OSLOMET

'How will it hit us?'

Introducing artificial intelligence (AI) in the Norwegian healthcare services: Three modes of actor mobilisation

Mari Serine Kannelønning

'How will it hit us?'

Introducing artificial intelligence (AI) in the Norwegian healthcare services: Three modes of actor mobilisation

Mari Serine Kannelønning

OSLOMET

Thesis for the degree of Philosophiae Doctor (PhD)
PhD programme in Social Sciences, with specialization in
Information, Library and Archival Studies
OsloMet – Oslo Metropolitan University

Spring 2024

Acknowledgements

Summary

This thesis explores the introduction of artificial intelligence (AI) in the Norwegian healthcare services. It does so by focusing on the role of expectations and technological promises related to a future with AI in healthcare and how such visions are collectively pursued in the present.

AI is widely expected to play an essential role in providing financially viable public healthcare services in the future. The expectations typically include promises of technologies that can automate repetitive tasks and support healthcare professionals in making decisions, thereby improving their work efficiency and enhancing treatment quality and patient safety. However, the current state of the introduction of AI in healthcare services worldwide is characterised by a so-called AI chasm between the expectations and the promise of AI technologies and their actual deployment in real-world clinical settings. To navigate this complex landscape, governments, health authorities, AI environments and other stakeholders perceive the need for a collective effort among those with relevant knowledge and expertise.

This thesis uses the AI chasm as an entry point to explore three Norwegian initiatives that mobilise heterogeneous sets of actors in attempts to enable AI in the Norwegian public healthcare services: a national inquiry process led by the Norwegian Directorate of Health, an AI procurement process at a hospital trust and a nationwide network of professionals. These initiatives are studied through a qualitative case study approach that includes data from digital meeting observations, interviews and document analysis. Each study represents one of the three articles in this article-based thesis and explores different aspects of expectations, promises and collective work with the future of AI in healthcare as the focal point. Altogether, the case studies enabled an investigation of the overall research question of this thesis:

> *What are the significance and implications of 'mobilisation' in the early phases of introducing ambiguous, complex and advanced technologies like AI in healthcare?*

The thesis analyses the findings from the case studies by drawing on concepts and perspectives from Science and Technology Studies (STS) dealing with innovation processes and the shaping of technological developments. With a primary focus on the aspect of actor

mobilisation and Michel Callon's (1986b) article 'Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay', the thesis develops a theoretical apparatus with three modes of actor mobilisation at its core: Mode 1: Steered outcome; Mode 2: Negotiated outcome; and Mode 3: Fragmented and distributed outcome.

As a whole, this thesis aims to contribute to a better understanding of the role of future-oriented representations (e.g., expectations and technological promises) and the mobilisation of heterogeneous actors in the processes of introducing advanced and complex classes of technology, such as AI, into society.

Sammendrag

Denne avhandlingen tar for seg innføringen av Kunstig Intelligens (KI) i den norske helsetjenesten ved å studere ulike aktørers forventninger og løfter om fremtiden med KI i helsetjenesten, og hvordan man kollektivt jobber mot disse fremtidsvisjonene i nåtid.

Verden over er det bred enighet om at KI vil spille en viktig rolle for å opprettholde økonomisk bærekraftige helsetjenester i fremtiden. Det forventes for eksempel at KI-teknologier skal kunne automatisere repetitive oppgaver, støtte helsepersonell i beslutningsprosesser, forbedre arbeidseffektivitet og øke behandlingskvalitet og pasientsikkerhet. Pågående arbeid for å introdusere KI i helsetjenesten er imidlertid preget av en såkalt «KI-kløft» [AI chasm] mellom forventningene og løftene om KI-teknologier og deres faktiske implementering i virkelige kliniske praksiser. Både regjeringer, helsemyndigheter, KI-miljøer og andre interessenter mener at en felles innsats fra aktører med ulik kunnskap og ekspertise er nødvendig for å kunne navigere i dette komplekse landskapet.

Avhandlingen bruker KI-kløften som inngang til å utforske tre norske initiativer som i sine forsøk på å innføre KI i norsk helsetjeneste har mobilisert heterogene aktørgrupper: en nasjonal utredningsprosess ledet av Helsedirektoratet, en KI-anskaffelsesprosess ved et større helseforetak og et landsdekkende nettverk av fagpersoner. Disse initiativene studeres gjennom en kvalitativ casestudie-tilnærming, og studien inkluderer data fra digitale møteobservasjoner, intervjuer og dokumentanalyse. Avhandlingens tre artikler tar for seg hver sin case, og utforsker ulike aspekter ved forventninger, løfter og kollektivt arbeid som har en fremtidig helsetjeneste med KI som fokusområde. Samlet utgjør disse tre studiene grunnlaget for å svare på avhandlingens overordnede forskningsspørsmål:

> *Hva er betydningen og implikasjonene av 'mobilisering' i tidlige faser av innføringen av flertydige, komplekse og avanserte teknologier som KI i helsetjenesten?*

Avhandlingen analyserer funnene fra casestudiene ved å ta i bruk begreper og perspektiver fra 'Science and Technology Studies' (STS) som fokuserer på innovasjonsprosesser og teknologisk utvikling. Med særlig vekt på aktørmobilisering og Michel Callons artikkel: 'Some Elements of a Sociology of Translation: Domestication of the Scallops and the

Fishermen of St Brieuc Bay' (1986b), utvikles det en teoretisk ramme som består av tre former for aktørmobilisering: 'Modus 1: Styrt utfall', 'Modus 2: Forhandlet utfall' og 'Modus 3: Fragmentert og distribuert utfall'.

Avhandlingens mål er å bidra til en økt forståelse for rollen til fremtidsrettede representasjoner (forventninger og løfter etc.) og mobiliseringen av heterogene aktører i innføringsprosesser av avanserte og komplekse teknologiklasser som KI i samfunnet.

Contents

List of articles

Article 1   Kannelønning, M. S. (2023). Contesting futures of artificial intelligence (AI) in healthcare: Formal expectations meet informal anticipations. *Technology Analysis & Strategic Management*, 1–12.
DOI: https://doi.org/10.1080/09537325.2023.2226243

Status: Published

Article 2 (book chapter)   Kannelønning, M. S., Grisot, M., & Williams, R. (forthcoming). Towards experimental implementations: Moving emerging AI technologies into real-world clinical settings. In P. Giardullo & F. Miele (Eds.), *Algorithmic care: STS perspectives on automation of care*. Palgrave Macmillan.

Status: Accepted for publication
[Article not attached due to copyright]

Article 3   Kannelønning, M. S. (2023). *Navigating uncertainties of introducing artificial intelligence (AI) in healthcare: The role of a Norwegian network of professionals*. Technology in Society, Volume 76 (Mars 2024), Artikkel 102432.
DOI: https://doi.org/10.1016/j.techsoc.2023.102432

Status: Published

PART I: COVER CHAPTER

Prologue

Around 2016, I started noticing an increase in media coverage of recent developments within the field of artificial intelligence (AI). More and more frequently, new achievements were mentioned, from AI technologies improving cancer diagnostics to potentially solving climate issues. Since the 1950s, there have been ups and downs in the development of AI, but now a new 'AI spring' with greater and more widely shared expectations of success seemed to be the case. However, more dystopic narratives of AI emerged, too. These included warnings of AI technologies replacing the human workforce or potentially leading to the extinction of humanity. This hype-disaster rhetoric intrigued me. I have always been fascinated by how the emergence of new technologies changes our societies and ways of being. With powerful AI models such as OpenAI's ChatGPT, the AI discussion has escalated enormously during the last year alone (from December 2022). It has even come so far that prominent AI researchers and tech leaders have called for a pause to its development as it is accelerating too fast for necessary management and risk-mitigating efforts to happen or keep up (Future of Life Institute, 2023). Suddenly, AI dominates conversations in all corners of our society: education, justice, media, social services, healthcare and so on. As a contribution to the ongoing discussions, my thesis addresses the latter: the healthcare sector.

As is the experience of most doctoral students, my PhD project changed tremendously during its first year. The first 'turn' happened only a few weeks after becoming a PhD student, as a reply to an email I had sent to a project manager of an AI project at a hospital included the word 'premature'. Before this response, I considered the project as a perfect case to explore the deployment of AI-based information systems within the work practices of clinicians. It was planned as a project positioned across the fields of Library and Information Science (LIS) and Science and Technology Studies (STS), with a theoretical approach based on crosscutting themes from information infrastructure research and scholars such as Karen Ruhleder, Susan Leigh Star and Geoffrey Bowker (e.g., Bowker & Star, 1999; Star & Ruhleder, 1994). In such a project, I saw an opportunity to explore what would happen as this hyped and much-discussed type of technology, loaded with high expectations and uncertainties regarding their materialisation, entered clinical practices. However, after I received the email reply and after initial investigations into other projects, I realised that I had failed to see the actual state of AI in healthcare. The stories circulating of AI technologies performing better than humans in areas such as clinical diagnostics did not necessarily

highlight the fact that this was mainly within the borders of research projects. Thus, one could say that my first insight as a researcher into AI in healthcare was that the promoted technologies were mainly constituted by claims and promises of certain capabilities rather than actual performance in real-world clinical settings.

In this situation in which AI technologies were not yet properly in use but existed or, rather, came into being through different articulations of their potential, I found a renewed focus for my PhD project. With the role of expectations and promises in the ongoing process of introducing AI in healthcare as a new focal point, my project turned into a prologue to the research project I initially had imagined. My choice of theoretical framework changed as well, and the STS literature became more prominent than the LIS literature. With these alterations, my motivation behind the PhD project also changed. As I dived into the matter, problems of another order became important to me. I especially became concerned that the technologies we choose to develop and take into use today, regardless of how harmless or useful they may seem, can have severe consequences for future generations. In line with the philosopher Hans Jonas (1982), I want to advocate for the responsibility we have in the present to do what we can to control the technologies we release into society. We need to render visible existing and emerging uncertainties related to a future with AI in healthcare in order to be able to understand, mitigate or monitor their many dimensions. Only then can a sustainable outcome for the future be achieved.

1 Introduction

Multiple studies show that AI is expected to play a crucial role in providing sustainable healthcare services in the future. Public healthcare services are already under pressure due to factors such as rising costs, workforce shortages and exponential growth in digitalised health data (Aung et al., 2021; Meskó et al., 2018; Topol, 2019). In this context, the expectations for AI typically include promises of technologies that can automate tasks, replace healthcare professionals or support professionals in decision making. AI is expected to improve diagnostic accuracy and efficiency in clinical workflows, facilitate better disease and treatment monitoring and improve patient treatment and safety (Kaul et al., 2020, p. 807). Currently, one of the most promising areas in which AI technologies have proven to have positive effects is within medical areas dealing with image analysis such as radiology, pathology, dermatology and ophthalmology (Rajpurkar et al., 2022, p. 722; Wang et al., 2019, p. 293; Yu et al., 2018).

Despite the advancements in the field of medical AI, research on AI implementation in healthcare shows that there is a 'last mile' to cover or a 'gap' or 'chasm' to bridge to achieve widespread deployment (Cabitza et al., 2020; Coiera, 2019; Keane & Topol, 2018; Seneviratne et al., 2020). This space between expectations and deployment (described in the following as the 'AI chasm') is constituted by issues related to areas such as data limitations, regulations, ethics, human trust and the lack of evidence of AI performance in operative clinical practices (see, for example, Alami et al., 2020; Keane & Topol, 2018; Parikh et al., 2019; Rajpurkar et al., 2022).

Such issues imply that the future with AI in healthcare is filled with uncertainties. As one of my informants pinpointed, people in the healthcare services are aware that AI will eventually become a part of clinical practices, but they do not know yet how it will hit them. For instance, there are uncertainties surrounding whether specific AI technologies will ensure accurate diagnoses and equal treatment for a hospital's total patient population, whether regulations for using AI in clinical practices will allow for less intervention or effort from clinicians or whether the clinicians will have the right balance of trust so they neither refuse to use the technology and stay at the *status quo* nor let the technology make the final decision and potentially cause severe and adverse effects on patient treatment and safety. In addition

to such concrete uncertainties, there are several more, both known and unknown, attached to the introduction of AI in healthcare.

Many of the uncertainties concerning the future with AI technologies are related to the characteristics that make these technologies different and potentially more disruptive than any other class of technology. This includes, for instance, the unknown short-term and long-term outcomes of algorithms that continuously learn during use as they are exposed to new data or the ability of AI to perform tasks previously requiring human cognition; but it goes beyond this as well. How do we ensure that the future with AI is more aligned with its optimistic expectations than with its dystopic counterparts? As knowledge of real-world clinical use is currently limited, the most crucial driving force behind the ongoing work of introducing AI in healthcare is based on anticipation.

As touched upon, many of the uncertainties posed by the issues within the so-called AI chasm are complex and consist of a combination of technical, juridical, ethical, medical and social factors. To deal with these issues, governments and other key actors (see, for example, Danish Ministry of Finance & Danish Ministry of Industry, 2019; Meld. St. 7 [2019–2020]; NHS, n.d.) see it as necessary to mobilise and establish collaborations between various actors from both public and private sectors (for similar arguments, see also Alami et al., 2020, p. 7; de Neufville & Baum, 2021). Bringing together actors from diverse fields allows for the sharing and merging of knowledge and expertise that is relevant to cast light on the issues challenging or problematising widespread deployment of AI. Through such collectivity, several aspects of the uncertainties regarding the future with AI can better be identified and considered to avoid compromising people's health and lives through the introduction of AI. This underscores the reasons it is especially critical to ensure a broad inclusion of actors: healthcare services are one of the most essential public services, dealing with decisions of life and death. However, the questions are what could or should such mobilisations of key actors look like, who are the actors already mobilised, what interests do the actors have, who are the most powerful actors and what do such mobilisations imply for the future with AI in healthcare.

Inspired by these less-explored questions in an emerging field, this article-based thesis aims to contribute to a better understanding of the significance and implications of such mobilisations in the processes of introducing advanced and complex classes of technology

into society. More precisely, the thesis uses the AI chasm as an entry point, while the aspect of 'mobilisation' is at the core of the investigation. The term 'mobilisation' is, in this context, to be understood in a broad sense, referring to how diverse actors are brought together to collectively contribute to the shaping of technological developments.

1.1 Research question

Based on three qualitative case studies that constitute the empirical foundation of this thesis as a whole, the present cover chapter explores the thesis' overall research question:

> *What are the significance and implications of 'mobilisation' in the early phases of introducing ambiguous, complex and advanced technologies like AI in healthcare?*

To answer this question, the findings from the case studies are analysed by drawing on concepts and perspectives from Science and Technology Studies (STS) dealing with innovation processes and the shaping of technological developments, with a primary focus on the aspect of actor mobilisation. Central in this context is the seminal article of the sociologist and prominent STS scholar Michel Callon (1986b): 'Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of St Brieuc Bay'.

1.2 The three cases in short

This thesis' case studies took a qualitative approach and included data from digital meeting observations, interviews and document analysis. Each study represents one of the thesis' three articles and explores early key initiatives in the ongoing process of introducing AI in Norwegian public healthcare services. Apart from being individual empirical studies that tell different stories related to the expectations, promises and uncertainties of the future with AI in healthcare, these studies are all examples of mobilisations of heterogeneous actors. While grounded in a specific national context, this thesis has relevance outside of the Norwegian borders as other countries worldwide are standing on the same doorstep, working with similar processes of introducing AI in healthcare (see, for example, AAAiH, 2019; AAiH, 2023; AI Sweden, n.d.; FCAI, n.d.; NHS England, n.d.).

The first study (article 1) was of a national inquiry process led by the Norwegian Directorate of Health discussing how to enable AI within radiology as a medical area at the forefront of

AI applications. The process was organised as part of a government policy implementation process and involved approximately 100 stakeholders with various backgrounds. The study of this initiative explored the characteristics of the formal expectations of AI in healthcare as articulated by government policies and how these are contested by informal anticipations of actors operating closer to clinical practices.

The second study (article 2) concerned a competitive dialogue-based procurement process at a large Norwegian hospital trust aiming to procure commercially available AI solutions for radiology. The project was the first of its kind in a Norwegian public hospital, and due to its size, the vendors involved also claimed that it was the first of its kind in Europe. This study investigated what happens as visionary claims of AI are concretised in relation to the specific needs and requirements of a user organisation and what role is played by expectations and promises during such early encounters with real-world clinical settings.

The third study (article 3) was of a nationwide, informally established network of approximately 200 professionals from different fields, organisations and sectors in Norway interested in AI and healthcare. The network aimed to bridge what they called the 'implementation gap' in clinical AI by facilitating knowledge and experience sharing among its members and influencing decision makers in the field in general. The study of this network explored how such informally established networks can contribute to enabling AI in healthcare.

1.3   Thesis structure

This doctoral thesis consists of two parts: 1) the present cover chapter and 2) the three articles. The cover chapter proceeds as follows: First, a background and context section is provided, covering a short history of AI, a review of the status of AI in healthcare today and selected STS work on AI. Second, the theoretical framework is introduced, including a selection of the concepts and perspectives from the three articles and an outline of the framework used to explore the overall research question of this thesis, in particular, in this cover chapter. Third, the research project's methodology, methods and material are described and discussed. Fourth, the three studies are summarised, including their findings and contribution to the thesis as a whole. Fifth, the findings from the three studies are analysed and discussed in light of the cover chapter's research questions. Inspired by Callon's work, I

develop a theoretical apparatus with three different modes of actor mobilisation at its core. This section also discusses the thesis' contributions to existing research and the theoretical field of STS. Sixth, a conclusion and final remarks are provided. The cover chapter is then followed by the thesis' second part: the three articles in the same order as presented herein.

2   Background and context

The purpose of this section is to introduce the reader to background information relevant to the thesis' overall topic and later discussion. It includes a brief account of selected highlights from the history of AI, some clarifications on the concept of AI, an outline of what AI in healthcare is today and what occupies the current chasm between expectations and deployment of AI in healthcare. Finally, to situate this study in the lineage of STS work on AI, a short review of literature relevant to my thesis is provided.

2.1   A brief history of AI

As Galanos (2023) pinpointed, it can be hard to ascertain when the field of AI really began. Ideas of imitating human intelligence or the functions of the human body can be found in ancient mythology, philosophy, religion and fiction. Within the sciences, cybernetics can be seen as a vital precursor (p. 54). From the field of mathematics, Alan Turing is known for establishing the early foundation of AI. In his seminal work *Computing Machinery and Intelligence* (1950), he explained how to create intelligent machines and how to test whether these have reached a certain intelligence. He called the test 'the imitation game', also known as the Turing Test.

Five years later, in 1955, the term and field of 'artificial intelligence' was launched in a funding proposal for the Rockefeller Foundation for a two-month long summer workshop. The proposal, written by the scientists John McCarthy, Marvin L. Minsky, Nathaniel Rochester and Claude E. Shannon, did not include a definition of 'artificial intelligence' but highlighted that the study to be carried out during the workshop was to proceed from the assumption 'that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it' (McCarthy et al., 2006, p. 13). Furthermore, the proposal stated that they would attempt to find out 'how to make

machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves' (p. 13).

At a BBC TV debate in 1973, McCarthy explained how he came to call the field 'artificial intelligence'. In the years before the funding proposal in the 1950s, he and his colleague, Shannon, tried to use a different term, 'automata studies', as their aim was to collect relevant studies to launch the field. They chose this term because Shannon thought 'artificial intelligence' sounded a bit 'too flashy'. However, McCarthy did not find the term 'automata studies' sufficient to cover the kind of studies he had in mind, as his long-term goal was to develop machines with human-level intelligence. Thus, as they attempted to obtain funding for the summer workshop, they tried to call it something other than 'automata studies', namely 'artificial intelligence' (BBC TV, 1973, time 46:00–47:45).

After years of great optimism in the 1950s and the beginning of the 1960s, the field of AI faced criticism and financial obstacles in the 1970s because many of the expected goals were never achieved due to unforeseen or underestimated challenges, including limited computational power and the issue of how to make the programmes solve more complex problems on a larger scale (Russell & Norvig, 2010, pp. 20–22). As the field failed to deliver on the promised outcomes, crucial research funding was withdrawn. This period is often called the first AI winter. Then, in the 1970s, systems were built which enabled more reasoning steps to be included by encoding domain-specific knowledge into the systems, making it possible to solve more complex problems than the previous simple problem-solving systems were capable of doing (Russell & Norvig, 2010, p. 22). This development can be seen as the beginning of the 'knowledge-based systems', of which 'expert systems' is an example (Galanos, 2023, p. 57).

With the development of expert systems, the field of AI entered a new spring in the 1980s. During the decade, this renewed spark also included a revival of the subfield of neural networks (Russell & Norvig, 2010, p. 16). This approach, often called connectionism, was inspired by the function of neurons in the brain rather than the reasoning process of the mind, which was the basis for the previously described approaches, often called symbolic AI. Despite this progress, the developments came to a halt once more as the goals that had emerged during the period failed to be met, and as the decade ended, a new AI winter approached (2010, p. 24). However, in the 2000s, some of the old goals, including the

development of systems that could solve more complex tasks, were finally achieved, enabled by enhanced computing power, increased amounts of digitalised data and cloud storage. These achievements led to the development of more sophisticated versions of AI, such as machine learning (ML) and its subfield deep learning (DL), which today are enabled for use in most sectors (Topol, 2019). Presently, AI technologies have the ability to solve a wide range of tasks, including complex game playing, such as Google's AlphaZero, the creation of art, such as OpenAI's DALL-E 2, advanced levels of conversation, such as OpenAI's ChatGPT, and disease detection, like ScreenPoint Medical's Transpara Breast Care.

Current AI technologies are often broadly defined as machines capable of performing tasks and roles mainly performed by humans that require a certain intelligence to complete (Morley et al., 2020, p. 1). A similar description can be found in the OECD Council's recommendation on AI (OECD, 2019), defining AI systems as follows: 'An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy' (p. 7).

In such broad definitions, AI becomes an umbrella term or a technology class covering a vast amount of technologies developed for various purposes using different approaches, typically within one of the two overall categories of rule-based AI or machine learning (ML). Rule-based AI includes expert systems, while neural networks (NN) and deep learning (DL) are subsets of ML-based AI. ML, NN and DL approaches involve a range of algorithms that enable independent learning by a system based on large amounts of data with a reduced need for explicit programming or human guidance. In contrast to rule-based systems, ML- and DL-based algorithms can enable the discovery of hitherto unknown patterns in data without preformulated decision rules for each individual task (Yu et al., 2018, p. 720). ML and DL are often used as synonyms for AI. Such technologies can either be trained on supervised or unsupervised datasets. The notion of 'supervised learning' means that the algorithms are trained on labelled datasets in which the input data and output data corresponding to each other are known, and the algorithm 'learns' from these already known and shown connections. Such algorithms can then make predictions based on the patterns learned during the training. 'Unsupervised learning' concerns algorithms that are set to find hidden patterns and structures in unlabelled data. Thus, they explore the data without being given specific prediction tasks (Alloghani et al., 2020, p. 4).

The outlined foundational events from the history of AI and the brief treatment of AI as a concept and technology class show the multifaceted nature of AI, both as an academic field (it was established based on the work of scholars from different fields) and as a term covering a vast number of technologies. This conceptual ambiguity plays a central part in my research, affecting the discussions that take place, the work undertaken and the people engaging in the initiatives studied in terms of who they are and what they work to achieve. As I initially touch upon in articles 1 and 3, this ambiguity also has practical consequences as the understanding or definition of AI that is used influences decisions made during various stages of innovation processes. This could be in policy development and implementation, as exemplified in article 1, or as different actors try to join forces to contribute to enabling AI in healthcare, as exemplified in article 3. With the many definitions of AI, questions like whether AI will be beneficial or not will lead to several different answers, depending on who is asked (Wang, 2019, p. 28). Such varying perspectives on and understandings of AI underscore that if one kind of actor is not involved or listened to in a particular innovation process, this could have consequences for the sustainability of the innovation (cf., especially article 3).

The many divergent perceptions of AI can be exemplified by a computer scientist for whom AI can be not only a DL model for detecting bone fractures in medical images but also algorithms that are intriguing to develop within a research project, not necessarily with specific application areas or users in mind. For a representative from the health authorities, AI can be something to investigate in terms of its potential, or it can be something about which to write reports and develop strategies. For hospital managers, AI can be viewed as a way to potentially save money and cope with staff shortages, and for physicians, AI can be a specific work tool for making faster decisions and reducing workloads or the opposite (for more on the multiplicity of what may appear as one thing, see Mol [2002]). These different views also exemplify that if or when AI is specified or concretised based on a selection of the many understandings of AI, a larger variety of definitions and interests may emerge.

Reflecting the lack of a widely accepted definition and in line with the terminology used in articles 1 and 3, this thesis uses the general terms 'AI' and 'AI technologies' interchangeably. These labels cover AI both broadly, for instance, when treating overall conceptions of AI, and more narrowly, including specific technologies.

2.2 AI in healthcare

Healthcare services, or, more precisely, medicine, have long been considered a promising area for AI applications (Yu et al., 2018, p. 719). Hence, the development of AI for use in healthcare has progressed much in parallel with the overall AI advancements outlined in the previous subsection. Ideas and developments of technologies simulating medical reasoning which were foundational for clinical decision-support systems can be traced back to the 1950s (Miller, 1994, p. 11), while rule-based systems (e.g., expert systems) for clinical problem-solving evolved especially during the 1970s (Szolovits et al., 1988, p. 80). In recent years, the developments within the field of AI, including ML, have led to several achievements in clinical AI. Currently, the preeminent approaches for developing AI technologies for clinical use are those utilising ML and DL methodologies (see, for example, Dicuonzo et al., 2023; Racine et al., 2019; Rajpurkar et al., 2022; Yu et al., 2018).

A simple search in PubMed, the world's largest search engine for biomedical literature, illustrates how the development and interest in AI in medicine have escalated during the last seven years. In 2016, 193 papers were indexed in PubMed with the term 'artificial intelligence' or 'AI' in the title field. In the following years and up to 2023, this number multiplied into 5,590 papers indexed for the year 2022. Even though these numbers indicate that much knowledge has been generated and the field is developing, the process of introducing AI technologies in healthcare can still be considered to be in its early stages. As mentioned in the introduction, AI technologies are expected to improve several areas within healthcare services, but the widespread deployment of such technologies is not yet a reality. This is despite the significant advancements in DL for image analysis, which is currently the most successful area for AI applications in medicine (Rajpurkar et al., 2022, p. 722; Wang et al., 2019, p. 293; Yu et al., 2018). This implies that numerous AI models are developed for narrow task solving, especially within medical specialities such as radiology, pathology, dermatology, ophthalmology and cardiology. Such technologies can aid in interpreting medical scans (e.g., X-rays or Magnetic Resonance Imaging [MRI] scans), digitised pathology slide images, digitised dermoscopic images, retinal scanning or electrocardiograms. For instance, the use of DL models trained on large amounts of data to analyse such modalities can support clinicians in detecting cancer, diabetic retinopathy or heart attacks (Topol, 2019).

For more complex medical processes, such as disease identification and treatment recommendations, more proprietary systems are required, encompassing several technologies and employing different AI techniques, for instance, for processing images (computer vision) and speech and text (natural language processing). However, such systems have proven more challenging to implement in existing practices (Davenport & Kalakota, 2019, p. 95). IBM's Watson serves as a clear example of how proprietary systems can fail to adapt to real-world conditions; it simultaneously provides an excellent example of how AI technologies, in general, can tremendously fail to deliver on the over-dimensioned promises attached to their early versions (Ross & Swetlitz, 2017, 2018; Strickland, 2019).

As already implied, there is a space or a chasm between theory and practice, on the one side, and expectations, development and deployment on the other side (see, for example, Aristidou et al., 2022; Bi et al., 2019; Cabitza et al., 2020; Coiera, 2019; Panch et al., 2019; Rajpurkar et al., 2022; Seneviratne et al., 2020). This chasm is evident regardless of whether it concerns AI technologies developed 'in-house', for instance, as research projects at a specific hospital in which the algorithm is designed to accommodate local needs and trained on local patient data, or whether it is AI technologies that are CE marked ('Conformité Européenne') commercial products intended for use in a broader range of contexts and trained on data other than from local patient populations. CE marking indicates that requirements from the European Union are met concerning safety, health and environmental protection. However, AI technologies used for humans in healthcare (e.g., for diagnostics, treatment or monitoring) are also considered to be medical devices. Thus, to obtain a CE mark, the manufacturer must ensure that the product complies with the European Medical Device Regulation as well. This again implies that a CE mark can only be obtained through a 'notified body approval' as opposed to other products for which a self-declaration from the manufacturer may be sufficient (Malvehy et al., 2022, p. 361; European Union, 2022).

Like in many other countries, the number of AI deployments in Norwegian public healthcare services is low. According to a report from October 2022 that was produced by the previously mentioned national coordination project (cf., study 1), each of the four Norwegian health regions and some municipalities has started using AI (Helsedirektoratet et al., 2022). However, the list of examples from the healthcare services is sparse. Out of the five projects mentioned, only two seem to involve products that have actually been taken into use in healthcare settings and not only as part of a trial. Furthermore, one of the five projects is the

procurement process studied within this doctoral project (study 2), which has not yet (as of August 2023) started using the AI product procured. The report also summarises the work of the regional health authorities, which for one of the regions includes trials and the use of algorithms within areas such as bone age measures, detection of anatomical structures in medical images, processing of MRI pictures of hearts and analyses of tests within medical laboratories. For the rest of the regions, no use of AI is reported, and most of the projects referred to are either research projects, innovation projects or trial projects testing out specific technologies (Helsedirektoratet et al., 2022, pp. 11–14). This status of the introduction of AI, which mainly consists of early development projects and tests of AI, is further confirmed by an overview of Norwegian AI projects made by the KIN network (cf., study 3). This overview underscores that the majority of the ongoing AI projects are either research and development (R&D) projects or projects that aim to test or validate AI technologies (KIN, 2023).

While the issues constituting the chasm may vary depending on the type of technology and its purpose and areas of application, there are several common characteristics. As such, the following section will treat the key components contributing to the AI chasm on a general level.

2.2.1   The AI chasm

To illuminate what constitutes the AI chasm, a review of current research addressing the chasm will be presented here. The literature was retrieved through searches in Google Scholar and using snowball sampling by following references in articles already retrieved. As I reached the number of 26 articles, I experienced a certain redundancy in the issues listed within the articles, which further delimited the inclusion of more articles in the review. Some of the most crucial issues from this analysis are organised into three categories: Technology and data; Humans and organisation; and Regulations, responsibility and ethics.

Technology and data

In a synthesis of key challenges posed by AI, Alami et al. highlighted the generalisability and reproducibility of AI technologies and their performance. This includes shortcomings regarding the datasets on which the algorithms are trained, such as data that are nonrepresentative for diverse use contexts and their patient population (Alami et al., 2020, p.

4; Rajpurkar et al., 2022, p. 34). This shows how challenging it can be to develop algorithms with performances or benefits that are transferable from one healthcare setting to another, that will work as intended on whatever patient group and that integrate with the differing information systems already in use (Keane & Topol, 2018, p. 1). Added to this issue are factors such as how data are decontextualised, meaning that algorithms developed for certain purposes cannot consider other contextual parameters that can be relevant for diagnostic processes, such as psychosocial factors or factors related to ethnic diversity (Alami et al., 2020, p. 4).

Another concern is the quality of the existing data, including the well-known problem of 'garbage data in, garbage data out'. In the case of AI, and especially for ML-based algorithms that 'learn' based on the input data, if the training data is of poor quality, this will only be amplified when processed in large quanta, leading to output data of even poorer quality (Alami et al., 2020, p. 4). Cabitza et al. (2020) similarly asserted 'the fact that no algorithm, no matter how smart or intelligent it is, can produce value if its input data lacks value in the first place' (p. 7). However, the data that exists in the healthcare services today are not necessarily curated, labelled and quality assured according to what is needed for algorithms to work properly or as intended (Bi et al., 2019, p. 148). Another related challenge is the lack of available datasets large enough for the algorithms to work sufficiently. In healthcare, a part of this problem is a lack of opportunity or reluctance to share data across hospital organisations and/or between the public and private sectors (Sun & Medaglia, 2019, p. 375).

Explaining how ML- or DL-based technologies have reached a particular result has proven to be extremely difficult and is often referred to as the 'black box' problem (Rajpurkar et al., 2022, p. 35). Arguably, the inner workings of other types of technologies, for instance, scanners for medical imaging or, for that matter, humans, can be difficult to understand, too. However, as exemplified by the 'garbage in, garbage out' issue, when AI technologies are implemented in real-world clinical settings, especially on a larger scale, the flaws of an algorithm could cause greater harm to a larger amount of patients in one day than a scanner or a clinician could do in years (Topol, 2019, p. 51). Without a certain level of transparency, it will be hard to identify where the error, bias or failure lies and subsequently adjust the technology to keep the problematic results from continuing (Alami et al., 2020, p. 4; Sun & Medaglia, 2019, p. 376). The lack of transparency is heightened by potential limitations in training datasets as these do not necessarily cover all aspects relevant to the case. The

transparency issue is reinforced by flaws in the code of the algorithm or future changes in work routines or patient treatment that cannot be accounted for at the time the algorithms are developed (Alami et al., 2020, p. 4). Bringing about the full potential of AI technologies requires that they be integrable and interoperable with other information systems from different providers already in use in clinical workflows. It is only then that a seamless and efficient sharing and flow of data can happen (Panch et al., 2019, p. 1; Zając et al., 2023, p. 22).

Lastly, it is important to add to the list of challenges regarding technology and data the risk of algorithms being hacked, potentially harming people on a large scale, or data breaches that reveal sensitive patient information to unauthorised persons (Topol, 2019, p. 52).

Humans and organisation

The human ability to trust AI recommendations is affected by a lack of transparency in how an AI technology performs its reasoning as well as the basis for its reasoning (Kelly et al., 2019; Xiao et al., 2018, p. 6). It is similarly important to understand how an algorithm reaches certain decisions and to know whether it is possible to make corrections in the system if errors are discovered (Ching et al., 2018, p. 12; Sun & Medaglia, 2019, p. 376). A related issue is the ability to assess the precision of AI in making diagnoses and recommendations. This may vary as different standards and opinions exist that inform clinicians' decisions and against which an AI technology's outcome can be measured (Alami et al., 2020, p. 5).

From the clinician's perspective, a crucial question is whether the AI technologies will be integrated with existing work practices and information systems (Wang et al., 2019, p. 293). For example, previous studies have shown that AI technologies can deliver the result too late according to the overall workflow, rendering the analyses useless (Hollander et al., 2004). As AI algorithms become a part of clinicians' work processes, another issue related to poor user experience is 'alert fatigue' (Wong et al., 2021). As Alami et al. (2020) summarised based on a review of four studies (Blease et al., 2019; Laï et al., 2020; McCullagh et al., 2014; Yu et al., 2018): '[I]f clinicians feel overloaded and workflows become more complex, AI may be rejected because of self-perceived inefficacy and performance, alert fatigue, cognitive overload, and disruption of interpersonal communication routines' (p. 5).

While issues such as these may lead clinicians to mistrust the algorithm and/or to reject the AI technology, it is also possible that clinicians might come to rely too much on the algorithm (Cabitza et al., 2020). Such overreliance may make them skip verification of the result and thereby overlook potential errors, also called automation bias (Arora, 2020, pp. 227–228). As AI technologies are developed by humans and trained on human-shaped data, some errors and inaccuracies are unavoidable. Regardless, they may still be helpful. Based on a study of human–AI collaboration in a healthcare setting, Bossen and Pine (2023) found that despite certain known flaws, AI technologies may become, as they put it, what Robin is to Batman. As long as the shortcomings are known and humans remain in control, such technologies can be useful helpers. They can even offer learning opportunities for their users (Bossen & Pine, 2023).

AI may also change how work is divided among different professions and require a different set of skills and expertise. This will again require the establishment of new rules and workflows, the redefinition of professional responsibilities and an evaluation of the types of professionals and competencies needed to work effectively with AI (Gillan et al., 2019). From a patient perspective, AI could result in less contact with clinicians, eventually leading to greater isolation for some patients. It could also create unrealistic expectations of outcomes or, conversely, make patients suspicious and, thus, not accepting of the AI result or the use of AI at all. These concerns underscore that user organisations, such as hospitals, that procure AI technologies need to consider whether the AI technology they aim to buy will provide the value they are seeking and whether they are ready to cope with the changes a deployment of AI will cause on an organisational and individual level as outlined (Alami et al., 2020, p. 5).

It will be necessary for the healthcare services to invest considerable amounts of money, time and effort to introduce AI in clinical practices (Sendak et al., 2019). However, such resources may not be available in the implementing organisation (Alami et al., 2020, p. 6). This includes resources to ensure a suitable technical set-up for the new integration, evaluate and secure the information, test and validate the AI performance, improve the healthcare professionals' competencies and adjust existing workflows or patient flows according to the new opportunities (He et al., 2019). Interestingly, as Alami et al. (2020, p. 6) pointed out, requirements such as these that are necessary to properly implement AI in local practices are typically downplayed in the literature on promises of AI performance or within the policies supporting AI development and implementation.

Regulations, responsibility and ethics

Other areas of importance for introducing AI technologies in healthcare are regulations, distribution of responsibility and ethics (Rajpurkar et al., 2022, p. 34). As learning algorithms for use in relation to humans in healthcare are not yet approved by either relevant European or American regulatory frameworks, the learning feature is 'locked'. In other words, AI technologies entering clinical workflows become more like existing static non-AI technologies, as they do not evolve while in use. In a study in 2018, Keane and Topol noted that it might take years before the field, including regulatory frameworks, has been developed to deal with learning algorithms 'on a case-by-case basis in a real-world setting' (p. 1). Five years later, on 30 March 2023, the United States Food and Drug Administration (FDA) published draft guidance concerning the possibility of receiving approval for medical AI technologies that change during use without seeking new clearance from the FDA. This could potentially lead to the use of continuous learning algorithms in clinical settings as well as more frequent updates of algorithms in the near future (FDA, 2023). In the European Union, however, such possibilities are not yet set on the agenda. According to the European Medical Device Regulation, manufacturers are prohibited from launching medical AI devices on the market that potentially change from their intended use or conditions of use (COCIR, 2020, p. 10).

An often-discussed topic regarding the use of AI in clinical settings is responsibility. Who is responsible if something goes wrong in patient diagnostics or treatment if AI is part of the process? Is it the AI manufacturer, the developer, the physicians or the hospital management? It is frequently argued, including in my data, that as long as AI is used for decision support, the clinicians are still responsible for verifying the result and making the final decision. But over time and as the AI result continues to be perceived as reliable, the clinicians may be inclined to more easily accept the AI decisions, despite potential doubts (Alami et al., 2020, p. 6). This, again, makes it difficult to place responsibility if mistakes are made. No matter how the issue of responsibility is resolved or is potentially shared, it will affect the roles of the actors that are part of it.

Ethically, AI technologies for use in healthcare present several challenges that must be carefully considered during development and deployment. This includes the risk of algorithms replicating human biases (Char et al., 2018, p. 981), which may lead to discriminatory healthcare outcomes, such as incorrect diagnoses or unequal treatment for

certain population groups (Parikh et al., 2019, p. 2377). Additionally, biases can be introduced through the development of the AI algorithms themselves. This could be due to poor code quality and representativeness in the training datasets (Challen et al., 2019, p. 233). Discovering potentially unfair algorithms or biased outcomes can be challenging due to the previously mentioned lack of transparency and interpretability of the algorithms' inner workings (Morley et al., 2020, p. 2). Several examples of severe consequences due to biased data and lack of transparency already exist in other sectors. One well-known example is from the Netherlands, where the tax authorities used an algorithm to discover potential benefits fraud. After years of use, the algorithm was exposed as biased. However, this happened after the lives of many low-income families, often from ethnic minorities, were ruined as they were indebted and driven deeper into poverty based on wrongful accusations of fraud (Heikkilä, 2022).

Thus, even though AI technologies can be used for good (e.g., helping humans to analyse vast amounts of data), they are also powerful tools with many potential flaws. Most importantly, the outcome of widespread use of these technologies is not something we can control. As this review shows, multiple issues and uncertainties are attached to the deployment and use of AI. It is necessary to somehow navigate this complex landscape while aiming to introduce AI in healthcare. In this regard, a final reference to the literature addressing the AI chasm is relevant. As Alami et al. (2020, p. 7) concluded, from a health technology assessment perspective, initiatives to stimulate early dialogue among stakeholders are crucial in order to identify the necessary evidence for informed decision making when dealing with the complexities of AI technologies and the uncertainties related to their performance and outcome of use in real-world settings.

2.3   AI in STS

Before I elaborate on the thesis' theoretical framework, which draws on STS concepts and perspectives, I will give a brief review of STS literature on AI, within which my thesis is situated. The works included are indicative of the directions typically taken by scholars within STS. Among the notable early STS research on AI is the work of Langdon Winner (1978), who in the 1970s, criticised the belief in autonomous technologies (or technics out of control). Within this work, he questioned the stance of computer scientists like Marvin L. Minsky, who viewed the effects of incomprehensible computer programmes like AI systems

as a sign of progress. In particular, Winner (1978, p. 304) drew attention to issues of moral agency in such complex systems by problematising Minsky's claim that a programmer cannot be fully responsible for the outcome of such autonomous technologies. A decade later, Steve Woolgar (1985) argued for a sociology of machine intelligence that takes the human–machine relationship into account or the language of both expert machines and machine experts. As an extension of this work, he also indicated that distinguishing between technology and sociology is not a fruitful dichotomy (Woolgar, 2012).

A similar argument is found in Harry Collin's works on AI and expert systems. Some of Collins's work has focused on what artificial experts can and cannot do (Collins, 1990), how expert systems can be used to understand the nature of knowledge and that machines cannot be understood separate from their users (Collins, 2012) and, spinning off the recent AI achievements, how our idea of AI as machines with incredible capabilities blurs our vision of AI's limitations. He argues that such shortcomings in human assessments will lead humans into a world where they have made themselves dependent on 'stupid' machines (Collins, 2018, p. 5). In the 1990s, Diana E. Forsythe (1993a, 1993b) explored both the construction of work and knowledge production in AI and, in particular, expert systems, while Lucy Suchman (1987) contributed to the understanding of human intelligence and interaction by problematising the construction of intelligent machines. In 2007 Suchman followed up by contributing to the understanding of aspects of humanlike machines and human–machine configurations, raising questions such as 'What figures of the human are materialized in these technologies? What are the circumstances through which machines can be claimed, or experienced, as humanlike?' (Suchman, 2007, p. 229). Recently, Suchman has also drawn attention to AI and the automation of military intelligence. Similar to Winner, Collins and others, she problematises in this case the human ideas or visions of machine capabilities (Suchman, 2022).

Other recent studies within or in the periphery of STS have found that particular AI systems for public sector accounting and healthcare are designed based on politically motivated managerial interests in cost efficiency rather than the needs and requirements of the actual users (Henriksen & Blond, 2023), that imaginaries of AI within national AI strategies contribute to 'talking AI into being' in certain ways (Bareis & Katzenbach, 2022) and, closer to the context of my study, that the Nordic national AI strategies include a more or less shared vision regarding the future with AI in healthcare (Tucker, 2023). While the latter two

studies revolve around powerful visions of AI included in national strategies, a study by Radhakrishnan (2021) questions in particular the narrative of 'AI for social good'. She argues that the excessive attention on AI technologies in healthcare comes at the expense of solving the actual problems of those underserved, sick and poor populations, in her case in India, and does not necessarily lead to sustainable healthcare solutions. Torenholt and Langstrup (2023) also drew attention to narratives of AI in healthcare as they argued that there exists alongside the beliefs that algorithms for automated clinical decision making potentially disrupt clinical practice a view of algorithms as a continuation or part of established practices. In any case, both perceptions give legitimacy to using algorithms in clinical practices. A study by Winter and Carusi (2022) explored the issue of trust in AI for clinical decision making, which is one of the most crucial hindrances to AI adoption (cf. section 2.2.1 on the AI chasm). They showed how collaborations between AI developers and clinical experts in the processes of AI development and validation are co-constituting trust in the technologies.

The selected and reviewed STS studies take various perspectives on AI in general and healthcare in particular. Additionally, the studies mention several of the issues constituting the 'AI chasm'. In doing so, they underscore that AI technologies are neither neutral nor straightforward to deploy, as they tendentiously are depicted by certain actors (e.g., policymakers and AI developers).

3   Theoretical framework

The choices of theoretical concepts and perspectives used within my thesis are empirically grounded; they have been driven by my data material and research questions. At the same time, the different theoretical frameworks of my three articles and the present cover chapter build upon elements from various strands within the interdisciplinary field of Science and Technology Studies (STS).

What is noticeable in STS is the particular interest in the shaping of technological development. Within this type of technology studies lies the acknowledgement that technology is not 'only' technology; it is shaped by and, thus, also embodies a combination of technical, social, economic and political factors (Bijker et al., 2012, p. xli; Bijker & Law, 1992, p. 3). Authors associated with this perspective perceive technologies as constructed by several actors and not by a single 'mastermind' (Bijker et al., 2012, p. xvii). They also deny

technology determinism, arguing that technologies do not have an inherent independent force that alone drives innovation processes or social change.

Three distinct approaches have been constitutive for this way of thinking: Large Technological Systems (LTS), Social Construction of Technology (SCOT) and Actor-Network Theory (ANT) (Bijker et al., 2012, p. xiv). Thomas Hughes (2012), the founder of LTS, defines technological systems as consisting of both physical and nonphysical components. These components range from artefacts like technical parts or books comprising scientific elements to organisational entities like firms and banks or juridical components such as regulatory frameworks. Within these systems consisting of a 'seamless web' between society and technology, human 'system builders' orchestrate the innovation. Overall, Hughes highlights that such technological systems 'are both socially constructed and society shaping' (Hughes, 2012, p. 45). Authors like Bijker and Pinch, who are associated with SCOT, similarly argue that technologies are socially constructed. At the centre of a SCOT analysis are the social groups for whom the technology in question has meaning (Pinch & Bijker, 2012, pp. 22–23). This could be groups such as users or consumers, policymakers or designers, and it is through the interaction and negotiations among these types of actors that technologies are developed and shaped. Finally, the third approach, ANT, adds another component to what shapes technological developments, namely nonhuman actors comprised of entities of technology and nature. Within ANT, nonhuman actors are perceived as equal to human actors. The important thing is not what or who the actor is, but that it is an entity made to act by others and, likewise, that can make others act (Latour, 2005, p. 46). Simply put, it is the relationships among such actors acting in concert that make up the 'network', and it is this 'working net' of actors that shapes technological developments (see Latour, 2005, pp. 128–133). As Michel Callon (1986b) puts it, networks are the 'mechanism by which the social and the natural worlds progressively take form' (p. 224). This alignment of actors highlights that within ANT investigations, technologies are viewed as constructed but not exclusively by the social (Latour, 2003).

Although these approaches have the constructivist perspective in common and agree that the activity of actors in relationships with others is essential in the shaping of technologies and society, this is also where they part. While LTS and SCOT see technology as socially constructed, focusing respectively on the human actors as 'system builders' or 'social

groups', ANT takes it further, viewing technology as constructed by an 'actor-network', without distinguishing between humans and nonhumans.

During my research process, I became increasingly aware of the role of the relational aspect in the ongoing shaping of the introduction of AI in healthcare; what I was studying was the construction of AI in healthcare based on mobilisations of actors who joined forces during a period of time. The AI technologies played a crucial part in this work or, more precisely, the expectations and claims related to their performances. Even though these technologies were not yet extensively *acting* as technologies in real-world clinical settings, they were brought into the conversation in various ways (e.g., through documents, presentations, statements and conversations) and set in relation to the other actors. Thus, they together constituted 'networks' that shaped the outcome of the processes. As this representation indicates, the main theoretical framework of this cover chapter is inspired by the ANT approach in studies of emerging technologies and technology development.

As I pinpointed in the thesis prologue, I became aware of another aspect quite early in my study; that is, the role of rhetoric or discourse within the construction of technological development. Actors initiating projects in technological development are typically in a position in which they have to work to convince others about their project and its underlying assumptions (Sismondo, 2010, p. 11). In my studies, such actors can be exemplified by the coordination project team that involved different stakeholders in the implementation of the National Health and Hospital Plan 2020–2023 (article 1), the AI vendors trying to convince the procurement team to procure certain AI solutions (article 2) or the KIN network's secretariat working to mobilise various members to increase their ability to influence the decision makers and, thus, contribute to enabling AI in healthcare (article 3). The rhetorical work of these actors includes bringing forth what Konrad et al. (2016) called future-oriented representations, such as articulated visions, expectations or promises. Within STS strands like the Sociology of Expectations, such future-oriented representations are perceived as a way of mobilising necessary resources innovation (Borup et al., 2006, p. 286). As this indicates, rhetorical work is especially crucial in the early phases of emerging technologies, which was evident in my studies, too. This is the reason I also find the dynamics of future-oriented representations, whether in addition to or as part of the relational aspect of technology development, relevant for the exploration of actor mobilisation in this cover chapter.

Dynamics of expectations, technological promises and promissory work as dealt with by STS scholars like Borup et al. (2006), Brown and Michael (2003), Joly (2010), Pollock and Williams (2010) and Van Lente (2000, 2012) came to play a central part in the analysis of articles 1 and 2. Additionally, as I argue in article 2, the work of bringing forward expectations and promises to convince others of a particular project resembles the process of 'problematisation' as described by Callon (1986b), one of the ANT founders. What unfolds during both the inquiry process (article 1), the procurement process (article 2) and the KIN network (article 3) can further be seen as processes of 'translation'. Callon characterises such processes as involving the already mentioned moment of 'problematisation' and the additional moments of 'interessement', 'enrolment' and 'mobilisation'. These concepts constitute what Callon called in 1986 the Sociology of Translation, a label that Bruno Latour has suggested as an alternative label for ANT (Latour, 2005, p. 106).

As this introduction to my choice of theoretical framework indicates, I not only repeat the theory already treated in the articles but focus on the extended part that includes elements from ANT or the Sociology of Translation. To set the scene, I will start with an introduction to future-oriented representations but will keep the focal point on their role in actor mobilisation in line with the scope of this cover chapter. As part of this outline, I will also include relevant or overlapping elements from other strains within STS in order to show how the theory used in the articles is part of a larger landscape. Thereafter, I will turn to the Sociology of Translation as the second and final part of the theory section, introducing the additional analytical concepts that expand on what is already used to answer the research questions of this cover chapter.

3.1   Mobilisation through future-oriented representations

As I highlight in article 1, future-oriented representations, such as certain technological expectations, may play a performative role in innovation processes (Borup et al., 2006, p. 289; Konrad et al., 2016). Expectations can be seen as framing a specific future, usually including a vision of progress or a solution to a problem. According to scholars associated with the Sociology of Expectation, such expectations bring to the present what to expect in the future and what to prepare for in the present, both in terms of opportunities and risks (Borup et al., 2006, p. 286). As such expectations become widely shared, a protected space is provided (Konrad, 2006, p. 439). Consequently, these shared expectations gain formative

strength. They legitimise, guide and coordinate work that is engaged in by different actors during innovation processes (Van Lente, 2012, pp. 773–774) (cf., article 1). Such forces also show that future-oriented representations like expectations have a relational character, not only by temporally connecting the present to the future but also by connecting groups of human actors (Robinson et al., 2021, p. 814). Expectations are drivers for mobilising resources and founding new organisations and networks (Borup et al., 2006, p. 286). As they do so, they bridge boundaries between otherwise separate individuals and organisations and create opportunities for broader coordination (Borup et al., 2006, p. 289).

Another form of future-oriented representations is technological promises. As referred to in article 2, which treats both expectations and promises of specific AI technologies emerging during a procurement process, at least two main ingredients affect the ability of technological promises to convince and mobilise actors (Joly, 2010, p. 4). According to Joly (2010), the first ingredient is legitimacy. Similar to the concept of expectations, promises are typically related to a problem shared by several actors and, at the same time, offer a solution to it. The other ingredient necessary to convince people to believe in the promise is credibility. The degree of the perceived credibility of a promise put forward may depend on who the promise makers are and whether there is convincing evidence supporting the promise makers' claims (Joly, 2010, p. 4).

A relatable perspective can be found in the STS phrasing 'co-production', which was introduced by Bruno Latour in 1990 (p. 147) and later picked up by Sheila Jasanoff using the term as part of the Sociotechnical Imaginaries framework (Jasanoff & Kim, 2015, p. 333). The concept of Sociotechnical Imaginaries is another variation of future-oriented representations. Similar to collectively shared expectations (cf., article 1), communally adopted Sociotechnical Imaginaries are crucial for co-production of technological development as they contribute to setting the agenda for action. In contrast to, for instance, the Sociology of Expectations, which focuses on the anticipatory aspects in investigations of how expectations about the future influence technological development, Sociotechnical Imaginaries are more centred around the values, ideologies and cultural norms embedded in institutionalised and widely-shared beliefs (the Sociotechnical Imaginaries) (Sovacool & Hess, 2017). Even though this framework is not part of any of my three articles, it is included in this theory section as it is relevant to the rationale behind this cover chapter's chosen focal point, namely why it is relevant to look at mobilisations of actors to understand technology

development and change. For this purpose, I find Jasanoff's (2004) explanation of the fundamental reasoning behind 'co-production' interesting as she explains how knowledge (or technology) and the social are two inseparable parts:

> *society cannot function without knowledge any more than knowledge can exist without appropriate social support [...]. It [knowledge] both embeds and is embedded in social practices, identities, norms, conventions, discourses, instruments, and institutions–in short, in all the building blocks of what we term the social. The same can be said even more forcefully of technology* (pp. 2–3).

As an example of how technology is shaped in such a two-way dynamic with the social, Jasanoff and Kim argue that artefacts like cars and computers would never be taken into use if it was not for the interference of different social actors, such as scientists, engineers, public authorities, commercial firms, regulators and users (Jasanoff & Kim, 2015, p. 2). This could sound like a principle similar to what is found within ANT. However, they part at least at one crucial point, namely the aspect of power or agency. In line with a well-established ANT critique, Jasanoff and Kim argue that the symmetrical alignment of humans and nonhumans in ANT, in which responsibility is diffused within the network of actors, leads to a depoliticisation of power 'by making its actions opaque or invisible' (Jasanoff & Kim, 2015, p. 17). In their ANT critique, Jasanoff and Kim underscore that imagination, which can be used as a tool of power and action (cf., expectations, promises and Sociotechnical Imaginaries), is reserved for human actors and institutions, although the implementation of the visions and imaginaries will require a network of several actors (Jasanoff & Kim, 2015, p. 17).

In summary, the frameworks of the Sociology of Expectations, technological promises and Sociotechnical Imaginaries all draw attention to how innovation happens through mobilisation or co-production of different social actors, institutions and practices, with future-oriented representations playing crucial initiating and agenda-setting roles. In short, the analytical focus within these two branches is the discourse that orients the actions of human subjects (Sovacool & Hess, 2017). Conversely, ANT is a 'hybrid' analytical framework applying attention across (and beyond) the more traditional sociological analytical triangle consisting of agency, structure and discourse (Sovacool & Hess, 2017, pp. 733–734). As Sovacool and Hess (2017) remarked, 'ANT examines the facts, machines, people and

bureaucracies that must be aligned, molded and disciplined to create technological development and acceptance' (p. 720).

3.2 Mobilisation through moments of problematisation, interessement and enrolment

What happens between invention (e.g., the initial ideas, projects or prototypes) and innovation when the invention is put in relation to users can often appear as 'a fate played out in accordance with a mysterious script' (Akrich et al., 2002a, p. 188). Through his article on the domestication of scallops and fishermen of St Brieuc Bay, Callon (1986b) offers an analytical framework to reveal such concealed scripts. Within this particular work, Callon shows how certain actors and their capacities can attract other actors, whether they are humans or nonhumans, to comply with their goals and potentially continue towards innovation (1986b, p. 201). He does so by paying attention to the establishment of relationships and the interaction occurring between actors relevant to the specific project. In this context, Callon describes a process characterised by the aforementioned four overlapping moments of 'problematisation', 'interessement', 'enrolment' and 'mobilisation'.

The point of departure for Callon's story is a steadily decreasing population of scallops in this French bay and a particular project initiated by three researchers attempting to restock the bay with scallops by involving both human actors (e.g., fishermen and scientific colleagues) and nonhuman actors (e.g., scallops, predators and collectors made for sheltering the scallops). From here, Callon examines how the project is shaped through translations happening during a 'simultaneous production of knowledge and construction of a network of relationships in which social and natural entities mutually control who they are and what they want' (Callon, 1986b, p. 203; for related work, see also, for example, Callon [1986a], Latour [1987, 2005] and Law [1991]).

The moment of problematisation concerns how something is made or becomes indispensable. In Callon's conceptual language, the idea or assumptions behind a project, such as the one of AI technologies being solutions to challenges in the healthcare services, are put forward by certain actors with a particular agency within the project (e.g., governments and AI vendors). As their ground assumptions are shared, these become an 'obligatory passage point', meaning that if, for instance, AI technologies do not perform according to the expectations, the project

will fail as the 'obligatory passage point' is proven weak or incorrect. Thus, for actors to become mobilised, they must be convinced of the project's fundamental assumption. Callon (1986b, p. 204) calls such work that is carried out to convince others of a project's rationale 'problematisation'. It is also in these contexts that future-oriented representations have a role to play.

In moments of problematisation, each actor relevant for a project to proceed (e.g., AI technologies, physicians and hospital managers) is identified and defined based on their characteristics and the problems necessary for them to overcome to reach their goals related to the project's aim. Based on such knowledge, the actors behind the project can 'translate' or adjust the presentations of the project according to the different actors' identities, relating it to their problems and thereby properly convincing them of its relevance. However, there is an 'obstacle' concerning the actors identified through moments of problematisation. As actors are real, and reality is a process, what is mobilised may change (Callon, 1986b, p. 207). Therefore, the mobilisation that happens during problematisation can never be solid or definite.

During moments of interessement, the actors who aim to keep the others interested in their project take a set of actions to impose and stabilise these other actors' identities as defined through the problematisation. These actions include the use of so-called 'trapping devices', which could be in the shape of different types of texts, visuals or verbal articulations. The actions of interessement are also performed to avoid interference by other actors who could 'threaten' the project by otherwise defining the actors' identities (Callon, 1986b, p. 208). For instance a project aiming to introduce AI in healthcare can define the properties and identity of AI technologies according to the interests of actors such as physicians or hospital managers.

Through such actions, AI becomes a 'result' of the associations that link the idea of AI or specific AI technologies to the other actors. This strategic work to capture the actors' interests and convince them to stay as allies in the network can also be viewed as a way of creating a favourable balance of power for those with a strong agency within the project (Callon, 1986b, p. 211). Finally, a successful result of the interessement will confirm the validity of the problematisation and the alliances between actors that the problematisation implies (Callon, 1986b, pp. 209–210).

Even though the process of interessement shows positive achievements in building relationships between actors and has convinced crucial actors to become or stay interested in a project, it does not always lead to successful alliances between the actors. There is no guarantee that all the essential actors will let themselves be enrolled as parts of the project (Callon, 1986b, p. 211). Regardless of how convincing the arguments to involve other actors are, the result is not necessarily a unison acceptance by all relevant parties. The moment of enrolment is about the different negotiations between actors that occur during moments of interessement, and how the strength of these negotiations is tested and, finally and potentially, leads to success: enrolment.

However, during the unfolding negotiations in which relationships between actors are potentially tied and untied, relevant transactions can also be non-existent or weak. In the case of AI in healthcare, this could be exemplified by the potential for transactions between different physicians and their representatives who are involved in the project. The physicians mobilised (e.g., those who identify with the problematisation and attend discussions and meetings) do not represent every type of physician, medical expertise or clinical work practice. Thus, those who do not attend or are not properly represented in the conversations will be enrolled through their representatives without any resistance (Callon, 1986b, pp. 213–214). This issue leads up to the last moment of the Sociology of Translation, namely 'mobilisation'.

According to Callon, to succeed with the project at hand, the question of representatives – who speaks in the name of whom – is essential. This also includes paying attention to the distinction between representatives and spokespersons, with the latter as the outcome of actors that have formed alliances and act as a mobilised unit of force. Hence, moving from representatives to spokespersons is a part of the process Callon calls 'mobilisation'. As touched upon, mobilisation happens through all of the outlined moments in which the entities that were not mobile beforehand are made mobile (Callon, 1986b, p. 216).

Mobilisation happens through different displacements or actions of transformation (=translations), as exemplified: AI technologies are transformed into visions, expectations and promises. In turn, these are transformed into forms like texts or other types of expressions. Through such displacements, actors are mobilised and participate in the

negotiations happening within the network of actors. If the relevant actors act as it was hoped and relationships between them are established, the mobilisation is successful, and the enrolment is transformed into active support (Callon, 1986b, p. 218).

3.3 Fragile mobilisations

Through Callon's four moments, a consensus about the project at hand might be achieved and a constraining network of relationships established. However, such consensus and alliances might be fragile, as they can be quickly challenged by new translations or unforeseen circumstances (e.g., as offers are launched during a process, including other technological options, see article 2) (Callon, 1986b, pp. 218–219). In the case of AI technologies, this can be exemplified by IBM's Watson for Oncology. After great hype, the technology was abandoned or sold by IBM as the promises of its performance did not materialise in real-world clinical settings (Miliard, 2022). If AI technologies fail to perform as expected, it might result in the technologies being detached from the project, which, in turn, might bring the representativity into question. If a technology is promoted within a network of actors based on a successful performance in one context but later performs less advantageously, this will result in a controversy over representativity (Callon, 1986b, p. 220).

As the fundamental assumption of a project can change with such controversies, so can the different actors' identities and characteristics. The actors who are involved can suddenly distance themselves from the project or the alliance and, thereby, distance themselves from the 'obligatory passage points'. Such a move may cause changes to the existing network of human and nonhuman actors and further alter the network's description (Callon, 1986b, p. 224). This again underscores that during a process of innovation or problematisation, interessement, enrolment and mobilisation, anything can happen.

4   Methodology, methods and materials

As I mentioned in the prologue of this thesis, my PhD project took quite a turn during the first year. The same did the world around us. After two weeks as a PhD student, the first lockdown of Norwegian society due to the COVID-19 pandemic occurred, and we were set to work from home for an indefinite time to minimise the spread of the virus. Many doors closed in the following period, and people lost important pillars in life (loved ones, jobs,

social life, income, sanity etc.). In academia, researchers lost their access to study fields, the possibility of face-to-face discussions with peers and in-person presentations at conferences. At the same time, other doors opened as most parts of the world adjusted to the new circumstances. Suddenly a greater part of our work life happened in the digital sphere. We somehow became used to participating in online meetings, seminars and conferences conducted on platforms such as Zoom and Teams.

The research design of my doctoral project took shape in parallel with these disruptions, adjustments and developments in our society and our new work–life habits, including our enhanced use of information technology, especially for communication. As I entered the field of AI in healthcare, opportunities for collecting data in ways I had not considered before emerged, both in terms of study objects and methods. Through the Internet and digital communication technology (Zoom, Teams etc.), I started recalibrating my research project by accessing the field through new as well as digital tools and, most importantly, with an open and explorative approach.

For almost a year, I carried out pre-fieldwork preparation to better understand the current state of the field of AI in healthcare. As I did, I focused on elements that could shed light on issues such as what the ongoing work to introduce AI in healthcare was about and, in a Norwegian context, who the key actors were. Through this work, I found my way from my early articulations of the research area of interest (the introduction of AI in the Norwegian public healthcare services) and of questions related to the kind of projects that were organised, who was involved and who seemed to be in charge to research problems with more distinct research questions for further investigation and development (cf., the three articles). This is similar to what Hammersley and Atkinson (2019, pp. 26–27) emphasised as an essential part of the initial phases of ethnographic fieldwork.

Through this process, I identified the three cases. These appeared as key AI initiatives within the Norwegian public healthcare sector conducted outside the borders of research projects, which altogether matched what I was looking for: non-research projects with a certain or, at least, potential importance for and influence on the ongoing AI introduction in the Norwegian healthcare services. Especially at the beginning, the activities of the three initiatives and the access I gained were clearly shaped by the ongoing pandemic and the fact that people worked from home. This, in turn, influenced how my research design developed. It remained a methodology inspired by ethnographic fieldwork but with another aim than that of my initial

idea of studying the deployment and use of AI in real-world clinical settings and without the same degree of participation.

4.1   Research design

This thesis includes three studies based on a qualitative, ethnography-inspired case study research design. I chose to do case studies as these are beneficial when the purpose is to understand a less explored field, when you as a researcher have little control over the developments of the events and when the interaction between actors is a focal point (Fangen, 2010, p. 187). Through this approach, I also acquired a thorough and comprehensive understanding of the complexities of the initiatives because they were carried out within their natural context (Crowe et al., 2011). By conducting three case studies, I was also able to draw a bigger picture of the current state of AI in healthcare as I studied different processes and actor constellations, each with varying goals and outcomes shaping the introduction of AI in the Norwegian healthcare services. This approach corresponds well with what Hyysalo (2021) noted concerning how emerging technologies typically come about: 'New technologies are shaped in multiple interlinked settings and processes and characterized by high contingencies and many different choices' (p. 15).

For instance, by carrying out multiple case studies, I discovered not only how 'top-down' initiated processes, such as the inquiry process, have certain formative effects on the introduction of AI. Through the study of the procurement process and the KIN network, I could also bring forth a better understanding of why such forces are problematic and insufficient to fulfil the formal expectations; expectations change and cause changes, and further explorations of challenging issues and uncertainties have to be made.

More concretely, the doctoral project is based on data from nonparticipant observations of digital meetings, semi-structured interviews and document analysis. All three studies included meeting observations. In the study of the inquiry process, observation was combined with document analysis, while the studies of the procurement process and the KIN network included semi-structured interviews. Figure 1 below illustrates the course of my overall research process during my three-year PhD project. The first year started with pre-fieldwork preparation, and approximately 11 months later, the data collection process of my first two cases took over. After five months, study 1 ended along with the project studied. Three

months later, I started following the case of study 2, which I completed approximately a half year before study 3 was finalised.

Figure 1. Timeline of PhD project and overall research process and design
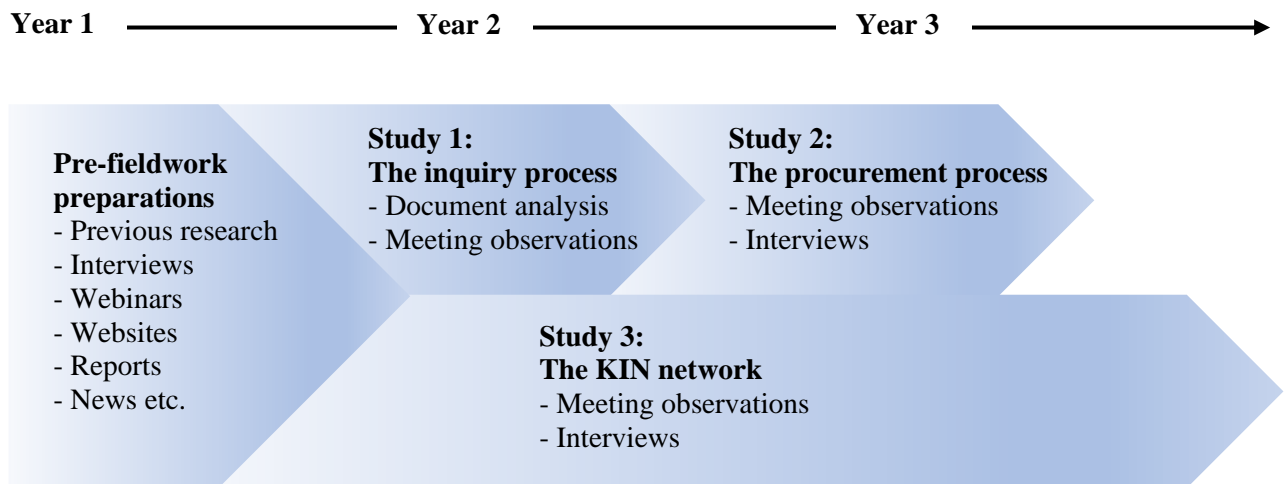
In section 4.2, I will start by expanding on the pre-fieldwork preparations. Thereafter, I will elaborate on the activities carried out within the three cases, the course of my parallel research process, the methods used, some limitations and issues concerning the scientific quality of the work and research ethics.

4.2   Pre-fieldwork preparations

During the first year of my PhD project, I conducted pre-fieldwork preparations. I started by exploratively approaching the field of AI in the Norwegian healthcare services to better understand what was taking place and to sample cases. As a part of this work, I conducted eight semi-structured background interviews, I listened to relevant podcasts and read previous research, newspaper articles, website information and reports. I also joined webinars and online conferences on the topic.

The people I interviewed during this phase were either doing work related to AI in healthcare or had a certain knowledge of or interest in the field. I recruited the informants as I came across their names in relation to AI projects I read about, through seminars I attended where they made presentations or after suggestions from other informants. Many of the informants came from the field of radiology, which is an area in which AI development and deployment have come quite far. Additionally, radiology was a focus area for some of the larger AI

healthcare projects in Norway (cf., both the inquiry process and the procurement process in this thesis). However, the particular selection of informants may also be reinforced as some of the informants were recruited using the technique of 'snowball sampling' (Vogt, 2005, p. 300) during the interviews, asking the informants for other potential people with whom to talk. The following informants were interviewed in the order listed:

- A communication officer from an AI research project at a Norwegian public hospital
- A division manager from a technology company developing AI technology for radiology
- A senior consultant radiologist and leader of the Norwegian Society of Radiology with knowledge and interest in AI
- A radiologist at a private image diagnostics company with experience in testing and using an AI technology
- A team leader of a research group within radiology and AI at a large hospital
- A neuroradiologist with knowledge of AI
- A radiologist and developer of AI algorithms
- A project manager of a planned procurement process of AI for image diagnostics at a Norwegian hospital trust (which later became one of the case studies)

As I was exploring the field to see what was deemed important to the people inhabiting the world I entered, my interviews were mostly made up of a few open-ended questions (see appendix A for an example of an interview guide). I asked them to tell me about their work and their experiences and/or interests related to AI in healthcare, typically followed by questions formulated based on what they told me and what was relevant to my research agenda. Thus, the interviews took a reflexive form (Hammersley & Atkinson, 2019, p. 122). Other questions asked was: What do you think about the future with AI, and what do you see as crucial for enabling AI in the Norwegian public healthcare services?

During this initial research phase, I learned more about the actual status of introducing AI in the Norwegian healthcare services and which topics seemed to be important to the people inhabiting the field. These investigations confirmed my initial impression of a field consisting of much hype, high hopes, expectations, claims and promises, with limited demonstrations and observable use of AI technologies in clinical settings. At the same time, I got an overview of organisations, people and projects that seemed central in the field as they were playing an important role in the ongoing work to enable AI in healthcare. From this insight, I

contacted persons with managerial positions in three initiatives that appeared to be some of the most significant in the Norwegian healthcare services (the inquiry process, the procurement project and the KIN network). These initiatives were referenced during the interviews I conducted, the texts I read and/or during the seminars and conferences I attended. They were the first of their kind in Norway, they were either in the early stages or about to start, they were of a certain size and they represented different types of approaches to enabling AI in healthcare. Fortunately, they all welcomed me as a researcher. Through these initiatives, I was given the opportunity to explore the process of introducing AI in healthcare 'in the making' through different projects and over different time spans.

4.3   Ethnography-inspired approach

In ethnographic fieldwork, several methods and modes of participation can be used during the study depending on the location explored and how 'deep' into a site it is relevant or possible to go (Marcus, 1995, p. 108). As meetings needed to be held digitally due to COVID-19, these events became the core activities of the initiatives I studied. Thus, they came to be of great importance for my data collection. In short, without the digital meetings, neither the initiatives nor my fieldwork would have been feasible to carry out. By using Zoom and Teams, meetings could typically be conducted with people working remotely due to COVID-19 or situated in different places in Norway or other countries. During these activities, the actors I was interested in observing and listening to came together, and their perspectives on AI in healthcare became possible to investigate in relation to each other. Thus, my main access to data was through digital meetings because it was often not possible to conduct in-person meetings.

In addition to observing these meetings, I conducted document analysis and interviews (also mainly through Zoom or Teams), all of which were feasible methods to use according to the circumstances. The following subsection presents the three empirical cases, with a focus on the activities executed within them and my parallel research activities.

4.4   Case details

When cases for investigation are sampled, additional sampling usually has to be done along dimensions such as contexts, time and people (Hammersley & Atkinson, 2019, p. 38). In my three cases, there were already a set of frames that limited the need for making decisions on elements such as where to observe, when to observe, whom to observe, with whom to talk

and how to record and document what was observed or heard during the fieldwork. As my access to data was primarily gained through digital meetings organised by the teams in charge of the initiatives I followed, I had no influence on when the meetings were set, who was invited to participate or the topics discussed. It also was not possible for me to ask questions for research purposes during the meetings, as this would have interfered with the conduct of the meetings and affected the outcomes in ways not in harmony with the organisers' intentions. However, I was in charge of deciding which documents to analyse for what purpose, whom to interview and which questions to ask.

4.4.1 Study 1

The inquiry process organised by the national coordination project established through the National Health and Hospital Plan (2020-2023) was completed over a period of five months, from January to May 2021. The approximately 100 stakeholders who participated included hospital managers, (AI) project managers, physicians (mainly radiologists), patient organisations, procurement officers, AI researchers and AI vendors. These stakeholders were involved through three types of digital meetings (insight meetings, workshops and feedback meetings) organised in three phases with 11 meetings in total (see article 1 for more details).

I attended all of the meetings except for the first three insight meetings because I was enrolled as an observing researcher after their completion. My main research activity during the meetings was paying attention to what was said and by which type of stakeholder and taking notes. I also read the material produced by the project team during the process and analysed the policy document initiating it all.

4.4.2 Study 2

The aim of the procurement process was to buy 1–6 commercial AI technologies for image diagnostics, and I followed it for a year and as long as the procurement process lasted. Before the proper dialogue started, the procurement project had carried out a pre-phase project, which included workshops with various types of radiologists working within the hospital trust to identify areas of work that could benefit from the introduction of AI. The project team had analysed the market of commercially available AI technologies for image diagnostics, sent out an invitation to tender and selected five vendors that qualified for participation in the dialogue process. I was enrolled as a researcher in time to attend the first of 14 dialogue

meetings between the procurement team and the five prequalified vendors (see article 2 for more details).

Along with one or two other researchers from a project at the Norwegian Centre for E-health Research, I attended the meetings as an observing researcher. The meetings were completed after three months, and the other researchers and I (forming a research group of 5-6 persons) then started having one-hour status meetings with the project manager every third week to receive updates on the project, discuss what we had seen and heard and so on. As a part of my (individual) study of this procurement process, I also conducted six interviews. I interviewed the hospital trust's project manager both before and after the dialogue meetings with the vendors, while the last four interviews were with the remaining vendors before they handed in their final offer and, thus, before the winner of the competition was announced.

4.4.3   Study 3

The KIN network consists of professionals working with or interested in AI and healthcare, and I followed it for two years. Contrary to the other initiatives, there was neither an end date nor an explicitly defined final result associated with this initiative. The aim was to facilitate discussions, knowledge and experience sharing across disciplines, organisations and institutions for as long as it was deemed relevant. Thus, I stopped collecting data as I approached the end of my PhD even though the network continued to exist. The network's main activities are quarterly digital meetings and three to four yearly seminars or conferences, in person if possible (see article 3 for more details).

My main research activity was observations of the network's digital events (meetings and seminars) and two in-person conferences. The observations of the digital events did not involve active participation, while the observations of the in-person conferences offered some opportunities for interaction. In the periods between the meetings, I went through my fieldnotes and the PowerPoint presentations for the meetings, which sometimes included notes taken by the organisers and thus acted as meeting minutes. Finally, as my data collection process was coming to an end, I interviewed the secretariat members that had been a part of the network from the beginning, which included six out of eight people.

4.5 Methods

As touched upon, my research design was shaped by the circumstances caused by the COVID-19 pandemic, which dictated that the main access for gaining data was through the Internet and platforms such as Zoom and Teams. These restrictions meant that there were methods I could not use, such as in-person participant observation. At the same time, the chosen methods had to be executed in certain ways. Thus, to better explore the field and adjust to the circumstances, I used a combination of methods during the case studies. These methods are presented in the following subsections in the same order as they first were conducted: nonparticipant digital meeting observations, document analysis and interviews.

4.5.1 Nonparticipant digital meeting observations

The most widespread activity of all three cases was digital meetings. Thus, attending meetings for data collection became what I did the most, which included 32 meetings totalling 133 hours. The meetings varied in length, aim and form (see the method sections of the three articles for more details). The meetings observed had in common that there were certain agendas to follow, goals to achieve, a time to keep and attendees expecting to contribute according to the meeting agenda. Thus, my role as an observer was restrained to being a nonparticipant. Apart from introducing myself and my research project at the beginning of the meetings to make people aware of my presence and inform them of my research project, I usually kept my camera and microphone turned off, making as little out of my presence as possible to avoid disturbing the intended and expected course of the meeting. Even though it can be hard to understand what is going on in a conversation in which you do not take part, nonparticipant observation can be advantageous in situations with a formal and fixed structure, such as the meetings I observed (Fangen, 2010, pp. 77–79).

As I started observing the meetings, I wanted to document different types of interaction. Apart from the dialogues, I wanted to keep an eye on people's facial expressions (e.g., if they seemed to disagree or were especially eager). I also imagined that I could observe their use of emojis enabled by the digital platforms during the meetings, such as clapping hands or thumbs-ups, when other people talked. However, I was surprised by the lack of such interaction and how little I was able to perceive as an observer through the digital meeting platforms' interface. During the meetings, there was typically a presentation that took up much of the space on the screen, only allowing a view of a few of the meeting participants'

faces and some of the upper bodies as thumbnails. Additionally, there was typically a certain number of people at most of the meetings who had their cameras off, either due to network issues or other unmentioned reasons, leaving a black square facing the other meeting participants. Lampa et al. (2021) observed similar issues in their observations of digital meetings. For instance, they saw that the digital meeting format resulted in nonverbal cues often going under the radar. Such cues would otherwise be interpretable by other meeting participants as indications of certain feelings, opinions or wishes to speak. In my case, it also turned out that as a single researcher occupied with taking notes and documenting the themes discussed, I had little time left to pay attention to what else might or might not be happening on the screen. This would probably have been different if I had been able to record the meetings and transcribe them afterwards instead of taking notes while observing.

As with much fieldwork, recording was not an option. First of all, getting consent from all meeting participants to make recordings was more or less impossible, as neither the organisers nor I fully knew who the attendees would be. Additionally, some people attended the meetings after they had started, which made a heads-up during the meeting introduction insufficient. Another issue was whether knowledge that I was making recordings would keep people from actively participating in the meetings, Finally, and especially for the procurement process, certain elements discussed during the meetings were confidential in relation to competitors, making it important for all parties to feel 'safe' to speak and avoid information leaks.

Despite the limitations, as my data collection proceeded and my early research ideas and questions changed into more concrete research problems, I became aware that my main research interest lay within the content of or came to the surface through the verbal communication taking place. This also meant that seeing people became less important. I did not use an observation protocol but took as many notes as possible while focusing on the verbal exchanges between the participants, and naturally, I missed some details of the conversations at each meeting. However, similar to Thedvall's experience (2013, p. 112), I gained a fuller understanding of the discussions over time as I attended several meetings. This nevertheless left at least two questions regarding whether my data would be different under other circumstances: Would the discussions unfolding during the digital meetings have different outcomes if they took place as in-person meetings? What would the data I collected

look like if I could do participant observations and ask questions during the discussions or in the breaks for further elaboration on certain topics?

As I combined the nonparticipant meeting observations with methods such as interviews, I had an opportunity to ask further questions about elements from the observations. For articles 2 and 3, the explorative approach taken during the observations (taking notes of most of what was said without using an observation protocol to see what was in the 'landscape' before closing in on it) enabled me to discover what I wanted to investigate further through interviews and with whom.

4.5.2   Document analysis

A document is something material (digital or physical) that is designed for a particular reason. It relates to something outside of itself and functions as a connection to the surrounding world (Asdal & Reinertsen, 2020, p. 15). For instance, policy documents can be seen as instruments of power used by governments, corporations and other organisations to govern spaces, people and ideas. They can tell us something about the paths that are open and those that are closed (Nyqvist, 2013, p. 97). Thus, policy documents can be viewed as ideological vehicles that advance 'a preferred and often idealised vision aimed at governing and impacting in certain directions' (Nyqvist, 2013, p. 92). In the case of my research project, document analysis was used in study 1 to investigate how the Norwegian Government's policy plan for healthcare services (the National Health and Hospital Plan [2020-2023]; NHSP) presented the future with AI in healthcare and the goals they aimed to achieve in this regard. It further exposed how certain areas were to be prioritised, thus shaping the trajectory of the processes employed in the healthcare services, ranging from relevant authorities to local hospitals.

As Nyqvist similarly emphasised, documents reach out and touch the actors they pass. They connect different organisations and bridge gaps between them, and they create 'interconnectivity across large-scale arenas' (Nyqvist, 2013, p. 92). From this point of view, policy documents are structural artefacts that can be intriguing objects of study. As my study shows, the inquiry process led by the Norwegian Directorate of Health was initiated by a policy plan (the NHSP) that had certain formative effects on the outcome of the process (cf., article 1). An example of how the plan further affected the field was the fact that the AI

procurement process studied (study 2) was prioritised above other projects applying for funding from the regional healthcare authorities. The procurement project was also one of three Norwegian projects to be supported by the Norwegian Directorate of Health and other national public agencies on legislation and data protection issues.

4.5.3 Digital semi-structured interviews

During my research project, I used interviews for two different purposes. As previously mentioned, I conducted eight semi-structured interviews during the first year as part of the pre-fieldwork preparation. Additionally, in the study of the procurement process and the KIN network, semi-structured interviews were used to supplement and dig further into areas relevant to the research problems and the research questions that emerged during the nonparticipant meeting observations. The interview guide used during these interviews consisted of overall topics for discussion with a selection of optional questions (Kvale & Brinkmann, 2009, p. 130). The questions were typically open-ended, and I followed up the informants' answers with new questions, not necessarily based on the guide but aiming to keep the conversation about topics relevant to the research problem. Hence, the interviews were not entirely explorative. The interviews can also be characterised as elite interviews; the informants were all managers or experts and knowledgeable about the topics discussed. An advantage of this type of interview is that the power balance between the informants and the interviewer is more symmetric than in other cases (Kvale & Brinkmann, 2009, p. 147).

The informants interviewed during the study of the procurement process were:
- Four representatives from the participating AI vendors, all in senior positions and working with AI and sales
- The project manager of the procurement process

The informants interviewed during the study of the KIN network were:
- a researcher from a research department at a private limited company
- a department manager from a public agency
- two managers from a national research centre
- a senior adviser from one of the four Norwegian regional health authorities
- a department manager from one of the four Norwegian regional health authorities

For more details on the interviews, such as their main topics and findings, see articles 2 and 3. For examples of typical 'interview guides' for each type of interview, see appendix B, C and D.

4.5.4   Methods summary

A triangulation of methods, can be advantageous in case studies and for the purpose of understanding complex phenomena (Fangen, 2010, p. 187). As I had limited access to the field apart from the meetings to which I was invited after reaching out to the organising project teams, interviews became a crucial additional source for in-depth exploration of some of the observations and to gain a better understanding of the field. The same counted for the document analysis and the less structured and systematic reading of documents serving as supplementary information, such as reports, newspaper articles and procurement documents (Fangen, 2010, p. 144). Before I summarise my articles and analyse and discuss their findings as part of answering the research question of this cover chapter, I will briefly give an overview of the study's data material and how it is analysed, followed by elaborations on how I have worked to ensure the scientific quality of my work and my considerations regarding research ethics.

4.6   Data material and analysis

The length of time of my three studies varied from five months (study 1) to one year (study 2) to two years (study 3) (cf., figure 1, p. 35). During this time, several meetings, interviews and the reading of documents occurred. Table 1 below provides an overview of the material upon which my three studies are based, with the approximate number of activities and the material these involved or resulted in.

Table 1. Volume: meetings, fieldnotes, interviews, transcriptions (all three studies)

	Study 1	Study 2	Study 3	Total
Number of meetings and approx. duration	8 meetings 17 h 30 m	14 meetings 58 h 30 m	10 meetings (incl. conferences) 57 h	32 meetings 133 h
Fieldnotes	47 p.	77 p.	70 p.	194 p.
Documents analysed	5 documents 152 p.	-	-	5 documents 152 p.
Interviews and approx. duration	-	6 interviews 6 h	6 interviews 6 h	12 interviews 12 h
Transcriptions (interviews)	-	64 p.	62 p.	126 p.
Meetings with project managers	3 meetings approx. 2 h	5 + meetings approx. 5 h		8 + meetings approx. 7 h

To analyse the NHSP policy plan in study 1, I used a version of the technique of 'directed content analysis' (Hsieh & Shannon, 2005, p. 1281). It was a structured process in which the document was analysed for content specifically matching concepts from the article's theoretical framework. The fieldnotes from the observed meetings in all three studies and the transcribed interviews from studies 2 and 3 were first analysed with open coding; they were read line-by-line while different themes were identified. These themes were later coded and grouped as the text was closely reread (Emerson et al., 2011, p. 172). Through such a process, significant observations relevant to each study's research problem could be distinguished from the rest of the material (Emerson et al., 2011, p. 175). Secondary material included in the studies, such as meeting summaries, reports and website information, were also read line-by-line for information to supplement the findings from the observations and interviews (for more details on the data analysis procedure of each study, see the three articles).

4.7 Limitations

The fact that most of the activities took place through digital platforms gave me a unique form of access to almost everything that occurred during the processes that were followed. It was easy to obtain both an overview of the activities and access; I never received a 'no' from the organisers to my requests to participate in the activities as an observing researcher. However, there were also less fortunate aspects of conducting this kind of research. For instance, when I analysed my fieldnotes from the meetings, I found that they were at times quite 'thin' as the conversations captured were not necessarily rich in terms of information or content. I believe this was related to the fact that I only captured the verbal communication unfolding during the meetings without the opportunity to ask questions along the way. Apart from occasionally noting a certain tone in the voices of a few of the meeting participants, it was hard to capture nonverbal signs indicating, for example, frustration or disagreements. Perhaps more importantly, I suspected that the engagement among the participants was marked by the digital meeting format. As I experienced myself during digital meetings I attended, the platform affected how active I was as a meeting participant. This made me question the quality of my findings: Were the participants distracted by something else during the meetings, did they hold something back and did they avoid going into complex explanations or difficult topics online – was it easier just to drop it? What about becoming aware of conflicts as an observer, as meeting participants could, for instance, silently switch off the camera instead of slamming the door? In relation to the studies of the procurement process and the KIN network, I conducted interviews with key participants that made it possible to ask follow-up and in-depth questions. However, immediate follow-up interviews after the meetings that could mainly be focused on the content of the particular meeting(s) might have given better and more nuanced data, especially in relation to the inquiry process that was only based on observations and document analysis. To mitigate this weakness in my studies, I aimed for another way of ensuring a certain quality to my findings, which I will outline in section 4.8.

4.8 Scientific quality

The philosopher Isabelle Stengers argues that aiming to risk one's research result is crucial for scientific knowledge to achieve reliability (Stengers, 1999). By taking risks, she implies that researchers must deliberately seek to test their findings by opening their results to refutation by other interested parties. As this happens, the research result can either be

strengthened or weakened (Stengers, 1999, p. 69). Stengers' criterion of risk-taking can to a certain degree be compared to Popper's principle of falsification, which concerns the process of falsifying hypotheses or claims through tests and experiments. However, as Stengers notes, this principle is too narrow as it distinguishes between sciences and non-sciences based on whether they include theories and hypotheses that can be refuted through observations and experiments (Crotty, 1998, p. 32). For Stengers (1999), such a perspective is too narrow as it excludes sciences, such as history and social sciences, that do not necessarily study 'situations where what is predicted can be compared to what occurs' (p. 13, my translation). For instance, when social scientists study complex societal situations, one cannot speak of a contradiction simply because the prediction does not turn out to be the case (Stengers, 1999, p. 13). Stenger's criterion remedies this problem for sciences that are excluded because the results cannot necessarily be refuted. She argues that if a researcher takes other practices and interests into account and is open for counterarguments through risk taking, the knowledge produced can achieve greater reliability.

Inspired by this line of thinking, I tried to strengthen the reliability of my data, not by 'risking' the findings but more mildly 'testing' them through different activities during my research process. This testing included presentations of diverse selections of my findings to different kinds of audiences in different forms, including:

- Presentation of findings at a digital meeting with members of the team organising the inquiry process (study 1).
- Informal sharing and discussions of findings with the various informants during the interviews (studies 2 and 3) and the project manager of the procurement process during our frequent status meetings (study 2).
- Presentations at conferences with peers both from STS and the field of AI.
- Presentation of findings from all three studies at the last conference I observed, which was organised by the KIN network (November 2022). Participants from each case attended the conference, and the audience had the opportunity to comment and ask questions. I stayed for the whole conference and talked to people during the breaks.
- Writing an article-based thesis which implied submitting three empirical research articles for peer review in academic journals.

Even though I never experienced comments that directly refuted or contradicted my findings, I experienced comments that both nuanced and confirmed them. This way of testing and

simultaneously sharing my findings with the field itself also tells a story of my situatedness in the field. I contributed with knowledge that the field itself could use in the continued work of introducing AI in healthcare. Thus, I not only took on the role of an observing or interviewing researcher with or without a face on the computer screen but also that of a researcher with an ethical obligation to involve others and share knowledge with the field and scientific community (Norwegian National Committee for Research Ethics in the Social Sciences and the Humanities [NESH], 2022).

4.9   Research ethics

My doctoral project follows the research ethics guidelines developed by the Norwegian National Committee for Research Ethics in the Social Sciences and the Humanities (NESH, 2022). Among several important topics ranging from obligations towards peers in the research community to protection of vulnerable groups in society, I find five aspects to be especially relevant to address in this cover chapter: consent to participate in research, anonymity, confidentiality, data storage and respecting private companies' needs for anonymity.

The data collected during the meeting observations did not include information that could compromise the meeting participants' privacy. The participants were typically informed at the beginning of each meeting about my participation as an observing researcher. Each interview informant was given an information and consent letter to sign before the interview. Apart from giving information about the research project and the overall topics for the interviews, the consent form also underscored that the participation was voluntary and that consent could be withdrawn without giving any reason or having any negative consequences. The informant was also informed that quotes from interviews could be used in publications and presented at conferences but would in all cases be anonymised. If there was a chance that the informant would be recognised through the material planned to be presented or published, the informant would be asked to approve the use of the quote (for more details, see an example of information letter and consent form in appendix E). Data collected that included information about the participants were stored in a secure and legal manner at the Oslo Metropolitan University's OneDrive cloud service requiring Feide, a solution for secure identification within the Norwegian education sector. Regarding companies' need for anonymity, I also signed a declaration of confidentiality as I was enrolled as a researcher in

the procurement process (study 2) due to the competition aspect of the procurement process and the private interests of the companies involved.

Finally, a description of my doctoral project, including the information and consent form, my anonymisation procedures and data storage plan, was approved by the Norwegian Centre for Research Data (NSD), now called Data Protection Services for Research provided by the Norwegian Agency for Shared Services in Education and Research (SIKT). See appendix F for NSD's ethics approval.

5 Article summaries

This section presents summaries of the three articles, with an emphasis on research problems, research questions, findings and contributions to the thesis as a whole. Further elaboration of how these results contribute to the research question of this cover chapter is found in section 6 'Analysis and discussion'.

5.1 Article 1: 'Contesting futures of Artificial Intelligence (AI) in healthcare'

Kannelønning, M. S. (2023). Contesting futures of artificial intelligence (AI) in healthcare: Formal expectations meet informal anticipations. *Technology Analysis & Strategic Management*, 1–12. https://doi.org/10.1080/09537325.2023.2226243

Status: Published

5.1.1 Research problem and position of the article

Governmental policies are designed to be performative. The visions and expectations within them initiate activities and guide and mobilise actors to contribute to their implementation (see, for example, Brown et al., 2000; Konrad et al., 2016). In the case of AI in healthcare, whether or how the expectations of AI will materialise and what the implications of widespread use will be is still unknown. Thus, such formal expectations can be seen as not only powerful but also problematic. With their formative strength, elements disputing them can easily be ignored, potentially leading to costly detours or even severe and irreversible societal consequences.

With this research problem as a point of departure, I used the inquiry process initiated through the Norwegian National Health and Hospital Plan (NHSP) 2020–2023 as a study site. Based on data from document analysis and nonparticipant meeting observations, I explored how government policies set the agenda for introducing AI in healthcare and how the

expectations included in the policy were disputed by more informal anticipations of stakeholders involved in the inquiry process.

5.1.2 Findings

The findings in article 1 show how the formal expectations included in the NHSP were contested by specific informal anticipations of actors operating closer to clinical practices. I see this as an example of how even strong and widely-shared expectations will meet resistance during their deployment. However, despite the contestations, the formal expectations continued to have strong formative effects showing elements of inevitability and path dependencies. Alternative pathways, including investigation of the contesting elements, were neither thoroughly addressed in the final report of the inquiry process nor considered as prioritised areas for future initiatives. This, in turn, raises questions about the fulfilment of the Norwegian Government's vision in the long run as there have been no attempts in the present to mitigate elements potentially challenging the future with AI as brought forward by those working closer to clinical practices. Based on this, a paradox emerged in the early phases of introducing AI in the Norwegian healthcare services, as alluded to in the following quote from one of the meeting participants: 'People in the clinical practices know that something is coming; they just don't know, yet, how it will hit them'.

The paradox arises between the expectations ('something is coming'), the AI chasm, including uncertainties of what might come ('how it will hit them'), and the formative forces of the formal expectations leading to many of the questions concerning the future stay unexplored. This includes questions such as how AI technologies will perform in real-world clinical settings and what a broad introduction of AI in healthcare will imply both in the short term and long term for physicians and patients, the healthcare services and society at large.

5.1.3 Contribution to the thesis

The study of the inquiry process was the first study I conducted and finalised during my doctoral project, and its findings became crucial for my further investigations. It made me aware of the role of the dynamics of expectations in introducing AI in healthcare and gave me a better understanding of the potential resistance processes steered by strong formative expectations that they sooner or later will meet. From here, I continued to pay attention to expectations of AI in healthcare but for the purpose of gaining insights into what may occur

as expectations and promises of AI technologies move closer to actual deployment (cf., article 2).

In terms of how this study contributes to answering the research question of this cover chapter, it revealed the particularities and implications of mobilisations of stakeholders in processes driven by strong formative expectations. This led me to define a certain mode of mobilisation, upon which I will elaborate as 'Mode 1' in the discussion section.

5.2 Article 2 (book chapter): 'Towards experimental implementations'

Kannelønning, M. S., Grisot, M., & Williams, R. (forthcoming). Towards experimental implementations: Moving emerging AI technologies into real-world clinical settings. In P. Giardullo & F. Miele (Eds.), *Algorithmic care: STS perspectives on automation of care*. Palgrave Macmillan.

Status: Accepted for publication

5.2.1 Research problem and position of the article

The process of introducing AI in healthcare is in its early phases. Thus, most of the initiatives currently under way are based on expectations and promises of AI performance rather than on experiences and knowledge from actual deployments and use. This implies that healthcare organisations, such as hospitals, that want to utilise AI in their services must decide whether or how to introduce the technologies based on various sorts of anticipations, including promises put forward by AI vendors and the user organisation's own expectations, needs and requirements. As the outcome of potential investments in AI is accordingly hard to predict, the question is whether the various promises of technological performance are plausible enough and how a user organisation can be convinced of such.

Based on these issues and by using the AI procurement process initiated by a Norwegian public hospital trust as an example, article 2 examined the role of expectations in the early stages of AI deployment. Through nonparticipant meeting observations and interviews, I explored this situation where various actors meet (AI vendors, the project team members and future users) and where the vendors try to convince others to choose their technologies, while the actors within the procurement team assess what they are offered and make decisions based on the prospects presented.

5.2.2 Findings

This article showed how the expectations of the procurement team and the technological promises of the AI vendors played a crucial role in the assessment and selection of technological options. This was exemplified by a shift in the aim of the procurement process, which initially was to procure between one and six AI technologies but ended up acquiring an AI platform, potentially making a larger number of AI technologies available. This shift showed that the promises of AI continued to have legitimacy among the different actors who assessed the technologies before deciding on what to procure. Even though they were not convinced to procure the first group of technologies (the specific AI technologies), the procurement team did not cancel the procurement. Instead, they chose to proceed with another AI technology option that had them convinced was a more sustainable solution, at least temporarily. Furthermore, this change showed how the promises of specific AI technologies lacked a certain credibility, while the promises of an AI platform were perceived as a more credible solution for enabling AI in clinical practices. As in the first study, a paradox emerged, this time in relation to this shift in the technological option for procurement. Even though procuring an AI platform was perceived as the best solution, this choice was still based on expectations and promises; limited evidence of the benefits of an AI platform and whether the onboarded area-specific technologies would apply to local contexts was still the case. Consequently, the procuring hospital was left on almost the same bare ground as when the journey started, and the only way to finally validate the technological promises seems to be to deploy AI and start experimenting in real-world clinical settings.

5.2.3 Contribution to the thesis

As I studied the procurement process, it became evident that expectations and promises of AI technologies both cause changes and change themselves as they are brought together in constellations with other actors. Thus, this study not only brought forth an understanding of how expectations and promises are unreliable and cause unexpected outcomes but also made me aware of the significance of the relational aspect in processes of innovation. As the dialogue between the vendors and the procurement team unfolded during the procurement process, the different actors expressed their interests, expectations and, on the vendors' part, promises in various forms (e.g., through conversations or PowerPoint presentations). This was also how the different actors made the others act in certain ways, leading to the particular procurement. Based on this insight, I continued to explore the relational aspect of introducing

AI. Combined with the increasing awareness of the many issues complicating and delaying AI deployment, this led me to focus my further investigations on how constellations of different actors are working to reach a future with AI in healthcare.

Furthermore, during this study, the characteristics of an actor mobilisation crystallised in which none of the actors could attain what they wanted without the others. For the procurement process to result in a sustainable outcome or an acquisition of AI at all, negotiations and reaching an agreement were necessary. Based on this insight, a 'Mode 2' of actor mobilisation took shape.

5.3 Article 3: 'Navigating uncertainties of introducing Artificial Intelligence (AI) in healthcare'

Kannelønning, M. S. *Navigating uncertainties of introducing artificial intelligence (AI) in healthcare: The role of a Norwegian network of professionals*. Technology in Society, Volume 76 (Mars 2024), Artikkel 102432.
DOI: https://doi.org/10.1016/j.techsoc.2023.102432
Status: Published

5.3.1 Research problem and position of the article

As previously mentioned, the current state of introducing AI is characterised by several issues complicating and delaying deployments of AI technologies in healthcare services worldwide (cf., the AI chasm). Due to these intricate issues involving elements from different expert fields and domains, it is perceived as necessary to mobilise heterogeneous actors with different knowledge and capabilities to navigate it. Through collective activities carried out by such hybrid actor constellations, crucial issues and the uncertainties these raise regarding the future of AI in healthcare can be unpacked and potentially lead to more widespread AI deployments.

With the landscape of complex issues that are somehow necessary to address as the point of departure, article 3 explored how a constellation of various actors can contribute to enabling AI in healthcare. It did so by drawing on data from meeting observations and interviews conducted during the study of the informally established Norwegian network of professionals, the KIN network.

5.3.2 Findings

Knowledge production and sharing happened through the KIN network between actors that were not necessarily previously linked. Even though the intention of the network is to mobilise actors from various domains, organisations and sectors, the study found that certain actors were not yet a noticeable part of the collective activity of the network. Actors such as future users of AI (e.g., healthcare professionals) and specific AI technologies with proven performance in real-world clinical settings were not sufficiently represented and spoken for. As long as AI is mainly treated on an overall generic level, the actors speaking on behalf of both AI technologies and the future users are AI researchers or other AI enthusiasts. Thus, it is the interests of the current spokespersons that shape the knowledge produced and shared within and beyond the network. As the network has tight connections to decision makers within the healthcare sector, the knowledge they produce is also their contribution to the enabling of AI in healthcare. In the end, the lack of proper representatives for the actors not yet mobilised could lead to an outcome that is later refuted as AI deployments are attempted but do not necessarily apply or as they become accepted in their current, unnuanced forms.

Constellations of actors like the KIN network have their strengths and weaknesses, which are influenced by the present phase of the introduction process to which they aim to contribute. However, the actors currently lacking within the network will not be possible to mobilise before more deployments occur. Thus, the third study was left with a similar challenge to that of the procurement process: a proper representativity leading to a 'strong' hybrid knowledge production can first be realised as someone starts to deploy AI, and actors not yet mobile are mobilised. Only then can a strong hybrid knowledge be produced and existing and arising uncertainties be explored, mitigated and monitored, thereby contributing to ensuring a good and sustainable future with AI in healthcare.

5.3.3 Contribution to the thesis

It was especially during the work on article 3 that I became more aware of how the particular constellations of actors mobilised within the studied initiatives affected both what happened during the projects and the project outcomes and implications. I started reflecting on what was the driving force within the initiatives, which actors were mobilised, who was not and so on. Regarding the KIN network, this seemed to be more fragmented than within the others. There was no clear end goal, actors came and went and the meeting agendas took shape along

the way. These characteristics led to the definition of 'Mode 3' of actor mobilisation, which, will be elaborated further and discussed in section 6.

It was during this study that the thesis' overall research question took shape. The importance of mobilising various actors in introducing complex technologies like AI became evident through all of my studies in different ways. A quite explicit prompt in this regard came from an informant who shared the following experience from participating in an AI development project:

> *We experienced that one of the great values of the project was to bring together people who worked with different perspectives [...] from completely different worlds, who otherwise never spoke to each other. Then, when we sat in the same room and discussed very specific issues, it was a bit like hallelujah. [...] Later, I started talking to others around the country and saw that everyone is struggling with all the same things, and everyone is trying to reinvent the wheel on their own* (a member of the KIN network secretariat, study 3).

Table 2 below gives an overview of the three articles' titles, research questions, analytical focus and empirical material.

Table 2. An overview of the thesis's three articles

	Article 1	Article 2 (book chapter)	Article 3
Title	Contesting futures of artificial intelligence (AI) in healthcare: Formal expectations meet informal anticipations.	Towards experimental implementations: Moving emerging AI technologies into real-world clinical settings.	Navigating uncertainties of introducing artificial intelligence (AI) in healthcare: The role of a Norwegian network of professionals.
Research questions	What characterises formal expectations of AI in healthcare articulated in official strategies and policies? How are these expectations contested by informal anticipations and future assessments of actors operating close to clinical practices?	What happens as the visionary claims of AI technologies are concretised in relation to the specific needs and requirements of a user organisation? What role do the promises and expectations play during this stage?	How can an informally established network of professionals, like the KIN network, contribute to enabling AI in healthcare? Sub-questions: How does the KIN network characterise its purpose and role, and how do they go about achieving what they aim for?
Analytical focus	Zooming out on: Policy implementation concerning innovative technologies. Zooming in on: The dynamics of expectations as formal government expectations of AI in healthcare meet informal anticipations of people working closer to clinical practice.	Zooming out on: Procuring innovative technologies in public sector. Zooming in on: The role of expectations and promises of AI for radiology as they are enacted by vendors and selected by a procuring public hospital trust.	Zooming out on: Mobilisation of actors collectively contributing to introducing emerging technologies. Zooming in on: The mobilisation of different actors to enable hybrid knowledge production to identify, explore, mitigate and/or monitor uncertainties related to the future of AI in healthcare.
Empirical material	- Documents - Fieldnotes from digital meeting observations	- Interviews - Fieldnotes from digital meeting observations - Documents (secondary)	- Interviews - Fieldnotes from digital meeting observations - Documents (secondary)
Status	Published	Accepted for publication	Under review

6 Analysis and discussion

In section 2 'Background and context', I showed how the concept and field of AI were established in the 1950s. As a 'flashy' synonym for terms like 'automata studies', the concept did its job. It became an umbrella term gathering peers from related fields and simultaneously appealing to research funders. These early accomplishments prove a central point in my thesis: the expectations of AI engage and mobilise necessary actors and resources, which, in turn, contribute to AI coming into being. Since the establishment of the field of AI, there have been both great achievements and drawbacks, or so-called AI winters, leading to what we today may call the warmest AI summer ever; a change in seasons caused by the achievements in recent decades in fields such as ML and DL, enabled by enhanced computer power, an increase in digitalised data and cloud storage.

Despite the accomplishments in the case of AI in healthcare, there is still a chilly breeze coming from the chasm between expectations and deployment. A key observation in this thesis is that mobilisations of a range of actors are happening seemingly because it is perceived as necessary to navigate the complexities of the current state of introducing AI. My three studies are examples of such mobilisations. The studies show how these actor constellations can take many shapes and have differing outcomes, ranging from processes driven by powerful actors, such as governments with a particular agenda, to practitioners coming together based on shared interests and a desire to collectively influence the developments in their own field.

Much is at stake in the ongoing processes aiming to reach widespread deployment of AI in healthcare. Matters of life and death, in fact. This underscores that the exact constellations of actors mobilised and the dynamic among them also matters. How and why do these particular actors come together, who are they, who are the most influential, who achieves what and what do these constellations and dynamics imply for the introduction of AI? Phrased differently, and to repeat this cover chapter research question: *What are the significance and implications of 'mobilisation' in the early phases of introducing ambiguous, complex and advanced technologies like AI in healthcare?*

6.1 The significance and implications of mobilisation of actors

The three studies have at least four common traits. First, the entry point of each study was the AI chasm. They all aimed to go beyond the hype rhetoric and selling points to better understand the current state of introducing AI. Second, the initiatives studied were all examples of different actor constellations. Third, moments of problematisation and interessement were a part of each initiative. The activities pursued were based on future-oriented representations of AI as solutions to particular problems, which in one way or another was of interest to the involved actors. As such, the belief in AI as a solution to healthcare challenges enabled the mobilisation of actors within the three studies. These mobilisations likewise confirmed that the overall expectations of AI technologies were widely shared among the actors participating in the initiatives. Thus, the future-oriented representations of AI in healthcare had undoubtedly reached a certain legitimacy (Joly, 2010; Van Lente, 2012). Fourth, AI technologies were key actors within the different actor constellations.

As AI technologies cannot speak for themselves and demonstrations of AI technologies in clinical settings were not a part of the initiatives, AI technologies were brought into the projects through other means. The idea of AI (i.e., the expectations and promises) and, in some cases, concrete stories of specific AI technologies became part of the projects through presentations, conversations, reports, requirements specifications and written offers from AI vendors. These forms were also working as 'trapping devices' to convince actors to stay interested in the project (Callon, 1986b). In such a way, AI technologies also became actors mobilised within the initiatives but only through the work of human actors speaking on their behalf (Akrich et al., 2002b; Callon, 1986b). As AI technologies were a part of the different constellations, they affected the actions of the other actors and were likewise affected themselves. One of the clearest examples of the latter became visible during the procurement process (cf., article 2) in which negotiations between AI in the shape of promises put forward by the vendors and AI as expectations held by the procurement team transformed the outcome of the procurement process. The transformation included a shift in the aim of procurement, going from procuring AI technologies for specific medical areas to procuring an AI platform.

As AI technologies were identified as entities with certain capabilities that can contribute to solving challenges in the healthcare services, they were connected to other actors with corresponding interests. Examples of such interests were physicians who need help to manage growing workloads or hospital managers who have to allocate resources according to the often-sparse budgets at hand. However, for actors like AI technologies, physicians and hospital managers to become allies and, thereby, enhance the possibilities of widespread AI deployment, the AI technologies must apply to local clinical settings and cause positive effects corresponding to the others' interests. This includes that the deployment of AI will lead to better treatment, more precise patient diagnostics and more efficient use of resources. In contrast, if the deployment leads to unequal patient treatment due to biased data or creates more examinations and work for the physicians due to too sensitive algorithms, actors like the AI technologies, physicians and hospital managers will most likely detach themselves from the project. This, in turn, implies that the enrolment of actors necessary for the project to continue will not happen. As for now, the assumption of AI technologies as solutions to healthcare challenges still works as a crucial 'obligatory passage point' (Callon, 1986b) for the mobilisations aiming to deploy AI in healthcare. The success of the mobilisations depends on whether AI technologies will perform according to the assumptions.

If we compare the overall project of introducing AI in healthcare with Callon's example of the domestication of scallops and fishermen, it is evident that the 'network' in the case of introducing AI is quite different. It is neither about one single group of actors bringing other actors together to achieve their aim (Callon's three researchers) nor is the 'obligatory passage point' based on the performance of a single kind of nonhuman actor (Callon's scallops). In the ongoing work of introducing AI in healthcare, everything is more ambiguous, both in terms of the many kinds of actors with an agency in the various mobilisations and the variety of key human and nonhuman actors implicated (e.g., physicians within different medical areas, their interests, expertise and work practices and the many kinds of AI technologies developed for such diversity). Furthermore, it concerns a global trend: AI is attempted to be introduced in healthcare services worldwide – it is not about a single research project. This complexity makes it similarly hard to take measures of the success of introducing AI and clearly see the short-term and long-term outcomes of such processes as a whole. However, this does not make the studies of actor mobilisations in the case of AI any less meaningful.

The ambiguous, complex and extensive 'network' of actors aiming to introduce AI in healthcare indicates that the process towards widespread deployment depends on more than the success of a single project carried out by a smaller selection of actors. Due to the current early phase of introducing AI, the answer to the question of whether the contributions of the multiple mobilisations of actors will succeed in reaching widespread AI deployment is still unknown. As of now, temporary measures of success can only be taken by investigating the outcome of particular mobilisations as they unfold and come to an end. My studies are examples of initiatives that were followed, and they represent different mobilisations of actors aiming to enable AI in healthcare through different approaches and with varying outcomes. As such, they do not solely represent the three 'modes of mobilisation' that I next present; they are also modes of 'reality-making', as they potentially contribute to shaping a world where AI technologies are a part of most clinical practices.

6.1.1   Three modes of mobilisation

Mode 1: Steered outcome

The first mode concerns actor mobilisations initiated by powerful actors like governments and other authorities having a particular agenda. The agency of these 'initiators' influences the outcome of the process in a strong sense, leaving the mobilised actors with limited or no actual influence. In the most extreme cases, this steered outcome mode may even take the form of autocracy, a 'top-down' command to be executed, which we can see examples of in countries such as China.

This mode is based on the inquiry process (article 1), which brought together AI researchers, AI vendors, AI project managers, physicians, hospital managers, procurement officers and patient organisations to contribute to an identification of obstacles and opportunities for the deployment of AI in the healthcare services. The project's goal was to develop a report with suggestions for future initiatives to be implemented on a national level by the organisations represented within the project team (the authors of the report), including representatives from health authorities and public agencies. Whether the project was to proceed with these initiatives was to be decided upon by the Norwegian Ministry of Health and Care through the yearly allocation letters.

Even though the project organising the inquiry process was able to interest the various actors to participate in the meetings and to contribute their knowledge and opinions, it can be argued that the invited actors had very little actual influence on the outcome of the process. The elements put forward by spokespersons representing actors operating closer to clinical practices raised issues related to the future use of AI in clinical settings and contested parts of the government's vision. However, these elements were not transformed into initiatives in the final report. This suggests that even though large sets of actors are mobilised, processes like policy implementation that at the outset have strong formal expectations may be too formative for their own good. As the expectations have reached a certain protection by being widely shared, they influence the interpretations of what is discussed during processes like the inquiry process. Consequently, elements that may cause resistance to the project in the future are overlooked (Konrad, 2006, p. 438). This represents a *steered* outcome mode.

The particular mobilisation of the inquiry process implied that elements threatening the expectations of AI and the 'obligatory passage point' were overlooked. Consequently, potential opportunities to assert effort in the present to reduce factors that may later challenge the overall project were bypassed. In other words, even though the initiators of the inquiry process were the most powerful in the current constellation, with an agency influencing the outcome of the process the most, the result of their work could end up turning against their aim of enabling AI in healthcare. For instance, the selected areas to prioritise for future work listed in the final report may lead to a weakening of the ties between physicians and AI technologies in the future as issues relating to these actors and their relationships are not properly explored and addressed in the present. From a Sociology of Translation perspective, this would imply that the interest of these actors in the introduction of AI will potentially fade if some of the contesting elements highlighted during the inquiry process manifest themselves. This includes issues such as whether the AI technologies introduced would be accurate enough or whether they would lead to fair outcomes and be trustworthy. Ultimately, it may not be possible to enrol some of the key actors if such questions are not explored in time.

Mode 2: Negotiated outcome
The second mode can be described as a mobilisation in which the actors involved are more aligned in terms of agency. Although there are actors initiating the mobilisations, these do not necessarily have a stronger influence on the outcome of the mobilisation than the other

participants. The outcome is shaped through negotiations between the actors and can be compared to a more democratic process because all actors have an opportunity to affect the outcome, but none can achieve what they want by themselves.

The procurement process consisted of a smaller actor constellation than those of the two other studies. The human actors involved were representatives from both regional offices and local departments within the initiating hospital trust. This included radiologists, procurement officers, IT officers, data protection officers, representatives from a collaborating hospital trust and the project manager. Despite this already mobilised group of actors, the AI vendors, in particular, had to work to keep the others convinced of the different AI technologies they put forward and spoke for (as the salespersons they are).

In contrast to the outcome of the actor mobilisation within the inquiry process, which was characterised by being shaped by the steering force or formative effects of the Norwegian government's vision, the actor constellation in the procurement process can be characterised as being based on *negotiations*. The result of the procurement process was shaped by the negotiations that happened as the vendors' promises of AI technologies met the expectations of AI held by the procurement team. Different versions of AI were presented during the procurement process to appeal to the interests of the actors within the procurement team. However, these were also discussed against the needs and requirements of the user organisation. Based on this dialogue, the aim of the procurement shifted from being about AI technologies for specific medical areas to revolving around AI as part of a platform setup. A crucial reason for this shift was that four of the five AI vendors not only presented specific AI technologies as requested but also included PowerPoint slides of another technological option, namely AI platforms. This option convinced the procurement team to change their mind, showing that the 'trapping devices' (Callon, 1986b) used by the vendors were quite efficient. However, the decision to procure an AI platform instead of specific AI technologies was still based on expectations as little evidence of the platforms in use existed.

The implications of this exact mobilisation are related to this aspect of making decisions based on expectations. As the study of the procurement process showed, many future-oriented representations will likely change and cause changes as they enter into relationships with other actors. This is why decisions based on expectations are problematic, as pinpointed in article 2. For projects like the procurement process, this predicament is especially clear; the

negotiations that occurred during the process and their outcome (the procurement of an AI platform) are based on mutable future-oriented representations. They are not based on demonstrations of AI technologies in real-world clinical settings. Even though demonstrations could have been possible, the performance of AI technologies in one setting is not necessarily transferable to other contexts (Keane & Topol, 2018, p. 1). If the new expectations of the AI platform and the opportunity to test and take into use a larger variety of AI technologies do not materialise, the alliances established through the procurement process may dissolve. This includes the alliances between the winning vendor, the different actors in the procurement team and the AI platform with the selection of onboarded AI technologies. However, whether or not this will happen will first be known as attempts are made to verify the promises of the AI technologies or, better still, the technologies are deployed in the clinical settings as they are intended.

Mode 3: Fragmented and distributed outcome
The third mode of mobilisation concerns an actor constellation that is more loosely defined. The actors involved are many, and the outcome is a patchwork based on the interests of this variety of members. The initiators may have an agenda, but their main aim is to bring forth the opinions or knowledge of the community. In the most extreme cases, such a mode could take an anarchistic course with limited management.

Of the three studies, the analysis of the KIN network indicated the least defined frames in terms of size, time and outcome. The network had no limitations regarding members who were allowed to join on an ongoing basis, nor did it have a specific goal to be fulfilled within a certain time frame. The network aimed to facilitate knowledge sharing among professionals, its members and crucial decision makers in the field. It did so by staying open for new members and developing the meeting agendas based on the members' interests on a meeting-to-meeting basis. Subsequently, the agendas came to address a large variety of topics related to AI, some more challenging to thoroughly discuss than others.

As a result of being this type of mobilisation, which is characterised by a fluid and almost flat structure, the KIN network and its outcome were similarly quite *fragmented* and *distributed* in character. The particular actor constellation raised questions as to whether the knowledge produced was nuanced or hybrid enough to contribute to a sustainable deployment of AI. After all, the actors mobilised did not represent a significant variety of disciplines, sectors or

organisations, which otherwise could have enabled a more hybrid knowledge production (see Callon et al., 2011; Gibbons et al., 1994) – a hybrid knowledge production that is perceived as essential to navigate the complexities of introducing AI in healthcare, both by the network and other actors in the field. The majority of the members were actors that could be characterised as AI enthusiasts (AI researchers and people working with AI projects related to healthcare), while the actors who were not yet possible to mobilise were the future users of AI and the AI technologies operating in real-world clinical settings.

The 'weak' hybrid knowledge production of the current network implied that the topics for discussion flagged by the network's members as most desirable continued to be insufficiently covered. Consequently, the details of these topics stayed hidden or concealed from the members, who could not learn from them. These less-addressed topics, which reflected many of the issues occupying the AI chasm, could have been explored if more widespread deployments happened and the not-yet mobilised actors were made mobile. Until then, AI enthusiasts remain the primary spokespersons for AI technologies. Thus, it can be argued that it is the more optimistic images of AI technologies that are shared with the essential decision makers, implying that it is these perspectives that influence future policies. As the perspective of AI enthusiasts influences policymakers, this may have further implications since the new policies will continue to have formative effects on the further introduction of AI in healthcare (cf., Mode 1 or article 1). In the end, the result of such processes affected by poor representativity among the mobilised actors may lead to a refusal of the AI technologies as they are finally attempted to be deployed. Thus, this mobilisation is left with a similar problem as the outcome of the procurement process: The problem of mobilisation within the KIN network can first be solved as more real-world deployments start happening. Only then will it be possible to mobilise the missing key actors necessary to produce 'strong' hybrid knowledge within the network and further share this knowledge with other stakeholders, such as decision makers.

Common features and some potential future implications
In all three modes of mobilisation, there are examples of actors aiming to attract others to become or stay interested in the projects at hand and reach the projects' aims. These actors include the members of the project team organising the national inquiry process, the AI vendors and the members of the procurement team, and the secretariat and some of the members and presenters of the KIN network (e.g., AI researchers or managers of AI-related

projects). However, if the actors attempting to mobilise others do not succeed in bringing the key actors of the future of AI in healthcare and their interests into play, the projects' contributions may end up lacking the necessary support to enable AI in healthcare. In this context, even a single question, such as whether AI technologies will be applicable and have benefits across local clinical settings, involves several types of actors who thereby become connected. This exemplifies how the social and the natural entities are related.

It is in this relationship between AI technologies and other actors within clinical settings that the actors aiming to enable AI in healthcare will have the most challenging negotiations. As indicated several times during this thesis, this is not a straightforward achievement. At this level, we cannot yet know the outcome of the pairing of the key actors, which the expectations have as a premise. Will a broad spectre of AI technologies attach themselves to real-world clinical settings? Will physicians trust the AI technologies entering their work practices enough to use them as intended? Will the hospital managers experience an effect in terms of more efficient use of resources? The uncertainties lying within these questions and many others further question the solidity of the arguments put forward during moments of problematisation by policymakers, healthcare authorities, AI vendors, AI researchers and other actors. As it stands today, with the process of introducing AI in healthcare, to find elements of answers to the many crucial questions posing uncertainties for the future as envisioned with AI, it is necessary to start experimenting and deploying AI in real-world practices.

6.2 Contributions to STS research and theory

This study follows in the footsteps of previous STS work on AI, which has addressed aspects such as uncritical beliefs in AI technologies, the human–AI relationship, the development of AI technologies and the potential consequences of their deployment. The findings of my doctoral project build upon and supplement these studies in several areas.

The issues of ethics and responsibility in the case of autonomous technologies as addressed by Winner in the 1970s is still a highly discussed topic. A central question in this concern is who to praise or blame for the outcome of such complex systems (Winner, 1978, p. 304). Accordingly, issues related to ethics and responsibility are some of the most critical aspects within the AI chasm (Rajpurkar et al., 2022, p. 34). This was also reflected within my studies,

as these topics were frequently discussed by the actors involved. For instance, in the case of AI in healthcare, the responsibility issue is especially crucial for the physicians who are users of the technologies and in charge of the diagnosis or treatment given to patients. Letting the machines alone make the final decisions regarding people's lives and health is still perceived as involving too high a risk, at least by regulators within the European Union (COCIR, 2020). This restriction in utilising AI technologies according to their full potential also exemplifies what I pinpoint in article 2: STS researchers should not take the AI vendors' promises of AI for granted, as these may change as they enter social contexts.

This argument, in turn, aligns with Woolgar's (1985) argumentation for a sociology of machine intelligence and the importance of taking the human–machine relationship into account while investigating AI. STS researchers have to examine what happens as AI technologies and humans are set in relation to each other, not only through spokespersons in theoretical discussions, as in my three studies, but also as AI technologies are deployed in real-world settings. As more and more deployments occur, so will the opportunity to finally verify claims of AI performance and illuminate elements of answers to uncertainties regarding AI introduction. This could be by shedding light on questions such as whether the physicians have the right balance of trust to use AI safely and responsibly and whether the technology is beneficial for hospitals and does not cause harmful outcomes.

Without such real-world demonstrations, the limitations of AI technologies will stay hidden in the shadows of the more unrealistic or fictional views of AI as machines with capabilities that can be compared to or exceed those of humans, as Collins (2018) emphasised. My thesis contributes to a better understanding of why such visions or expectations are problematic and how they may be challenged based on empirical data. The three studies bring forth elements such as how concerns and anticipations of actors working closer to clinical practices contest the vision of the Norwegian Government (article 1), how the lack of evidence of the actual performance of specific AI technologies affects the credibility of the AI promises and how taking AI technologies into use in real-world clinical settings seems to be the only way to verify them (article 2) as well as mobilise necessary actors for sustainable deployments not yet properly represented (article 3). Collins argues that it is essential not to become enchanted by the supposed intelligence of AI and allow 'ourselves to become the slaves of *stupid* computers' (Collins, 2018, p. 5). As an extension of this, it can be argued that a widespread AI deployment may lead to humans becoming facilitators of the machines, playing a

monitoring function while staying in the background of the decisions being made and, subsequently, potentially becoming more and more 'slaves' of the machines, as Collins pinpointed.

In light of this potential change in the role of humans, the questions raised by Suchman (2007, p. 229) concerning what in machines can be claimed or experienced as humanlike could be turned around: even though machine intelligence cannot be compared to human intelligence in terms of, for instance, cognitive abilities and the understanding of social contexts, humans might perceive them as having humanlike capabilities. Thus, the machines may not have many humanlike features, but they might be given some in the eyes of humans.

In their study of developments of two AI systems for the public sector (including healthcare), Henriksen and Blond (2023) found that politically motivated managerial interests in cost efficiency rather than the needs and requirements of the actual users influenced the design of particular AI systems. This exemplifies how visions of AI articulated by powerful actors have a formative effect, which is in line with one of the main arguments in my article 1. However, as Bareis and Katzenbach (2022) argued, such forces of future-oriented representations will not necessarily lead to successful fulfilment. Resistance will sooner or later occur. Bareis and Katzenbach merely indicated that such resistance will occur; my first study takes it further, showing what such resistance or scepticism may look like. This study also exemplified how powerful visions, despite the resistance, continue to have strong formative effects with elements of inevitability and path dependency. Radhakrishnan (2021) gave another example of the consequence of such formative effects as she showed how the attention paid to the idea or narrative of 'AI for social good' may lead to actual problems of specific populations being overlooked. Thus, she argues that the expectations of AI as a solution for reaching sustainable healthcare services may not become entirely true, at least not for certain population groups.

Closer to the topic of this cover chapter are Winter and Carusi's (2022) findings regarding how trust in AI technologies can be established through collaborations between AI developers and clinical experts during AI development processes. This finding shows how the mobilisation of actors implicated in phases of AI deployment and validation can better ensure that the technologies deployed will be accepted and taken into use. This collaborative aspect supports the focus of this thesis concerning the mobilisation of actors as a way to potentially reach a more sustainable, widespread AI deployment. It is also in this area that the

present work not only builds upon but also supplements the existing body of STS work on AI by focusing on the less addressed topic of 'mobilisation' in the case of introducing AI in healthcare.

My doctoral project also contributes to the existing body of interdisciplinary research on the AI chasm, development and deployment beyond STS. As AI technologies have entered a new phase, a spring or a hot summer, the hype, discussions and even deployments of AI have reached most corners of society. However, most of the existing research is based on either theoretical discussions or cases of AI development, deployment and use happening within the borders of research projects or other kinds of temporary experiments due to the current early phases of AI technologies. In this context, my studies offer empirical insights into what happens as AI technologies, or the ideas thereof, move into real-world social contexts. Finally, apart from contributing to the existing STS work from before and after the recent AI boom and other studies taking at the outset in the current state of AI in healthcare, my thesis also contributes theoretically to the field of STS.

The analysis and discussion of the three modes of mobilisation address in many ways what is at the heart of STS. For decades, STS scholars have been concerned with how emerging technologies are shaped through relationships between heterogeneous actors. Essential in this context is that such processes are not perceived as linear. The analysis carried out by STS scholars is concerned with the unpacking of the mechanisms behind the technological developments, as underscored by the three central STS approaches, Large Technological Systems (LTS), Social Construction of Technology (SCOT) and Actor-Network Theory (ANT). My thesis also contributes to such an analysis through the perspectives taken during the three articles and the focus on the concept of mobilisation in this cover chapter. In the case of introducing ambiguous and complex technologies like AI, examining various constellations of actors is a way of understanding and rendering visible how technologies come into being, as pinpointed before me by STS scholars from the three approaches. Based on this cover chapter's analysis, it can be argued that Callon's depiction of mobilisation in the case of the domestication of scallops and fishermen falls short in describing the various mobilisations I observe in my material. Thus, the three modes of actor mobilisation developed based on the studies of my thesis are my theoretical contributions to the concept of mobilisation in STS.

7 Conclusion and final remarks

This cover chapter set out to explore the significance and implications of actor mobilisation in early phases of introducing AI technologies in the Norwegian public healthcare services. It did so by using the current chasm between expectations and deployments of AI in real-world clinical settings as an entry point.

The initiatives studied had several common traits, such as being established based on expectations of AI in healthcare and mobilising several various actors to contribute to fulfilling these expectations. However, the mobilisations varied in terms of the particular constellations of actors and these actors' actual influence on the outcome of the mobilisation itself. These different characteristics also had different implications for the contributions of the initiatives to the overall introduction of AI. Based on such findings, three modes of actor mobilisations were identified and developed in this cover chapter.

The first mode exemplifies how actors mobilised as a part of a process conducted by powerful actors with strong formative expectations as the driving force will lead to an outcome that can be characterised as *steered*. Due to this characteristic, this type of mobilisation implies that the knowledge or opinion of the less influential actors involved is overlooked, even though they may include elements that will later challenge the materialisation of the expectation. The second mode represents a mobilisation in which the influence of the actors involved is more aligned. As none of the actors can achieve the goal of the project alone, the outcome of this mobilisation is characterised by being based on *negotiations*. However, whether the actors brought together at this point are representative enough to speak on behalf of actors, such as the AI technologies, according to how these technologies will perform in real-world settings may later determine the success of this particular mobilisation. Thus, until deployment of the procured solution is attempted, the final implications of this mobilisation can be characterised as postponed. The third mode of actor mobilisation concerns a more loosely defined constellation. By being large in size, open for actors to come and go, and with a more or less flat structure and fluid agenda, this mode of mobilisation contributes to more *fragmented* and *distributed* outcomes. However, similar to the second mode, it will not be until more AI deployments occur and actors in the shape of AI users and technologies in use are possible to mobilise that such mobilisations can contribute to knowledge leading to more sustainable deployments of AI. Until then, there are

too many issues that may affect the final outcome of the mobilisations that are hidden or concealed.

The findings of the thesis' three studies and cover chapter indicate that starting to validate and deploy AI technologies in real-world clinical settings is crucial to begin revealing elements of answers to many of the current uncertainties regarding the short-term and long-term outcomes of introducing AI in healthcare. Such actions will enable the mobilisation of essential actors not yet adequately represented or mobilised (e.g., future users and technologies in use). An increase in the emergence of these actors will make it possible to initiate and establish more heterogenous actor mobilisations to better explore, identify, mitigate and monitor both existing and emerging uncertainties.

The many questions related to how a widespread deployment of AI in the public healthcare services will hit us are currently difficult to answer empirically. However, as the introduction of AI is still 'in the making', actors within the practice field (e.g., AI projects and early AI users), STS researchers and others must continue to bring forth the specifics of the ongoing processes and the uncertainties attached to the future. In short, there is a need for continued investigations and monitoring of what happens as more and more AI deployments occur. Only then can the courses of the AI introduction in healthcare services be adjusted to ensure that the technologies we introduce in society today are also responsible and sustainable solutions for future generations.

8 References

AAAiH. (2019). *The Australian Alliance for Artificial Intelligence in Healthcare.* Retrieved 29 June 2023 from https://aihealthalliance.org

AAiH. (2023). *The Alliance for Artificial Intelligence in Healthcare*. Retrieved 29 June 2023 from https://www.theaaih.org

AI Sweden. (n.d.). *Information-driven healthcare*. Retrieved 27 July 2023 from https://www.ai.se/en/projects-9/information-driven-healthcare

Akrich, M., Callon, M., & Latour, B. (2002a). The key to success in innovation part I: The art of interessement (A. Monaghan, Trans.). *International Journal of Innovation Management*, *6*(02), 187-206. https://doi.org/10.1142/S1363919602000550

Akrich, M., Callon, M., & Latour, B. (2002b). The key to success in innovation part II: The art of choosing good spokespersons (A. Monaghan, Trans.). *International Journal of Innovation Management*, *6*(02), 207-225. https://doi.org/10.1142/S1363919602000562

Alami, H., Lehoux, P., Auclair, Y., Guise, M. d., Gagnon, M.-P., Shaw, J., Roy, D., Fleet, R., Ahmed, M. A. A., Fortin, J.-P., de Guise, M., & Ag Ahmed, M. A. (2020). Artificial intelligence and health technology assessment: Anticipating a new level of complexity. *Journal of Medical Internet Research*, *22*(7), 1-12. https://doi.org/10.2196/17707

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science. In: Berry, M., Mohamed, A., Yap, B. (eds), *Supervised and Unsupervised Learning for Data Science* (pp. 3-21). Unsupervised and Semi-Supervised Learning. Springer. https://doi.org/10.1007/978-3-030-22475-2_1

Aristidou, A., Jena, R., & Topol, E. J. (2022). Bridging the chasm between AI and clinical implementation. *The Lancet*, *399*(10325), 620. https://doi.org/10.1016/S0140-6736(22)00235-5

Arora, A. (2020). Conceptualising artificial intelligence as a digital healthcare innovation: An introductory review. *Medical Devices: Evidence and Research*, 223-230.

Asdal, K., & Reinertsen, H. (2020). *Hvordan gjøre dokumentanalyse : en praksisorientert metode* (1. utgave. ed.). Cappelen Damm akademisk.

Aung, Y. Y. M., Wong, D. C. S., & Ting, D. S. W. (2021). The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare. *British Medical Bulletin*, *139*(1), 4-15. https://doi.org/10.1093/bmb/ldab016

Bareis, J., & Katzenbach, C. (2022). Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology, & Human Values*, *47*(5), 855-881. https://doi.org/10.1177/01622439211030

BBC TV. (1973). The Lighthill debate on Artificial Intelligence: "The general purpose robot is a mirage". *Controversy* [TV show]. https://www.youtube.com/watch?v=03p2CADwGF8

Bi, W. L., Hosny, A., Schabath, M. B., Giger, M. L., Birkbak, N. J., Mehrtash, A., Allison, T., Arnaout, O., Abbosh, C., & Dunn, I. F. (2019). Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA: A Cancer Journal for Clinicians*, *69*(2), 127-157. https://doi.org/10.3322/caac.21552

Bijker, W. E., Hughes, T. P., & Pinch, T. (2012). *The social construction of technological systems, anniversary edition: New directions in the sociology and history of technology*. MIT Press.

Bijker, W. E., & Law, J. (1992). *Shaping technology/building society: Studies in sociotechnical change*. MIT Press.

Blease, C., Kaptchuk, T. J., Bernstein, M. H., Mandl, K. D., Halamka, J. D., & DesRoches, C. M. (2019). Artificial intelligence and the future of primary care: Exploratory qualitative study of UK general practitioners' views. *Journal of Medical Internet Research*, 21(3), e12802. https://doi.org/10.2196/12802

Borup, M., Brown, N., Konrad, K., & Van Lente, H. (2006). The sociology of expectations in science and technology. *Technology Analysis & Strategic Management*, 18(3-4), 285-298. https://doi.org/10.1080/09537320600777002

Bossen, C., & Pine, K. H. (2023). Batman and Robin in healthcare knowledge work: Human-AI collaboration by clinical documentation integrity specialists. *ACM Transactions on Computer-Human Interaction*, 30(2), Article 26. https://doi.org/10.1145/3569892

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. MIT Press.

Brown, N., & Michael, M. (2003). A sociology of expectations: Retrospecting prospects and prospecting retrospects. *Technology Analysis & Strategic Management*, 15(1), 3-18.

Brown, N., Rappert, B., & Webster, A. (2000). *Contested futures: a sociology of prospective techno-science*. Routledge. https://doi.org/10.4324/9781315259420

Cabitza, F., Campagner, A., & Balsano, C. (2020). Bridging the "last mile" gap between AI implementation and operation: "Data awareness" that matters. *Annals of Translational Medicine*, 8(7), 501. https://doi.org/10.21037/atm.2020.03.63

Callon, M. (1986a). The sociology of an actor-network: The case of the electric vehicle. In M. Callon, J. Law, & A. Rips (Eds.), *Mapping the dynamics of science and technology: Sociology of science in the real world* (pp. 19-34). Springer.

Callon, M. (1986b). Some elements of a sociology of translation: Domestication of the scallops and the fishermen of St Brieuc Bay. In J. Law (Ed.), *Power, action and belief. A new sociology of knowledge* (pp. 196-233). Routledge & Kegan Paul.

Callon, M., Lascoumes, P., & Barthe, Y. (2011). *Acting in an uncertain world: An essay on technical democracy*. MIT Press.

Challen, R., Denny, J., Pitt, M., Gompels, L., Edwards, T., & Tsaneva-Atanasova, K. (2019). Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety*, 28(3), 231-237.

Char, D. S., Shah, N. H., & Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. *The New England Journal of Medicine*, 378(11), 981. https://doi.org/10.1056/NEJMp1714229

Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P. M., Zietz, M., Hoffman, M. M., Xie, W., Rosen, G. L., Lengerich, B. J., Israeli, J., Lanchantin, J., Woloszynek, S., Carpenter, A. E., Shrikumar, A., Xu, J., . . . Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface*, 15(141). https://doi.org/10.1098/rsif.2017.0387

COCIR. (2020). *Artificial intelligence in Eu medical device legislation*. https://futurium.ec.europa.eu/sites/default/files/2020-09/COCIR_Analysis_on_AI_in_medical_Device_Legislation_-_Sept._2020_-_Final_2.pdf

Coiera, E. (2019). The last mile: Where artificial intelligence meets reality. *Journal of Medical Internet Research*, 21(11), e16323. https://doi.org/10.2196/16323

Collins, H. M. (1990). *Artificial experts: Social knowledge and intelligent machines*. MIT Press.

Collins, H. M. (2012). Expert systems and the science of knowledge. In W. E. Bijker, T. P. Hughes, T. Pinch, & D. G. Douglas (Eds.), *The social construction of technological*

systems, anniversary edition: New directions in the sociology and history of technology (pp. 321-340). MIT Press.

Collins, H. M. (2018). *Artifictional intelligence: Against humanity's surrender to computers*. John Wiley & Sons.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. Sage.

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach. *BMC Medical Research Methodology*, *11*(1), 100. https://doi.org/10.1186/1471-2288-11-100

Danish Ministry of Finance, & Danish Ministry of Industry, B. a. F. A. (2019). *National strategy for artificial intelligence*. https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf

Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, *6*(2), 94-98. https://doi.org/10.7861/futurehosp.6-2-94

de Neufville, R., & Baum, S. D. (2021). Collective action on artificial intelligence: A primer and review. *Technology in Society*, *66*, 101649. https://doi.org/10.1016/j.techsoc.2021.101649

Dicuonzo, G., Donofrio, F., Fusco, A., & Shini, M. (2023). Healthcare system: Moving forward with artificial intelligence. *Technovation, 120*, 102510. https://doi.org/10.1016/j.technovation.2022.102510

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic fieldnotes* (2 ed.). University of Chicago Press.

European Union (2022). *CE marking*. Your Europe (europe.eu). https://europa.eu/youreurope/business/product-requirements/labels-markings/ce-marking/

Fangen, K. (2010). *Deltagende observasjon* (2. utg. ed.). Fagbokforlaget.

FCAI. (n.d.). *AI for health*. Retrieved 27 July 2023 from https://fcai.fi/ai-for-health

FDA, t. A. F. a. D. A. (2023). *CDRH Issues draft guidance on predetermined change control plans for artificial intelligence/machine learning-enabled medical devices*. FDA, the American Food and Drug Administration. Retrieved 8 June 2023 from https://www.fda.gov/medical-devices/medical-devices-news-and-events/cdrh-issues-draft-guidance-predetermined-change-control-plans-artificial-intelligencemachine

Forsythe, D. E. (1993a). The construction of work in artificial intelligence. *Science, Technology, & Human Values*, *18*(4), 460-479. https://doi.org/10.1177/0162243993018004

Forsythe, D. E. (1993b). Engineering knowledge: The construction of knowledge in artificial intelligence. *Social Studies of Science*, *23*(3), 445-477. https://doi.org/10.1177/0306312793023003002

Future of Life Institute. (2023). *Pause giant AI experiments: An open letter*. Retrieved 30 July from https://futureoflife.org/open-letter/pause-giant-ai-experiments/

Galanos, V. (2023). *Expectations and expertise in artificial intelligence: Specialist views and historical perspectives on conceptualisation, promise, and funding*. The University of Edinburgh. https://era.ed.ac.uk/bitstream/handle/1842/40420/Galanos2023.pdf

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. SAGE Publications, Limited. http://ebookcentral.proquest.com/lib/hioa/detail.action?docID=1024114

Gillan, C., Milne, E., Harnett, N., Purdie, T. G., Jaffray, D. A., & Hodges, B. (2019). Professional implications of introducing artificial intelligence in healthcare: An

evaluation using radiation medicine as a testing ground. *Journal of Radiotherapy in Practice*, *18*(1), 5-9.

Hammersley, M., & Atkinson, P. (2019). *Ethnography: Principles in practice* (4th ed.). Taylor & Francis Group.

He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature Medicine*, *25*(1), 30-36.

Heikkilä, M. (2022, 29 March). Dutch scandal serves as a warning for Europe over risks of using algorithms. *Politico*. https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/

Helsedirektoratet, Direktoratet for e-helse, Helsetilsynet, Statens legemiddelverk, Helse Nord RHF, Helse Vest RHF, Helse Midt-Norge RHF, Helse Sør-Øst RHF, Folkehelseinsituttet, & Kommunesektorens interesseorganisasjon. (2022). *Kunstig intelligens i helsetjenesten. Status og veien videre for det nasjonale koordineringsarbeidet.* https://www.helsedirektoratet.no/tema/kunstig-intelligens/Kunstig%20intelligens%20i%20helsetjenesten%20-%202022.pdf/_/attachment/inline/7290856c-450a-47d6-aaa7-d660c205d4d4:f6c3e75605639f3d2c4750f2a6d8196a6e1a40d1/Kunstig%20intelligens%20i%20helsetjenesten%20-%202022.pdf

Henriksen, A., & Blond, L. (2023). Executive-centered AI? Designing predictive systems for the public sector. *Social Studies of Science*, *0*(0), 03063127231163756. https://doi.org/10.1177/03063127231163756

Hollander, J. E., Sease, K. L., Sparano, D. M., Sites, F. D., Shofer, F. S., & Baxt, W. G. (2004). Effects of neural network feedback to physicians on admit/discharge decision for emergency department patients with chest pain. *Annals of Emergency Medicine*, *44*(3), 199-205. https://doi.org/10.1016/j.annemergmed.2004.02.037

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*(9), 1277-1288. https://doi.org/10.1177/1049732305276687

Hughes, T. P. (2012). The evolution of large technological systems. In W. E. Bijker, T. P. Hughes, T. Pinch, & D. G. Douglas (Eds.), *The social construction of technological systems, anniversary edition: New directions in the sociology and history of technology* (pp. 45-76). MIT Press.

Hyysalo, S. (2021). The biographies of artifacts and practices methodology for the study of sociotechnical change. In *Citizen Activities in Energy Transition* (pp. 15-33). Routledge. https://doi.org/10.4324/9781003133919

Jasanoff, S. (2004). *States of knowledge: The co-production of science and the social order*. Routledge.

Jasanoff, S., & Kim, S.-H. (2009). Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea. *Minerva*, *47*(2), 119. https://doi.org/10.1007/s11024-009-9124-4

Jasanoff, S., & Kim, S.-H. (2015). *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*. University of Chicago Press.

Joly, P.-B. (2010). On the economics of techno-scientific promises. In M. Akrich, Y. Barthe, F. Muniesa, & P. Mustar (Eds.), *Débordements. Mélanges offerts à Michel Callon* (pp. 203-222). Presse des Mines. https://doi.org/10.4000/books.pressesmines.747

Jonas, H. (1982). Technology as a subject for ethics. *Social Research*, *49*(4), 891-898. http://www.jstor.org/stable/40971222

Kaul, V., Enslin, S., & Gross, S. A. (2020). History of artificial intelligence in medicine. *Gastrointest Endosc*, *92*(4), 807-812. https://doi.org/10.1016/j.gie.2020.06.040

Keane, P. A., & Topol, E. J. (2018). With an eye to AI and autonomous diagnosis. *NPJ Digit Med*, *1*, 40. https://doi.org/10.1038/s41746-018-0048-y

Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC medicine*, *17*, 1-9. https://doi.org/10.1186/s12916-019-1426-2

KIN. (2023). *Prosjekter*. Kunstig intelligens i norsk helsetjeneste, Nasjonalt senter for e-helseforskning. Retrieved 3 August from https://ehealthresearch.no/kin/prosjekter

Konrad, K. (2006). The social dynamics of expectations: The interaction of collective and actor-specific expectations on electronic commerce and interactive television. *Technology Analysis & Strategic Management, 18*(3-4), 429-444. https://doi.org/10.1080/09537320600777192

Konrad, K., Van Lente, H., Groves, C., & Selin, C. (2016). Performing and governing the future in science and technology. In U. Felt, R. Fouché, C. A. Miller, & L. Smith-Doerr (Eds.), *The handbook of science and technology studies* (pp. 465-493). MIT Press.

Kvale, S., & Brinkmann, S. (2009). *Interviews: Learning the craft of qualitative research interviewing* (2nd ed.). Sage.

Laï, M.-C., Brian, M., & Mamzer, M.-F. (2020). Perceptions of artificial intelligence in healthcare: Findings from a qualitative survey study among actors in France. *Journal of translational medicine*, *18*(1), 1-13. https://doi.org/10.1186/s12967-019-02204-y

Lampa, E., Sonnentheil, B., Tökés, A., & Warner, G. (2021). What has the COVID-19 pandemic taught us about conducting patient and public involvement remotely? Insights from a series of digital meeting observations. *Research Involvement and Engagement*, *7*(1), 1-8. https://doi.org/10.1186/s40900-021-00315-9

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Latour, B. (1990). Postmodern? No, simply amodern! Steps towards an anthropology of science. *Studies in History and Philosophy of Science Part A*, *21*(1), 145-171. https://doi.org/10.1016/0039-3681(90)90018-4

Latour, B. (2003). The promises of constructivism. In D. Ihde & E. Selinger (Eds.), *Chasing technoscience: Matrix for materiality* (pp. 27-46). Indiana University Press. https://hal-sciencespo.archives-ouvertes.fr/hal-01027765/document

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford University Press.

Law, J. (1991). *A Sociology of monsters: Essays on power, technology and domination* (Vol. 38). Routledge.

Malvehy, J., Ginsberg, R., Sampietro-Colom, L., Ficapal, J., Combalia, M., & Svedenhag, P. (2022). New regulation of medical devices in the EU: impact in dermatology. *Journal of the European Academy of Dermatology and Venereology*, 36(3), 360-364. https://doi.org/10.1111/jdv.17830

Marcus, G. E. (1995). Ethnography in/of the world system: The emergence of multi-sited ethnography. *Annual review of anthropology*, *24*(1), 95-117. https://doi.org/10.1146/annurev.an.24.100195.000523

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, 31 August 1955. *Ai Magazine*, *27*(4), 12. https://doi.org/10.1609/aimag.v27i4.1904

McCullagh, L., Sofianou, A., Kannry, J., Mann, D., & McGinn, T. (2014). User centered clinical decision support tools. *Applied Clinical Informatics*, *5*(04), 1015-1025. https://doi.org/10.4338/ACI-2014-05-RA-0048

Meld. St. 7 (2019–2020). *National Health and Hospital Plan 2020–2023. Summary*. Retrieved from https://www.regjeringen.no/contentassets/95eec808f0434acf942fca449ca35386/en-gb/pdfs/stm201920200007000engpdfs.pdf

Meskó, B., Hetényi, G., & Győrffy, Z. (2018). Will artificial intelligence solve the human resource crisis in healthcare? *BMC Health Services Research*, *18*(1), 545. https://doi.org/10.1186/s12913-018-3359-4

Miliard, M. (2022, 21. January). IBM to sell Watson Health assets to Francisco Partners. *HealthcareITNews*. https://www.healthcareitnews.com/news/ibm-sell-watson-health-assets-francisco-partners

Miller, R. A. (1994). Medical diagnostic decision support systems—past, present, and future: A threaded bibliography and brief commentary. *Journal of the American Medical Informatics Association*, *1*(1), 8-27. https://doi.org/10.1136/jamia.1994.95236141

Mol, A. (2002). *The body multiple: Ontology in medical practice*. Duke University Press.

Morley, J., Machado, C. C. V., 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. https://doi.org/10.1016/j.socscimed.2020.113172

National Committee for Research Ethics in the Social Sciences and the Humanities (NESH). (2022). *Guidelines for research ethics in the social sciences and the humanities*. National Research Ethics Committees. Retrieved 20 July from https://www.forskningsetikk.no/en/guidelines/social-sciences-humanities-law-and-theology/guidelines-for-research-ethics-in-the-social-sciences-humanities-law-and-theology/

NHS. (n.d.). *The national strategy for AI in health and social care*. NHS England - Transformation Directorate. Retrieved 21 February 2023 from https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/the-national-strategy-for-ai-in-health-and-social-care/

NHS England. (n.d.). *The NHS AI Lab*. Retrieved 27 July from https://transform.england.nhs.uk/ai-lab/

Nyqvist, A. (2013). Access to all stages?: Studying through policy in a culture of accessibility. In C. Garsten & A. Nyqvist (Eds.), *Organisational Anthropology. Doing Ethnography in and among Complex Organisations* (pp. 91-105). Pluto Press.

OECD. (2019). *Recommendation of the Council on Artificial Intelligence*. OECD/LEAGL/0449. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Panch, T., Mattie, H., & Celi, L. A. (2019). The "inconvenient truth" about AI in healthcare. *npj Digital Medicine*, *2*(1), 77. https://doi.org/10.1038/s41746-019-0155-4

Parikh, R. B., Teeple, S., & Navathe, A. S. (2019). Addressing bias in artificial intelligence in health care. *JAMA*, *322*(24), 2377-2378. https://doi.org/10.1001/jama.2019.18058

Pinch, T., & Bijker, W. E. (2012). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. In W. E. Bijker, T. P. Hughes, & T. Pinch (Eds.), *The social construction of technological systems, anniversary edition: New directions in the sociology and history of technology* (pp. 11-44). MIT Press.

Pollock, N., & Williams, R. (2010). The business of expectations: How promissory organizations shape technology and innovation. *Social Studies of Science*, *40*(4), 525-548. https://doi.org/10.1177/0306312710362275

Racine, E., Boehlen, W., & Sample, M. (2019). Healthcare uses of artificial intelligence: Challenges and opportunities for growth. *Healthcare Management Forum*, *32*(5), 272-275. https://doi.org/10.1177/084047041984383

Radhakrishnan, R. (2021). Experiments with social good: Feminist critiques of artificial intelligence in healthcare in India. *Catalyst: Feminism, Theory, Technoscience*, *7*(2). https://doi.org/10.28968/cftt.v7i2.34916

Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine*, *28*(1), 31-38. https://doi.org/10.1038/s41591-021-01614-0

Robinson, D. K., Audétat, M., Joly, P.-B., & Van Lente, H. (2021). Enemies of the future? Questioning the regimes of promising in emerging science and technology. *Science and Public Policy*, *48*(6), 814-817. https://doi.org/10.1093/scipol/scab055

Ross, C., & Swetlitz, I. (2017). IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close. *Stat*. https://www.statnews.com/2017/09/05/watson-ibm-cancer/

Ross, C., & Swetlitz, I. (2018). IBM's Watson supercomputer recommended 'unsafe and incorrect'cancer treatments, internal documents show. *Stat*. https://www.statnews.com/wp-content/uploads/2018/09/IBMs-Watson-recommended-unsafe-and-incorrect-cancer-treatments-STAT.pdf

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach*. Pearson Education, Inc.

Sendak, M., Gao, M., Nichols, M., Lin, A., & Balu, S. (2019). Machine kearning in health care: A critical appraisal of challenges and opportunities. *EGEMS (Wash DC)*, *7*(1), 1. https://doi.org/10.5334/egems.287

Seneviratne, M. G., Shah, N. H., & Chu, L. (2020). Bridging the implementation gap of machine learning in healthcare. *BMJ Innovations*, *6*(2), 45-47. https://doi.org/10.1136/bmjinnov-2019-000359

Sismondo, S. (2010). *An introduction to science and technology studies* (2nd ed.). Wiley-Blackwell.

Sovacool, B. K., & Hess, D. J. (2017). Ordering theories: Typologies and conceptual frameworks for sociotechnical change. *Social Studies of Science*, *47*(5), 703-750. https://www.jstor.org/stable/48590466

Star, S. L., & Ruhleder, K. (1994). Steps towards an ecology of infrastructure: complex problems in design and access for large-scale collaborative systems. Proceedings of the 1994 ACM conference on Computer Supported Cooperative Work. https://doi.org/10.1145/192844.193021

Stengers, I. (1999). *For en demokratisering av vitenskapene* (Vol. nr 4). Spartacus.

Strickland, E. (2019). IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. *IEEE Spectrum, 56*(4), 24-31. https://*doi:10.1109/MSPEC.2019.8678513*

Suchman, L. (2007). *Human-machine reconfigurations: Plans and situated actions*. Cambridge University Press.

Suchman, L. (2022). Imaginaries of omniscience: Automating intelligence in the US Department of Defense. *Social Studies of Science*, *0*(0). https://doi.org/10.1177/03063127221104938

Suchman, L. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, *36*(2), 368-383. https://doi.org/10.1016/j.giq.2018.09.008

Szolovits, P., Patil, R. S., & Schwartz, W. B. (1988). Artificial intelligence in medical diagnosis. *Annals of Internal Medicine*, *108*(1), 80-87. https://doi.org/10.7326/0003-4819-108-1-80. PMID: 3276267

Thedvall, R. (2013). Punctuated entries. Doing fieldwork in policy meetings in the European Union. In A. Nyqvist & C. Garsten (Eds.), *Organisational anthropology: Doing ethnography in and among complex organisations* (pp. 106-119). Pluto Press.

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, *25*(1), 44-56. https://doi.org/10.1038/s41591-018-0300-7

Torenholt, R., & Langstrup, H. (2023). Between a logic of disruption and a logic of continuation: Negotiating the legitimacy of algorithms used in automated clinical decision-making. *Health*, *27*(1), 41-59. https://doi.org/10.1177/1363459321996741

Tucker, J. (2023). The future vision(s) of AI health in the Nordics: Comparing the national AI strategies. *Futures*, *149*, 103154. https://doi.org/10.1016/j.futures.2023.103154

Turing, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, *LIX*(236), 433-460. https://doi.org/10.1093/mind/LIX.236.433

Van Lente, H. (2000). Forceful futures: From promise to requirement. In N. Brown, B. Rappert, & A. Webster (Eds.), *Contested futures: A sociology of prospective techno-science* (pp. 43-63). Routledge. https://doi.org/10.4324/9781315259420-12

Van Lente, H. (2012). Navigating foresight in a sea of expectations: Lessons from the sociology of expectations. *Technology Analysis & Strategic Management*, *24*(8), 769-782. https://doi.org/10.1080/09537325.2012.715478

Vogt, W. P. (2005). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences*. SAGE.

Wang, F., Casalino, L. P., & Khullar, D. (2019). Deep learning in medicine—Promise, Progress, and Challenges. *JAMA Internal Medicine*, *179*(3), 293-294. https://doi.org/10.1001/jamainternmed.2018.7117

Winner, L. (1978). *Autonomous technology: Technics-out-of-control as a theme in political thought*. MIT Press.

Winter, P., & Carusi, A. (2022). 'If you're going to trust the machine, then that trust has got to be based on something': Validation and the co-constitution of trust in developing artificial intelligence (AI) for the early diagnosis of pulmonary hypertension (PH). *Science & Technology Studies*, *35*(4), 58-77. https://doi.org/10.23987/sts.102198

Wong, A., Otles, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Cooley, O., Pestrue, J., Phillips, M., Konye, J., Penoza, C., Ghous, M., & Singh, K. (2021). External validation of a widely implemented proprietary sepsis prediction model in hospitalized patients. *JAMA Internal Medicine*, *181*(8), 1065-1070. https://doi.org/10.1001/jamainternmed.2021.2626

Woolgar, S. (1985). Why not a sociology of machines? The case of sociology and artificial intelligence. *Sociology*, *19*(4), 557-572. https://doi.org/10.1177/0038038585019004005

Woolgar, S. (2012). Reconstructing man and machine: A note on sociological critiques of cognitivism. In W. E. Bijker, T. P. Hughes, & T. Pinch (Eds.), *The social construction of technological systems, anniversary edition: New directions in the sociology and history of technology* (pp. 303-320). MIT Press.

Xiao, C., Choi, E., & Sun, J. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. *Journal of the American Medical Informatics Association*, *25*(10), 1419-1428. https://doi.org/10.1093/jamia/ocy068

Yu, K.-H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, *2*(10), 719-731. https://doi.org/10.1038/s41551-018-0305-z

Zając, H. D., Li, D., Dai, X., Carlsen, J. F., Kensing, F., & Andersen, T. O. (2023). Clinician-facing AI in the wild: Taking stock of the sociotechnical challenges and opportunities

for HCI. *ACM Transactions on Computer-Human Interaction*, *30*(2), 1-39. https://doi.org/10.1145/3582430

PART II: THE ARTICLES

Article 1

Status: Published

Routledge
Taylor & Francis Group

🔓 OPEN ACCESS | Check for updates

Contesting futures of Artificial Intelligence (AI) in healthcare: formal expectations meet informal anticipations

Mari S. Kannelønning

Faculty of Social Sciences, Oslo Metropolitan University, Oslo, Norway

ABSTRACT

Artificial Intelligence (AI) technologies are expected to play an essential role in future healthcare systems for saving resources, improving treatment quality and enhancing patient safety. Governments worldwide are preparing for such a future by developing strategies and policies. Drawing on concepts and perspectives from Science and Technology Studies (STS), this article explores how government policies set the agenda for introducing AI in healthcare, using the Norwegian National Health and Hospital Plan 2020–2023 (NHSP) as an example. The article further investigates how the formal expectations included in the NHSP are met with informal anticipations expressed by actors working closer to clinical practices involved in an inquiry process initiated by the policy. Taking a qualitative approach, the article explores what characterises formal expectations of AI in healthcare and how these expectations are contested. The study finds that there are tensions between the different assumptions and that crucial issues concerning the future usage of AI are not yet on the Norwegian Government's agenda. Pairing the forces of the formal expectations with the ambiguous concept of AI, the current chasm between AI development and deployment, and the uncertainties constituted by the contesting elements, the article concludes by introducing a paradox of inevitability.

Introduction

Artificial Intelligence (AI) is expected to play an essential role in future healthcare systems, providing economic viability and improving treatment quality and patient safety (Aung, Wong, and Ting 2021; Morley et al. 2020). These expectations typically concern AI technologies that can automate repetitive tasks and support healthcare professionals in making decisions during the whole clinical pathway: from diagnostic processes to treatment and follow-up care.[1] However, there is still a chasm between commercially available AI technologies and their deployment in real-world clinical settings (Cabitza, Campagner, and Balsano 2020; Coiera 2019; Seneviratne, Shah, and Chu 2020). Meanwhile, governments worldwide are developing strategies and policies on AI, including visions of a future with AI in healthcare to guide and accelerate implementation processes (see, for example, the European Commission's policy brief Knowledge4Policy 2021; or the UK's National Strategy for AI in Health and Social Care n.d.).

This article explores how such policies, and the expectations of the future they articulate, set the agenda for introducing AI in healthcare, using the Norwegian National Health and Hospital Plan (NHSP) 2020–2023 as an example. Closely following an inquiry process initiated through the NHSP, involving different stakeholders and producing various outputs, the article also examines how formal expectations are met with informal anticipations expressed by actors situated closer to clinical practices (including hospital managers, procurement officers, and physicians).

Policies like the NHSP are designed to be performative; they initiate activities and actions in the present (Eames et al. 2006, 362). The visions or expectations of the future they bring forth can thus be seen as strategic tools that inform decision-making processes (Beckert 2016, 35–36), mobilise relevant actors (Konrad et al. 2016, 466), and set the political and technological agenda for others to follow (Brown, Rappert, and Webster 2000, 17). Due to this formative strength, they can be seen as both powerful and problematic. For instance, elements disputing them might easily be ignored, potentially leading to costly detours or even severe and irreversible societal consequences.

The field of AI is well-known for its hype, and we are frequently exposed to the rhetoric of AI as technologies of great promise and potential (Bareis and Katzenbach 2022; Verdegem 2021). However, whether or how the expectations of AI will materialise or cause unwanted outcomes is still unknown. This article aims to go beyond the selling points and examines both formal expectations and viewpoints disputing them. The article addresses the following research questions: what characterises formal expectations of AI in healthcare articulated in official strategies and policies? How are these expectations contested by informal anticipations and future assessments of actors operating close to clinical practices? Finally, in the discussion, the article elaborates on a paradox constituted by the formative force of expectations and the uncertainties related to what AI really is and what an introduction of AI might imply. The study refers to a Norwegian case, but the article has relevance outside its Norwegian context, as AI technologies are discussed and policies are developed and implemented worldwide.[2]

The article starts by presenting the conceptual framework, followed by a description of the material and methods, including a presentation of the context, data collection and analysis. It then proceeds by identifying essential aspects of the Norwegian Government's vision, i.e. the formal expectations of AI in healthcare. With this as a backdrop, the article continues by bringing forth the contestations that emerged during the inquiry process studied. Finally, the insights provided by the analysis are discussed in relation to the conceptual framework, followed by concluding remarks.

Forces and dynamics of expectations

This work's theoretical framework builds on concepts and perspectives from the field of Science and Technology Studies (STS), dealing with forces and dynamics of expectations in innovation and technological change (Borup et al. 2006; Brown and Michael 2003; Pollock and Williams 2010).

Scholars within the Sociology of Expectations see expectations as having a performative role in innovation processes (Borup et al. 2006, 289; Konrad et al. 2016). Van Lente (2012) draws attention to such formative strengths by highlighting three essential forces of expectations. The first, 'legitimation', is about how expectations raise awareness and justify investments in early phases of emerging technologies, typically including a solution to a perceived problem. The second, 'heuristic guidance', concerns how expectations provide direction and guide choices and decisions during an innovation process. The third force is a 'coordination effect', as expectations indicate work that has to be done by various actors to fulfil the goals (2012, 773–774). From this point of view, expectations are articulations that 'do things'; as they build agendas and create work in the present, they have performative effects (2012, 772).

Similarly, Geels and Smit (2000) pinpoint that expectations are used to convince relevant stakeholders, foster investments, shape agendas and establish 'protected spaces' (2000, 882). As such, expectations are often exaggerated and optimistic (2000, 881). Brown, Rappert, and Webster

(2000, 3–4) elaborate accordingly that to increase the chances of fulfilling specific envisioned futures, anyone who wants to promote a particular expectation must involve themselves in rhetorical, organisational and material activities aiming to 'colonise' their version of the future. As the expectations become widely shared, they will gain certain protection or enter a 'protected space' created by the collective belief (Konrad 2006, 438). Consequently, the formative force of the expectations is strengthened by being shared by many and, thus, further paired with an element of inevitability (Van Lente and Rip 1998, 222). This sense of inevitability may result in neglecting alternative solutions, potentially contributing to further path dependencies and 'lock-in' effects (Konrad 2006, 430). Such 'lock-in' effects also imply irreversibility, as the allocated resources used to follow a specific direction cannot be reused (e.g. money and people's time). In other words, as Tutton states: 'Each future followed is another future not taken' (2017, 487). However, shifts in widely shared expectations may occur due to real-world events, and consequently, the 'protected space' may dissolve (Konrad 2006, 439). Bareis and Katzenbach stress similarly that despite the power of stakeholders behind certain expectations, their strategies and policies will meet resistance and scepticism as they are implemented into society (2022, 874). The potential shifts in expectations and the opposition they might encounter emphasise further that the outcome of strong formative expectations can be hard to calculate (Beckert 2016, 9).

The fact that even widely shared expectations may not necessarily materialise is why Pollock and Williams (2010) problematise the idea of strong expectations as self-fulfilling prophecies. As unforeseen factors may challenge shared beliefs, they argue for the need to consider what expectations or promises consist of and how they are constructed and distributed (2010, 529). In a similar vein, Brown draws attention to the importance of considering both the temporal and spatial 'situatedness' of expectations to prevent misdirections during innovation (2003, 10). The temporal situatedness is a question of 'when' and what happens as processes driven by certain expectations evolve and different forms of material and social issues appear. Considering what to expect on both shorter and longer terms can function as a caution and enable adjustments of directions before it is too late. The spatial situatedness of expectations concerns the questions of 'where' and 'who'. It emphasises the importance of seeing expectations not as singular representations of the future but as something that needs interpretation and is acted upon differently across actor groups and practices. Uncovering spatial differences can give insights into various agencies of differing power, how expectations are contested, and how some end up setting the agenda (Brown 2003, 13).

Complicating factors in the case of AI

A definitional vagueness characterises the concept of AI (Wang 2019), which, at best, is an overall label covering a large variety of technologies. This ambiguity complicates both the conversations and the outcome of the work related to AI technologies (Krafft et al. 2020; Wang 2019). Whatever definition is used, it will have practical consequences, such as influencing decisions made during various stages of innovation processes (e.g. in policy development and implementation). Without specifying AI, questions like whether AI will be beneficial or not will give several different answers (Wang 2019, 28). Thus, the lack of precise definitions can have quite far-reaching consequences. In the context of national AI strategies, Bareis and Katzenbach (2022) finds that, even though the visions included in policies are based on vague definitions of AI, they are powerful rhetorical devices. Their force lies in how they allocate resources and set rules, turning the visions into certain inevitable technological pathways (2022, 863).

In addition to how the vague and diverging definitions of AI might complicate the process of introducing AI in healthcare, the previously mentioned chasm between available AI technologies and real-world deployments may also affect the process. Apart from limited evidence of actual use and benefits in real clinical settings, the most common explanations for this chasm are issues related to data limitations, regulations, human trust, ethical data use, equity and bias (see, for example, Aung, Wong, and Ting 2021; Freeman et al. 2021; Rajpurkar et al. 2022). Thus, through

the research literature, factors potentially hindering or causing resistance to the many visions of AI technologies in healthcare and their deployment are already known and circulating.

As a final clarification before continuing, reflecting the lack of a widely accepted definition of AI, this article uses the general terms 'AI' and 'AI technologies' interchangeably. These labels cover both AI in a broad sense, for instance, when treating overall conceptions of AI and in connection with the statements of the participants involved in the meetings of the inquiry process. In this context, 'AI technologies' may allude to more specific technologies, for example, AI-based applications to be integrated into existing information systems (e.g. systems for radiology like PACS – Picture Archiving and Communication Systems). In these settings, the term 'algorithm' is additionally used for variation, reflecting how the meeting participants used it.

Material and methods

The context

In 2019 the Norwegian Ministry of Health and Care Services published the NHSP policy plan, through which a coordination project as a national cross-public agency collaboration was established. The project, called *Better use of Artificial Intelligence*, was led by the Directorate of Health and included members from the Directorate of eHealth, the Norwegian Medicines Agency and the four Norwegian regional health authorities. The inquiry process followed in the present study was carried out by the project team and entitled *Good clinical practice and the need for standardisation using Artificial Intelligence in Radiology*[3](my translation). The overall aim of the inquiry process was to identify gaps and suggest national initiatives to ensure safe and proper use of AI in healthcare services. Different digital meetings (insight meetings, workshops and feedback meetings) were carried out in three rounds and involved stakeholders such as hospital managers, project managers, radiologists, patient organisations, AI researchers, AI vendors, and procurement officers. The meetings took place over five months (January–May 2021), and approximately 100 people participated altogether (Helsedirektoratet et al. 2021, 2). After each round of meetings, written summaries were published on the project's website for whoever was interested. Additionally, the summaries were emailed to the stakeholders participating in the upcoming meetings. The final outcome of the process was a report addressed to the Norwegian Ministry of Health and Care Services in the autumn of 2021.

Data collection and analysis

The overall approach of this study is based on a combination of document analysis and meeting ethnography. The document analysed was the NHSP, while the meetings observed were digital, and the observations were centred around the verbal communication unfolding. I attended the meetings as a non-participant observer and kept my camera and microphone off, except for introducing myself if prompted by the chairperson. The primary data from these meetings were fieldnotes written during the observations. I wrote as much as possible, but naturally, I missed some bits and parts of the conversations. However, similar to Thedvall's experience (2013, 112), I gained a fuller understanding of

Table 1. The volume of documents, meetings, and participants.

	Insight meetings	Workshops	Feedback meetings and beyond
Documents (volume)	Project summaries (35 p.) Project analysis (9 p.) Fieldnotes (22 p.)	Project summary (21 p.) Fieldnotes (15 p.)	Project summary (13 p.) Fieldnotes (10 p.) Final report (74 p.)
Number of meetings and duration[a]	2 × 2 h 1 × 1.5 h	3 × 3 h	2 × 2 h
Number of participants (approx.)	45	43 (invited)	40

[a]I did not attend the first three insight meetings, as I was enrolled as an observing researcher after their completion. However, I received and analysed the notes taken by the project team.

the discussions over time as I attended several meetings. As a supplement to the fieldnotes, I used the project team's summaries and reports.

Table 1 below shows the volume of the material and meeting details included in the study.

The NHSP was analysed using a technique similar to 'directed content analysis' (Hsieh and Shannon 2005, 1281). It was a structured process where the document was analysed for content specifically matching Van Lente's three forces of expectations: legitimation, guiding heuristics and coordination (2012). This work was done by thorough readings, while relevant pieces of text were annotated and categorised. The data from the inquiry process, on the other hand, were first subjected to open coding. The fieldnotes, summaries, and reports were read line-by-line, while diverse and disparate themes were identified. These were later coded as the text was closely reread (Emerson, Fretz, and Shaw 2011, 172). As this inductive process became more focused, topics emerged, reflecting what was perceived as important to the participants. Thus, significant observations regarding the future of AI in healthcare, seen from the perspective of actors operating close to clinical practices, could be distinguished and identified from the rest of the material from the inquiry process (2011, 175). This selection of observations was then elaborated on and categorised.

Findings

This section presents the findings in three parts. First, the findings from the analysis of the NHSP are presented, using Van Lente's three dimensions as points of reference. This analysis serves as a backdrop for the second part, which explores the informal anticipations and future assessments of stakeholders involved in the inquiry process, contesting the formal expectations of the NHSP. Lastly, the final report from the inquiry process is shortly outlined.

The NHSP: the formal expectations of AI in healthcare

Legitimation

Like many countries, Norway faces the problem of an aging population with an increasing need for care while the number of taxpayers and the workforce of healthcare professionals is decreasing. After the development of the Norwegian welfare state during the last century, where the so-called Norwegian or Nordic model with equal rights to healthcare was established (Meld. St. 14 (2020–2021), p. 3), public healthcare services have become a significant charge to the national budget. Maintaining the standards of the Norwegian healthcare system is thus seen as a considerable challenge for the future economy (Finansdepartementet 2021). Based on this background, in the NHSP, the Norwegian Government presents AI technologies as a part of the solution for upholding the welfare state and mitigating healthcare struggles (e.g. by reducing the constantly growing workloads in the sector) (Meld. St. 7 (2019–2020)a). Thus, referring to such societal challenges can be seen as attempts to legitimise the government's vision of introducing AI in healthcare.

Heuristic guidance

In addition to legitimisation, the NHSP contains an explicitly formulated vision with clear elements of heuristic guidance. The vision provides direction for the ongoing introduction of AI in healthcare by highlighting specific goals and future roles and activities which may guide the choices and decisions made in the present processes:

> Artificial intelligence makes it possible to utilise national medical databases to provide faster and more accurate diagnostics, better treatment and a more effective use of resources. The regulations make it possible to use medical data to bring maximum benefit to the community, the directorates provide guidance on legal restrictions, and ethical problems associated with the use of artificial intelligence are handled in cooperation with other European countries. The health and care service has established a culture of innovation and knowledge-sharing, and works closely with businesses to develop the tools needed by the service and patients. (Meld. St. 7 (2019–2020)b, 26)

In this quote, the NHSP proclaims that AI will ensure (1) faster and more accurate diagnostics, (2) better treatment, and (3) more effective use of resources. These goals are to be achieved by a set of (guiding) activities such as: (a) providing national access to data, (b) offering support on legal issues (by the directorates), (c) solving ethical issues on a national level in cooperation with the EU, (d) establishing a culture of innovation and knowledge-sharing across the healthcare services, and (e) enabling the development of technologies in a close relationship between public and private sector.

Coordination

Finally, the NHSP describes concrete work to be coordinated and carried out by various stakeholders to reach the future with AI in healthcare. For instance, the plan states that it is crucial to adjust relevant national frameworks, such as the legal framework for sharing and using health data, to secure a safe introduction of AI technologies (Meld. St. 7 (2019–2020)a, 98). It also highlights that improving data quality is important, as, in their words, AI does not become more intelligent than the quality of the data allows (96). Furthermore, working on better access to data and storage capacity is acknowledged as essential for achieving optimal effects of the use of AI. The need to make the public specialist health services responsible for contributing to the realisation of a suitable infrastructure is further emphasised (96–97). Finally, the NHSP concludes that relevant government agencies are to assess the national framework conditions and coordinate further work to enable the use of AI in healthcare, i.e. the national coordination project (98).

The inquiry process: contesting informal anticipations and assessments

The above analysis shows how the NHSP set an overall direction for introducing AI in the Norwegian healthcare sector. However, as the inquiry process unfolded, elements that contested the formal vision emerged especially challenging the goals of using AI for more efficient use of resources, better treatment quality and patient safety. These contestations, presented in five categories below, were mainly based on anticipations and assessments of participants operating in or near clinical practices and concerned issues related to the specific contexts and conditions of future use of AI.

Inaccurate diagnostic outcome

The participants in the inquiry process were concerned about whether inaccurate algorithms might lead to over-diagnosing (finding too much), under-diagnosing (finding too little) or misdiagnosing. Even though the risk of making imprecise diagnoses is not unique to AI technologies, they worried that this would scale to another level of magnitude when complex algorithms were involved. The same concern also applied to algorithms trained on somehow biased datasets or from patient groups different from those belonging to a particular hospital. The meeting participants were concerned that such inaccurate results would produce more rather than less work (e.g. more tests and screenings of patients) and negatively affect the patients' trust.

Deskilling physicians

A second concern was that the use of AI technologies could potentially cause a deskilling of physicians such as radiologists. If AI is used for decision support, for instance, in image diagnostics, and the reasoning behind a diagnosis accordingly becomes 'black boxed', it was anticipated that the radiologists' expertise would decline. This issue especially concerned the expertise of young radiologists and whether they would suffer from no longer learning from the reasoning of and discussions with more experienced colleagues. The potential deskilling was additionally seen as something that could make the hospital system quite fragile and, over time, even more dependent on AI technologies. Some meeting participants also argued that a new kind of expertise was necessary, as physicians will have to know how AI technologies work before trusting them properly. They worried that a lack of knowledge and potential mistrust could lead to either no use or misuse. This issue, in turn,

raised another dilemma related to how physicians would have time to gain a basic understanding of AI in their already busy workdays.

Other competencies

The need for enhancing hospital managers' competencies was also brought up during the discussions: 'One of the greatest needs is to give the managers a more sober insight into what AI will bring in the medium-term future' (participant, feedback-meeting with radiology environments). The need for AI developers with knowledge of the work of physicians in clinical practices was also deemed important. Moreover, the question of whether introducing AI in clinical workflows would make it necessary to include other competencies in the clinics was raised. A meeting participant compared this with the automation of the bank sector, which allegedly led to even larger IT departments. Essential considerations mentioned in this regard were whether, for instance, the radiology resources spared are to be replaced by other types of staff and whether this is desirable for decision-makers aiming to save resources.

Overreliance on algorithms

During the meetings, AI technologies' (alleged) capabilities were criticised for 'presuming' that people are 100% rational. Participants used the phenomenon of decision fatigue as an example of the opposite, asking questions such as: what happens during the day as physicians become tired – will this affect how they use the AI technology? Will they become more capable of trusting it at the end of the day, or after months or years of experiencing certain reliability – would this lead to an overreliance, letting the AI technology make the final decisions? And what if the algorithms become skewed over time as they are exposed to new data – what happens then? These questions, taking as a starting point a view of AI as continuously learning technologies, evoked a perceived need for establishing work tasks to monitor algorithms.

Relevance and applicability of AI technologies

The importance of evaluating AI technologies in relation to local contexts and conditions was emphasised during some of the meetings, and several questions were raised in this regard: are there areas that could benefit from AI – is there a need? How will AI technologies affect existing workflows, and does the work really become more efficient for those using AI? The following statement exemplified a general concern that such aspects were not dealt with: 'It [the capabilities of AI] is exaggerated. We see the big headlines but forget to see what fits into clinical practices' (participant, feedback-meeting with radiology environments). Finally, the participants worried about the potential lack of ability to adjust CE-marked algorithms to local contexts and data due to regulations restricting such modifications of commercial products.

The final report: continued enactment of formal expectations

Despite the informal anticipations contesting and potentially threatening the materialisation of the formal expectations, the final report did not convert these uncertainties into areas for prioritised future investigations. The report suggests, however, several other activities to be carried out by the Directorate of Health and other public agencies, much in line with the activities highlighted in the NHSP. These activities included providing guidelines for procurement of CE-marked AI technologies in the healthcare sector, giving courses to enhance healthcare professionals' competencies within digitalisation and AI, stimulating collaboration across the public sector, ensuring sufficient access to quality data, and supporting hospitals on juridical and ethical issues (Helsedirektoratet et al. 2021, 48–57). Finally, based on the report and its recommendations for future work, the Norwegian Ministry of Health and Care Services stated in the yearly allocation letter for 2022 addressed to the Directorate of Health that they are to 'continue the work of adapting national framework conditions so that the health and care services can safely use Artificial Intelligence for patient treatment.

In 2022, especially interdisciplinary guidance related to legal issues is to be prioritised' (Helse- og omsorgsdepartementet 2022, 36, my translation). Thus, the focus area noted in the allocation letter also aligned with activities highlighted in the NHSP.

Discussion

Overall, the study shows that there are tensions between the formal expectations of AI in healthcare articulated by the Norwegian Government and the informal anticipations of those operating closer to clinical practice. The issues contesting the government's main goals give an example of how formal expectations will meet resistance during policy implementation, as pinpointed by Bareis and Katzenbach (2022). However, where Bareis and Katzenbach, within the scope of their article, purely stress that resistance will occur, this study takes it further, showing what such resistance or scepticism may look like. The present study also exemplifies how powerful visions, despite the resistance, continue to have a strong formative effect with elements of inevitability and path dependency. This conversion of expectations, based on vague definitions of AI into a seemingly inevitable technological pathway, offers an example of expectations provided with a protected space (Konrad 2006, 438).

The NHSP vision of AI in healthcare seems to have a strong influence on the interpretations of the inquiry process's outcome, confirming the vision's strength (Konrad 2006, 438). Thus, the three forces of expectations, as they appear in the NHSP, continue to have performative effects in Norway (Van Lente 2012). AI is still perceived as a solution to challenges in the healthcare sector (legitimation), and both the listed activities (heuristic guidance) and work to be coordinated (coordination) to reach the goals are repeated in slightly different shapes in the reports and plans developed in the wake of the inquiry process. Consequently, the outcome of the inquiry process contributes to 'lock-in' effects, too (Konrad 2006). Alternative paths are not yet considered, including an investigation of the contesting issues put forward by the participants in the inquiry process. Avoiding such explorations of potentially challenging factors in the early phases of policy implementation could lead to less successful outcomes in the long run.

However, implementing the NHSP is a work in progress, and the tensions surfacing in the findings may still be addressed in the future. At the same time, upcoming and unforeseen factors such as regulative changes or disappointing technological performances can still cause a weakening of the formal expectations, leading to a loss of the protected space (Konrad 2006, 441–442). Such a loss of protection, involving a diminishing formative strength, underscores that even the strongest expectations are not necessarily self-fulfilling prophecies (Pollock and Williams 2010). Nevertheless, resources that cannot be reused have already been invested in developing and implementing the NHSP, including organising the inquiry process, writing reports and carrying out other activities based on the report's recommendations. Thus, with the work already done, alternative routes for mitigating present and future challenges in the healthcare sector, with or without AI, may already be passed (Tutton 2017, 487).

This article has focused on an inquiry process carried out in the early phases of introducing AI in healthcare and some immediate consequences of the process' outcome. Therefore, insights into the formal expectations' 'temporal' and 'spatial' situatedness are limited (Brown 2003). However, regarding the temporal situatedness, the issues contesting the formal vision can be seen as cautions of potential future obstacles to consider to avoid taking the wrong course. The tension between the formal expectations and informal anticipations also provides examples of spatial situatedness. Several interpretations of the future with AI in healthcare exist, some more powerful than others. As the study shows, the formal expectations continue to have a performative effect, demonstrating that they have a more substantial power than the informal anticipations. However, as mentioned, despite the sense of inevitability this power indicates, changes may occur as the NHSP implementation proceeds or as new and updated policies for AI in healthcare are developed. Until then, a statement made by a participant in the inquiry process stands strong as an example of the current situation:

> People in the clinical practices know that something is coming; they just don't know, yet, how it will hit them.

The quote pinpoints that people *know* AI will become a part of the clinic, acknowledging AI as something destined for the healthcare sector. As such, it confirms the argument that expectations are 'performative' or even 'constitutive' (Borup et al. 2006; Konrad et al. 2016; Van Lente 2012). However, the second part of the statement underscores that *how* AI will hit the healthcare sector is still unknown. Thus, a paradox seems to emerge: Paired with the sense of inevitability is an element of fundamental uncertainty. To get a better understanding of how the future with AI can turn out to be, it is necessary to go beyond the hype and selling points. It requires more knowledge of what AI for healthcare is or can be and how concrete AI technologies perform in real-world clinical settings. As Bareis and Katzenbach (2022) indicated in their study, the introduction of vast AI programms requires a certain rhetorical force. The technological promise seems vast but is still vague. Thus, Bareis and Katzenbach address the need for powerful actors to 'talk AI into being', narratively constructing AI as both inevitable and disruptive. The rhetorical strategies of grand legacies and international competition that the authors identify in Chinese, US, French and German strategy documents may be less grandiose in the Norwegian documents, such as the NHSP. However, they still push a certain agenda forward.

As the above alludes to and previous studies have shown, the paths of emerging technologies typically start with high expectations without much knowledge of what the future will bring (van Merkerk and van Lente 2005, 1096). This is, without doubt, the case when it comes to AI in healthcare, which in recent years has been subject to extreme hype and, at the same time, advocated for by governments worldwide. Taken together with the ambiguous definitions of AI, affecting how a future with these technologies is perceived and the decisions drawn based on such differing interpretations, the current situation of inevitability may lead to several misguided actions. Combined with the existing chasm between AI development and deployment addressed in several research papers, where limited evidence of actual benefits in real-world settings is a highlighted barrier, it is hard to tell what the future will bring. The same goes for whether or how the formal expectations, such as those included in the NHSP and similar policies, will materialise.

Before anything materialises, the direction of introducing AI in the Norwegian healthcare sector will be guided by what is implied by the three forces of the expectations in the NHSP. This is how the mentioned paradox arises: AI in healthcare is forcing its way forward with tremendous impact, but what is the phenomenon really about beyond the different expectations and definitions? Phrased differently, as long as the introduction process follows the current path, the element of inevitability stays strong based on the present and powerful formal expectations of the Norwegian Government. We may term this an *inevitability paradox*, i.e. that a consensus about the value of AI in healthcare exists parallel to a fundamental uncertainty about what AI technologies for use in real-world settings will look like, when a broader deployment will occur and what will happen in its wake.

Concluding remarks

The Norwegian National Health and Hospital Plan 2020–2023 (NHSP) articulates an overall vision of a future where three main goals are fulfilled: the introduction of AI has ensured faster and more accurate diagnostics, better treatment, and more effective use of resources. To reach this future situation, the NHSP highlights the importance of access to quality data and storage, adapting or developing legal and ethical frameworks, and establishing collaborations across the public and private sectors. The policy plan also initiated a project where relevant government agencies were to assess the national framework conditions regarding the use of AI and coordinate work to enable healthcare services to start using AI technologies. The inquiry process informing the present study was a part of this national coordination work, through which the formal expectations and informal anticipations 'met' and elements contesting the future envisioned emerged. These elements included worries about inaccurate algorithms, a deskilling of physicians, lacking competencies, a potential overreliance on algorithms, and AI technologies not applying to local contexts.

The issues contesting the NHSP vision exemplify how current expectations of AI in healthcare can meet resistance as the implementation of AI strategies and policies proceeds. As of now, despite the varying perceptions of the future with AI, only one of them seems to set the agenda in the present Norwegian context. The inquiry process' final report and the allocation letter from the Norwegian Ministry of Health and Care Services did not address the issues potentially limiting the efficacy of the formal vision, showing that the initial expectations keep having formative effects. Thus, the formal expectations have the momentum and power to continue the process towards the desired and articulated future. This continuation exemplifies the presence of a sense of inevitability and path dependency, where alternative directions are easier to neglect. Finally, from the analysis of the NHSP vision, AI technologies are expected to solve several large-scale problems. But the vague definitions, the current chasm between the development and implementation of AI in healthcare, and the contesting elements highlighted in this study indicate that the road is paved with uncertainties. Will the contesting issues, or other obstacles emerging, eventually cause the formal expectations to collapse? Will negotiations be the case, or will the formal expectations and the elements of resistance gradually intertwine?

These and similar questions call for additional investigations of implementations of policies like the NHSP and other initiatives aiming to deploy AI in healthcare. Such examinations can illuminate further whether or how the expectations of AI will materialise and whether or how the contesting elements will assert themselves (and thus challenge the early impression of something inevitable).

Notes

1. For examples see: https://www.ibm.com/resources/watson-health/artificial-intelligence-impacting-healthcare/; https://www.siemens-healthineers.com/digital-health-solutions/digital-solutions-overview/clinical-decision-support/ai-rad-companion
2. For examples of other enactments of AI in healthcare visions see: https://www.ai.se/en/node/81535/information-driven-healthcare, https://fcai.fi/ai-for-health, https://www.nhsx.nhs.uk/ai-lab/
3. Radiology was prioritised as focus area as it was perceived as one of the most mature areas for AI adoption. Despite this focus, the project team argued that many of the issues uncovered during the inquiry process was relevant for other disciplines as well (Helsedirektoratet et al. 2021).

Acknowledgements

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Mari S. Kannelønning is a PhD student in Social Sciences within the area of Information Studies at Oslo Metropolitan University, Norway. Her research interests revolve around emerging information technologies and the introduction of artificial intelligence (AI) in healthcare in particular. In her doctoral project, she specifically explores the expectations of AI in healthcare, the promissory work of AI suppliers, and the formation of networks of professionals working to enable AI in healthcare.

References

Aung, Y. Y. M., D. C. S. Wong, and D. S. W. Ting. 2021. "The Promise of Artificial Intelligence: A Review of the Opportunities and Challenges of Artificial Intelligence in Healthcare." *British Medical Bulletin* 139 (1): 4–15. https://doi.org/10.1093/bmb/ldab016.

Bareis, J., and C. Katzenbach. 2022. "Talking AI Into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics." *Science, Technology, & Human Values* 47 (5): 855–881. https://doi.org/10.1177/01622439211030007.

Beckert, J. 2016. *Imagined Futures: Fictional Expectations and Capitalist Dynamics*. Cambridge: Harvard University Press.

Borup, M., N. Brown, K. Konrad, and H. Van Lente. 2006. "The Sociology of Expectations in Science and Technology." *Technology Analysis & Strategic Management* 18 (3–4): 285–298. https://doi.org/10.1080/09537320600777002.

Brown, N. 2003. "Hope Against Hype - Accountability in Biopasts, Presents and Futures." *Science & Technology Studies* 16 (2): 3–21. https://doi.org/10.23987/sts.55152.

Brown, N., and M. Michael. 2003. "A Sociology of Expectations: Retrospecting Prospects and Prospecting Retrospects." *Technology Analysis & Strategic Management* 15 (1): 3–18. https://doi.org/10.1080/0953732032000046024.

Brown, N., B. Rappert, and A. Webster. 2000. "Introducing Contested Futures: From Looking Into the Future to Looking at the Future." In *Contested Futures: A Sociology of Prospective Techno-Science*, edited by N. Brown, B. Rappert, and A. Webster, 3–20. Aldershot: Ashgate.

Cabitza, F., A. Campagner, and C. Balsano. 2020. "Bridging the "Last Mile" Gap Between AI Implementation and Operation: "Data Awareness" That Matters." *Annals of Translational Medicine* 8 (7): 501. https://doi.org/10.21037/atm.2020.03.63.

Coiera, E. 2019. "The Last Mile: Where Artificial Intelligence Meets Reality." *Journal of Medical Internet Research* 21 (11): e16323. https://doi.org/10.2196/16323.

Eames, M., W. Mcdowall, M. Hodson, and S. Marvin. 2006. "Negotiating Contested Visions and Place-Specific Expectations of the Hydrogen Economy." *Technology Analysis & Strategic Management* 18 (3–4): 361–374. https://doi.org/10.1080/09537320600777127.

Emerson, R. M., R. I. Fretz, and L. L. Shaw. 2011. *Writing Ethnographic Fieldnotes*. 2nd ed. Chicago: University of Chicago Press.

Finansdepartementet. 2021. "Hovedpunkter i Perspektivmeldingen 2021." https://www.regjeringen.no/no/tema/okonomi-og-budsjett/norsk_okonomi/perspektivmeldingen-2021/hovedpunkter-i-perspektivmeldingen-2021/id2834526/.

Freeman, K., J. Geppert, C. Stinton, D. Todkill, S. Johnson, A. Clarke, and S. Taylor-Phillips. 2021. "Use of Artificial Intelligence for Image Analysis in Breast Cancer Screening Programmes: Systematic Review of Test Accuracy." *BMJ* 374:n1872. https://doi.org/10.1136/bmj.n1872.

Geels, F. W., and W. A. Smit. 2000. "Failed Technology Futures: Pitfalls and Lessons from a Historical Survey." *Futures* 32 (9-10): 867–885. https://doi.org/10.1016/S0016-3287(00)00036-7.

Helsedirektoratet, Direktoratet for e-helse, Statens legemiddelverk, Helse Nord RHF, Helse Vest RHF, Helse Midt-Norge RHF, & Helse Sør-Øst RHF. 2021. "Tilrettelegging for bruk av kunstig intelligens i helsetjenesten." https://www.helsedirektoratet.no/rapporter/tilrettelegging-for-bruk-av-kunstig-intelligens-i-helsesektoren-ny-01.10.2021.

Helse- og omsorgsdepartementet. 2022. "Tildelingsbrev til Helsedirektoratet for 2022." Oslo, Norge https://www.regjeringen.no/globalassets/departementene/hod/tildeling-oppdrag-og-arsrapporter/2022/tildelingsbrev-til-helsedirektoratet-for-2022.pdf.

Hsieh, H.-F., and S. E. Shannon. 2005. "Three Approaches to Qualitative Content Analysis." *Qualitative Health Research* 15 (9): 1277–1288. https://doi.org/10.1177/1049732305276687.

Knowledge4Policy. 2021. "How Can Europe Become a Global Leader in AI in Health? Knowledge for Policy." https://knowledge4policy.ec.europa.eu/file/how-can-europe-become-global-leader-ai-health_en.

Konrad, K. 2006. "The Social Dynamics of Expectations: The Interaction of Collective and Actor-Specific Expectations on Electronic Commerce and Interactive Television." *Technology Analysis & Strategic Management* 18 (3-4): 429–444. https://doi.org/10.1080/09537320600777192.

Konrad, K., H. Van Lente, C. Groves, and C. Selin. 2016. "16 Performing and Governing the Future in Science and Technology." In *The Handbook of Science and Technology Studies*, edited by U. Felt, R. Fouché, C. A. Miller, and L. Smith-Doerr, 465–493. Cambridge: MIT Press.

Krafft, P., M. Young, M. Katell, K. Huang, and G. Bugingo. 2020. "Defining AI in Policy Versus Practice." Paper Presented at the Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.

Meld. St. 14. 2020–2021. "Long-term Perspectives on the Norwegian Economy 2021." https://www.regjeringen.no/contentassets/91bdfca9231d45408e8107a703fee790/en-gb/pdfs/stm202020210014000engpdfs.pdf.

Meld. St. 7. (2019–2020)a. "Nasjonal Helse- og Sykehusplan 2020–2023." https://www.regjeringen.no/contentassets/95eec808f0434acf942fca449ca35386/no/pdfs/stm201920200007000dddpdfs.pdf.

Meld. St. 7. (2019–2020)b. "National Health and Hospital Plan 2020–2023." Summary. https://www.regjeringen.no/contentassets/95eec808f0434acf942fca449ca35386/en-gb/pdfs/stm201920200007000engpdfs.pdf.

Morley, J., C. C. Machado, C. Burr, J. Cowls, I. Joshi, M. Taddeo, and L. Floridi. 2020. "The Ethics of AI in Health Care: A Mapping Review." *Social Science & Medicine* 260: 113172. https://doi.org/10.1016/j.socscimed.2020.113172.

The National Strategy for AI in Health and Social Care. n.d. https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/the-national-strategy-for-ai-in-health-and-social-care/.

Pollock, N., and R. Williams. 2010. "The Business of Expectations: How Promissory Organizations Shape Technology and Innovation." *Social Studies of Science* 40 (4): 525–548. https://doi.org/10.1177/0306312710362275.

Rajpurkar, P., E. Chen, O. Banerjee, and E. J. Topol. 2022. "AI in Health and Medicine." *Nature Medicine* 28 (1): 31–38. https://doi.org/10.1038/s41591-021-01614-0.

Seneviratne, M. G., N. H. Shah, and L. Chu. 2020. "Bridging the Implementation Gap of Machine Learning in Healthcare." *BMJ Innovations* 6 (2): 45–47. https://doi.org/10.1136/bmjinnov-2019-000359.

Thedvall, R. 2013. "Punctuated Entries. Doing Fieldwork in Policy Meetings in the European Union." In *Organisational Anthropology: Doing Ethnography in and among Complex Organisations*, edited by A. Nyqvist and C. Garsten, 106–119. London: Pluto Press.

Tutton, R. 2017. "Wicked Futures: Meaning, Matter and the Sociology of the Future." *The Sociological Review* 65 (3): 478–492. https://doi.org/10.1111/1467-954X.12443.

Van Lente, H. 2012. "Navigating Foresight in a sea of Expectations: Lessons from the Sociology of Expectations." *Technology Analysis & Strategic Management* 24 (8): 769–782. https://doi.org/10.1080/09537325.2012.715478.

Van Lente, H., and A. Rip. 1998. "Expectations in Technological Developments: An Example of Prospective Structures to be Filled in by Agency." In *Getting New Technologies Together: Studies in Making Sociotechnical Order*, edited by C. Disco and B. van der Meulen, 203–230. Berlin/Boston: De Gruyter, Inc.

van Merkerk, R. O., and H. van Lente. 2005. "Tracing Emerging Irreversibilities in Emerging Technologies: The Case of Nanotubes." *Technological Forecasting and Social Change* 72 (9): 1094–1111. https://doi.org/10.1016/j.techfore.2004.10.003.

Verdegem, P., ed. 2021. *AI for Everyone? Critical Perspectives*. London: University of Westminster Press.

Wang, P. 2019. "On Defining Artificial Intelligence." *Journal of Artificial General Intelligence* 10 (2): 1–37. https://doi.org/10.2478/jagi-2019-0002.

Article 3

Navigating uncertainties of introducing artificial intelligence (AI) in healthcare: The role of a Norwegian network of professionals

Mari S. Kannelønning

Faculty of Social Sciences, Oslo Metropolitan University, Postboks 4, St. Olavs plass 0130, Oslo, Norway

ARTICLE INFO

ABSTRACT

Artificial Intelligence (AI) technologies are expected to solve pressing challenges in healthcare services worldwide. However, the current state of introducing AI is characterised by several issues complicating and delaying their deployments. These issues concern topics such as ethics, regulations, data access, human trust, and limited evidence of AI technologies in real-world clinical settings. They further encompass uncertainties, for instance, whether AI technologies will ensure equal and safe patient treatment or whether the AI results will be accurate and transparent enough to establish user trust. Collective efforts by actors from different backgrounds and affiliations are required to navigate this complex landscape. This article explores the role of such collective efforts by investigating how an informally established network of professionals works to enable AI in the Norwegian public healthcare services. The study takes a qualitative longitudinal case study approach and is based on data from non-participant observations of digital meetings and interviews. The data are analysed by drawing on perspectives and concepts from Science and Technology Studies (STS) dealing with innovation and sociotechnical change, where collective efforts are conceptualised as actor mobilisation. The study finds that in the case of the ambiguous sociotechnical phenomenon of AI, some of the uncertainties related to the introduction of AI in healthcare may be reduced as more and more deployments occur, while others will prevail or emerge. Mobilising spokespersons representing actors not yet a part of the discussions, such as AI users or researchers studying AI technologies in use, can enable a 'stronger' hybrid knowledge production. This hybrid knowledge is essential to identify, mitigate and monitor existing and emerging uncertainties, thereby ensuring sustainable AI deployments.

1. Introduction

Artificial Intelligence (AI) technologies are perceived as having a great potential for solving existing and future challenges within healthcare services, including rising costs, shortages in the healthcare workforce, and the exponential growth of digitalised health data necessary to process and manage [1–3]. Existing AI technologies are, however, not a clearly defined group of technologies but are developed for various use areas with different capabilities and outcomes [4,5]. AI technologies currently being developed for healthcare and slowly taken into use (as of 2023) are typically based on Machine Learning (ML) or Deep Learning (DL) approaches, with image analysis as one of the most promising areas of application ([6]; p. 722 [7]; p. 293; [8]).

The Norwegian Government's vision of AI in healthcare gives a typical example of the current expectations, stating that, in the future, AI

technologies will 'provide faster and more accurate diagnostics, better treatment and a more effective use of resources' ([9]; p. 26).[1] The vision further emphasises that mobilising and establishing collaborations between various actors from the public and private sectors is crucial to enable AI in healthcare (for similar arguments, see [10]; p. 7; [11]). According to research on AI implementation in healthcare, such constellations of actors will need to cover a 'last mile' or bridge a 'gap' to progress towards widespread deployment [12–14]. Problematic issues to address in this context relate to topics such as ethics, regulations, data access, human trust, and limited evidence of AI performance in real-world clinical settings [6,10,15,16]. Some of the key challenges involve uncertainties regarding whether AI technologies will secure equal and safe patient treatment or whether the machine learning decisions have the necessary transparency and explicability essential in diagnostic processes and to ensure user trust.

In this article, I explore how a specific constellation of actors attempts to address and navigate the many issues and uncertainties characterising the current situation of introducing AI in the context of the Norwegian public healthcare services. Specifically, the research takes a qualitative case study approach to examine the so-called KIN network (Kunstig Intelligens i Norsk Helsetjeneste – Artificial Intelligence in the Norwegian Healthcare Services), an informally established network consisting of professionals with different backgrounds and affiliations with interests in AI and healthcare. The network aims to contribute to the ongoing work of enabling AI in real-world clinical settings by facilitating knowledge and experience sharing among its members and interacting with decision-makers, such as politicians, different authorities and hospital managers.

The article draws on concepts from Science and Technology Studies (STS) dealing with innovation and sociotechnical change as coming about through collective efforts of heterogeneous actors (see, for example, [17–20]. With the KIN network as an object of study, the article suggests that an entry point to understanding the challenges of introducing AI in healthcare is to look at the issues concerning AI deployment addressed by such actor constellations. The rationale behind this approach is inspired by Callon et al., who argue that exploring controversies 'allows an inventory to be made of the different dimensions of what is at stake in a project' ([21]; pp. 29–30). Similarly, in the case of this article, exploring how the KIN network addresses the challenges and uncertainties of AI deployment can provide an inventory of what is at stake in the current introduction process. Such an analysis may also render visible elements not taken into account by more overall and generic expectations of AI, like the one of the Norwegian Government. Exploring uncertainties related to emerging complex technologies is crucial to delineate what may be overlooked in innovation processes but nevertheless are essential to identify, mitigate and monitor in order to achieve sustainable solutions.

More specifically, the article addresses the following research questions: *how can an informally established network of professionals, like the KIN network, contribute to enabling AI in healthcare?* To discuss this topic, I seek answers to the following sub-questions: *How does the KIN network characterise its purpose and role, and how does it go about achieving its aims?*

The article starts with an outline of the theoretical framework, followed by a presentation of the research methodology, including a description of the case, the research process and the data analysis. Subsequently, the article presents the findings, followed by a discussion and final remarks.

2. Innovation processes as collective activities

Scholars within Science and Technology Studies (STS) have argued that innovation and sociotechnical change result from the interactions among actors with different characteristics (see, for example, [17,18,20, 22]. Through the KIN network, human actors from different backgrounds and affiliations come together to share their knowledge and experience. Such heterogeneously organised knowledge production, happening across otherwise typically distinct boundaries between disciplines, sectors, and organisations, is what Gibbons et al. [23] refer to as a 'hybridisation'. This hybridity may have an effect leading to the establishment of actor constellations called 'hybrid forums', which again 'reflects the need of different communities to speak in more than one language in order to communicate at the boundaries and in the spaces between systems and subsystems' ([23]; p. 38). Similarly, Callon et al. [21] describe hybrid forums as 'open spaces where groups can come together to discuss technical options involving the collective' and where 'the groups involved and the spokespersons claiming to represent them are heterogeneous, including experts, politicians, technicians, and laypersons who consider themselves involved' (p. 18). Callon et al. emphasise further that the aspect of hybridisation also relates to the type of questions and problems discussed in such forums, which are related to a variety of domains and addressed at different levels. Moreover, a

common trait of these hybrid forums is that they often emerge due to unpredictable and messy advancements in science and technology, and they are suitable ways to manage or accommodate the uncertainties generated by such advancements ([21]; p. 18). In this context, uncertainties may be seen not only as the reasons for certain group constellations being established but also as a great motivation for seeking and gaining more knowledge ([24]; p. vi).

Aiming for heterogeneity in group constellations established for influencing societal transition processes is not uncommon in Nordic countries. It aligns well with 'consensus-building' being one of the key pillars of their political systems ([25]; p. 18). This consensus orientation places the Nordic countries as reformers somewhere in between the slower-moving systems of, for instance, Germany and systems which are able to keep a higher speed, such as in the UK. Even though the Nordic system of ensuring consensus takes time, and the result is not necessarily radical, the chances of implementing more sustainable solutions are often higher [26]. Aiming for consensus and sustainability corresponds, furthermore, well with the Nordic countries' AI strategies, which have a particular focus on AI for sustainable societies [27,28].

Within political science, another type of group formation is described as 'interest groups' [29]. These are, like hybrid forums, positioned between systems and subsystems. However, such constellations are typically not open spaces, nor do they include the same variety of member types. The members of interest groups are mainly spokespersons for the members of the organisations they represent; they do not necessarily represent themselves or their like. The kind of 'in the middle' position taken by interest groups reminds us how they are intermediaries, acting between 'networks' in the political science understanding of the term (see, for example [30], and the subgroups that have specific interests within the particular policy area. Thus, interest groups are typically associated with advocacy organisations (e.g., Doctors without Borders) or professional associations and/or trade unions (e.g., country-specific medical associations). Hence, interest groups aim not to create knowledge across, for instance, institutional boundaries, such as in hybrid forums, but help to 'facilitate relationships between actors that would otherwise have difficulty relating to one another' ([29]; p. 433).

Scholars associated with one of the most well-known STS frameworks, Actor-Network Theory (ANT), offer another perspective on collective activities. Akrich et al. [17] argue that innovation processes and their outcome are shaped by the number of actors that the project at hand is able to mobilise and by the character of the interactions between these actors. In this context, the actors can be both human and non-human (e.g., AI technologies). Thus, an essential element for mobilising the necessary actors and gaining enough power to proceed with innovation is to make and keep relevant actors interested in the project ([17]; p. 205). If the project, exemplified by the KIN network in the present study, succeeds in keeping the actors interested and establishing the necessary alliances, this confirms the validity of the project's aim or foundational principle.

Callon's [18] influential depiction of the domestication of scallops and fishermen of St Brieuc Bay gives several examples of how alliances between actors are established but also threatened and how this affects an innovation process. An essential point in Callon's story is how non-human actors (e.g., AI technologies) are entities important to involve in the ongoing processes and keep as allies, in the same way as human actors (e.g., physicians as AI users). However, as non-human actors like AI technologies cannot speak for themselves, they are given a voice through other actors who bring them into the conversations (e.g., AI researchers, vendors, users or researchers studying AI in use). Thus, considering who speaks in the name of whom is essential, which also includes paying attention to the distinction between spokespersons and representatives ([18,31]; p. 216). For instance, in the case of this study, AI researchers or AI vendors, who are currently the most knowledgeable regarding the capabilities of AI technologies, may appear as the AI technologies' main spokespersons. As these spokespersons have their own interests or agencies, such as promoting their research or selling

their products, they will most likely represent AI in a certain (optimistic) way. These interests influence, furthermore, which kind of information is shared and circulated within the project at hand–the KIN network and beyond. This distribution of power raises questions about which versions of AI the existing alliances that collectively aim to enable AI in healthcare are built upon and how this assemblage of AI versions and their beholders affects the outcome of the ongoing introduction process.

The KIN network's aim of facilitating the sharing of knowledge and experience among professionals with a focus on the issues and uncertainties complicating the introduction of AI in healthcare does not only underscore the network's *raisons d'être* but also that AI, as a sociotechnical phenomenon, is comprehensive and challenging to grasp, or worse still, to master. One significant consequence of an ambiguous concept such as AI [5] is that it complicates the ongoing conversations as people talk about AI from various perspectives and perceptions without knowing whether they are 'on the same page'. For instance, questions such as whether AI will make healthcare services more efficient will give several different answers, depending on whose opinion is asked and in what context ([5]; p. 28). Another outcome of the vast AI concept is that the flexibility of the term enables multiple actors to form their own expectations and develop their expertise accordingly. If this expertise, building on certain expectations, is broadly and convincingly shared, this could set an agenda other stakeholders will follow ([32]; p. 139).

These examples of potential consequences of the vague and flexible AI term show that there are rhetorical aspects of importance to be aware of while studying AI conversations. As an apropos, and perhaps slightly ironic, as AI is a broadly defined concept and treated accordingly in the conversations observed and the interviews conducted during this study, the generic and overall terms 'AI' and 'AI technologies' are used interchangeably in this article. These terms cover AI in both a broad and narrow sense and simultaneously reflect how it is used within the network studied and by the informants.

3. Research methodology

This article is based on a qualitative longitudinal case study approach, including data from non-participant observations of the KIN network's meetings, conferences and seminars, and semi-structured interviews with the network's secretariat.

3.1. Case description

In late 2020, the KIN network was established in Norway as a nationwide initiative to increase AI deployments in real-world clinical settings. The initiative came about through informal conversations among peers in relation to the first national conference on AI in healthcare for managers and clinicians in the Norwegian healthcare services, organised in Bodø in 2019. A year later, some of the initiators became a part of a secretariat of eight people, managing the network and facilitating its activities. The secretariat consisted of:

- a researcher from a research department at a private limited company
- a department manager from a public agency
- two managers from a national research centre
- a senior adviser from one of the four Norwegian regional health authorities
- a department manager from one of the four Norwegian regional health authorities

The activities carried out by the network are quarterly meetings organised by the secretariat and three to four yearly seminars or conferences organised by network members representing different geographical areas and institutions. The secretariat sets the meeting agendas inspired by input from the network members, while the local organisers develop the conference programmes.

The KIN network has no limitation regarding the number of members. As such, everyone interested can join the network, including, but not limited to, professionals with backgrounds in medicine, mathematics/statistics/physics, machine learning, health economics, healthcare research, pharmacy, and social sciences [33]. The members are typically recruited through colleagues, other acquaintances or a sign-up option on the network's website. During the two years this study was carried out, the member list grew from approximately 20 participants at the first official meeting to 160 members at the end of 2022. The largest group of members are researchers from fields within the hard sciences (e.g., informatics, machine learning in particular and medical physics) from universities, research centres or hospitals. Other groups are hospital employees and bureaucrats with backgrounds in medicine or technology, working at the intersection between technology and healthcare. Finally, some of the smaller groups are people from the industry, mainly from software companies, interest organisations, and municipalities. Additionally, in February 2023, the network had 32 observers, including the author of this article and project managers or senior advisers from affiliations such as the Norwegian Directorate of Health, the Directorate of E-health, the Board of Health Supervision, and the Board of Technology which supports political decision-making processes.

3.2. Research process

I was enrolled as an observer of the KIN network just in time to participate in the second official meeting in February 2021. For two years, until December 2022 and with a short revisit at a meeting in May 2023, I observed their quarterly digital meetings (eight altogether) and three of six conferences/seminars (two in-person). As I attended these activities, I especially paid attention to the people participating in the discussions and giving presentations, who they were and what they talked about. During these observations, I mainly observed the ongoing verbal communication unfolding and the visual presentations. As I did not know beforehand who would attend the meetings, getting consent from all participants to record the discussions was impossible. Therefore, the primary source of documentation was fieldnotes taken during the meetings. At two of the conferences I attended in person, I also conversed with participants during coffee breaks and meals, writing fieldnotes afterwards when by myself.

Furthermore, in Autumn 2022, I interviewed six of the eight secretariat members who had been part of the secretariat from the very beginning. The interviews lasted between 30 and 60 min and revolved around topics such as how the network was established, the aim and role of the network, who the members were, whom the network collaborated with, how the two years with the network had been, which kind of AI they saw as mediated through the network, and what they perceived as current achievements, challenges, and relevant work to carry out in the future. These interviews were recorded and fully transcribed.[2]

As secondary material, I examined the PowerPoint presentations from the presentations given during the meetings. These also acted as the network's meeting minutes, as the secretariat added a few comments to the presentations after the meetings. I also studied the conference programs and the information and documents published on the website (for website URL see, [34]). The website provided information about the network, the secretariat, the members and the membership policy. It also included a list of ongoing AI projects in Norway (primarily research and development projects at hospitals) and an overview of the network's activities.

_____

[2] The research project and its data management practice, including anonymisation procedures and data storage, was approved by the Norwegian Centre for Research Data (now called SIKT – the Norwegian Agency for Shared Services in Education and Research). The interview informants were informed about the procedures and signed a consent form.

3.3. Data analysis

The transcribed interviews were analysed in three phases. First, they were subject to explorative open coding, where they were read line-by-line and different themes were highlighted. Subsequently, the text was reread and coded with more specific codes ([35]; p. 172). As similar codes were grouped, overall categories emerged, such as 'the network's purpose', 'role of network', 'members', 'AI in healthcare challenges', 'topics discussed', 'agendas', 'challenges for network', 'interaction with other stakeholders' and 'influence in the field'. Lastly, a more abductive process was conducted as the categorised data were considered and refined based on the issues foregrounded by the research questions, which took shape after the first two analysis phases. This process resulted in two main categories: 'the network according to the network' and 'the network's activities, agendas and influence'. The fieldnotes, meeting minutes, conference programmes and website information were further conferred for elements supporting or supplementing the findings from the interviews. An example of such elements is the description of the network from the website included below.

4. Findings

The findings presented in this section are organised after the two categories resulting from the data analysis, which also reflects the topics of the research sub-questions.

4.1. The network according to the network

On the KIN network's website, the network is described as follows:

KIN is a national network for artificial intelligence in the healthcare service, which consists of various professional communities from all over the country. The network takes a bottom-up approach and aims to share experiences and put important issues concerning the clinical implementation of artificial intelligence on the agenda. *We connect professional communities by establishing meeting places for joint discussion and exchanging knowledge about implementing artificial intelligence in the healthcare sector. The network is open to anyone who wants to participate and share their work* (excerpt from the KIN network's website, my translation)

This quote describes the network's intentions as three-folded: to share experience and knowledge, set central issues on the agenda and connect people from different fields, all based on a 'bottom-up' approach.

Apart from being open to those who want to participate, the network's membership policy underscores that membership is personal, meaning that the participants represent themselves and not their affiliated organisations [33]. During the interviews, it was argued that being a personal member made the work and discussions easier as the members did not need permission from their managers to participate in the network or have a particular opinion on the topics discussed. The conversations unfolding during the meetings could thereby proceed more freely, without too many restrictions and reservations. It was further argued that this made the membership and the network more informal and, at the same time, aligned with the 'bottom-up' approach the network wished to take.

The use of the 'bottom-up' term to describe the approach was repeatedly emphasised during the introduction of the meetings. The logic behind the approach was explained by an informant as follows:

It's quite simple: keeping the focus on what the people with knowledge of AI in healthcare perceive as difficult or useful, what the solutions to the problems are, and so on, will result in a more professional-oriented agenda *than a political one. If we [the network] find different potential [in AI], which the top management either doesn't see, doesn't include in strategies or doesn't make decisions about … then you have to do what*

you normally do: try to let it [the knowledge] trickle upwards in the system, talk to your bosses and so on (a secretariat member, my translation)

This statement shows not only that individualised expert knowledge is essential to the network but also that the knowledge and expertise should be channelled (or trickle) further 'up' the system, letting the knowledge of the professionals inform the agenda in these circles, too. The informant argued moreover that this way of sharing information is the best way to establish proper foundations for important and strategic decisions in the case of AI:

To the extent that we stumble upon gold, right, we must tell our managers what it is so that they can act on it. In the boardrooms and in top management positions, you don't automatically know everything that happens at the grassroots level (a secretariat member, my translation)

Through these two latter quotes, we find examples of perceptions of how decision-makers need help from the knowledge of experts to act and make decisions related to the introduction of AI in healthcare. Similarly, another informant remarked that 'the bureaucracy', which the informant claimed was too distant from the dynamics in the professional communities, needed knowledge from the professionals to develop proper policies: 'It is hard to imagine knowledge-based policy in this area, without the professional communities being very actively mobilised'. Furthermore, this quote underscores that the network perceives the mobilisation of several actors as an essential factor in the early phases of introducing novel technologies.

There was also another reason why a network like the KIN network was perceived as necessary. As an informant argued, no healthcare organisation had come very far in deploying AI, which again made it important that the professionals had the ability to discuss and share knowledge in 'informal grassroots networks'. The use of the term 'grassroots' in this context can also be seen as a way of positioning the network as a counterpart to the more 'top-down' initiatives in the healthcare sector and, again, aligning with their 'bottom-up' approach. The characteristic of being an informal network was further stressed by referring to the fact that they did not have a budget or mandate from elsewhere: 'We don't have a budget or anything, right … people do it almost on a voluntary basis' (a secretariat member).

Prompted by a question of what the KIN network had achieved in the two years it had existed, an informant stated that it had become a sort of 'gravity centre' for many of the ongoing processes introducing AI in the healthcare services. Another informant remarked that: 'it has become a way of having a dialogue with the authorities, it has become a channel where people can give input [to the authorities], and it is easier for them [the authorities] to ask the KIN network [for input or feedback on certain topics]'. Thus, the network was described as a 'dialogue partner' but also as an 'expert group' or 'catalyst'. All of these descriptive terms, ranging from 'grassroots' to 'gravity centre' and 'catalyst', give quite different interpretations of the network's role: from being something that covers the ground, ensuring a solid fundament for knowledge to grow, to a centre from where different projects are created, start evolving and accelerating.

4.2. The network's activities, agendas and influence

The means to achieve the network's goals of knowledge and experience sharing were mainly the organisation of 'a series of focused seminars/conferences where professionals from relevant communities meet to share experiences through presentations and discussions' (excerpt from meeting minutes, 27. November 2020, my translation). On these occasions, both network members, observers and invited non-members contributed as presenters or participants in panel discussions. These contributors were typically from the research and hospital environments but also from the industry, legal experts, the cancer registry, the four regional health authorities, the data protection authority, the

Directorate of Health, the Directorate of e-health, and the Ministry of Health and Care services.

An informant explained that the meetings should reflect what most members agreed upon as necessary topics to discuss, adding the question: 'What are the most important and difficult topics that must be resolved in this area?'. Thus, to a certain degree, the meeting agendas were informed by the result of digital polls conducted at the end of each meeting. Through these polls, called 'temperature checks', the members could vote for topics they perceived necessary to address or crucial to discuss in future meetings. The polls consisted primarily of pre-defined topics for which the meeting participants could vote. However, at some meetings, suggesting other issues through a free-text option was also possible. A third way for the members to influence the meeting agendas was to contact the secretariat directly with ideas or wishes. Incoming suggestions were typically transformed into pre-defined topics, which were added to the polls for the upcoming meetings. Thus, the list of categories could change slightly from meeting to meeting.

The predefined topics of the polls typically included categories such as:

- 'Validation and adjustment of AI solutions to local conditions (incl. Norwegian patient groups)'
- 'Ethics and legislations regarding AI (bias in data, black box, responsibility)'
- 'Clinic and users – decision support and communication of uncertainties [the system's output provided for the clinicians, informing them about the accuracy or quality of the AI result]'
- 'How to make AI solutions as beneficial for the clinics as possible'
- 'Make or buy [develop AI in-house or buy commercially available products]'
- 'Competence development within the sector'
- 'Cloud solutions for data sharing'
- 'Harmonization/data quality'
- 'Policy and financial incentives'
- 'Validation of continuous learning technologies'
- 'Infrastructure and cyber security'
- 'Consequences of new rules for certification and approval of AI solutions [how to obtain CE-marking following the updated Medical Device Regulations, 2021]'. (a synthesis of categories from the different meeting presentations, my translation and additional explanations in brackets)

The upper three categories of this list were rated as the top three of the most desired topics for discussion during the two years the meetings were observed. As I attended a meeting five months after I completed the data collection, this ranking was still the case. Although the topic of this particular meeting was 'Quality assurance and validation of AI', the same topic was voted for as the number one topic for future discussion at the end of the meeting. Based on this ranking, it can be argued that areas perceived by the network members as involving most uncertainties are all related to the deployment of AI in real-world clinical settings and key issues particular to AI technologies for healthcare. The uncertainties included questions such as: how can or should healthcare providers validate or test AI technologies before deployment in clinical practices,[3] will it be possible to adjust the technologies according to local conditions such as particular clinical workflows, procedures and patient data, will the outcome of use be fair and safe for all patients, who is responsible if errors occur and how can the clinicians be sure of the accuracy and quality of the AI result?

Apart from the meetings and conferences, the network members and

especially the secretariat attended or carried out additional activities to influence politicians, authorities and other decision-makers. These were activities organised outside the borders of the defined network (the regular meetings and conferences), such as workshops on AI adoption organised by a national coordination project also aiming to enable AI in healthcare initiated by a policy plan of the Norwegian Government [9] (cf., [36]). The secretariat also wrote a letter to the Ministry of Health and Care Services stressing the need for a continued focus on enabling AI adoption, and they organised an event at the largest and most important yearly political gathering in Norway, Arendalsuka, in 2022. At this event, they reached out to politicians and health authorities as they argued for a need for a national 'roadmap' to overcome many of the current issues of introducing AI in healthcare services.

Another type of interaction between the network and the authorities happened as representatives from the authorities gave presentations at several of the network meetings about topics such as access to data for AI use in healthcare or about activities carried out on a national level to ease the introduction of AI in clinical practices, including guidance on juridical issues. Based on the authorities' presence at various meetings either as presenters or observers, an informant claimed that the network had become valuable for the authorities as they could both 'inform and obtain knowledge for their own part'.

However, some of the informants also questioned the actual influence of the KIN network. For instance, regarding the final decision-making, one informant remarked that no matter what knowledge or recommendations the network shared with the health authorities, 'the government will do as it pleases'. Another informant questioned whether the knowledge shared and accumulated through the network would reach the clinical practices and further benefit the patient treatment. The informant elaborated on this issue by relating it to the Norwegian tradition of organising hospitals into what the informant called 'silos'; one silo for research and another for patient care, between which few bridges exist for knowledge transactions. The informant explained this as follows:

> … one of the great weaknesses of the Norwegian public hospital sector is that it is rigged, as in the 1970s. It's rigged for streamlining patient care on the one hand and research on the other, separately. The latter produces knowledge, preferably by studying patient care. When such knowledge is established, to a sufficient extent, there are fragile mechanisms for deploying it into the clinic and further changing the clinic (a secretariat member, my translation).

The informant continued to elaborate on how these silos also challenge the introduction of AI. As the informant stated, AI is a kind of information technology which is even farther away from patient care than the research carried out in the research silo. The research is at least related to specific patient groups and, therefore, has a connection to the hospitals' clinical practices. In the case of AI, however, the informant perceived this distance or lack of access to clinical settings as a reason why it is hard to find out whether or how AI technologies can benefit clinical work.

5. Discussion

5.1. Hybrid knowledge production to enable AI in healthcare

As the KIN network is open and aims to mobilise different actors interested in AI and healthcare from across disciplines, sectors and organisations, conditions are established for enabling hybrid knowledge production, as described by Gibbons et al. [23]. However, despite the network's attempts to mobilise members representing diverse expertise and affiliations, most of them were AI researchers from the research and university sectors and hospitals' research units, with a background in fields like informatics, machine learning and medical physics. This group was followed by smaller groups of bureaucrats working within areas of healthcare and technology, hospital employees working on

---

[3] For insights on what a validation of AI for use in healthcare might imply, see [44]; which includes an overview of the British standard for a validation framework, informed by multiple and various experts and public consultations.

technology-related projects and representatives from the AI or software industry. This distribution of members raises questions about whether the actors mobilised through the network are heterogeneous enough; are they able to produce the hybrid knowledge necessary to gain a better understanding of the many uncertainties concerning the introduction of AI in healthcare and, subsequently, contribute to ensuring sustainable AI deployments? It also raises questions about who speaks in the name of whom [18,31] and what the current selection of spokespersons might imply for the knowledge shared within the network and the network's contribution to the introduction of AI in the Norwegian public healthcare services.

From the current member list, it can be argued that the majority of the members who speak on behalf of AI technologies have interests in promoting such technologies in certain positive ways; they are AI enthusiasts. Thus, it can also be argued that these actors' expertise, grounded in their expectations of AI and communicated within the network, set the agenda for other members and associated stakeholders to follow [32]; p. 139). As of now, what seems to be missing in the network, and therefore not influencing the agenda nor the knowledge production, are members representative of actors such as AI users and specific AI technologies in use in real-world clinical settings (e.g., actual users or researchers studying AI in use). Such spokespersons could potentially provide more knowledge of the uncertainties identified by the network members as crucial to discuss in order to proceed with the introduction of AI in healthcare.

However, as new and more actors emerge, new uncertainties will surface, too [21]. Consequently, as the missing actors become mobile and become a part of existing actor constellations, previously overlooked, deemed unimportant or unknown issues can be identified and explored. That the lack of representative spokespersons may result in crucial issues being overlooked underscores the importance of continuous work to mobilise new actors as they appear to enable a more hybrid knowledge production.

The issue of crucial actors not yet properly mobilised within the KIN network questions further the network's self-declared 'bottom-up' approach. Rather than representing the 'bottom', it can be argued that the majority of the current members are positioned in the middle, between the system (the government and health authorities) and the subsystems or the actual 'bottom' (including actors such as AI users, patients, and specific AI technologies in use). The network may also be seen as taking a middle position in the tension between industry advocates pushing for a market approach and more top-down governmental processes focusing on developing national frameworks and regulations to avoid or mitigate potential harms (cf., [36]). This position is underscored by the variation in members representing both poles who collectively, through the network, seek to contribute to enabling AI in healthcare. This perspective, positioning the network in the middle, suggests that the role of the KIN network resembles more the intermediary role of interest groups rather than the hybrid knowledge-producing role of hybrid forums.

Conversely, it can be argued that as long as the discussion of AI stays on an overall, mostly theoretical level where AI is treated in general terms and not as specific technologies with evidence of real-world performances, the current network members are the 'bottom'. Thus, paradoxically, this is also where the network differs from interest groups. The present network members mainly represent themselves and their interests; they are the 'grassroots' from which knowledge grows and 'trickles upwards' in the system. Moreover, as they have no budget or formal mandate, the term 'informal grassroots network' may not be as farfetched after all, positioning them as a counterpart to the more 'top-down' initiatives.

As long as the number of AI deployments in healthcare continues to be limited, the number of users and AI technologies in use will be limited, too. However, as soon as the number of deployments increases, the new 'bottom', including spokespersons representative for actors such as AI users and specific AI technologies in use, can be mobilised

into the network. Eventually, with more knowledge of AI technologies' actual performance and evidence of immediate impact, new uncertainties can be identified and explored, and new knowledge can be shared within the network and beyond. Furthermore, as the new 'bottom' is mobilised, it can be argued that a broader consensus can be built within the network. However, such consensus-building will take time as the number of network members has increased and become more heterogeneous. This broader participation and heterogeneity may also require that the secretariat work actively to avoid the establishment of fractions within the network as a result of the network becoming more distributed in character. However, if a broader consensus is achieved and the hybrid knowledge produced as a result is shared with decision-makers and other stakeholders, the chances of a more sustainable outcome might be enhanced. On the other hand, failing to mobilise the new set of actors emerging could mean that the network eventually dissolves; if the knowledge produced has no relevance to the new situation with AI in healthcare, the members' and other stakeholders' interest in the network will probably decline.

5.2 The KIN network as a hybrid forum

Introducing AI involves multiple uncertainties, many of which seem to be constitutive of the KIN network and their meeting agendas, similar to the hybrid forums of [21]; p. 18). The uncertainties related to AI deployment identified and prioritised by the network members as essential to achieve more knowledge about can furthermore be seen as incentives for continuing the meetings [24]; p. vi). Elements of answers to the current uncertainties will become visible as more and more AI technologies are deployed and taken into use. However, it will not be possible to predict or get a complete overview of either short-term or long-term outcomes of the various and widespread AI deployments that will take place in different local clinical practices, their particular workflows, procedures and patient data. Thus, as the introduction of AI enters new phases, a myriad of new uncertainties connected to the many variations of AI technologies and use contexts will appear.

The prospects of a continued landscape of known and unknown uncertainties call similarly for a continuation of exploring and addressing emerging uncertainties. Even though such explorations will never lead to exhaustive inventories of all possible uncertainties, they will make visible some 'means to take measure' as AI technologies are introduced, taken into use and used over time. Subsequently, the inventories will enable the involved actors to anticipate and monitor critical issues and uncertainties and discover new ones as others are reduced ([21]; p. 22). Arguably, in cases concerning the introduction and use of ambiguous and unpredictable technologies like those based on AI, such continuous work seems more relevant than ever.

As Callon et al.'s hybrid forums, constellations like the KIN network can serve as an 'apparatus of elucidation' ([21]; p. 35). Today, such elucidation of the inventory of the present uncertainties can be exemplified by the list of topics identified through the 'temperature checks' as important to the network members to discuss. The meetings addressing topics such as AI validation and data access for AI use are further examples. As the members share their knowledge and this knowledge 'trickles upwards' to decision-makers in the system, it can be argued that they affect the current state of introducing AI after all. As such, they may even contribute to reducing some of the present uncertainties and, thus, to a certain extent, contribute to enabling AI in healthcare. This could, for instance, be as the KIN network draws the authorities' attention to the healthcare services' need for a national 'roadmap' or hospitals' need for support to start validating AI in their local clinical settings, despite the organisational silos between research activities and patient care.

As the discussion above indicates, the ongoing 'elucidation' and contribution to reducing the present uncertainties are mainly based on knowledge produced by a 'weak' hybrid or heterogeneous actor constellation. Actors such as AI users, patients and AI technologies in use are not represented by spokespersons who can speak of how AI

technologies actually perform in real-world clinical settings or how they affect the lives of those inhabiting this world. As touched upon, the network's existence depends on the constellation of spokespersons it is able to mobilise ([31]; p. 218). If spokespersons representative for the now missing actors become a part of the network, the network can also better ensure that what is spoken for is not later refuted by the actors they speak in the name of [18,37]. Phrased differently, if a constellation of actors, like the KIN network, wants to continue bringing forth knowledge that contributes to a sustainable introduction and future with AI in healthcare, the new 'bottoms' of actors that emerge must be mobilised. If such mobilisations happen, a broader consensus-building and 'stronger' hybrid knowledge production can also be ensured within the network.

However, if wider participation and heterogeneity are achieved, it will introduce a new layer of complexity to the network constellation, which at the same time emphasises the problematic nature of AI as an umbrella term for different types of technologies. With the diversity of AI technologies and the many users and patients affected by these technologies, all of whom may be represented by a variety of spokespersons (no longer just the AI enthusiasts), the complexity of the network seems limitless. Consequently, as the network becomes increasingly heterogenous, securing consensus among the members and preventing the formation of factions or 'silos' within the network, where members gather based on shared interests and expertise, becomes more challenging.

Hence, as various AI technologies are deployed, and new sets of actors are made mobile, it may become necessary to define some boundaries for the network. This may include making decisions concerning which specific type of AI technologies or medical area the network should focus on, as well as which users and patient groups should be included and represented by whom. Ultimately, who should be a part of the network's new 'bottom' will have to be negotiated.

6. Summary and final remarks

Through this study, the characteristics of a network of professionals aiming to contribute to enabling AI in the Norwegian public healthcare services have been explored. So has their ability to lessen the issues and uncertainties currently complicating and delaying the deployment of AI in healthcare.

Through the KIN network, knowledge production and sharing between actors not necessarily previously linked happens. However, the heterogeneity of the network can be questioned and problematised. As of now, certain actors cannot be sufficiently mobilised and are therefore not yet spoken for within the network. These actors, lacking representative spokespersons, include human actors, such as physicians as users of AI and patients as recipients of services supported by AI, as well as non-human actors, such as different but specific AI technologies in use, spoken for by, for instance, researchers studying AI use. In the current network, AI users, patients and AI technologies are mainly spoken for by actors who can be characterised as AI enthusiasts. This representativity, or the lack thereof, affects the knowledge produced and shared within and beyond the network. Thus, based on the network's characteristics, it can be argued that its contribution to enabling AI in healthcare has its natural limitations. The network cannot produce the hybrid knowledge necessary to reduce currently known uncertainties regarding AI deployments in real-world clinical settings.

As more and more AI deployments occur and the actors not yet properly mobilised start interacting, more concrete issues and uncertainties will emerge. Thus, as introducing AI progresses into more widespread deployment, such actors can better be mobilised. Subsequently, a stronger hybrid knowledge that contributes to a more sustainable introduction of AI can be produced. For now, in the early phases of introducing AI in healthcare, this study shows that the KIN network is an 'apparatus of elucidation', bringing forth uncertainties necessary to explore and address in order to progress in deploying AI. These

uncertainties are made visible through the votes given by the network members concerning the topics they deem most important to discuss and the further activities carried out within the network. As the current members interact and share their knowledge with other stakeholders, they draw decision-makers' attention to these elements of importance for enabling AI in healthcare. Which role informally established networks, like the KIN network, will play as AI becomes more widely deployed will depend on whether the spokespersons representative for the new actors that emerge are mobilised. It also depends on whether they are able to stay as a united constellation. Only then can a strong hybrid knowledge of existing and arising uncertainties be produced.

Finally, although this study is limited to a single case within a Norwegian context, it highlights a global trend. Numerous initiatives worldwide are working towards enabling widespread deployment of AI in healthcare, exemplified by the Alliance for Artificial Intelligence in Healthcare [38], the Canadian Association of Radiologists Artificial Intelligence Working Group [39] and the Australian Alliance for Artificial Intelligence in Healthcare [40]. Thus, to enhance our understanding of the role played by different constellations of actors in contexts of emerging complex technologies, this study calls for further investigations of such mobilisations. Moreover, the study calls for investigations of what occurs as AI technologies are increasingly deployed in healthcare and new actors and uncertainties appear. Further research in these areas will contribute to a more comprehensive understanding of the introduction and future of AI in healthcare.

Author statement

The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Data availability

The data that has been used is confidential.

References

[1] Y.Y.M. Aung, D.C.S. Wong, D.S.W. Ting, The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare, Br. Med. Bull. 139 (1) (2021) 4–15, https://doi.org/10.1093/bmb/ldab016.

[2] B. Meskó, G. Hetényi, Z. Győrffy, Will artificial intelligence solve the human resource crisis in healthcare? BMC Health Serv. Res. 18 (1) (2018) 545, https://doi.org/10.1186/s12913-018-3359-4.

[3] E.J. Topol, High-performance medicine: the convergence of human and artificial intelligence, Nat. Med. 25 (1) (2019) 44–56, https://doi.org/10.1038/s41591-018-0300-7.

[4] P.M. Krafft, M. Young, M. Katell, K. Huang, G. Bugingo, Defining AI in policy versus practice, in: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 2020.

[5] P. Wang, On defining artificial intelligence, Journal of Artificial General Intelligence 10 (2) (2019) 1–37, https://doi.org/10.2478/jagi-2019-0002.

[6] P. Rajpurkar, E. Chen, O. Banerjee, E.J. Topol, AI in health and medicine, Nat. Med. 28 (1) (2022) 31–38, https://doi.org/10.1038/s41591-021-01614-0.

[7] F. Wang, L.P. Casalino, D. Khullar, Deep learning in medicine—promise, progress, and challenges, JAMA Intern. Med. 179 (3) (2019) 293–294, https://doi.org/10.1001/jamainternmed.2018.7117.

[8] K.-H. Yu, A.L. Beam, I.S. Kohane, Artificial intelligence in healthcare, Nat. Biomed. Eng. 2 (10) (2018) 719–731, https://doi.org/10.1038/s41551-018-0305-z.

[9] Meld. St. 7, National Health and Hospital Plan 2020–2023. Summary, 2019–2020. Retrieved from, https://www.regjeringen.no/contentassets/95eec808f0434acf942fca449ca35386/en-gb/pdfs/stm201920200007000engpdfs.pdf.

[10] H. Alami, P. Lehoux, Y. Auclair, M.d. Guise, M.-P. Gagnon, J. Shaw, D. Roy, R. Fleet, M.A.A. Ahmed, J.-P. Fortin, M. de Guise, M.A. Ag Ahmed, Artificial intelligence and health technology assessment: anticipating a new level of complexity, J. Med. Internet Res. 22 (7) (2020) 1–12, https://doi.org/10.2196/17707.

[11] R. de Neufville, S.D. Baum, Collective action on artificial intelligence: a primer and review, Technol. Soc. 66 (2021), 101649, https://doi.org/10.1016/j.techsoc.2021.101649.

[12] F. Cabitza, A. Campagner, C. Balsano, Bridging the "last mile" gap between AI implementation and operation: "data awareness" that matters, Ann. Transl. Med. 8 (7) (2020) 501, https://doi.org/10.21037/atm.2020.03.63.

[13] E. Coiera, The last mile: where artificial intelligence meets reality, J. Med. Internet Res. 21 (11) (2019), e16323, https://doi.org/10.2196/16323.

[14] M.G. Seneviratne, N.H. Shah, L. Chu, Bridging the implementation gap of machine learning in healthcare, BMJ Innovations 6 (2) (2020) 45–47, https://doi.org/10.1136/bmjinnov-2019-000359.

[15] P.A. Keane, E.J. Topol, With an eye to AI and autonomous diagnosis, NPJ Digit Med 1 (2018) 40, https://doi.org/10.1038/s41746-018-0048-y.

[16] R.B. Parikh, S. Teeple, A.S. Navathe, Addressing bias in artificial intelligence in health care, JAMA 322 (24) (2019) 2377–2378, https://doi.org/10.1001/jama.2019.18058.

[17] M. Akrich, M. Callon, B. Latour, The key to success in innovation part I: the art of interessement (A. Monaghan, Trans.), Int. J. Innovat. Manag. 6 (2) (2002) 187–206, https://doi.org/10.1142/S1363919602000550.

[18] M. Callon, Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay, in: J. Law (Ed.), Power, Action and Belief. A New Sociology of Knowledge, Routledge & Kegan Paul, 1986, pp. 196–233.

[19] M. Callon, Society in the making: the study of technology as a tool for sociological analysis, in: W.E. Bijker, T.P. Hughes, T. Pinch (Eds.), The Social Construction of Technological Systems, Anniversary Edition : New Directions in the Sociology and History of Technology, MIT Press, 2012, pp. 77–97.

[20] B. Latour, Reassembling the Social : an Introduction to Actor-Network-Theory, Oxford University Press, 2005.

[21] M. Callon, P. Lascoumes, Y. Barthe, Acting in an Uncertain World: an Essay on Technical Democracy, MIT press, 2011.

[22] W.E. Bijker, T.P. Hughes, T. Pinch, The Social Construction of Technological Systems, Anniversary Edition : New Directions in the Sociology and History of Technology, MIT Press, 2012.

[23] M. Gibbons, C. Limoges, H. Nowotny, S. Schwartzman, P. Scott, M. Trow, The New Production of Knowledge : the Dynamics of Science and Research in Contemporary Societies, SAGE Publications, Limited, 1994.

[24] H. Nowotny, The Cunning of Uncertainty, Polity, 2016.

[25] S. Kuhnle, M. Alestalo, The modern Scandinavian welfare state, in: P. Nedergaard, A. Wivel (Eds.), The Routledge Handbook of Scandinavian Politics, Routledge, 2017.

[26] P. Lægreid, Nordic administrative traditions, in: P. Nedergaard, A. Wivel (Eds.), The Routledge Handbook of Scandinavian Politics, Routledge, 2017.

[27] J. Dexe, U. Franke, Nordic lights? National AI policies for doing well by doing good, Journal of Cyber Policy 5 (3) (2020) 332–349, https://doi.org/10.1080/23738871.2020.1856160.

[28] A. Theodorou, J.C. Nieves, V. Dignum, Good AI for Good: How AI Strategies of the Nordic Countries Address the Sustainable Development Goals, 2022 arXiv preprint arXiv:2210.09010.

[29] M.T. Heaney, J.M. Strickland, A network approach to interest group politics, in: J.N. Victor, A.H. Montgomery, M. Lubell (Eds.), The Oxford Handbook of Political Networks, Oxford University Press, 2018, pp. 433–452.

[30] A.P. Joosse, H.B. Milward, Health policy networks, in: J.N. Victor, A.H. Montgomery, M. Lubell (Eds.), The Oxford Handbook of Political Networks, Oxford University Press, 2018, pp. 629–650.

[31] M. Akrich, M. Callon, B. Latour, The key to success in innovation part II: the art of choosing good spokespersons (A. Monaghan, Trans.), Int. J. Innovat. Manag. 6 (2) (2002) 207–225, https://doi.org/10.1142/S1363919602000562.

[32] V. Galanos, Expectations and Expertise in Artificial Intelligence: Specialist views and Historical Perspectives on Conceptualisation, Promise, and Funding University of Edinburgh], 2023. Edinburgh, https://era.ed.ac.uk/bitstream/handle/1842/40420/Galanos2023.pdf.

[33] KIN, Medlemskap. Nasjonalt Senter for E-Helseforskning, 2023. Retrieved 29. June from, https://ehealthresearch.no/kin/medlemskap.

[34] KIN, Kunstig intelligens I Norsk Helsetjeneste (KIN). Nasjonalt Senter for E-Helseforskning, 2023. Retrieved 29. June from, https://ehealthresearch.no/kin.

[35] R.M. Emerson, R.I. Fretz, L.L. Shaw, Writing Ethnographic Fieldnotes, 2 ed., University of Chicago Press, 2011.

[36] M.S. Kannelønning, Contesting futures of artificial intelligence (AI) in healthcare: formal expectations meet informal anticipations, Technol. Anal. Strat. Manag. (2023) 1–12, https://doi.org/10.1080/09537325.2023.2226243.

[37] B. Latour, Science in Action: How to Follow Scientists and Engineers through Society, Harvard university press, 1987.

[38] AaiH, The Alliance for Artificial Intelligence in Healthcare, 2023. Retrieved 29. June from, https://www.theaaih.org.

[39] A. Tang, R. Tam, A. Cadrin-Chênevert, W. Guest, J. Chong, J. Barfett, L. Chepelev, R. Cairns, J.R. Mitchell, M.D. Cicero, M.G. Poudrette, J.L. Jaremko, C. Reinhold, B. Gallix, B. Gray, R. Geis, T. O'Connell, P. Babyn, D. Koff, W. Shabana, Canadian association of Radiologists white paper on artificial intelligence in radiology, Can. Assoc. Radiol. J. 69 (2) (2018) 120–135, https://doi.org/10.1016/j.carj.2018.02.002.

[40] AaaiH, The Australian alliance for artificial intelligence in healthcare, Retrieved 29 June from, https://aihealthalliance.org, 2019.

[41] NHS. (n.d.). The National Strategy for AI in Health and Social Care. NHS England - Transformation Directorate. Retrieved February 21st from https://transform.england.nhs.uk/ai-lab/ai-lab-programmes/the-national-strategy-for-ai-in-health-and-social-care/.

[42] Danish Ministry of Finance, & Danish Ministry of Industry, National Strategy for Artificial Intelligence, B. a. F. A, 2019, https://en.digst.dk/media/19337/305755_gb_version_final-a.pdf.

[43] J. Tucker, The future vision(s) of AI health in the Nordics: comparing the national AI strategies, Futures 149 (2023), 103154, https://doi.org/10.1016/j.futures.2023.103154.

[44] M. Sujan, C. Smith-Frazer, C. Malamateniou, J. Connor, A. Gardner, H. Unsworth, H. Husain, Validation framework for the use of AI in healthcare: overview of the new British standard BS30440, BMJ Health Care Inform 30 (1) (2023), https://doi.org/10.1136/bmjhci-2023-100749.

Appendices

Appendix A: Interview guide, pre-fieldwork preparations

Appendix B: Interview guide, procurement project manager (study 2)

Appendix C: Interview guide, vendors (study 2)

Appendix D: Interview guide, the KIN network (study 3)

Appendix E: Information letter/informed consent form

Appendix F: Ethics approval

Intervjuguide – innledende intervju

1. **Bakgrunn**
   - Bakgrunn/nåværende stilling/arbeidsoppgaver

2. **Om prosjektet/arbeidet**
   - Kan du fortelle om de prosessene du/dere står i nå – hva har dere gjort og hvor skal dere hen? Mål/fremdrift/utfordringer/usikkerheter

   - Kan du fortelle litt om hvordan du ser på innføringen av KI i helsetjenesten – hva handler det om sett fra ditt perspektiv?
     Nåværende situasjon/utvikling/milepæler/utfordringer/fremtiden

3. **Hvis aktuelt: om teknologien**
   - Kan du fortelle mer om teknologien dere har utviklet/jobber med å utvikle?
     Tiltenkt bruksområde/brukere/mål/utviklingsprosessen/status/kommersialisering/implementering/utfordringer/usikkerheter

   - Kan du fortelle om hvordan brukskontekst og bruk tenkes inn i prosjektet?
     Brukerinvolvering/arbeidsflyt/informasjonsflyt/samarbeid/use cases?

4. **Avslutningsvis**
   - Er det noen problematikker/utfordringer du/dere opplever/har støtt på, som du kunne tenke deg ble belyst?

   - Er det noen rapporter/dokumentasjon du synes er nyttige?

   - Har du forslag til andre jeg kan snakke med?

Intervjuguide – prosjektleder, anskaffelse

[Oppfølgingsintervju etter dialogprosess med leverandører]

1. Om anskaffelsesprosessen og valg av AI plattform-løsning

- Kan du fortelle om erfaringene fra anskaffelsesprosessen så langt?
  - Var det noe som overrasket underveis?
  - Hva var avgjørende for at dere valgte å gå for en AI plattform fremfor enkeltløsninger?
  - Hva ser dere som fordeler og ulemper med en plattform-løsning?

2. Om innføringen av AI løsningen

- Hva tenker du om å integrere plattformen i den eksisterende infrastrukturen med de ulike PACS'ene som sykehusene deres har i dag? Utfordringer/krav/behov
- Hva tenker du må til for å integrere løsningen i radiologenes arbeidspraksis? Passe inn i eksisterende arbeidsflyt?/nye flyt/planer/konfigurasjoner/involvering
- Hvordan blir prosessen videre generelt? Validering/utfordringer/involvering osv.

3. Diverse oppfølgingsspørsmål/avklaringer

- Hvordan opplever dere dette med partnerskap og samarbeid etter implementering som leverandørene snakket om under dialogmøtene?
- Hva kan et sånn partnerskap innebære?/Hvordan ser du for deg at forholdet blir mellom dere og plattform-leverandøren?
- Opplever du at dere har ulikt syn på hvor tette samarbeidspartnere dere skal være utover at de (leverandørene) er mellomledd?
- Hva er risikoen for begge parter ved en implementering, slik du ser det? På kort og lang sikt
- I forhold til kravspesifikasjonen, kan du si noe om hva som har vist seg å være umulige krav å få innfridd (som dere kunne ønske var mulige)?
- Har dere fått nyttige innspill til den endelige kravspesifikasjonen fra leverandørene som har hatt betydning? Hvilke?

Intervjuguide – leverandører

[Intervju etter dialogprosess med helseforetak]

1. **Bakgrunn**

2. **Om innføringen av KI i helsetjenesten generelt**

   - Hvordan opplever du/dere dagens situasjon hvor det jobbes fra flere kanter med å få til en utstrakt bruk av KI i helsetjenesten, men hvor det fortsatt er en vei å gå?

   - Kan du fortelle om gjennombrudd dere har opplevd i løpet av de siste årene?

   - Hva tenker du skal til for at innføringen av KI blir mer utbredt?

   - Hvordan jobber dere med å få det til? Hva tenker du er deres rolle?

   - Hva tenker du at helseforetakene må gjøre for å få det til?

3. **Om anskaffelsesprosessen**

   - Kan du fortelle om hvordan det har vært å være en del av anskaffelsesprosessen? Læringer/overraskelser/endringer

   - Helseforetaket har gjennom prosessen kommet med behov, ønsker og krav. Ikke alt kan innfris for begge parter. På hvilke områder ser du at det kan være nødvendig å inngå kompromiss?

   - Helseforetaket ønsker å vite kost/nytte-verdien ved å innføre de spesifikke teknologiene, men, slik jeg har forstått det, er det vanskelig å fremskaffe slike «bevis» fordi effekten og hvordan den måles vil variere fra sykehus til sykehus/praksis til praksis. Hva tenker du om dette? Gjør dere noe for å få frem flere/andre typer «bevis», eller gjør dere noe annet?

   - Er det noe du ville gjort annerledes om dere skulle delta i en ny anskaffelsesprosess?

4. **Om plattformteknologien**

   - En plattform eller markedsplass for KI i helse er jo noe nytt og forholdvis uprøvd. Kan du fortelle om hvordan du ser at en slik «markedsplass» med KI modeller vil fungere i kliniske praksiser på offentlige sykehus? Fjerne eller legge til nye modeller/endre arbeidsflyt osv.

Ser du noen utfordringer i denne sammenhengen? Er det noe sykehusene må gjøre for å omstille seg?

- Når vi snakker om konkrete løsninger: slik jeg forstår det er det reguleringer som gjør at dynamiske/lærende KI modeller ikke kan tas i bruk i klinisk praksis (de må hentes ut og oppdateres i «lab»). Hva tenker du om dette?

- Helseforetaket var opptatt av at det etableres tillit til den nye teknologien blant legene for at de tar den i bruk på en god måte. Hvordan tenker du at dette best kan gjøres?

5. Avslutningsvis

- Hvis du skulle nevnt noe som må gjøres ute i helsesektoren for å få til innføring av KI akkurat nå – hva skulle det være?

- Er det noen spesifikke aktiviteter eller planer dere vil jobbe med fremover?

Intervjuguide – KIN-nettverket

1. **Bakgrunn**
   - Faglig bakgrunn/nåværende stilling/arbeidsoppgaver

2. **Om KIN nettverket**
   - Kan du fortelle om hvordan nettverket ble til og hvordan du ble en del av det?

   - Kan du fortelle om nettverkets formål?

   - Hvem er medlemmene (er det noen som ikke er med?)?

   - Kan du fortelle om «midlene» dere tok i bruk/bruker for å oppnå målene med nettverket? Aktiviteter/samarbeid med andre osv.

3. **Om nettverkets arbeid frem til nå**
   - Kan du fortelle om hvordan disse årene med KIN-nettverket har vært? Milepæler/erfaringer/utfordringer

4. **Om KI i helsetjenesten**
   - Hva tenker du at KI i helsetjenesten handler om? Nåtid/fremtid

   - Hvilke typer KI-teknologi opplever du at det er fokus på i KIN-nettverket?

   - Hvordan tenker du at Norge kan få til en bred innføring av KI i helsetjenesten? Hva skal til, og hva kan nettverket bidra med?

5. **Avslutningvis**
   - Noen tanker for fremtiden? For nettverket/KI i helsetjenesten mm.

   - Er det noe du lurer på som du synes det hadde vært interessant om jeg stilte spørsmål til i mine samtaler med andre eller så etter i analysen av materialet mitt?

Informasjon om deltakelse i forskningsprosjekt og samtykkeskjema

"Introduksjonen av kunstig intelligens (KI) i helsetjenesten"

Dette er et spørsmål til deg om å delta i et doktorgradsprosjekt som studerer introduksjonen av KI-baserte løsninger i helsetjenesten. I dette skrivet gir jeg deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål
Doktorgradsprosjektets hovedfokus er kompleksiteten i den pågående introduksjonen av KI i klinisk praksis. Prosjektets overordnede formål er å 1) synliggjøre dynamikkene som er med til å forme den retningen introduksjonsprosessen tar (visjonene, barrierene, interessene osv.), og 2) gjennom denne synliggjøringen bidra til at realiseringen av KI i klinisk praksis kan skje på best mulig måte med et bærekraftig resultat.

Den delen av studien du er invitert til å bidra til vil gjennomføres med intervju via Zoom. Fokus for intervjuet er dine tanker og erfaringer knyttet til bestemte tematikker relatert til en innføring av KI i klinisk praksis.

Hvem er ansvarlig for forskningsprosjektet?
Mari S. Kannelønning, stipendiat ved OsloMet - storbyuniversitetet er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?
Du kontaktes fordi du er fagperson med relevant kunnskap.

Hva innebærer det for deg å delta?
Som deltaker i prosjektet vil du involveres gjennom intervju. Intervjuets varighet er ca. 45-60 minutter. Intervjuet tas opp på diktafon.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Ditt personvern – hvordan jeg oppbevarer og bruker dine opplysninger
Sitater fra intervju kan bli brukt i publikasjoner og presenteres på konferanser, fortrinnsvis uten navn og andre kjennetegn som f.eks. stillingstittel. Ved bruk av navn eller andre kjennetegn, avtales dette
nærmere med deg.

Jeg behandler personopplysninger konfidensielt og i samsvar med personvernregelverket.

- Utover meg, stipendiaten, vil min hovedveileder Terje Colbjørnsen samt medveileder Miria Grisot, kunne få tilgang til dataene jeg samler inn (inkl. lydopptak).
- Navnet ditt og kontaktopplysningene dine vil jeg oppbevare adskilt fra datamaterialet. I datamaterialet vil navnet ditt være erstattet med en kode. Datamaterialet vil oppbevares i egen mappe på min personlige datamaskin som synkroniseres med en skytjeneste med databehandleravtale (begge deler vil være låst med passord/FEIDE). Slik vil ingen uvedkommende får tilgang til dine personopplysninger.

Hva skjer med opplysningene dine når jeg avslutter forskningsprosjektet?
Dine personopplysninger vil som sagt anonymiseres gjennom hele prosjektet. Opptak gjort i løpet av intervjuene, hvor stemmen din kan identifiseres, vil slettes når forskningsprosjektet avsluttes og doktorgradsprosjektet er godkjent høsten 2023.

Dine rettigheter
Hvis du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om deg, og å få utlevert en kopi av opplysningene,
- å få rettet personopplysninger om deg,
- å få slettet personopplysninger om deg, og
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger.

Hva gir meg rett til å behandle personopplysninger om deg?
Jeg behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra OsloMet - storbyuniversitetet har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan du finne ut mer?
Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:
- OsloMet - storbyuniversitetet ved Mari S. Kannelønning, epost: marika@oslomet.no, mobil:
  46 91 98 10
- Vårt personvernombud: Ingrid S. Jacobsen, epost: personvernombud@oslomet.no, telefon:
  67 23 55 34

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:
- NSD – Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Mari Serine Kannelønning
(Stipendiat/forsker)

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet «Introduksjonen av kunstig intelligens i helsetjenesten», og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i intervju
- ☐ at opplysninger som stillingstittel og arbeidsoppgaver, som kan føre til at jeg blir gjenkjent, kan publiseres
- ☐ at opplysninger som stillingstittel, arbeidssted og arbeidsoppgaver, som kan føre til at jeg blir gjenkjent, kan lagres etter prosjektslutt, til videre forskning – hvis aktuelt. Dette gjelder kun skriftlige opplysninger som transkripsjon av intervju (opptak vil slettes ved prosjektslutt).

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

----
(Signert av prosjektdeltaker, dato)

Sikt

Vurdering av behandling av personopplysninger

Referansenummer	**Vurderingstype**	**Dato**
397412	Standard	14.10.2022

Tittel
Introduksjonen og bruken av kunstig intelligens-baserte informasjonssystemer i helsetjenesten

Behandlingsansvarlig institusjon
OsloMet – storbyuniversitetet / Fakultet for samfunnsvitenskap / Institutt for arkiv, bibliotek- og informasjonsfag

Prosjektansvarlig
Mari Serine Kannelønning

Prosjektperiode
01.05.2020 – 01.12.2023

Kategorier personopplysninger
Alminnelige

Lovlig grunnlag
Samtykke (Personvernforordningen art. 6 nr. 1 bokstav a)

Behandlingen av personopplysningene er lovlig så fremt den gjennomføres som oppgitt i meldeskjemaet. Det lovlige grunnlaget gjelder til 01.12.2023.

Meldeskjema ↗

Kommentar
Personverntjenester har vurdert endringen registrert i meldeskjemaet.

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg. Behandlingen kan fortsette.

Perioden for behandling av personopplysninger ble utvidet til 01.12.2023. Opprinnelig prosjektslutt var 30.09.2023.
Utvalg 3 ble lagt til. Utvalget mottar informasjon og samtykker til behandling av sine personopplysninger i forskningsprosjektet.

OPPFØLGING AV PROSJEKTET
Vi vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Kontaktperson: Simon Gogl
Lykke til videre med prosjektet!