how central the concepts of knowledge, explainability, and transparency are to the literature on responsibility. One cannot be considered responsible for something one knew nothing about or something one had no control over (Noorman, 2020). An interesting remark is that the literature generally considers explainability and transparency a problem when discussing AI (Pasquale, 2015), and this was a concern for some of the interviewees as well. However, some of the interviewees argued that the issue of explainability and transparency might be equally present without AI and that it might in fact increase transparency when AI is introduced.

5. Making sure the human is in the loop

The data suggests that it is important to keep the human in the loop in order to distribute responsibility and to see the AI as a complement rather than a responsible agent. This is also reflected in the literature as discussed in section 2.3.1 of the theoretical framework. Currently, meaningful human control (Santoni de Sio & van den Hoven, 2018) is necessary since AI is not fully autonomous or conscious and therefore cannot be held accountable for any wrongdoings. But even if AI becomes fully autonomous and conscious in the future, the data suggests that we still might want to keep the human perspective since that is central to a field like healthcare where trust and communication are valued highly.

5.2 Chapter Summary

Figure 1: The model visualizes the five aspects central to managerial learning and managing responsibility when introducing AI-DSS in healthcare.

In this chapter, the researchers have analyzed and discussed the gathered data in light of the theoretical framework provided in chapter two, with the aim of responding to the research question. The researchers identified five factors that facilitate managerial learning regarding responsibility when introducing AI-DSS in healthcare (see Figure 1): 1) diverse and cross-functional teams, 2) critically assessing and staying up to date with legal and ethical frameworks and best practices, 3) comprehensive cost-benefit analysis, 4) educating and motivating everyone about their responsibilities, and 5) making sure the human is in the loop. All these factors allow managers to better deal with the complicated issue of their own and others' responsibilities when introducing AI-DSS in healthcare.

6. Conclusion

6.1 Summary of Main Insights

The purpose of this thesis was to explore and understand various factors that can facilitate managerial learning when introducing AI-DSS in public clinical healthcare in terms of responsibility. The research began by framing the research question in relation to the purpose and was followed by a two-step process: the first step comprised of a thorough investigation into the existing literature around relevant aspects in topics such as AI, responsibility, healthcare management, and organizational learning, and concluded them with reference to the existing research gap that relates to management. The second step was to interview experts in AI, ethics, and healthcare and obtain their expert opinions on different aspects of their understanding of each of the relevant topics along with potential managerial learnings that can alleviate the research gap.

The concluded findings from the theoretical framework and data from the interviews culminated in five factors that can facilitate managerial learning regarding responsibility when introducing AI-DSS in healthcare. The first factor concerned the importance of creating diverse and cross-functional teams since both the literature and the data strongly highlighted the need to include all stakeholders in the whole process of development and implementation of AI-DSS, in order to deal with varying interests and to be able to distribute responsibility.

The second factor concerned the importance of critically assessing and staying up to date with ethical and legal frameworks since this helps managers to attribute and distribute responsibilities properly, in a field that is constantly changing. Because of a few societal challenges related to this point, the researchers also concluded that although individual managers cannot resolve the issues themselves, they can be part of informing the development of ethical and legal frameworks.

The third factor concerned the need for a comprehensive cost-benefit analysis of implementing AI-DSS as a managerial responsibility, which includes defining what and who bears the costs or receives the benefits. This factor is important because it facilitates managerial learning on the issue, and lets the manager make informed decisions with the relevant stakeholders in mind, which in the case of healthcare is foremost the patients.

The fourth factor pointed to the central role that education and regulations play in creating awareness, capability, and motivation to act on one’s responsibilities. Although education (and regulations) is partly a societal matter, the researchers concluded that individual managers have a responsibility to keep their teams literate in terms of both ethics and AI, since having knowledge is central to being held responsible as suggested by the literature.

The fifth and final factor of relevance to managers highlighted, in line with both the data and the literature, how important it is to keep humans in control of the AI since the AI at least currently cannot be held responsible on its own. Furthermore, it is particularly important in a field like healthcare, where human perspectives and communication is central to the practice.

6.2 Future Research

In addition to this study, the researchers want to shine a light on avenues of study that could expand on the theoretical framework and potentially be addressed for the purposes of future research. First, how to best teach ethics and thereby improve ethical literacy for non-medical stakeholders involved in designing AI-DSS could be studied. Second, how to address the challenges faced when adapting legal regulations that help govern AI-DSS to different socio-cultural settings. Lastly, there is a need to delve deeper into the design aspects of AI-DSS in a way that prioritizes user and societal interests, specifically how AI may benefit developing countries and help address issues of inequality and injustice. The above suggestions would require a multi-faceted study of interdisciplinary topics that cover for example AI, ethics, law, and management.

References

Ahrens, J., Hendrickson, B., Long, G., Miller, S., Ross, R., & Williams, D. (2011). Data-intensive science in the US DOE: case studies and future challenges. *Computing in Science & Engineering*, *13*(6), 14-24.

Amann, J., Blasimme, A., Vayena, E., Frey, D., & Madai, V. I. (2020). Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*(1), 1-9.

Anderson, C., Day, K., & McLaughlin, P. (2006). Mastering the dissertation: Lecturers' representations of the purposes and processes of master's level dissertation supervision. Studies in Higher Education, 31(2), 149–168.

Baker-Brunnbauer, J. (2021). Management perspective of ethics in artificial intelligence. *AI and Ethics*, *1*(2), 173-181.

Bertot, J., Estevez, E., & Janowski, T. (2016). Universal and contextualized public services: Digital public service innovation framework. *Government information quarterly*, *33*(2), 211-222.

Bell, E., Bryman, A., & Harley, B. (2019). Business research strategies. *Business Research Methods*, 17-37.

Bleher, H., & Braun, M. (2022). Diffused responsibility: attributions of responsibility in the use of AI-driven clinical decision support systems. AI and Ethics, 1-15.

Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications.*Artificial Intelligence in Healthcare*,25–60.

Bolman, L. G., & Deal, T. E. (2017). *Reframing organizations: Artistry, choice, and leadership*. John Wiley & Sons.

Bowens, F. M., Frye, P. A., & Jones, W. A. (2010). Health information technology: integration of clinical workflow into meaningful use of electronic health records. *Perspectives in Health Information Management/AHIMA, American Health Information Management Association*, *7*(Fall).

Braun, M., Hummel, P., Beck, S., & Dabrock, P. (2021). Primer on an ethics of AI-based decision support systems in the clinic. *Journal of medical ethics*, *47*(12), e3-e3.

Bryman, A. (2012). *Social Research Methods* 5th Edition Oxford University Press.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530-1534.

Canada Directive. (2021). *Directive on Automated Decision-Making*. The Government of Canada. URL: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592

Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376*(2133), 20180080.

Castelvecchi, D. (2016). Can we open the black box of AI?. *Nature News*, *538*(7623), 20.

Coeckelbergh, M. (2020). Artificial intelligence, responsibility attribution, and a relational justification of explainability. *Science and engineering ethics, 26*(4), 2051-2068.

Coeckelbergh, M. (2010). Moral appearances: emotions, robots, and human morality. *Ethics and Information Technology*, *12*(3), 235-241.

Coiera E (1997). *Guide to medical informatics, the Internet, and telemedicine*. Chapman & Hall, Ltd.

Coiera, E. W. (1996). Artificial intelligence in medicine: the challenges ahead. *Journal of the American Medical Informatics Association, 3*(6), 363-366.

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Journal of British Surgery*, *102*(3), 148-158.

Connelly, L. M. (2013). Limitation section. *Medsurg Nursing*, *22*(5), 325.

Crigger, E., & Khoury, C. (2019). Making policy on augmented intelligence in health care. *AMA journal of ethics, 21*(2), 188-191.

Curșeu, P. L., & Schruijer, S. G. (2017). Stakeholder diversity and the comprehensiveness of sustainability decisions: the role of collaboration and conflict. *Current Opinion in Environmental Sustainability*, *28*, 114-120.

Denzin, N. K. (2008). *The landscape of qualitative research* (Vol. 1). Sage.

Deursen, A. V., Dijk, J. V., & Ebbers, W. (2006). Why e-government usage lags behind: Explaining the gap between potential and actual usage of electronic public services in the Netherlands. In *International Conference on Electronic Government* (pp. 269-280). Springer, Berlin, Heidelberg.

Duffy, J. R., & Lemieux, K. G. (1995). A cardiac service line approach to patient-centered care. *Nursing Administration Quarterly*, *20*(1), 12-23.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative science quarterly, 44*(2), 350-383.

EU Commission, T. (2021) *Proposal for a regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.* COM (2021), 206.

Eshleman, A. (2016). Moral responsibility. *The Stanford Encyclopedia of Philosophy.* Edward N. Zalta (ed.).

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion*, *58*, 82-115.

Finlay, L. (2006). 'Rigour','ethical integrity or 'artistry'? Reflexively reviewing criteria for evaluating qualitative research. *British Journal of Occupational Therapy*, *69*(7), 319-326.

Floridi, L. (2016). Faultless responsibility: On the nature and allocation of moral responsibility for distributed moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2083), 20160112.

Floridi, L. (2013). Distributed morality in an information society. *Science and engineering ethics*, *19*(3), 727-743.

GAO. (2021). *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities. Including Insights from the Comptroller General's Forum on the Oversight of Artificial Intelligence.* U.S Government Accountability Office

GAO Fast Facts. (2021). *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities.* U.S Government Accountability Office. URL: https://www.gao.gov/products/gao-21-519sp

Goh, S. C., Chan, C., & Kuziemsky, C. (2013). Teamwork, organizational learning, patient safety and job outcomes. *International journal of health care quality assurance*.

Gijsberts, C. M., Groenewegen, K. A., Hoefer, I. E., Eijkemans, M. J., Asselbergs, F. W., Anderson, T. J., ... & Den Ruijter, H. M. (2015). Race/ethnic differences in the associations of the Framingham risk factors with carotid IMT and cardiovascular events. *PLoS One*, *10*(7), e0132321.

Gray, B., & Schruijer, S. (2010). Integrating multiple voices: Working with collusion in multiparty collaborations. In *Relational practices, participative organizing*. Emerald Group Publishing Limited.

Gualdi, F., & Cordella, A. (2021). Artificial intelligence and decision-making: The question of accountability. In *Proceedings of the 54th Hawaii International Conference on System Sciences* (p. 2297).

Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. *Handbook of qualitative research*, *2*(163-194), 105.

Halligan, A., & Donaldson, L. (2001). Implementing clinical governance: turning vision into reality. *Bmj*, *322*(7299), 1413-1417.

Higgins, D., & Madai, V. I. (2020). From bit to bedside: a practical framework for artificial intelligence product development in healthcare. *Advanced intelligent systems*, *2*(10), 2000052.

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *science*, *332*(6025), 60-65.

Info-comm Media Development Authority and Personal Data Protection Commission. (2020) *Model Artificial Intelligence Governance Framework*, *Second Edition*. Singapore.

Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, *33*(3), 371-377.

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise: A flaw in human judgment*. Little, Brown.

Kiseleva, A. (2020). AI as a Medical Device: Is It Enough to Ensure Performance Transparency and Accountability?. *EPLR*, *4*, 5.

Kiseleva, A. (2021). COMMENTS ON THE EU PROPOSAL FOR THE ARTIFICIAL INTELLIGENCE ACT. *Available at http://dx.doi.org/10.2139/ssrn.3949585*.

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications policy*, *44*(6), 101976

Köhl, M. A., Baum, K., Langer, M., Oster, D., Speith, T., & Bohlender, D. (2019). Explainability as a non-functional requirement. In *2019 IEEE 27th International Requirements Engineering Conference (RE)* (pp. 363-368). IEEE.

LeCompte, M. D., & Goetz, J. P. (1982). Problems of reliability and validity in ethnographic research. *Review of educational research*, *52*(1), 31-60.

Lingman, M.., Engström, M., Clarin, M., Lövenvald, O., Berg, S., Sigridsson, C. … Andersson, K. (2021). *En handbok för information-driven vård*. [A handbook for information-driven healthcare - Insights from the inside.] AI Sweden

Ludwin, S. K., & Murray, T. J. (2017). Dilemmas in medical ethics in the age of big data. *Multiple Sclerosis Journal*, *23*(10), 1306-1308.

Lyman, B., Hammond, E., & Cox, J. (2018). Organizational learning in health systems: A concept analysis. *Journal of Nursing Management*.

Lysaght, T., Lim, H. Y., Xafis, V., & Ngiam, K. Y. (2019). AI-assisted decision-making in healthcare. *Asian Bioethics Review*, *11*(3), 299-314.

Mason, J. (2002). *Qualitative researching* (2nd ed.). London: Sage.

Matthias, A., 2004. "The responsibility gap: Ascribing responsibility for the actions of learning automata," *Ethics and Information Technology*, 6: 175–183.

Mazmanian, P. E., Davis, D. A., & Galbraith, R. (2009). Continuing medical education effect on clinical outcomes: effectiveness of continuing medical education: American College of Chest Physicians Evidence-Based Educational Guidelines. *Chest*, *135*(3), 49S-55S.

Mehr, H., Ash, H., & Fellow, D. (2017). Artificial intelligence for citizen services and government. *Ash Cent. Democr. Gov. Innov. Harvard Kennedy Sch., no. August*, 1-12.

Miliard M. (2018) AI funding tops $1 billion mark as mobile investments wane. *Healthcare ITNews*. URL: http://www.healthcareitnews.com/news/health-it-funding-surpasses-7-billion-first-time-ai-and-analytics-biggest-winners

Miller, S. M. (2018). AI: Augmentation, more so than automation. *Asian Management Insights. 5*,(1), 1-20. Asian Management Insights.

Ministry of Enterprise and Innovation. (2018). *National approach to artificial intelligence*. Swedish Government

Ministry of Infrastructure. (2021). *Remiss av Europeiska kommissionens förslag till förordning om harmoniserade regler för artificiell intelligens.* Swedish Government. Available at https://www.regeringen.se/remisser/2021/05/remiss-av-europeiska-kommissionens-forslag-till-forordning-om-harmoniserade-regler-for-artificiell-intelligens/

Montavon, G., Samek, W., & Müller, K. R. (2018). Methods for interpreting and understanding deep neural networks. *Digital Signal Processing*, *73*, 1-15.

Morley, J., Machado, C. C., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: a mapping review. *Social Science & Medicine*, *260*, 113172.

Müller, V.C. (2021) Ethics of Artificial Intelligence and Robotics. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.).

Mårtensson, M. (2000). A critical review of knowledge management as a management tool. *Journal of knowledge management*.

Nigel Gilbert, G. (1977). Referencing as persuasion. *Social studies of science*, *7*(1), 113-122.

Nolan, J., McNair, P., & Brender, J. (1991). Factors influencing the transferability of medical decision support systems. *International journal of bio-medical computing*, *27*(1), 7-26.

Noorman, M. (2020). Computing and Moral Responsibility. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.)

Nyholm, S. (2018). Attributing agency to automated systems: Reflections on human-robot collaborations and responsibility-loci. *Science and engineering ethics*, *24*(4), 1201-1219.

Pasquale, F. (2015). *The black box society*. Harvard University Press

Papadopoulos, I., & Lees, S. (2002). Developing culturally competent researchers. *Journal of advanced nursing*, *37*(3), 258-264.

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 33-44).

Rakich, J. S., Longest Jr, B. B., Darr, K., & Guerrieri, L. R. (1993). Managing health services organizations. *Journal for Nurses in Professional Development*, *9*(3), 162.

Reddy, S., Allan, S., Coghlan, S., & Cooper, P. (2020). A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association*, *27*(3), 491-497.

Reypens, C., Lievens, A., & Blazevic, V. (2021). Hybrid Orchestration in Multi-stakeholder Innovation Networks: Practices of mobilizing multiple, diverse stakeholders across organizational boundaries. *Organization Studies*, *42*(1), 61-83.

Reed, C. (2018). How should we regulate artificial intelligence?. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2128), 20170360.

Richter, J. P. (2013). *Organizational Factors of Safety Culture Associated with Perceived Success in Patient Handoffs, Error Reporting, and Central Line-Associated Bloodstream Infections*. The Ohio State University.

Riege, A., & Lindsay, N. (2006). Knowledge management in the public sector: stakeholder partnerships in the public policy development. *Journal of knowledge management*.

Rigby, M. (2019) Ethical Dimensions of Using Artificial Intelligence in Healthcare. *AMA J. Ethics 21*, E121–E124.

Roff, H. M., & Moyes, R. (2016). Meaningful human control, artificial intelligence and autonomous weapons. In *Briefing Paper Prepared for the Informal Meeting of Experts on Lethal Au-Tonomous Weapons Systems, UN Convention on Certain Conventional Weapons*.

Roloff, J. (2008). Learning from multi-stakeholder networks: Issue-focussed stakeholder management. *Journal of business ethics, 82*(1), 233-250.

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. *Ieee Access*, *8*, 42200-42216.

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

Russell, Stuart J.; Norvig, Peter (2009). *Artificial Intelligence: A Modern Approach* (3rd ed.). Upper Saddle River, New Jersey: Prentice Hall. ISBN 978-0-13-604259-4

Sadeghifar, J., Ashrafnejad, F., Mousavi, S. M., Nasiri, A. B., Vasokolaei, G. R., Zadeh, K. N., ... & Mosavi, M. (2014). The relationship between organizational learning and staff empowerment in hospital: A correlational study in Iran. *Asian Social Science*, *10*(16), 27.

Salathé, M., Wiegand, T., & Wenzel, M. (2018). Focus group on artificial intelligence for health. *arXiv preprint arXiv:1809.04797*.

Sangki, J. (2018). Vision of future e-government via new e-government maturity model: Based on Korea's e-government practices. *Telecommunications Policy*, *42*(10), 860-871

Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, 1-28.

Santoni de Sio, F., & Van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*, 15.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research methods for business students*. Pearson education.

Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach*. John Wiley & Sons.

Shenton, A. K. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for information*, 22(2), 63-75.

Sioli, L. (2021). *A European Strategy for Artificial Intelligence*. CEPS webinar - European approach to the regulation of artificial intelligence.

Sloane, M., & Moss, E. (2019). AI's social sciences deficit. *Nature Machine Intelligence*, *1*(8), 330-331.

Smiley, M. Collective Responsibility. (2017) *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.)

Stahl, B. C. (2021). *Artificial intelligence for a better future: an ecosystem perspective on the ethics of AI and emerging digital technologies* (p. 124). Springer Nature.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques.*

Sullivan, Y. W., & Fosso Wamba, S. (2022). Moral judgments in the age of artificial intelligence. *Journal of Business Ethics*, 1-27.

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science*, *361*(6404), 751-752.

Talbert, M. Moral Responsibility. (2019) *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.).

Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence*. Vintage.

Thompson, D. F. (1980). Moral responsibility of public officials: The problem of many hands. *American Political Science Review*, *74*(4), 905-916.

Tigard, D. W. (2021). There is no techno-responsibility gap. *Philosophy & Technology*, 34(3), 589-607.

Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, *25*(1), 44-56.

Tsang, L., Kracov, D. A., Mulryne, J., Strom, L., Perkins, N., Dickinson, R., ... & Jones, B. (2017). The impact of artificial intelligence on medical innovation in the European Union and United States. *Intellect Prop Technol Law J*, *29*(8), 3-12.

UK Guidance. (2021). *Ethics, Transparency and Accountability Framework for Automated Decision-Making*. The Central Digital and Data Office and the Office for Artificial Intelligence. UK Government.

von Eschenbach, W. J. (2021). Transparency and the black box problem: Why we do not trust AI. *Philosophy & Technology*, *34*(4), 1607-1622.

World Health Organization. (2014). *Health systems governance for universal health coverage action plan: department of health systems governance and financing* (No. WHO/HSS/HSF/2014.01). World Health Organization.

WHO. (2021). *Ethics and governance of artificial intelligence for health*: *WHO guidance*. World Health Organization.

Wienert, J. (2019). Understanding health information technologies as complex interventions with the need for thorough implementation and monitoring to sustain patient safety. *Frontiers in ICT*, *6*, 9.

Wiersma, W. (2000). *Research methods in education: an introduction*.

WIPO (2019). *WIPO Technology Trends 2019: Artificial Intelligence*. Geneva: World Intellectual Property Organization.

Yoon, S.; Lee, D. (2019) Artificial Intelligence and Robots in Healthcare: What are the Success Factors for Technology-based Service Encounters? *Int. J. Healthc. Manag. 12*, 218–225.

Appendix A

Disclaimer

Terms and conditions for the interview consist of the following points:

1. The interviewee’s right to be anonymous.
2. The interviewee's consent to be recorded.
3. Informing the interviewee that he/she will get the opportunity to read and accept a summary of the interview.

All interviewees were asked and informed of this before the start of the interview.

Appendix B

Interview Guide

No. #	Leading Question	Research Theme
1	Please introduce yourself and what is your role.	Introduction
2	How do you understand the role of AI in healthcare? a) How do you understand the role of AI in clinical healthcare in particular?	Role of AI-DSS in Healthcare
3	From a technical perspective, can you tell us anything about configuring and using AI in healthcare?	Role of AI-DSS in Healthcare
4	How can explainability be applied to AI in healthcare?	Role of AI-DSS in Healthcare
5	What potential problems and benefits do you see of implementing AI in healthcare?	Benefits of AI-DSS in Healthcare Challenges with AI-DSS in Healthcare
6	How do you strike a balance between the risks and benefits of implementing AI in healthcare?	Managerial Learning
7	Who do you think can be relevant stakeholders responsible for the implementation and use of AI in healthcare?	Relevant Stakeholders Distributing Responsibility
8	What could be some ethical dilemmas with the implementation of AI in healthcare?	Challenges with AI-DSS in Healthcare
9	What can be a potential chain of decision making when using AI in clinical healthcare?	Relevant Stakeholders Distributing Responsibility
10	How do you think we can distribute responsibility among these stakeholders? a) If they think regulations and frameworks are the answer - how would such regulations or frameworks need to be structured?	Distributing Responsibility Managerial Learning
11	If you are a key stakeholder in this process, what inputs do you think you could provide to a healthcare manager to ensure that AI is working in the best interest of the patient?	Managerial Learning

12	How can AI be managed in clinical healthcare? b) What procedures do you have in place to manage responsibility when using AI in healthcare?	Distributing Responsibility Managerial Learning
13	How do you deal with a lack of awareness, capability, or motivation to see and act according to moral obligations towards the behavior of the systems that are designed, controlled, or used?	Lack of Moral Awareness, Capability, & Motivation
14	How can AI follow a regulatory framework while living up to its purported purpose in clinical healthcare?	Regulations and AI-DSS
15	What could be the role of regulation in the implementation of AI in healthcare?	Regulations and AI-DSS Managerial Learning
16	How can legal delimitations apply to different stakeholders in the chain of decision-making of clinical healthcare?	Regulations and AI-DSS Distributing Responsibility

Table 3: Interview questions with the corresponding themes.