

# **Human-AI Collaboration in Public Services**

The case of sick leave case handling in the Norwegian Labour and Welfare Administration

ERIKSSON, CHRISTER  
OLSEN, KRISTIAN

## **SUPERVISORS**

Professor - Pappas, Ilias  
PhD Research Fellow - Schmager, Stefan  
Professor - Vasilakopoulou, Polyxeni

**University of Agder, 2023**

Faculty of Social Science  
Department of Information Systems



# Preface

The results presented in the master's thesis have been carried out in connection with the subject Master's thesis in information systems at the University of Agder (UiA), spring 2023. Christer Eriksson and Kristian Olsen have carried out the study as a final master's degree study in Information Systems at UiA.

We thank our supervisors, Ilias Pappas, Stefan Schmager, and Polyxeni Vasilakopoulou, for their valuable guidance and feedback throughout this project.

To all the informants participating in the interviews, thank you for your time and valuable thoughts and insight.

Thank you to our fellow students for their encouragement throughout the master's program. We appreciate you all and hope to see you around in the future.

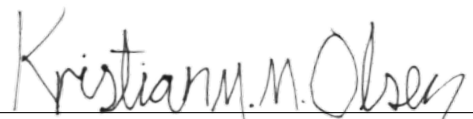
We thank our families for their motivation and support throughout the study.

Kristiansand 26.05.2023



---

Christer Eriksson



---

Kristian M.M. Olsen

# Abstract

Public service delivery has seen a surge in demand from society in recent years, especially after the Covid-19 pandemic. It is crucial for public service organizations to keep up with digitization efforts to meet these demands. The most important information systems innovation for the public sector is artificial intelligence (AI). Not all processes in public service delivery can be fully automated by AI, requiring human presence to ensure human discretion and fair judgment. This study investigates the needs for human-AI collaboration in public services. To further narrow the study, it focuses on the case of sick leave case handling in the Norwegian Labour and Welfare Administration (NAV). Human-AI collaboration in public services is a field of research that has gained increased interest recently, but there is a need for further research in this area. To help shed more light on this topic, we formulated the following research question:

*"What are caseworkers' needs for Human-AI collaboration in public services?"*

To help us answer the research question, we formulated three sub-questions related to the caseworkers' needs for AI when handling sick leave cases, their expectations for the future working with AI, and how our contributions can facilitate meeting their needs. A case study was conducted, with semi-structured interviews and focus group interviews with 16 public service practitioners as data-gathering methods. The research is set in the organization NAV, specifically in the area of sick leave case handling. We conducted a systematic literature review on the topic of transparency in human-AI interaction, which we found to be a central topic in the human-AI collaboration literature. We found multiple aspects of AI to encompass the topic of transparency, which forms the basis for the background theory used in this study.

Findings show that caseworkers have quite similar needs relating to human-AI collaboration. They think that AI could mostly assist in the internal processes of the organization. Bias in AI's decisions, based on data available, was a concern amongst the caseworkers. The caseworkers suggested that humans should be in the loop to allow more trust in the systems, and thought that not all processes should be automated. Findings point towards simplifying and streamlining caseworkers' work processes, by using AI-assisted tools in their systems.

Based on the findings and recommendations, we suggest further research into the responsible development and deployment of AI, different government legislations, and how human-AI collaboration differs cross-culturally. From this study, we propose implications for practice in NAV. To leave the human in control, and to improve caseworkers' decision-making processes, a human-centered approach to human-AI collaboration is suggested. Public sector organizations could streamline and simplify work processes through human-AI collaboration, which could help reach some of the UN's sustainable development goals.

# Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Research problem and motivation . . . . .	2
1.2	Research questions . . . . .	3
1.3	Research approach . . . . .	3
1.4	Thesis outline . . . . .	3
<b>2</b>	<b>Theoretical framing &amp; related literature</b>	<b>5</b>
2.1	Theoretical framework: Joint cognitive systems . . . . .	5
2.2	Related literature . . . . .	6
2.2.1	Literature search procedure . . . . .	6
2.2.2	Findings . . . . .	10
<b>3</b>	<b>Case background &amp; method</b>	<b>15</b>
3.1	Case background . . . . .	15
3.2	Research approach . . . . .	16
3.3	Interview design . . . . .	17
3.4	Data gathering . . . . .	18
3.4.1	Preliminary data collection . . . . .	18
3.4.2	Case study of caseworkers at NAV based on sick leave case handling . . . . .	19
3.5	Data analysis . . . . .	20
3.6	Validity and reliability . . . . .	21
3.7	Ethical issues . . . . .	21
<b>4</b>	<b>Findings from the data analysis</b>	<b>23</b>
4.1	Caseworkers' needs for AI . . . . .	23
4.2	Caseworkers' expectations for the future with AI . . . . .	27
4.2.1	Preferred future . . . . .	27
4.2.2	Worst-case future . . . . .	28
4.2.3	How their preferred future could be achieved . . . . .	29
4.3	Challenges of AI in public services . . . . .	30
4.3.1	Machine bias . . . . .	31
4.3.2	Loss of human contact . . . . .	31
4.3.3	Multiple systems . . . . .	32
4.3.4	Communication . . . . .	33
4.4	Summary of findings . . . . .	35
<b>5</b>	<b>Discussion</b>	<b>36</b>

5.1	Model for human-AI collaboration for caseworkers working with sick leave cases in NAV . . . . .	36
5.1.1	Current processes . . . . .	38
5.1.2	Digitization and new work practices . . . . .	39
5.1.3	Bias and fairness . . . . .	40
5.2	Limitations . . . . .	41
<b>6</b>	<b>Conclusions</b>	<b>43</b>
6.1	Implications . . . . .	43
	<b>References</b>	<b>45</b>
	<b>Appendices</b>	<b>55</b>
A	Consent form . . . . .	55
B	Interview guide . . . . .	57
C	Systematic Literature Review . . . . .	61
D	Interview guide: preliminary interviews . . . . .	62

## List of Figures

2.1	Hypothetical computer consultant in a Joint Cognitive System (Woods, 1986). . . . .	5
2.2	Documents by year(Eriksson & Olsen, 2022) . . . . .	7
2.3	Prisma flow diagram of the literature screening process (Eriksson & Olsen, 2022) . . . . .	8
2.4	Concept matrix based on Webster and Watson, 2002 (Eriksson & Olsen, 2022) . . . . .	9
2.5	Key concepts related to transparency in AI . . . . .	10
3.1	The research process (Oates et al., 2022) . . . . .	17
3.2	A visual representation of the data analysis process. . . . .	20
5.1	Hypothetical human-AI collaboration in the context of NAV based on Joint Cognitive Systems (Woods, 1986). . . . .	37

## List of Tables

2.1	Search query used for the literature search (Eriksson & Olsen, 2022). . . . .	7
3.1	Table of interviews. . . . .	19

# 1 Introduction

Artificial intelligence (AI), a research area initiated in the 1950s (McCarthy et al., 2006), has received significant attention in science and practice (Meske et al., 2022). Research shows that AI applications are currently the most important information systems innovations for the public sector (Benbunan-Fich et al., 2020). According to McKinsey's "The state of AI in 2022" global survey, AI adoption in organizations has more than doubled during the past five years (McKinsey, 2022b). The most popular use cases for AI span a range of different functional activities, including service operations, product and/or service development, marketing and sales, and risk. Among these use case functions, service operations optimizations are the most popular use case for AI (McKinsey, 2022b). Citizens are continuously demanding more transparency, efficiency, and responsiveness from the public sector (Lynn et al., 2022). As Vassilakopoulou et al., 2023 mentions, digitally mediated requests from citizens have increased in public service organizations, especially during social distancing. AI-powered systems can partially or entirely automate tasks exclusively performed by humans (Park et al., 2019), and can increase the efficiency of processes across sectors, including public services (AI4users, 2020).

The Norwegian Labour and Welfare Administration (NAV) is central to the Norwegian public administration. "NAV administers benefit schemes and pensions, providing services tailored to citizens' needs and circumstances, supporting a well-functioning job market, keeping people active and overall, enduring comprehensive and efficient labour and welfare administration" (Vassilakopoulou et al., 2023). NAV's research and development department focuses on building a good basis for knowledge-based practice and service development, decisions, advice, and recommendations in labour- and welfare-related areas (NAV, 2021). NAV collaborates with different universities on specific domains of interest, gaining more knowledge on particular topics for their research and development projects. Through a collaboration with researchers from the AI4Users research project and NAV, this thesis focuses on the domain of Human-AI collaboration in public services.

Recent research has focused a lot on Human-AI Interaction (HAI) and Explainable AI (XAI) (Meske et al., 2022; D. Wang et al., 2020). However, interaction is not the same as collaboration (D. Wang et al., 2020). "Collaboration involves mutual goal understanding, preemptive task co-management, and shared progress tracking. Most of the human activities today are done collaboratively; thus, to integrate AI into the already complicated human workflow, it is critical to bring the Human-AI perspective into the root of the algorithmic research and plan for a Human-AI Collaboration future of work" (D. Wang et al., 2020). Despite the potential benefits of AI, there are some complex organizational changes and social consequences it presents. These challenges include biased algorithms that benefit certain subgroups (e.g., using race to predict treatment). The potential negative effects AI could have in areas such as NAV, public services, and society includes potentially amplifying human biases and inequality (Whittaker et al., 2018). There is also a po-



tential risk of over-automating processes with AI. The involvement of humans in incorporating the systems and the human labour that underlies the production and maintenance of AI systems could easily be overlooked (Park et al., 2019).

## 1.1 Research problem and motivation

Research shows that AI applications can increase the efficiency and quality of public services (Pencheva et al., 2020; Schmager et al., 2023; van Noordt & Misuraca, 2022; Vassilakopoulou et al., 2023). Humans, however, cannot understand AI applications' inner workings, making it difficult to explain how they turn data input into output. "This is known as the "black box" problem, which can impede the involvement of humans in shaping, operating, and monitoring the use of AI in service delivery" (AI4users, 2020). Thus, it is important to design human-friendly and trustworthy solutions that foster the responsible use of AI applications. This is crucial for public service practitioners, as the decisions they make often have a direct impact on human lives.

McKinsey, 2022a states that machines working together with humans, rather than replacing humans, leads to the tasks being performed better together than either could do separately; "Companies that design and plan for machine and human qualities to become complementary, rather than oppositional, will have the most effective teams". Vassilakopoulou and Grisot, 2020 mentions that public sector organizations would benefit from a process that rewards disruption from within, which could be a step toward incorporating a human-AI collaboration in the organization. By involving public sector practitioners and researchers in joint innovation efforts, the organization can reorganize the way they work with technology. This way, organizations can adopt new work practices which involve humans and AI working together, rather than just simply implementing new technologies (Velsberg et al., 2020).

This thesis focuses on Human-AI collaboration in public services; more specifically, we investigate the needs and understanding of AI from caseworkers in the Norwegian Labour and Welfare Administration (NAV) that work with sick leave cases. The caseworkers in NAV help citizens with sick leave, promoting the transition back to work, as well as other activity needs (NAV, 2023a). The research is based on the case of sick leave case handling in NAV. This is an interesting topic for practice, but we identified a lack of research in this field. When performing the literature review on transparent AI, an important part of human-AI collaboration, we found there to be few publications on the topic of human-AI collaboration in public services specifically. This motivated us to work towards filling this research gap. The findings in the literature review show that research in AI transparency has only had a significant focus since 2017, even though artificial intelligence has been around since the early 40s (Haenlein & Kaplan, 2019). There is also a research gap between citizens' expectations and government abilities regarding AI within the scope of arising challenges (Mehr, 2017). Wirtz et al., 2019 proposed four major dimensions of AI challenges. These challenges consist of AI technology implementation, AI law and regulation, AI society, and AI ethics. AI can reduce administrative burdens and encourage resource allocation (Eggers et al., 2017). There is little empirical research concerning the level of use of AI within organizations,

especially in the public sector (Mikalef et al., 2019). Research has been done with citizens, but more research should be done by exploring both citizens' and public servants' stances regarding digital discretion in AI-supported public services (Schmager et al., 2023).

## 1.2 Research questions

Public services have digitalized services in order to make the services more efficient, with the goal often being to offer better products and services (Babar & Yu, 2019; Loonam et al., 2017; Morakanyane et al., 2017). Despite increasing investments in AI research and research contributions, AI in the public sector is still a young field of research that falls short in describing the challenges that come with it (Wirtz et al., 2019). A part of our research motivation is to research caseworkers' needs for human-AI collaboration, their expectations and understanding of AI, and how we could contribute to facilitating their needs. To further investigate this, we need insight from the caseworkers, which forms the basis for the research question for this thesis:

*What are caseworkers' needs for Human-AI collaboration in public services?*

Sub-questions used to break down and further explore the research question:

- *What are caseworkers' needs for AI when handling sick leave cases?*
- *What are caseworkers' expectations for the future working with AI?*
- *How can we contribute to facilitating meeting the needs of the caseworkers?*

## 1.3 Research approach

We conducted a qualitative case study as the research approach to gain insight into the research question and supporting sub-questions. The research is set in the NAV organization, specifically in the area of sick leave case handling. Semi-structured interviews were conducted with 16 caseworkers in the case study, and three caseworkers in a pilot study. The interviews were conducted at different offices in Agder, ranging from small to large offices. Some interviews were focus groups, and some were one-on-one interviews. Insight and clarification on the caseworkers' daily tasks and challenges were gathered through semi-structured interviews. This insight allows us to understand their needs for collaborating with AI, and their expectations for the future with AI. We also identified different challenges the caseworkers could face when working with AI.

## 1.4 Thesis outline

The structure for the rest of the report is presented here. Each main chapter has its own subsections.

Chapter 2 (Theoretical framing & related literature) gives a background in theory and literature that fit our research problem. This chapter aims to highlight important related literature findings on which we have based our thesis.

Chapter 3 (Case background & method) includes the research design and perspective and how data

was collected and analyzed. The interviews are explained, and background information relevant to the thesis is presented.

Chapter 4 (Findings from the data analysis) presents findings from the data collection process. These findings are categorized into different themes relating to our research questions. The results give insight into the caseworkers' thoughts, daily work processes, and challenges.

Chapter 5 (Discussion) uses findings from chapter 4 and related theory to present recommendations NAV could use to implement or improve human-AI collaboration in a trustworthy way for caseworkers. Limitations for this thesis are also presented.

Chapter 6 (Conclusion) includes the most interesting findings from the thesis and concludes what the findings contribute to NAV, public service organizations, and research. Implications for further research and practice in NAV are also presented, which could be useful to investigate for public service organizations and other researchers.

## 2 Theoretical framing & related literature

This chapter provides an overview of the framework used to guide the research. The systematic literature review used for literature synthesis is presented, as well as the related theory used in the study.

### 2.1 Theoretical framework: Joint cognitive systems

To guide this research, we used a theoretical framework. A theoretical framework is "a structure that guides research by relying on a formal theory; that is, the framework is constructed using an established, coherent explanation of certain phenomena and relationships" (Eisenhart, 1991). Eisenhart, 1991 argues that a theoretical framework can facilitate communication, encourage systematic research programs, and demonstrate progress among like-minded scholars working on similar or related research problems.

We based our theoretical framework on the works of Woods, 1986, from his research on Joint Cognitive Systems. We argue that his theory on joint cognitive systems is a good framework for our research on human-AI collaboration, as it explains how a typical computer consultant with the perspective on decision support would work. In his example, he created a hypothetical machine expert that would offer some sort of problem-solving, as shown in figure 2.1: The user initiates a session; the machine controls data gathering; the machine offers a solution; the user may ask for an "explanation" if some capability exists; and the user accepts or overrides the machine's solution (Woods, 1986).

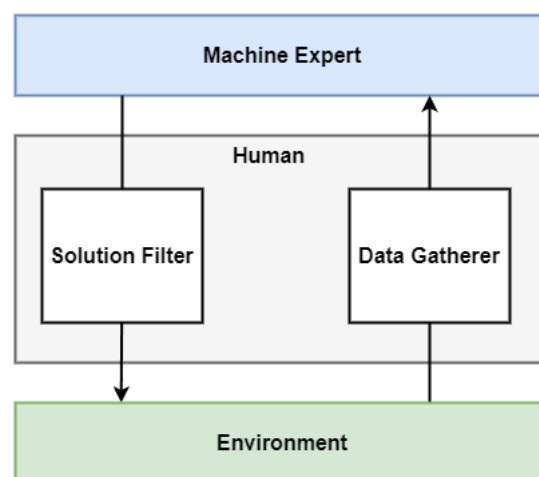


Figure 2.1: Hypothetical computer consultant in a Joint Cognitive System (Woods, 1986).

The classical interface design for the hypothetical consultant is not how to interface the machine to the user, but rather, "how to use the human as an interface between the machine and its environment" (Woods, 1986). "This emphasis results in a user interface design process focusing on features to aid the user's role as a data gatherer and on features to help the user accept the ma-

chine's solution" (Woods, 1986). As a result, the machine controls the interaction in this type of joint system. Woods, 1986 argues that a machine controlling the solutions can negatively affect the user and total system performance.

Contrary to the classical interface design previously mentioned, Woods, 1986 argues that planning, designing, modeling, and evaluating joint human-machine cognitive systems is key to applying technology effectively. He also argues that a problem-driven approach, rather than a technology-driven approach, is required to develop joint human-machine cognitive systems to aid the user in the process of reaching a decision, rather than recommending a solution Woods, 1986. This interface design on joint human-machine cognitive system might be a suitable approach to accountable human-AI collaboration.

Even though the research from Woods, 1986 is from the 80s, his theory on human-computer interaction is still highly relevant today. Recent research has been undertaken to extend and empirically test his theory. Balakrishnan and Dwivedi, 2021 built on this theory, where they researched the role of cognitive absorption in building user trust and experience through human-machine interaction with modern technology such as chatbots. IJtsma et al., 2019 also built on this theory in their research, where they proposed a methodology for making informed design decisions when determining the allocation of work and the interaction modes for human-robot teams (IJtsma et al., 2019). Stowers et al., 2017 built on the joint cognitive system theory in their research on the impact of agent transparency on operator performance, response time, perceived workload, perceived usability of the agent, and operator trust in the agent, in the context of military human-machine collaboration. Marathe et al., 2018 built on Woods, 1986 theory in their research on bidirectional communication for effective human-agent teaming, which is an approach that fosters communication between human and intelligent agents to improve mutual understanding and enable effective task coordination (Marathe et al., 2018).

## **2.2 Related literature**

### **2.2.1 Literature search procedure**

A literature review was conducted in a pilot study, establishing a good foundation for this thesis. The reasons for performing a systematic literature review are to summarize existing empirical evidence regarding technology and identify gaps in the current research. The summary is then used to suggest areas for further research. Another reason is to provide background to appropriately position new research activities (Kitchenham & Charters, 2007). The literature review was performed to examine what previous research has contributed to the topic of transparency in AI-human interactions, research gaps, and which areas need further research. We only included the essential parts of the literature review in the thesis; a link to the entire literature review can be found in appendix C: Literature review.

## Formulating search query

When searching for related literature, we formulated keywords relevant to our topic in the preliminary study. These keywords were also shown to be relevant to this thesis. We limited the number of keywords to ensure that the search results would be adequate. "AI" and "artificial intelligence" were chosen because they are central keywords for this study. To find previous research that was linked together with AI and transparency, we chose to use "transparency" and "transparent". Behavioral research is often combined with research and studies on human-computer interactions and is an important aspect of AI and transparency. The search query resulted in articles that seemed relevant to the study and gave us a good foundation for the theoretical background. There were some topics and articles that we did not consider during the literature search procedure, which we added later.

Search Query
TITLE-ABS-KEY(AI OR "Artificial intelligence" AND transparency OR transparent AND "Behavioral research" OR behavior OR behavioral) AND PUBYEAR >2016 AND PUBYEAR <2024 AND ( LIMIT-TO ( OA,"all" ) ) AND ( LIMIT-TO ( DOCTYPE,"ar" ) OR LIMIT-TO ( DOCTYPE,"cp" ) ) AND ( LIMIT-TO ( SUBJAREA,"COMP" ) )

Table 2.1: Search query used for the literature search (Eriksson & Olsen, 2022).

## Screening process

The literature screening process was done in three steps. We limited different search criterias in the first screening step. We limited the year of publication on the articles to a range from the year 2017 to the year 2023 to get the newest and most relevant documents. AI transparency is a relatively new area in information systems research, so most of the relevant articles related to this have been published from 2017 onward. Limiting the document year resulted in excluding 188 articles older than 2017. Figure 2.2 shows the increase in research on AI transparency from 2017.

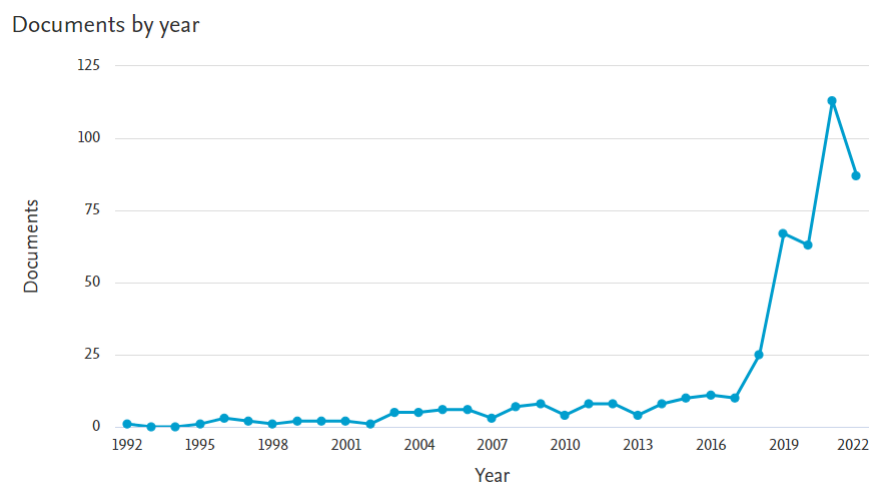


Figure 2.2: Documents by year(Eriksson & Olsen, 2022)

Next, we wanted to find document types related to articles and conferences to find mostly relevant empirical research articles. The document criteria were set to open access so that only documents

with free access were found. The subject area was set to computer science to find articles relevant to the topic of AI, and we also saw that most of the research on AI was done in the computer science area. The first screening step resulted in excluding 377 articles, leaving us with 93 articles for the next step.

In the second step of the screening process, we read through the titles and the abstracts of the 93 articles from the first screening step. After reading the abstracts, articles that did not fit our topic area were excluded. This screening step resulted in excluding 54 articles, leaving us with 39 relevant articles.

In the final step of the screening process, we did a thorough read-through of each of the remaining articles. After reading through 39 articles, we found that some were based on non-empirical research, and some were too technical to be used for the thesis. After excluding those articles, we were left with 13 relevant articles, which was insufficient for the thesis. Therefore, a backward search of the references was conducted on the 13 relevant articles, which resulted in finding seven more highly relevant articles. The result of the screening process and the backward search was 20 highly relevant articles that we included for the literature review. A visual overview of the screening process from identification to inclusion can be seen in figure 2.3. Additional theory related to human-AI collaboration, joint cognitive systems, public services, and other key characteristics of AI was added later.

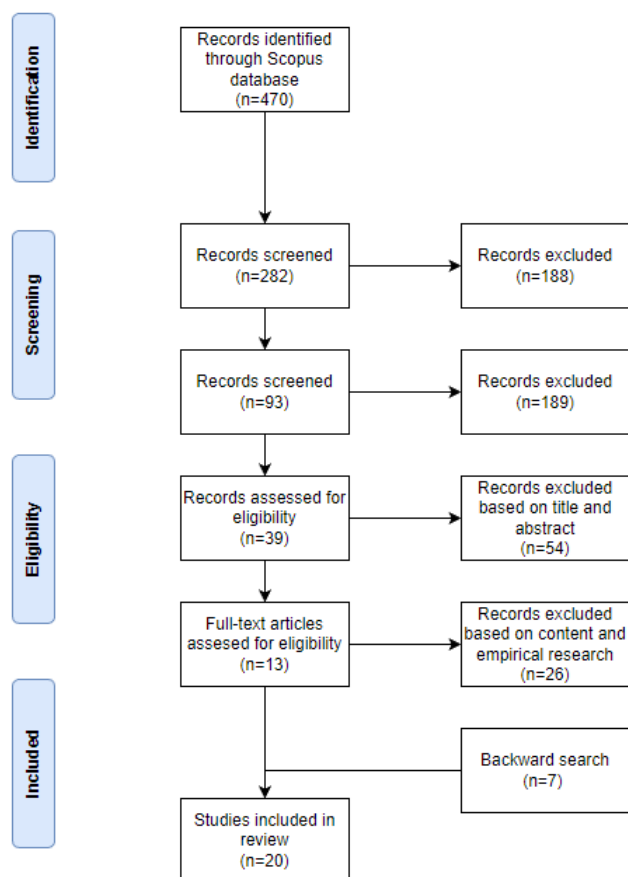


Figure 2.3: Prisma flow diagram of the literature screening process (Eriksson & Olsen, 2022)

The list of selected papers was mostly completed after the literature search using a backward and forward snowballing strategy. This strategy consists of identifying additional papers through the reference list from the selected papers and their citations (Wohlin, 2014).

## Results

From the resulting articles found in the literature search, we developed a concept matrix. The concept matrix was inspired by Webster and Watson, 2002 guide on writing a literature review. We identified numerous concepts that are important for this study. An article by Vassilakopoulou et al., 2023 researched the organization NAV and their service agents and chatbots. This article fascinated us, as they researched the same organization as we do in this thesis, and it sheds light on human-computer interactions, trust, and interpretability related to AI in public services. Through other articles, we also found many interesting challenges related to transparent AI, such as explainability, interpretability, accountability, black-box AI, trust, and ethics. The main technologies identified which are being used by public organizations are intelligent agents and chatbots. The intelligent agents mainly consisted of recommendation agents and decision support systems (Androutsopoulou et al., 2019; Aoki, 2020; Neururer et al., 2018; Vassilakopoulou et al., 2023). These types of technologies are also used by NAV, which was expressed in the interviews.

			Concept									
			Enablers		Challenges						Technology	
			Human Computer Interactions	Explainable AI	Black-box AI	Explainability	Accountability	Trust	Interpretability	Ethics	Intelligent Agents	Chat Bots
#	Authors	Year										
1	Daronnat et al.	2021	x					x			x	
2	Van Berkel et al.	2021	x	x	x	x			x			
3	Ehsan et al.	2021	x	x		x	x	x	x			
4	Setzu et al.	2021		x	x	x						
5	Segkouli et al.	2021					x	x		x		
6	Köbis, N., Mossink, L.D.	2021	x				x					
7	Maier et al.	2020				x		x	x			
8	Hepenstal, S., McNeish, D.	2020	x	x		x	x	x	x			
9	Oppold, S., Herschel, M.	2020					x			x	x	
10	Bigras et al.	2019	x					x			x	
11	Tubella, A et al.	2019			x	x	x	x	x	x	x	
12	Ehsan, U et al.	2019	x	x	x	x		x	x		x	
13	Neururer, M et al.	2018	x					x		x	x	x
14	Meske et al.	2020		x	x	x	x	x	x	x	x	
15	Vassilakopoulou et al.	2022	x					x	x		x	x
16	Androutsopoulou et al.	2019					x	x	x	x	x	x
17	Aoki	2020	x				x	x			x	x
18	Cheng et al.	2019	x	x	x	x	x	x	x			
19	Hind et al.	2019	x	x		x		x	x	x		
20	Liao et al.	2020	x	x	x	x		x	x	x		

Figure 2.4: Concept matrix based on Webster and Watson, 2002 (Eriksson & Olsen, 2022)



From the literature review, we found the concept of human and AI behavior. This term derives from behavioral science, and its definition is "the way in which one acts or conducts oneself, especially towards others", "the way in which an animal or person behaves in response to a particular situation or stimulus", "the way in which a machine or natural phenomenon works or functions" (Oxford-Dictionary, 2023). The concept of human and AI behavior goes into how they behave or react when interacting with one another. The concept matrix in figure 2.4 shows the concepts we found in our systematic literature review and how we categorized them into different themes, as explained above.

### 2.2.2 Findings

Our systematic literature review identified three key concepts for human-AI collaboration; transparency, accountability, and explainability. Additionally, the theory on human-in-the-loop and bias was found to be important and relevant to human-AI collaboration. This section will go through these concepts and theories and explain why they are important for human-AI collaboration. Transparency in the context of AI can be described as an umbrella term encompassing several overlapping and associated concepts. Figure 2.5 shows how these interrelated concepts relate to transparency. These various concepts speak to making AI more understandable and human-compatible both individually and societally (Biran & Cotton, 2017; Haresamudram et al., 2023; Larsson, 2019; Larsson & Heintz, 2020).



Figure 2.5: Key concepts related to transparency in AI

Recital 58 in the European general data protection regulation (GDPR), expresses that technological complexity makes transparency especially important in public services and states: "The principle of transparency requires that any information addressed to the public or to the data subject be con-

cise, easily accessible and easy to understand, and that clear and plain language and, additionally, where appropriate, visualization be used" (GDPR, 2019; Vollmer, 2022).

Informally, transparency is the opposite of opacity or black box-ness (Lipton, 2018). Due to the growing complexity of underlying models and algorithms, AI appears as a "black box" because the internal learning processes, as well as the resulting models, are not entirely comprehensible (Meske et al., 2022). The opacity of "black box" AI systems poses significant challenges from an ethical perspective, especially regarding questions of trust (Eschenbach & Warren, 2021). Relying on black box systems is becoming increasingly risky for their lack of transparency and the systematic bias they have shown in real-world scenarios (Mattu et al., 2020; Setzu et al., 2021). An AI model is considered to be transparent if it by itself is understandable (Arrieta et al., 2019). "By transparent, we mean that the different factors that influence the decisions made by algorithms should be visible, or transparent, to the people who use, regulate, and are impacted by systems that employ those algorithms" (Lepri et al., 2018; Tubella et al., 2019).

In contrast to black box algorithms, which explain how the inputs relate to the outputs without showing the internal workings of the model, there are also "white box" algorithms, which explain the model's inner workings. Cheng et al., 2019 researched non-expert stakeholders' understanding of these algorithms through interactive explanations and white-box explanations. They found that interactive explanations and white box explanations can improve users' comprehension, but it comes with a trade-off of taking more time to comprehend. They also found that users' trust in algorithmic decisions was not affected by the explanation interfaces or their level of comprehension of the algorithm (Cheng et al., 2019). Thus, finding the right balance between algorithmic accuracy and transparency is important. There is still no consensus on how much transparency is mandatory and at what level more risks can increase for the organization or people instead of more benefits (Benam, 2022). Transparency is also seen as the, and is often taken as, a requisite for algorithmic accountability (Bryson & Winfield, 2017).

## **Accountability**

Transparency does not imply accountability nor responsibility, but complements and extends these (Tubella et al., 2019). "Accountability refers to being answerable to somebody else, to be obligated to explain and justify action and inaction" (Olsen, 2014). Accountability in the context of AI refers to an expectation that organizations and individuals ensure the proper functioning of AI systems they design, develop, operate, or deploy. It is also expected that they are in accordance with the roles and regulatory frameworks applicable, and demonstrate this through actions and decision-making processes. ("OECD AI Principle - OECD.AI," 2023).

AI can produce outcomes that do not reflect relevant factual circumstances. Galdi and Cordella, 2021 mentions that accountability issues arise when these outcomes are used as inputs to support public sector decision-making. AI can bring change and benefits to the public and private sectors. Bataller and Harris, 2016 underlines that, more specifically, virtual workforces, so-called intelligent automation, labour and capital augmentation where AI can complement the skills of existing

workforces, can lead to cost efficiency and savings. A lack of accountability in AI decision-making systems has raised concerns about using AI to support public sector service delivery (De Fine Licht & De Fine Licht, 2020). These concerns seem to revolve around ethical and legal aspects (Helbing et al., 2017).

A realistic but extreme example of why accountability is an important topic and challenge for artificial intelligence is the case from Busuioc, 2021 of self-driving vehicles: AI is anticipated to encode ethical life-and-death decisions. Should the self-driving vehicle prioritize saving its driver/passengers or pedestrians if a crash occurs? Who would be accountable for the decisions of the self-driving vehicle in this case? To mitigate the limitations of algorithms, human-in-the-loop systems have been proposed to increase algorithmic accountability (R. Wang et al., 2020; Zhu et al., 2018).

### **Human-in-the-loop**

Keeping humans in the loop helps to monitor and adjust the system and its outcomes (Rahwan, 2018). Human-in-the-loop aims to train an accurate prediction model with minimum cost by integrating human knowledge and experience (Wu et al., 2021). As Rahwan, 2018 mentions, human-in-the-loop (HITL) can be a powerful tool for regulating the behavior of AI systems. A human can identify misbehavior by an otherwise autonomous system and take necessary corrective action. In case the system misbehaves, the human can be involved to provide an accountable entity. Benedikt et al., 2020 states that automation problems in governmental applications that require high accuracy require HITL to be used. They also advocate for automation as a means to make efficiency savings and to speed up processing times. Some situations might require human intervention to maintain data quality. Human-machine teams are believed to offer superior results and build trust by inserting human oversight into the AI life cycle (Middleton et al., 2022). The ideas of HITL have been studied within the field of supervisory control for decades (Allen et al., 1999; Rahwan, 2018; Sheridan, 2012). After some time, these ideas made their way into the field of human-computer interaction (HCI), where developments of systems that can make intelligent decisions about how and when to engage the human began (Horvitz, 1999; Rahwan, 2018).

The combination of HITL with artificial intelligence and machine learning has interested researchers for some time. Training the algorithms in these types of systems could be explained as using the human workers to label the data, which the machine learning algorithms base their decision on (Dellermann et al., 2019; Ostheimer et al., 2021). HITL machine learning attempts to leverage the benefits of categorization skills and human observation and machine computation abilities to create improved prediction models (Fails & Olsen Jr, 2003; Schirner et al., 2013; Shih, 2018). HITL suggests that once we put human experts within the loop of an AI system, the regulation problem is solved (Rahwan, 2018). This is not always the case, as bias from humans and machines can be a problematic factor in regulation.

## **Bias**

Bias and fairness are complex human notions. While "bias" can refer to any form of preference, fair or unfair, this thesis uses the term to mean "unfair," "unwanted," or "undesirable" bias (Silberg & Manyika, 2019). Unfair bias can be described as systematic discrimination against specific individuals or groups of individuals based on the inappropriate use of particular traits or characteristics (Friedman & Nissenbaum, 1996). Perhaps the most discussed forms of "unfair bias" in the literature relates to specific groups or attributes, such as disabilities, race, gender, and sexual orientation (Silberg & Manyika, 2019). Osoba and Welser, 2017 lists real-world examples of applications and the ways AI systems affect our daily lives with their inherent biases, such as the existence of bias in AI chatbots, employment matching, flight routing, automated legal aid for immigration algorithms, and search and advertising algorithms (Mehrabi et al., 2021). These applications directly affect our lives and can harm society if not designed and engineered correctly, with considerations to fairness (Mehrabi et al., 2021).

Biases can be seen along a spectrum of human attitudes spanning our values, beliefs, allegiances, opinions, preferences, interests, stereotypes, prejudices, and misunderstandings. This broader definition shows how deeply biases are built into the human psyche and why they can never be entirely eliminated (Moschella, 2022). While biases in humans span across different human attitudes, biases in machine-learning systems are only as objective as its underlying dataset (Moschella, 2022), often referred to as "algorithmic bias". "If the dataset reflects different types of biases, so will the AI system. If specific populations are not sufficiently represented statistically, then the system will be weak in those areas. If a survey sample is not representative of the real world, or if the wording of a survey's questions is slanted in some way, these deficiencies will affect the survey's results" (Moschella, 2022).

Kleinberg et al., 2019 argues that there are two separate "algorithms" in an AI model; the "trainer", which can be biased by the underlying data and training process, and the "screener", which makes the predictions based on the trainer (Kleinberg et al., 2019; Silberg and Manyika, 2019). Kleinberg et al., 2019 notes that the distinction between the two separate algorithms is often underappreciated and that they are quite important in practice. They state that "people often worry that the algorithm could do anything, including unforeseen things that introduce bias. Recognizing there are two algorithms helps clarify that conditional on the choices that go into constructing the trainer, the screener cannot do "literally anything"; it is mechanically the result of whatever human decisions were made for the trainer" (Kleinberg et al., 2019).

For decades, society has seen unfortunate results from human biases, which stem from a lack of understanding and professional expertise (Moschella, 2022). However, "machine learning can establish a software and data foundation on which better and more consistent decision-making capabilities can be built" (Moschella, 2022). "While definitions and statistical measures of fairness are helpful, they cannot consider the nuances of the social context into which an AI system is deployed, nor the potential issues surrounding how the data were collected" (Silberg & Manyika, 2019). Therefore, it is important to consider where human judgment is needed and in what form,

besides providing definitions and applying statistical techniques (Silberg & Manyika, 2019). This is especially important to consider in the context of public services, where human lives are directly affected, and proper judgment is crucial.

### **Explainability**

"Explainability is associated with the notion of explanation as an interface between humans and a decision maker that is, at the same time, both an accurate proxy of the decision maker and comprehensible to humans" (Arrieta et al., 2019). Explainability in the context of AI can be achieved through explainable AI (XAI) methods and processes. Through our systematic literature review, we observed that XAI has had a resurgence in the AI/ML research field in the last few years. "The re-emergence of this research topic is the direct result of the unstoppable penetration of AI/ML across industries and its crucial impact in critical decision-making processes, without being able to provide detailed information about the chain of reasoning that leads to certain decisions made by it" (Adadi & Berrada, 2020). Adadi and Berrada, 2020 argues that XAI is essential if users are to understand, trust, and manage AI results to meet the social, ethical, and legal pressure of making AI decisions explainable and understandable.

Therefore, XAI is especially important in public service delivery, such as healthcare and finance, where there are significant consequences for the decisions the AI systems make. In these domains, users of the AI systems must be able to understand and interpret the decisions made by the AI system to make informed decisions based on the system's output. As XAI methods explain why an AI system arrived at a specific decision, it also increases transparency in how AI systems operate and can lead to increased trust (Pawar et al., 2020). Gilpin et al., 2018 argues that the transparency of a models behavior is not enough by itself to satisfy the goals of gaining the users' trust or producing insights about the cause of the decisions. Instead, explainability requires capabilities such as providing responses to user questions and the ability to be audited (Ehsan et al., 2021).

## 3 Case background & method

This chapter explains the reason for the design choices made and describes the research method used for this thesis. The case background for the study is presented, followed by the research approach. Further, the interview design and data gathering is described. The data analysis approach is then described, followed by validity, reliability, and ethical issues related to the research.

### 3.1 Case background

The Norwegian Labour and Welfare Administration (NAV) is an organization that administers approximately a third of the Norwegian national budget through schemes such as work assessment allowance, sickness benefits, and unemployment benefits. The main goal of NAV is to get more people active and in work, thus having fewer people relying on benefits (NAV, 2023a). NAV created an AI lab in 2017 to ensure specialist expertise in the area. One of their core values is to operate with artificial intelligence responsibly. There are many relevant application areas for AI in NAV, one of which is the area of sick leave (Jensen & Lyngstad, 2019).

Reducing workplace absenteeism due to illness is a government priority in Norway (Fineide et al., 2019). In the spring of 2021, a sandbox project that dealt with NAV's AI tool to predict the development of sickness absence started. This project aimed to clarify the legality of using AI in these areas and explore how citizens on sick leave could be profiled in a fair and transparent way (Datatilsynet, 2022). It also highlighted major challenges for public services that want to use artificial intelligence and opened up more areas in the AI field to research. NAV has strategic cooperation agreements with many universities and colleges in Norway (NAV, 2023b). One of these universities is the University of Agder (UiA), which is also a part of the AI4users research project.

The AI4users research project aims to address the "black box" AI problem and contribute to the responsible use of AI in the digitization of public services. The project takes a human-centered perspective for the development of tools and design principles to help users of AI to trust the solutions (AI4users, 2020). AI4users primarily targets non-experts to expand the scope of research into the responsible use of AI beyond experts in the field. In collaboration with our supervisors, we actively engaged with a group of researchers involved in AI4users. This partnership led us to become a part of their project, enabling us to collaborate with a designated employee from NAV who served as a research champion. The research champion played an important role in facilitating interviews with caseworkers from various offices within NAV Agder. By conducting interviews at both smaller and larger offices in the Agder region, our intention was to explore potential variations in caseworkers' perceptions of AI and their daily work routines while mitigating potential biases that could arise from exclusively interviewing caseworkers from specific offices.

## 3.2 Research approach

This study aimed to get insight into caseworkers' needs for human-AI collaboration and their expectations and challenges regarding AI. To achieve this, we used a qualitative method. A qualitative method was chosen to give a better background for understanding the caseworkers' thoughts and to bring out the best possible future for AI in public services based on the insights from the caseworkers. In addition, the method provides a better understanding of caseworkers' experiences against the research questions. Combined, this gives a broad perspective on decisions, with more detailed insight into their different situations and daily tasks (Barrett & Twycross, 2018). Barrett and Twycross, 2018 explains that achieving a broad understanding requires holistic, rich, and nuanced data to allow themes and results to arise through the analysis. Interviews can then be asked with open or closed questions related to different types of interview structures. These interviews can be unstructured, structured, or a combination of both (Creswell & Creswell, 2017; Saunders et al., 2019).

A case study is an empirical investigation that explores a contemporary phenomenon within the context of reality, using several sources as a basis (Noor, 2008). The purpose of a case study is not to study the whole organization, but to provide context and a backdrop for the phenomenon that occurs in the organization (Noor, 2008). This method gives researchers the opportunity to explore the "how" and "why" of real-time events, situations, and problems that do not require control of events or problems (K. Yin, 2013).

We conducted an in-depth case study of the particular case of how caseworkers at NAV both understood and interacted with AI in their work, and their needs and expectations for future human-AI collaboration. Oates et al., 2022 states that case studies are suitable for both theory building and theory testing, which is ideal for this study, as there is a lack of prior research in the area of human-AI collaboration (Lai, 2021). The principle purpose of the case study is "to shed light on a decision or a series of choices; why they were made, how they were done, and the result of these" (Schramm, 1971). Case studies can be divided into three types: descriptive, exploratory, and explanatory. The descriptive case study will lead to a detailed and rich analysis of a phenomenon in a given context. The exploratory case study is used for defining questions and hypotheses for potential new studies, while the explanatory study aims to answer the "how" and "why" questions about a particular phenomenon (Chopard & Przybylski, 2021; Oates et al., 2022). A short-term, contemporary study is most suitable for this thesis, as it examines and tries to explain what is occurring at the present moment. As for the type of case study, we chose an explanatory case study, as it gave us the opportunity to compare our findings to theory from existing literature (Oates et al., 2022). To fully decide which research strategy and design we wanted to use for this thesis, we used the research process model from Oates et al., 2022 for inspiration.

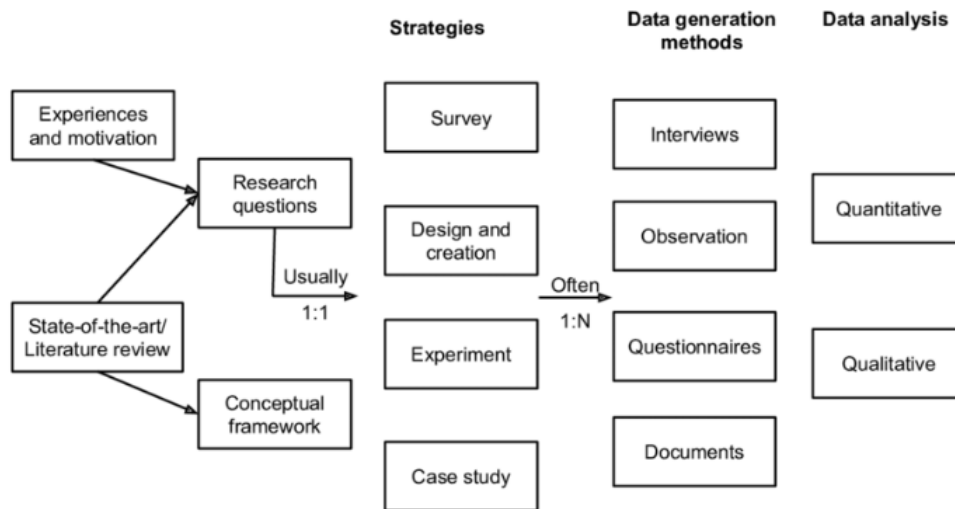


Figure 3.1: The research process (Oates et al., 2022)

Based on the research process in figure 3.1, we started by doing a literature review and a pilot study, which created the basis for the research question and supporting sub-questions for this thesis. The ideal research strategy, a case study, was chosen based on the research question and supporting sub-questions. We conducted interviews as our data generation method and used a qualitative data analysis method as a result. The formulation of the research questions was an iterative process, and they were slightly changed throughout the duration of the thesis.

### 3.3 Interview design

Interviews are often used as data generation methods in case studies. They can be suitable data generation methods when a researcher wants to obtain detailed information, i.e., "ask complex questions, need explanation or follow-up, or open-ended, or whose order and logic might need to be different for different people" (Oates et al., 2022). The interviews in this research were designed as semi-structured, with questions related to themes we wanted to investigate. Since it was semi-structured, there was also room for additional questions or themes depending on the flow of the conversation.

The interview guide was designed in collaboration with a researcher from the AI4Users project. The interview guide can be found in Appendix B, and was split into three themes; demographic and tech literacy, caseworker's current processes, and a foresight method with a fictive scenario. We also had discussion prompts if needed, as some participants might need guidance to get into the discussions.

The first theme was general demographic questions about the participants, such as their age, official job title, and how long they had worked in their respective roles. We also asked about their self-reported prior knowledge about AI and their self-reported frequency of technology use, on a scale from 1 to 5. These questions were asked to give us a better context when analyzing the data later



and to engage the participants in the interview process.

The second theme was questions about their current processes as caseworkers in NAV. We asked if they could tell us what their individual daily work tasks entailed, e.g., decisions they had to make, dialogue meetings, type of support, or other decisions. We also asked a follow-up question about how they thought AI/ML could be used for their decision-making processes. These questions were asked to better understand what the caseworkers do in their work, their needs for AI, and how AI could help them in these processes.

The third theme was questions related to what the caseworkers wanted or not, to happen in the future. We did this through a fictive scenario using a foresight method. The foresight method aimed to explore possible futures in relation to AI in their workplace. The fictive scenario was based on a tool that could predict the duration of a citizen's sick leave, and based on the scenario we wanted the participant to think about how the different futures might look like. The different futures were split into four directions; best-case/worst-case, probable, preferred, and possible futures. As non-experts in the field of AI, the fictive scenario made it easier for the caseworkers to think and describe how AI could assist them or make things worse in the future. This gave us more insight into what caseworkers might want from an AI system in the future, and what they are worried could go wrong utilizing AI in their work.

## **3.4 Data gathering**

The data gathering will be further explained in terms of background and how the data collection was done. The explanation gives an overview of the interviews and the basis for the results.

### **3.4.1 Preliminary data collection**

Before conducting the main interviews for this research, some preliminary data collection was conducted (Eriksson & Olsen, 2022). It allowed us to find out what needed to be studied and what should be taken into account for the results to be reliable (Ylikoski & Zahle, 2019). By conducting this preliminary data collection, we had the opportunity to evaluate our analysis method before the case study to ensure that all aspects of the process were taken into account. For the preliminary data collection, we interviewed managers of NAV, primarily because it was difficult to contact caseworkers directly without an established collaboration. We interviewed three managers at a NAV office in Agder, and all of them had prior experience or currently did some of the same tasks as current caseworkers in NAV (Eriksson & Olsen, 2022). They also had some knowledge of the topic of artificial intelligence. We conducted one-on-one semi-structured interviews through the messaging and collaboration application Microsoft Teams. Before the interviews, we sent the managers an email with relevant information regarding the interviews they would participate in. This included the interview topic and duration, and we scheduled a suitable time to meet. The preliminary data collection focused on transparency in AI/Human interaction and its impact on caseworkers at NAV.

The first questions from the interviews were related to their tasks and responsibility as managers because we wanted some knowledge about what their general tasks were and to make the interviewees more interested and comfortable sharing information with us. We further asked the interviewees about their knowledge of AI, with the intent of helping us get their baseline understanding of what AI is. The final set of questions we asked were related to their experience of AI in their work, their understanding of any AI systems they might have used, their trust in AI, and if they received any form of explanation from the systems they used.

### 3.4.2 Case study of caseworkers at NAV based on sick leave case handling

The case study was conducted through semi-structured interviews with a total of 16 caseworkers at NAV Agder. The interviews were conducted in person at three different NAV offices in Agder. At the first office, the interviews were structured as group interviews, where each interval of interviews consisted of 2 interviewees and the researchers. This was because we wanted to get the caseworkers to interact with each other and have a group discussion from which new insights might arise (Oates et al., 2022), and because of the time the caseworkers had available. The reason for interviewing caseworkers specifically was the importance of knowledge about non-experts' experience and understanding of AI in their work. Each interview was recorded and then transcribed in a Word document stored on a private drive only accessible by us and the collaborative researchers.

The first interviews took place in November 2022. These interviews were a part of the preliminary data collection and were included as they had similar aims as this thesis. The interviews in December were group interviews where two caseworkers participated together. We wanted to see if it would help the participants answer by allowing them to discuss with each other. The interviews in January and March were one-on-one interviews. Alongside these interviews, work on theoretical background and literature was done, assuring that the interviews had relevance to our study. The duration of the interviews was between 30 and 45 minutes approximately, as can be seen in table 3.1. This table also shows the method, study type, and number of participants.

The interviews might give answers that could give a wider empirical foundation for the human-AI collaboration literature. They might also help with design principles in the development of AI solutions for the purpose of helping users to understand and trust AI solutions. As stated above, the interviewees are non-experts, and by interviewing them, the research area is expanded to those who are affected by these AI systems.

Date	Method	Duration per interview	Total no. of participants	Study type
14.11.2022	Digital	30 min	1	Preliminary
18.11.2022	Digital	30 min	1	Preliminary
23.11.2022	Digital	30 min	1	Preliminary
07.12.2022	Face-to-face group interview	45 min	8	Case study
26.01.2023	Face-to-face	45 min	4	Case study
23.03.2023	Face-to-face	45 min	4	Case study

Table 3.1: Table of interviews.

### 3.5 Data analysis

After conducting the interviews, we analyzed the audio recordings resulting from the interviews. We started by transcribing the interviews to convert the audio data into text format to extract the data we had found. Initially, we started by transcribing a couple of the interviews manually, but since this was time-consuming, our supervisors introduced us to a new transcribing software called Autotekst. Autotekst is an auto-transcribing software developed by the University of Oslo, utilizing OpenAI to transcribe audio files into text format (Universitetet i Oslo, 2023). This tool saved us a lot of time while providing fairly accurate transcriptions, with a neat structure to the data.

After running the audio files through the software, we read through all the transcriptions while listening to the audio files. This was done to ensure everything was correctly transcribed, increasing the validity and reliability of our data. We made adjustments in areas where Autotekst made translation errors.

After all the interviews were transcribed, we began coding the transcriptions in the software NVivo. Coding the interviews made it efficient for us to categorize the findings into different themes. This was an iterative process in which we found new themes while working with the findings, and new codes were added accordingly. The themes were later used for the findings and discussion part of the thesis. The categories for the codes we used were based on the questions we asked in the interviews, themes related to the research questions, and related theory.

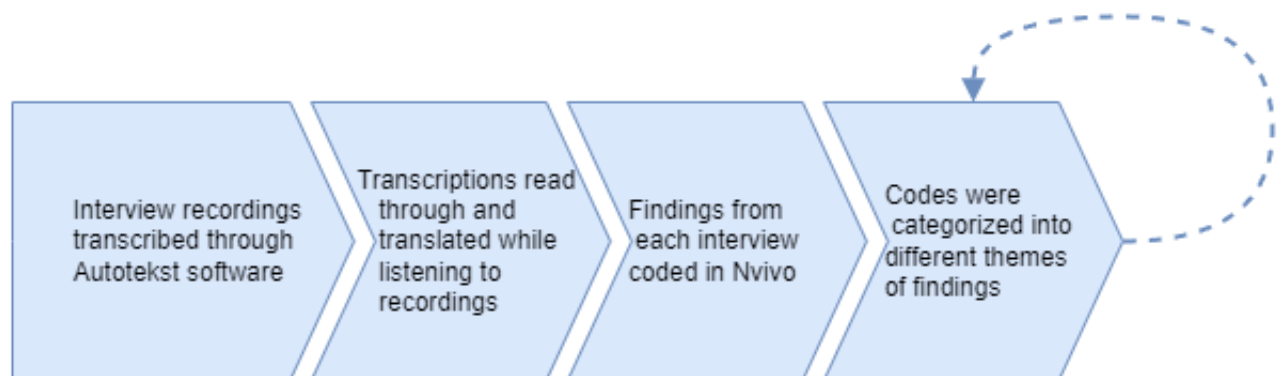


Figure 3.2: A visual representation of the data analysis process.

We used a thematic analysis method to find the background for the caseworkers' needs and challenges related to AI. Thematic analysis is a method for identifying, analyzing and reporting themes within the data collected (Braun & Clarke, 2006). When using this method for open ended responses as we did, it allowed flexibility and interpretation when analyzing the data. To ensure confidence in the findings, it was needed to be attentive to transparency of the method (Braun & Clarke, 2006). As mentioned above, the raw data from the transcriptions were formed into themes and codes (Castleberry & Nolen, 2018). We used the thematic analysis to explore and organize ideas and thematics to increase understanding, and to link the ideas to empirical data (Guest & McLellan, 2003; Richards, Richards, et al., 1995). The thematic analysis were used to find the needs and challenges of AI for the caseworkers, explore the cause, and code them into differ-

ent themes relating to one another. A thematic analysis is purely an indicative process, and not concrete rules for how to proceed (Braun & Clarke, 2006; Williams & Moser, 2019).

### **3.6 Validity and reliability**

Validity is used to describe if the data collected is relevant in relation to the problem to be elucidated (Halvorsen, 2008). It describes the extent to which a measure accurately represents the concept it claims to measure (Punch, 2013; Roberts & Priest, 2006). When doubts can be raised if variables that emerge from the research reflect the theoretical concept from which one works, it is a question of validity (Halvorsen, 2008).

We ensured that we had sufficient validity in our research by utilizing multiple techniques. We used multiple sources of evidence, investigators, and methods during the data collection phase to enhance credibility (Riege, 2003). We did this by interviewing different people at different offices, both with one-on-one and group interviews, and being accompanied by different researchers in the interviews. Additionally, researchers within the same research area reviewed our findings, increasing the validity. We also critically reviewed the sources used, and tried to use multiple sources from the literature to create a chain of research to compare (Riege, 2003).

Reliability says something about how reliable the measurements that have been made are (Halvorsen, 2008). A high degree of reliability means that almost no matter how the measurements are done, it would give roughly the same results. Reliability is the degree to which measures are free of errors, and therefore yield consistent results (Halvorsen, 2008; Lakshmi & Mohideen, 2013). To ensure reliability, it is especially important that the measurement process is as precise as possible. High reliability is an important requisite for data to be used in connection with the problem that has been raised.

Sufficient reliability was ensured through different methods. We conducted a pilot study, which ensured that we asked relevant questions regarding the research problem in this thesis (R. K. Yin, 1994). We used a semi-structured case study protocol, which ensured that we asked the same questions to everyone participating in the interviews. We recorded data digitally to capture everything that was said, eliminating researcher bias and ensuring nothing was left out in the analysis. The coding process were done to ensure that the analysis was defined carefully and consistently to maintain validity and reliability of the coding results (Williams & Moser, 2019). The validity and reliability were ensured through the presence of researchers and awareness linked to the dynamic interaction between qualitative data and researchers, that was further structured in the coding process (Green & Johns, 2019).

### **3.7 Ethical issues**

Ethical thinking could help society to trust the research being done, for example, that humans act honestly. New methods and technology within the social sciences and major political, economic,

and social changes in society are making ethics increasingly more complex (Israel & Hay, 2006).

There could be ethical issues if the caseworker wants to retract their statements/answers, which may damage the outcome of the report, but something we have to comply with because it is their right to withdraw from the research at any time they like.

Participants shall have the right to privacy. We always assessed the data's sensitivity and ensured we complied with the Personal Data Act. The possibility of identifying individuals was also taken into account, which is why we have not stated specifically which offices the interviews were conducted at. We made sure to transcribe the interviews as thoroughly as possible, to ensure that we reproduced conclusive results in the correct context. This was done so that we did not discuss something the participants believed was taken completely out of context (Jacobsen, 2015).

In this research, we always relied on informed consent, where the people participated voluntarily and could stop participating at any time. We informed the participants of this several times throughout the interviews. Participants in the project must always be aware of the consequences that the study may entail (Jacobsen, 2015). The participants were all given a consent form, which had information on the purpose of the study, and how the data were to be used. It is important to avoid building up a sense of fear or various negative suspicions so that the trust researchers have built up is not weakened or lost (Israel & Hay, 2006).

## 4 Findings from the data analysis

In this chapter, we present the findings from the data analysis (see table 3.1 for an overview of the interview subjects). In total, there are 15 semi-structured interviews; 11 one-on-one interviews, and 4 focus group interviews. Three interviews were conducted as preliminary data collection. The names of the interview subjects are anonymous and will therefore be referred to as (S1 to S16) and (P1 to P3). The quotes from the interviews are only a few examples of what was expressed and are there to bring out the most important findings.

### 4.1 Caseworkers' needs for AI

When asked about how AI could help the caseworkers in their daily work processes in the question about "current processes" from the interview guide, they also expressed some of their current needs for AI. Their needs were subjective, and varied from person to person.

One thing that was pointed out was that the employees at NAV still work with old systems, in an "old-fashioned" way towards the users. It was also mentioned that regulation and cooperation with employers needs to change, something that the caseworkers thought machines could do.

*"I think that many of the decisions we make today can disappear. I think that the way we work today have to be changed a lot. We are behind, we work on old systems, and we work on the old-fashioned way towards the users and the people on sick leave. Also regulations and cooperation with employers and that bit needs to change. I think that machines can do something to make things easier every day" (S11).*

Some of the caseworkers expressed their need for a way to sort out unnecessary dialogue meetings, so that they could spend more time and focus on the people that really needed these meetings.

*"I think it is an exciting thought, because how do you sort out the ones who need the most attention, the ones who need the most follow-up? Because there is a large number of people who come into NAV when they are on sick leave. It needs to be focused on finding the people who needs us the most, and i think that is something AI could help with" (S14).*

When we asked a follow-up question about how the caseworkers would utilize their time if they did not have to do the tasks they deemed as unnecessary, some of them mentioned that they would focus more on the people with low resources that cannot take care of themselves. The citations illustrate how many of the caseworkers had similar needs in regards to their focus on those who needed help the most.

*"We could use our time on the most important cases, those who really need it. Those with low resources who can not take care of themselves" (S4).*

*"We have lots of people (users) who are just strolling around because we do not have time to be on the case" (S3).*

*"If I did not have to have a dialogue meeting, I could have spent more time with those who need it" (S15).*

*"Maybe we could be more ON in the cases we have, have fewer and better cases, and be more often in them. As long as a meeting is not requested, and we have the information we need, and their workplace does not request a meeting, and they are working like 50%, they are rolling in the system, then maybe we could be more on them to push them a bit faster back to work. Or what we could do to make that happen, and what they can do for themselves. We could use our time on those who needs us the most" (S3).*

*"I just think we have so many people on sick leave to attend to, and we would get more time with the person on sick leave or the employer to advise" (S10).*

The caseworkers hoped that artificial intelligence could simplify their daily work, and they could avoid any unnecessary work that comes with not having all of the needed information readily available.

*"I would hope that AI simplifies my daily work. That it gives me profit in relation to the fact that I can do a better job. That I can see clearer in a situation, because I got information. I get to free up more working hours to do the tasks which is important" (S14).*

*"We are connected up to payment of sickness benefits, and there are big problems there because they have to wait up to six months before they get their money... I am sure something could be better there... I also think something could be done around assessing the dialogue meeting. It could be done more like assessment and information going out in that area" (S10).*

Most of the caseworkers were positive to digitization, because it could save them time that they normally would use on unnecessary tasks. In terms of what they think should be digitized, there are some differences.

*"More digitization is really good, and saves us a lot of time so that we can spend our time on what is important, like having meetings with people(the users)" (S5).*

*"Previously there have been many dialogue meetings that have been unnecessary. There has been a policy that all dialogue meetings should be held, and then there were some meetings that were unnecessary for both the doctors and employers" (S14).*

*"I do not think we have to be less people in NAV who work with what we work with, because I think that is not the way it should be. I think that we should work in a more streamlined way" (S11).*

*"I wish the meeting notices could be simpler. And maybe decision could be easier, what we send out to our patients and users. That it would be easier to understand, and fewer questions received in return" (S15).*

*"To have a tool that can tell you that this should actually collect information, tell you what should happen here in a very short time. That's what we want. We want to have people out at work again, and to get help to do that" (S9).*

*"I would like a good combination of talking to people and having support in the form of an artificial intelligence. Where you can have an assembly line, it is very sensible. You must never forget that you must talk to people" (S9).*

The common feature for these cases was how the caseworkers believed artificial intelligence and digitization could give support, and help change the way that they work. At the same time, they did not want technology to prevent anyone from getting the help that they need, which is a challenge we will go into more detail on.

The reason for wanting to digitize the process of talking to users was mostly because of time and efficiency.

*"It is mostly because of time, like efficiency. I feel that I save a lot of time just by answering questions in writing, because as soon as you pick up the phone, the conversation can last for a long time. It is also easier to see what they are actually asking in writing, and then you can just answer that concrete question" (S1).*

*"If you look at history, reports and meeting reports could have been much more automated. That would save us a lot of time, and certainly cooperation partners too" (S16).*

The subjects from the preliminary interviews believed that artificial intelligence could make their jobs easier, by receiving recommendations or information automatically, which they normally would have done manually. The systems where there has currently been implemented AI has decreased their workload considerably.

*"AI can help us in cases of doubt by giving a recommendation, so we do not have to think about it so much more if it comes with a certain outcome. Then it will be easier for us to make a decision" (P3).*

*"That the interaction between those who need us and what we want to help people with could flow a little easier. I would guess artificial intelligence will be able to help us with that in the future" (P2).*

*"That it could retrieve information from the doctor's report. If an AI could enter all the parameters we consider, it would have helped a lot. Then we could avoid having to write everything... and it would save us time." (P3).*



*"The amount of work in the departments that have been dealing with mail distribution has decreased considerably, because they are helped by the artificial intelligence" (P2).*

*"In relation to work capacity assessment, if a system could set up the effort required so that it is correct, it would make my assessment easier. It can be quite difficult to choose between one or another input needs, and there is a lot I have to decide on. A robot of some sort might be able to help me with that" (P3).*

When asked about what they reasonably would expect to happen with such an AI system in the future, many of the caseworkers expressed a need for a human to take a final decision if an AI system were to give recommendations, or if an AI system were to do certain tasks.

*"I think there must be a human advisor. It is fine with simple functions like moving tasks and such, but if you have more complex situations, there is a lot of discretion here. In my head a machine cannot assess discretion. So I think there must be a human being in the final process. ...the final piece of quality assurance is where a human is needed" (S14).*

*"There must always be people involved in the process. A robot can give information and read information, but it is not always logical" (S9).*

When asked about how they would feel if a machine acted as a coworker or a decision support tool, it was expressed that it sounded exciting as long as they had the control, and that humans direct the AI.

*"I think that would be a little more exciting, if it was still me who had the control. You can get new ideas or encyclopedias and use them as help, so to speak" (S9).*

*"If it is about people who can manage themselves and just want to look for a job and do not need anything else other than applying for daily allowance and knowing that you have to fill out a report card and maybe a job course or two, then it is much more natural to use automated processes. But people who come to us who are sick, they have low concentration. You cannot expect them to be able to do as much, and they need help in a completely different way" (S7).*

*"It can be a certain help, but NAV must be person-based I think, because of what we do"(S7).*

*"It must be controlled so that we see that things work, that is is directed" (S14).*

*"I do not know really, I like to speak with people. I do not think everything can be done by technology, because I have to speak with them" (S5).*

It was expressed that NAV has already tried to automate some decision making processes, and for one case they were told that the process was stopped because they were not allowed to send a

decision without a human being involved in the decision-making process. Several caseworkers has stated that laws and regulations are involved in delaying the development of AI systems.

*"They (NAV) tried to achieve an automatic decision that made the I4A decision, which was stopped. But they did not get far as to be allowed to send a decision without a human being involved... we depend on good questions being asked to the user, and that the user has good self-awareness. So if we fail at one of the two things there, it doesn't help us much" (P1).*

## **4.2 Caseworkers' expectations for the future with AI**

During the interviews with the caseworkers, we presented a fictional case based on a sick leave support system utilizing AI. After we presented the case, we asked them what their thoughts were on best-case and worst-case futures using such a system.

### **4.2.1 Preferred future**

A common denominator for the subjects was that they thought that the systems they have today is not satisfactory. This goes for the interaction systems with the doctors, but also for the internal systems they work in on a daily basis. They preferred a future in which they work in one integrated seamless system, where some tasks are automated, and where they could easily interact with different actors.

*"One system to work in or a few would help, more interaction with doctors in real-time would help a lot because we spend a lot of time just waiting, and also more interaction with the people who need us quicker. Some of the small tasks could be taken away that maybe a machine could do. It is hard to say what tasks, but of course there could be something there" (S4).*

*"A better interaction system for us with the users and for us with the doctors, because I think the interaction system now with the doctors is very time consuming" (S2).*

*"You know at the time that I have worked here, we just keep getting more systems to work in, but we do not phase out the old ones... And now I don't know how many there is. It's hard to put everything into one system, but that would make our day-to-day work easier just to have that one system" (S1).*

*"Everything in one system and with easier interactions, maybe for both users and doctors. Some kind of better communication systems within the systems to different actors. Because we do talk to activity leaders, users, doctors, employers. So to make those communications easier for everyone" (S2).*

Having time to focus on the most important tasks, and having automation of some tasks were something the caseworkers preferred. One of the tasks that could be assisted with AI was sorting

out who needs the most help or personal connection.

*"If I had time to actually interact with users, do meetings, phone calls and all these other things that are just click, read and finish. Like automating some of the tasks that still needs to be done by me, cause it takes a lot of time during the day and during the week" (S1).*

*"How do you sort out those who need the most attention, and the ones who need the most follow-up? It needs to be focused on finding the people who needs us the most. That is something that artificial intelligence could help with" (S14).*

*"Technology where we could have one good system so we can read information that we need from one UI, and not search in lots of systems to find the information because it is very easy to miss something" (S3).*

*"I could wish for a system that could call in and find a time for the candidate, the supervisor and the doctor, when it comes to dialogue meetings. With this, we would not have to spend half an hour to sit on the phone to find when the doctors are available" (S15).*

Because of tasks governed by laws and regulations, the caseworkers expressed that artificial intelligence potentially could recommend what to do, where the caseworkers make the final decision in the process.

*"Some of our tasks are governed by laws and regulations, and these assessments could maybe be done by machines. Based on regulations and experience, a machine could take a decision based on enough information, if I understood AI correct" (S6).*

*"One system to work in. More interaction with doctors in real-time would help a lot because we spend a lot of time just waiting... Some smaller tasks could be taken away that maybe a machine could do." (S4).*

#### **4.2.2 Worst-case future**

The subjects also expressed what they thought could be the worst possible future with such an AI system. Many of the subjects started off by jokingly saying they did not want to lose their job, but this also seemed like a somewhat genuine concern. There were also concerns of AI "missing" the people that could not take care of themselves, missing the people that could not lift their own case up because of low resources, missing the people that are not able to use technology, and generally missing the people that needs the caseworkers help the most. There were also concerns for not digitizing and keeping up with the rest of society with digitization. They did not want to be stuck with old systems that were slow and difficult to support, which also goes to show how important digitization is for public services.

*"I do not want to lose my job. No, I'm just kidding" (S16).*

*"We could miss the people that are not able to take care of themselves or lift their own case up because they have low resources. So they can be missed and not get the help they need from us" (S3).*

*"...there are big human consequences if we do not automate and keep up" (S7).*

*"Some people are also not able to use technology at all, they could even be in coma. There are all types of situations, so they are to some degree dependent on others to help them. We meet everyone from the ones who are really sick in hospital on their last days of their lives, to the ones who have an argument in their workplace and cannot get along with their coworkers or leaders.. Everything" (S4).*

*"It is completely ridiculous that we are operating with a system that is almost 50 years old, so the worst that I think could happen is that the system stops. We have to continue operating with even older systems that are even more difficult to support and update because people do not know them anymore, and the code does not exist anymore. ...Society is developing and we have to try to keep up, and we do not do that today" (S7).*

### **4.2.3 How their preferred future could be achieved**

We asked the subjects how they thought they could achieve their preferred futures. Some were not sure how this could be done, while some had thoughts on how they might achieve their preferred future. It was mentioned how AI could be implemented in the process of a potential reorganization.

*"One of our chiefs says that they want a review of how NAV is organized... It is clear that in that process where it is possible to reorganize the whole way of working, I think there is a huge potential for AI" (S13).*

It was also expressed that if AI were to help them with dialogue meetings, they needed to get information from the doctor in a different way than they do today. As every person is different and could be on a completely different scale of a diagnosis, they expressed the need for more parameters than just a standard diagnosis from the doctor. As mentioned, people are different, and human discretion comes into the picture. Making sure that AI can evaluate the people in a similar way as the caseworkers were expressed as important for the implementation of AI systems to be a viable option.

*"It is just that there is so much unforeseen in peoples lives. You and I can have the same diagnosis. You can be in full work and go work out every other day, but I am lying in bed trembling, right? We have one and the same diagnosis, and then the diagnosis system must work differently than it does today, and that could be an opportunity" (S13).*

*"I think there will always be a moment of uncertainty, because humans are not square, we live and things change. But AI could be a piece of guidance to be able to evaluate*

*the current path forward, or dialogue meetings, or whatever, could be a good idea. But then I also think that it requires that good parameters are included in the decision, not just the diagnosis there is today" (S14).*

*"Today the doctor just puts in a diagnosis when you get sick. If on the other hand the doctor had written how well you work on a scale from one to ten, right? Both physical and mental health and such, then maybe a machine could have helped us. ...I think you have to say a lot more about the users functions than just a diagnosis, as we get today" (S13).*

*"There are many parameters that play into it... If you get cancer, for example... for someone over 60 years, this can be very demanding, because the body does not have so much resources to get back to the previous functional level. But if you are 25 and get cancer, then maybe the situation is completely different, and it takes maybe two years, and then you are back in full work" (S13).*

*"...That is why I say that information must come from the doctor, right? Because the doctor will see, hear, smell, and feel this person in front of him, as we do when they come in here" (S13).*

The caseworkers mentioned they have information from previous cases on each person, which AI could potentially use to automate processes regarding sick leave cases in the future. A way for AI to read and make an assessment based on information the doctor writes was mentioned as a possibility to simplify some of their daily tasks. One caseworker wanted robots (artificial intelligence) to be implemented in as many areas as possible.

*"We have access to all the previous history from the same person, and use that in some cases" (S1).*

*"It can be a potential that the doctor write something that a machine can read and understand, for example if you should try to work, or that you still can not work. So the machine should not give out information, but it should receive information and make an assessment based on that information" (S13).*

*"I would very much like to have a robot that automatically does simple things, for example, paying out sick pay... The robot should be put in as much as possible" (S9).*

### **4.3 Challenges of AI in public services**

When talking about AI and automating tasks, some of the subjects expressed their thoughts on different challenges they could face when implementing or using AI applications for public services. They also mentioned some of their current issues that could potentially be a challenge while working with AI in the future.

### 4.3.1 Machine bias

Some of the subjects expressed concerns about machine bias, where AI would be biased in its decisions based on data, regardless of individual human emotions and other information. There were also concerns that a future where everything involves AI would have large margins of error.

*"We meet so many different individuals that have their own case, and no case is alike. A machine would treat them alike, and they are never alike. It is always different cases, different histories, and different issues, and some of them are not able to take care of themselves, so they need help. Someone have a doctor that help them, someone have a leader at work that help them, someone does not. So when it all falls apart, it really falls apart" (S3).*

*"There are so many different people who have to answer. If in the future everything happens with AI, then I think there will be large margins of error. There is a big difference between the people who answer and the additional things with a person" (S15).*

### 4.3.2 Loss of human contact

A challenge that was expressed through multiple interviews was that AI systems could potentially remove the human-human contact, which is important for caseworkers to be able to give certain people the help they really need. In some cases, it is important with human-human contact where there is a need for discretion and reading the person based on their emotions.

*"I think that during a period, we need to consider if a face to face meeting should be held. We should not lose the human perspective. You do not read people when you have meetings on Teams, that is completely wrong. We must have the human perspective" (S11).*

*"We can of course just have video meetings, but what happens when they come into our office, and you sit down and talk with them, and you see, do you get eye-contact? Do they smell alcohol? Can they read? There are a lot of things like that about people that a machine would never be able to catch" (S13)*

*"Sometimes, the users really want to talk to a person because they feel that it's easier to have someone on the phone where they ask for the meetings. And we try to keep it digital, but still, they always want that kind of connection" (S1).*

*"We have a lot of digital meetings every day, but when we sit around a table (physically) we often have better meetings. I do not know why, but we have better contact and body language. Maybe the user is sitting alone at home, sometimes it could be difficult, they cry and are very alone, and I do not like that" (S5).*

*"...we are going to be a more digitized society, which is fine, but I do not think we*

*should stop having that human connection. ... I notice that often if I have met a person once in a meeting, they also feel safer chatting digitally later as they know who they are talking to on the other end. So I believe we should keep talking to people" (S5).*

Another challenge that caseworkers found concerning is that AI might not be able to differentiate people based on their feelings, physical or mental state, or other human factors. It was also mentioned that the citizens' situations might not always be clear and that there could be things that are hard to read from the information displayed on a machine.

*"There are a lot of conditions, after what I can understand, that you have the machine fed with. You will not get the hang of the physical and mental health of that person who answers" (S15).*

*"All the areas you work with people, whether it is in the health sector, elderly sector, school or NAV, you have the same problems. It is that people are much more complex than working with woodwork or production, and how do you think AI takes the height of all the human factors?" (S13).*

*"There are so many human things that play into these situations. It is often not clear, there can be many things behind the scenes. Back pain can be because you have issues at home, it is not very A4" (S9).*

*"We are almost a bit there, that we are moving away from people. It is very compartmentalized, ...it is getting too much controlled by money, economy, time frames, templates, etc. It is a danger that not everything comes forward maybe, because there is a lot going on between the lines" (S9).*

### **4.3.3 Multiple systems**

Throughout most of the interviews it was mentioned that NAV has hundreds of systems, and that caseworkers usually interact with five to ten different systems on a daily basis. Dealing with going back and forth between so many systems were explained to be cumbersome, and it was also a huge time sink in their everyday work.

*"We work so cumbersome, as we work in five different systems. So if you have to check a case, you at least go through four of these systems" (S12).*

*"They can simplify and have fewer systems. One system would have been fantastic. Now I think I am in five different systems. When you know the systems, it is fine, but it is a lot of clicking and we should improve that" (S15).*

*"We use a lot of time switching systems and searching for information" (S4).*

*"I think I would have liked logging into one system and collected all the information there, rather than have to work on a case and having to open many different systems and go back and forth all the time" (S12).*

#### 4.3.4 Communication

Some of the people interviewed expressed that there was a need for better communication. Some of the caseworkers said that they do not get enough information from the general practitioners (GP), and that there needs to be more clear communication between employee (citizen), employer, NAV, and the GP.

*"I think NAV is often in a bigger position than they should be... I think that many employers are lacking that communication with the GP.. I think there should be more flow in the dialogue between the employer and the employee, more direct dialogue. But I think using a system for example, as we do with the dialogue now... which makes communication easier and you can clarify things quickly. You can write to them and they can answer in the process, then you get things into place quickly" (S11).*

*"I think it depends a lot on the doctor that gives the sick leave because some just put like a short sentence and it doesn't give a lot of information... the more I know, the more I can prepare and ask questions that touch on subjects without actually asking direct... So the more context we have, the better, but that depends mostly on the doctor I feel" (S1).*

They stated that there is a lack of communication between the employer and the GP, which could be improved upon. There should also be a better flow in the communication between employees and employers in sick leave cases, where the employer should take more initiative in making the communication flow more clearly.

*"We have to be more direct in the dialogue with the person on sick leave. We need more direct dialogue with the GP.. I think that being direct and asking the user the right questions... We kind of have to make them responsible for their own sick leave. We let them kind of be behind the scenes while the GP decides, and the employer does not dare to ask the employee about anything" (S11).*

*"The person on sick leave has to be more aware of their situation and responsibility as a person on sick leave... it is not a priority for the employer to do what they are meant to do for some reason. We have things out on NAV.no for employers, but very few of them go in to read and look there. So when they come here, they have questions for us, and then I have to research and answer them when they could have done it themselves. Because it is available for them... so it is more work for me because I have to send them links" (S10).*

*"Many of the dialogue meetings that are held are bad... I think this is because employers feel that they have to have a meeting. They think it is a meeting that NAV has requested, because NAV is supposed to offer something, but NAV is a facilitator. It is talking about the facilitation, a plan to get the employee back to work, or other work, and things like that... That NAV must be the one to start the meetings is a bit strange I*



*think, when it is the employers responsibility to follow up" (S12).*

Without a standard for the descriptions of each sick leave from the doctors, the caseworkers expressed that they sometimes got mixed signals from the doctor, and had to read between the lines to understand the actual needs of the users.

*"Sometimes you can get a sense that what the GP are not saying is also information for us. Maybe I get the feeling that they did not really want to give another sick leave, but it is a tricky situation for the doctor also. If they started giving the sick leave and it just continues. And maybe they are like giving us a little bit of information and they want us to like go deeper into the case. And maybe, you know, try to stop it" (S1).*

*"I have a case now where there was a conflict at work, and the person has been sick for almost a year because of it. That turned into a real diagnosis like depression. But the doctor writes that the patient should start looking for new job, and at the same time he puts that P76 depression diagnosis. The doctor gives the person sick leave, and tells us that it is time to push the user to let go of his old employer, and that he is healthy enough to do another job. That is when we have to go in and set certain demands" (S1).*

The doctors were seemingly only required to write more information about the sickness at certain intervals. And it did not seem like there were any limitations to how little or how much that information was.

*"Because also on this sick leave notices from the doctor, they are only required to write more information on certain intervals. After eight weeks they are supposed to write more information so we know about the activity demands. But sometimes they write one sentence, and sometimes they write more. It all comes down to the doctors" (S2).*

We asked if they could maybe use a list of standard questions for efficient communication between actors. There were mixed feelings about this, but in general they mentioned that standard questions could work internally between coworkers. One subject in particular mentioned that they actually developed a function in Teams in-house, where they used standard texts to quickly find what they might need for a case, which they said saved them a lot of time. Other subjects mentioned that standard texts would not work for communication between caseworkers and the GP, because the communication between them were so varied and complex. They also expressed that if they were to use standardized questions, it would have to in the form of text, as images or models would not be able to convey the complex tasks that caseworkers do.

*"A way to communicate with the doctors, both to ask and answer questions... It could be all sorts of questions, and I don't think you could make a list of questions" (S4).*

*"Some standard text and answers to the workplace or employees would make things easier" (S3).*

*"You know Wiki in Teams? It is a function in Teams where you can put in standard texts that we use again and again. They are put into one document where you just give them a number, so it is very easy to find back to them because you only see the heading or the title. So I have developed that here in-house so that it saves a lot of time for a few" (S13).*

## **4.4 Summary of findings**

We categorized the findings from our analysis into four main themes that were discussed throughout multiple interviews that we found to be important and related to human-AI collaboration.

The caseworkers pointed out that AI could change how they work, as today, they work with old systems in an "old-fashioned way" towards the user. Many of the caseworkers expressed that they thought AI could help simplify their daily work, and they were positive that digitization through AI could make their work simple and more efficient.

Most of the caseworkers wished they only needed to work in one integrated system, rather than a set of multiple different systems. Some caseworkers would also wish for some of the smaller tasks to be automated with AI, such as going through laws and regulations for different cases or finding available time to have a meeting with different actors in a case.

The caseworkers expressed that there could be big human consequences by not automating and keeping up with society. Automation saves both the users and the caseworkers a lot of time, resulting in efficient support in a short amount of time, which could be crucial for many people. They were also worried that a fully automated system might miss the people that are not able to take care of themselves or the people that might not be able to use technology at all, resulting in the people that need help the most not getting the help they need. The caseworkers also shared their thoughts on how they think their preferred future could be achieved.

Human-in-the-loop was also a theme that was brought up by most of the caseworkers. They expressed the importance of having a human in control of machines acting as coworkers or decision support systems, and to have a human take the final decision if an AI system were to give a recommendation.

Lastly, we found that there were multiple challenges for caseworkers related to collaborating and incorporating AI applications to their work. These challenges included a risk of machine bias, the potential loss of human contact, too many systems to juggle between, and a need for better communication between different actors.

## 5 Discussion

In this chapter, the findings from chapter 4 will be discussed in relation to our research questions. The discussion draws parallels between our findings and the literature. A hypothetical model for human-AI collaboration and recommendations for addressing NAV's challenges are presented.

### 5.1 Model for human-AI collaboration for caseworkers working with sick leave cases in NAV

Many potential areas where AI could be implemented in NAV would need human interaction, to ensure the final results are valid and correct. Inserting human oversight into the AI life cycle would build trust in the system's validity. While AI could optimize performance, human-in-the-loop would be highly important in increasing safety, transparency, and accountability (Benedikt et al., 2020; Middleton et al., 2022). If an AI system would come with conclusions directly to the citizens on sick leave, it could be hard to understand what these conclusions mean. This is also a factor for including humans in the loop, as the caseworker can explain their decision in a natural language understandable for the citizen (Middleton et al., 2022). A major challenge in the world of AI is trust. The caseworkers and the citizens must understand how and what the AI systems do, which could be coupled with admitting AI mistakes, and understanding why these mistakes happened (Middleton et al., 2022).

The findings stated that AI systems could help with simpler functions like moving tasks and would not work well for more complex situations. Caseworkers deal with quite a lot of discretion in their work, where it is important to see the person they are helping and understand their needs and situation. Artificial discretion could offer significant advantages in accuracy, scale, and efficiency. There are also some challenges to artificial discretion, as it can be quite difficult to understand and manage it in a transparent and accountable manner (Bullock, 2019; Young et al., 2019). As for the quality assurance for algorithmic decision-making in case management, it could be challenging to address this risk in the public sector, as there might be limited capacities of government workers (Guenduez et al., 2020), including a policy discourse that is fast-moving and skeptical publics (Ingrams et al., 2021; Wilson & van der Velden, 2022). There is always a challenge regarding discretion and human bias or cognitive limitations. A bias that arises from prejudice against individuals based on their congenital characteristics, religious beliefs, or other affiliations could sometimes prohibit an agent from making the correct decision (Kahneman, 2011).

Research by Di Vaio et al., 2022 shows a growing interest in human-AI collaboration in public sector organizations. Their findings suggest that human-AI collaborative qualities are linked to improving overall decision-making processes and will boost efficiency in the long term for public sector organizations (Di Vaio et al., 2022). McKinsey, 2022a states that "modern machine superpowers seem to be almost the opposite of some of today's most sought-after human qualities:

creativity, empathy, critical thinking, and emotional intelligence. Companies that design and plan for machine and human qualities to become complementary, rather than oppositional, will have the most effective teams" (McKinsey, 2022a). Building on Woods, 1986 joint cognitive systems theory, a human-AI collaboration could be established in NAV, fostering these human qualities. Figure 5.1 shows a hypothetical model of how a human-AI collaboration could be established through a joint human-AI cognitive system designed for complementary machine and human qualities.

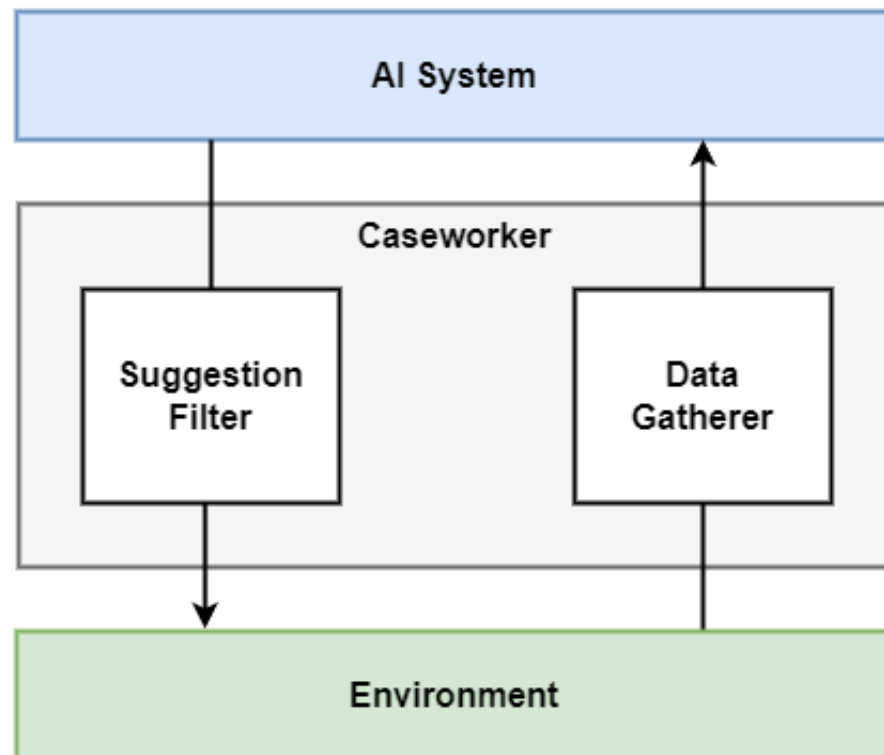


Figure 5.1: Hypothetical human-AI collaboration in the context of NAV based on Joint Cognitive Systems (Woods, 1986).

The "environment" in this context is the citizen, the doctor, or the employer. An example could be when a meeting is initiated between a caseworker and a citizen. Rather than fully automating, there is a human between the AI system and the environment, so the citizen interacts with the caseworker rather than the AI system directly. The AI system performs some of the data gathering related to citizens (sick leave data and other related data), and the case worker can add additional data from dialogue meetings and verify data. The AI system analyzes the data based on algorithms and parameters and gives results in the form of suggestions to aid the caseworker in reaching good decisions. The results should be understandable for humans through transparency and XAI methods, so the caseworker can understand what the results mean and how the AI developed the resulting suggestions. Based on the caseworkers' human qualities and professional experience, combined with their understanding of the AI systems' suggestions, the caseworker can filter through different suggestions given by the AI system and accept, deny, or override the suggestions given. The caseworker uses the best possible suggestion to aid them in reaching a decision and give the citizen the appropriate service based on their discretion. The caseworker should also be able to relay the resulting information to the citizen in a clear and comprehensible way, if needed, to maintain trust and accountability.

### 5.1.1 Current processes

The dialogue meetings for sick leave were identified as sometimes being unnecessary, resulting in work and time spent on unnecessary tasks that could have been omitted. These dialogue meetings are a part of NAVs policy, and it might also be hard to know when they are necessary. It seems important for the caseworkers to have time to focus on people with low resources, and those who need the dialogue meetings. With access to the necessary data and the correct parameters, artificial intelligence could help find the best time to hold the dialogue meeting or recommend whether to hold the meeting or not. The caseworkers send meeting notices to the parties involved, which is a process that could be assisted with AI. The meeting notices could be generated by AI, where the caseworkers decide whether to send the notice or not. Finding a good balance between AI and human interaction is important for this to be viable, especially when users of these tools and systems are in vulnerable situations. Vulnerable members of society may be among the last to benefit from AI (Ventures, 2019). In addition to artificial intelligence assisting the dialogue meeting summoning process, findings suggest a tool that collects information from the person on sick leave and recommends what the person should do going forward. With such a tool, caseworkers could work more efficiently and help more people than before.

One of the caseworkers' daily tasks is to retrieve information for doctor reports. This process is often tedious, and they have to evaluate the information to consider what to do with it. The way the doctors submit information is not specified in the findings, but it was stated that it was not any templates in place for this. The findings suggest an AI that could retrieve this information from the doctors, and with some pre-established parameters, evaluate and recommend what to do with this information. If an AI system like this should be implemented, there must be certain policies for what the doctor can and has to submit. Even though artificial intelligence might not need the specific parameters to give a recommendation, it will allow NAV to know what information the AI has considered. A common denominator for these findings is that caseworkers have tasks that artificial intelligence could assist with and make more efficient. If the AI systems are to make recommendations, they could be based on patterns unknown to the decision-makers. Therefore the caseworkers will have to judge and justify if they do not follow the AI recommendations, but not if they agree with it (Gualdi & Cordella, 2021).

Some of the systems NAV uses are old, which forces caseworkers to work "old-fashioned" towards the citizens. Extracting information from these systems, and collecting everything in one place using artificial intelligence could streamline the workload and possibly allow for more human connection with citizens that need it more. There is a discrepancy between the systems' language and compatibility, which is why a merge of systems has not been effectuated yet (Regjeringen, 2015). It was mentioned that the caseworkers were working in multiple systems at once. These systems consist of legacy systems, such as Infotrygd, which has been running since the 70s (Ringnes, 2018). There are issues with phasing this system out, as there are crucial sensitive data in this system, and there are issues in transitioning this data to a newer type of data that would be compatible with newer systems (Regjeringen, 2015). Another factor is that individual employees in

NAV need different access to sensitive information in the organization. This has been addressed by giving employees access to only the systems or information they need to do their respective job, splitting the systems related to what access an employee has. If a caseworker does not have access to a system or information important for a case, they might have to ask a coworker for this information. This could be solved by implementing a module-based platform architecture. Such a platform could create a work environment for each employee, implementing only the modules needed for their respective roles or accesses. This could create a feeling of working in one system for each employee, and they would not have to spend time going back and forth between systems. They could also implement an AI to this, doing smaller tasks such as finding laws and regulations, resolutions, and finding time for meetings. This could save caseworkers a lot of time, which they stated that they would spend on better quality dialogue meetings with the citizens that need their help the most.

Caseworkers also wanted a better way of communicating with different actors involved with a case, mainly employers, citizens, general practitioners (GP), or other coworkers in NAV. In one interview, it was mentioned that a caseworker developed a function in-house for standard texts in MS Teams called Wiki for their coworkers to use. They stated that this function helped them a lot when searching for keywords they repeatedly use in cases, rather than asking a coworker about a specific subject. A standard text function like this could save caseworkers a lot of time, and it could be improved upon by utilizing AI. Using AI solutions such as chatbots, which NAV already have, could efficiently give them accurate standard text results. This could save the caseworker time when communicating with coworkers, citizens, or employers. However, in another interview, it was also mentioned that standard text solutions would not work as well with communication with the GP. The information from the GP was often too detailed or complex to use standard texts. There are also citizens with special needs, and the information needs to be easier to understand (Sosialdepartementet, 2014). For this reason, there needs to be an overall simplified communication process between actors in sick leave cases that is easy to understand for all parts. It was also mentioned in the interviews that caseworkers could see AI as a valuable asset in sorting out the people in need of their help the most. Cases of citizens that go through NAV to apply for jobs are not as critical as citizens who need social or sickness benefits. An AI could help caseworkers find these critical cases faster, and the caseworkers could provide them with the help they need quicker.

### **5.1.2 Digitization and new work practices**

It was expressed that the caseworkers were worried they could lose their job due to digitization in the future. Organizations are more efficient if humans collaborate with AI, rather than if AI replaces humans (McKinsey, 2022a). Their role as human advisors seems far too important and complex to be fully replaced by a machine. Humans are complex, and human discretion is hard for a machine to comprehend (Young et al., 2019). Hence, in their role as advisors, there must be a human in the loop for human discretion in the final quality assurance. Rather than replacing caseworkers, AI should become a complementary quality, in a human-AI collaborative endeavor to enhance their decision-making capabilities and efficiency. Having a human in the loop would

reassure the caseworkers that it is possible for them to explain particular circumstances, or even contest algorithmic suggestions if needed (Schmager et al., 2023).

It was expressed throughout some interviews that one of the worst things that could happen in the future was for NAV to stop digitizing. The demand from society keeps increasing every year for caseworkers in public services, especially after the Covid-19 pandemic (NAV, 2023c), and they expressed the importance of keeping up with digitization to meet these demands. Today they work with a system from the 70s, and they do not want this to worsen in the future. However, the caseworkers stated that the current digitization efforts in NAV are improving and that NAV is going in the right direction with digitization. NAV is also focusing more on AI innovations (Jensen & Lyngstad, 2019), which is shown to be one of the most important IT innovations in the public sector (Benbunan-Fich et al., 2020). In Norway, NAV seems to be one of the most important organizations in digitization efforts. NAV seems to be heading in the right direction with its digitization efforts, collaborating with researchers for joint innovation efforts (NAV, 2023b; Regjeringen, 2015).

The findings from the data collection suggest that regulation and cooperation with employers of people on sick leave need to change. It is not clearly stated in what way it needs to change, but to achieve change in regulation and cooperation, there is a need for disruption within the organization. The involvement of public sector practitioners is a key factor in achieving this disruption from within (Vassilakopoulou et al., 2023). During one of the interviews, it was brought up that a chief at NAV wants a review of how NAV is organized. Not only is this an opportunity to assess their internal processes and their way of working, but it is also an opportunity to assess their current AI solutions, their current AI capabilities, and the potential for AI in NAV. Vassilakopoulou et al., 2023 mentions that public sector organizations would benefit from a process that rewards disruption from within due to public service organizations having a moderate track record of IS innovation results. Velsberg et al., 2020 suggests that the realization of smartness in the public sector requires a willingness to explore and adopt new work practices instead of simply implementing emerging technologies. To achieve disruption from within, involving public sector practitioners and researchers in joint innovation efforts is key (Vassilakopoulou et al., 2023). A joint innovation effort could encourage researchers to find new areas, work practices, or frameworks to responsibly incorporate AI into public sector organizations. Akbari et al., 2023 developed a framework for Responsible AI principles (RAI) based on ten interrelated principles for responsible AI. If a public service organization were to reorganize, they could use such a framework as a roadmap towards increasing their maturity for developing, deploying, and managing AI responsibly in the organization (Akbari et al., 2023).

### **5.1.3 Bias and fairness**

The individual citizens that use NAV's services are all different from each other, and no case is alike. They all have different histories, issues, and needs. The findings show that some citizens might not, in many ways, be able to take care of themselves. Some might have a doctor or a leader at their workplace to help them, and some might not have anyone to help them. Bias can

be introduced through artifacts from past human biases embedded in data that were required to "train" the system (Bellamy et al., 2018; Veale et al., 2018). Even though artificial discretion could improve tasks and eliminate certain biases present in human agents, there are still challenges with artificial discretion (Young et al., 2019). The risk of bias is present in humans and machines (Kahneman, 2011), and can occur if models and algorithms are oversimplified and rely on sparse, incomplete, or inaccurate datasets. A good example is from Buolamwini and Gebru, 2018, where facial recognition programs trained on images of mostly white people are biased against those with other skin tones. Algorithms have well-documented biases that could be damaging to minority groups and poor populations, which exacerbates the equity concerns of the government (Angwin et al., 2016; Bellamy et al., 2018; Garcia, 2017; O'Reilly-Shah et al., 2020; Park et al., 2019; Veale et al., 2018; Young et al., 2019).

Artificial intelligence can improve administrative decision-making by being more accurate, less corruptible, and more consistent in ways that can reduce bias if trained and used properly (Young et al., 2019). To counter the risk of machine bias and unfair decisions, NAV must ensure fairness and inclusiveness in their AI systems and human interaction (Feuerriegel et al., 2020). As machine learning models predict exactly what they have been trained on, it could be recommended that biases are considered already when selecting the models' training data (DeBrusk, 2018). Rooting out data that are not directly relevant to the model calculations could prevent bias from the early stages, and give the organization a solid model for their predictions (DeBrusk, 2018; Hellström et al., 2020).

## **5.2 Limitations**

The information gathered for this thesis is from caseworkers working in the area of sick leave at NAV. It could be argued that views and perspectives from citizens would be important to investigate too. We did not include citizens as our stakeholders. We acknowledge the need for a citizen-centric view of public services, and the need to include citizens' and users' perspectives in research of public services. In our study, we focused on the challenges and needs of caseworkers relating to AI and human-AI collaboration. Therefore we had to focus on the stakeholders that use and regulate artificial intelligence in the public sector, and every participant we interviewed was in some way a part of the sick leave department in NAV. There could be some bias over the results regarding their similar needs and thoughts, mostly structured around sick leave.

The context of sick leave itself could be seen as a limitation. There might be sensitive information that the caseworkers are unable to share, which could influence the results of the study. As there is a right of privacy for citizens using NAVs services, our data generation method could be seen as a limitation because observations would not be easily carried out. The context of the case of sick leave case handling is also quite specific, where there are requirements in that area that might not be applicable in other areas of public services.

Another limitation concerns the generalizability of the findings. We focused on public services,



specifically the Norwegian Labour and Welfare Administration because AI is being rapidly adopted into the public sector. However, our findings might not be reflective of other sectors, or even other areas of the public sector. Future research could adopt the approach of this study to investigate and compare human-AI collaboration in other public sector areas.

From the findings, we learned that many decisions can not be made without management's involvement. It could be interesting to delve deeper into the topic of this thesis, by including research from a managerial perspective. It would provide a more complex picture of strategy development in public services. Several top-level organizations see that artificial intelligence will impact the labour market in the time ahead (Norwegian Ministry of Local Government and Modernisation, 2020), so it would be interesting to look deeper into their strategy for operating and administrating AI in an accountable and trustworthy way.

## 6 Conclusions

In this study, we have investigated the research question "*what are caseworkers' needs for human-AI collaboration in public services*". To guide us towards answering the research question, we made sub-questions related to the caseworkers' expectations for the future working with AI, and how we can contribute to facilitating meeting their needs. After interviewing caseworkers working with sick leave case handling at NAV, we analyzed and discussed their needs, expectations, and challenges related to human-AI collaboration. The data from the analysis has shown several caseworkers' needs relating to AI and their experience and understanding of AI. In the discussion, we have considered the various needs for AI, and how to facilitate these needs in one overall context.

The study identified six key factors related to responsible human-AI collaboration for public service practitioners: Human-in-the-loop, bias and fairness, transparent AI, explainable AI, accountable AI, and exploring new work practices. The caseworkers' needs for AI are quite similar to each other. The areas where AI could assist caseworkers seem mostly related to internal processes. Processes that are in direct contact with users of NAV's services concern several challenges. We identified machine bias and loss of human contact as major challenges related to collaborating and incorporating AI applications in the caseworkers' processes. There were concerns amongst the caseworkers that AI could be biased in its decision based on the data available. To prevent machine bias in these processes, biases have to be considered when training the algorithms or models. Fairness and inclusiveness in AI systems and human interaction are a factor for preventing machine bias and unfair decisions, and the data available for the machine need to be carefully validated.

Fully automated systems are not necessarily the best approach, as having a human in the loop would allow more trust in the system, knowing that a human makes the final decision. Because human discretion is hard for a machine to comprehend, caseworkers should make the final decisions, to ensure the quality and fairness of the decisions made. As with systems that connect directly with citizens, having the caseworkers as a link between the AI and the citizens would allow them to explain their decisions in a natural language that the citizens can understand. We identified that the caseworkers are positive to digitization and think AI could help simplify and streamline their daily work. By collaborating with AI, caseworkers can do complex tasks where human discretion is necessary, leaving the smaller and simpler tasks to be done by AI. This would increase the overall efficiency of sick leave case handling processes.

### 6.1 Implications

The findings of this study highlight the importance of balancing between speedy public service (which can be facilitated through automation), and accountability (which requires control mechanisms and humans in the loop). This balance has long been a challenge in public services (Gayialis et al., 2016), and with the advancements in AI, it has become even more crucial. Further research

is required to delve deeper into this topic. Public organizations need to take a comprehensive view on AI accountability being able to answer for and justify actions related to AI, ensuring the ability of stakeholders to interrogate about AI and sanctioning when AI systems work in unacceptable ways (Kempton et al., 2023). Further research is needed on how this can be achieved in practice. Overall, the responsible development and deployment of AI is an area of research that requires significant further development, and Information Systems researchers, with their sociotechnical understanding, are well-suited to contribute to this crucial endeavor (Vassilakopoulou et al., 2022). As this study focused on the case of sick leave case handling, an area of the public sector that has a direct connection to the citizens, further research is needed on those areas of the public sector that has a different context. There might be different needs for human-AI collaboration for those who work in other areas, even if they work in other departments or organizations. Further research could extend across countries and different government legislations, as cultural identity and overall trust in government might influence humans' stance on AI (Schmager et al., 2023). This is an exciting research opportunity to see how human-AI collaboration differs cross-culturally.

This study identified a few implications for practice in NAV. By reviewing NAV's internal processes, NAV can explore and adopt new work practices for caseworkers that facilitate incorporating AI into their work practices. The study shows that "soft" AI suits smaller tasks, such as looking up laws and regulations. For complex tasks that caseworkers do daily, such as reviewing sick leave cases, it is crucial to establish a responsible use of AI applications through transparency and explainability. A human-centered approach to human-AI collaboration, designed for AI and human qualities to become complementary, could improve caseworkers' decision-making processes and efficiency and leave the human in control.

This study also identified implications for the social aspects of sustainability. "Social sustainability refers to equality, well-being, and balance across quality of life indicators between sociocultural groups over time and from one generation to the next" (Ross, 2013). Implementing human-AI collaboration into public service deliveries could help the Norwegian government in reaching some of the United Nations sustainable development goals (SDGs) in the future. Public sector organizations such as NAV could streamline and simplify many work processes through the implementation of responsible human-AI collaboration. Doing so could result in meeting some of the UN's SDGs, such as reduced inequalities by reducing human bias; decent work and economic growth by helping more people get back to work faster or facilitating other work; climate action by reducing carbon dioxide emissions with more effective solutions; innovation by implementing and innovating more AI technology into public service deliveries; and good health and well-being by helping more citizens and guiding them getting the help they need (UN, 2020).

# References

- Adadi, A., & Berrada, M. (2020). Explainable AI for Healthcare: From Black Box to Interpretable Models. *Springer eBooks*, 327–337. <https://doi.org/10.1007/978-981-15-0947-631>
- AI4users. (2020, December 14). *Project aim - ai4users*. <https://ai4users.uia.no/research/>
- Akbari, P., Pappas, I., & Vassilakopoulou, P. (2023). Justice as fairness: A hierarchical framework of responsible ai principles.
- Allen, J. E., Guinn, C. I., & Horvitz, E. (1999). Mixed-initiative interaction. *IEEE Intelligent Systems and their Applications*, 14(5), 14–23.
- Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through ai-guided chatbots. *Government Information Quarterly*, 36(2), 358–367. <https://doi.org/https://doi.org/10.1016/j.giq.2018.10.001>
- Angwin, J., Larson, J., Mattu, S., Kirchner, L., & ProPublica. (2016). There’s software used across the country to predict future criminals. and it’s biased against blacks. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Aoki, N. (2020). An experimental study of public trust in ai chatbots in the public sector. *Government Information Quarterly*, 37(4), 101490. <https://doi.org/https://doi.org/10.1016/j.giq.2020.101490>
- Arrieta, A., Díaz-Rodríguez, N., Kasabov, N., Benetot, A., Tabik, S., Barbado, A., García, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2019). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Babar, Z., & Yu, E. (2019). Digital transformation – implications for enterprise modeling and analysis. *2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW)*, 1–8. <https://doi.org/10.1109/EDOCW.2019.00015>
- Balakrishnan, J., & Dwivedi, Y. K. (2021). Role of cognitive absorption in building user trust and experience. *Psychology & Marketing*, 38(4), 643–668. <https://doi.org/https://doi.org/10.1002/mar.21462>
- Barrett, D. A., & Twycross, A. (2018). Data collection in qualitative research. *Evidence-Based Nursing*, 21(3), 63–64. <https://doi.org/10.1136/eb-2018-102939>
- Bataller, C., & Harris, J. (2016). Turning artificial intelligence into business value. <https://files.stample.co/browserUpload/fc3be0d1-906c-4db4-b572-649edf4c73ac>
- Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2018). Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.1810.01943>

- Benam, B. (2022). How Much Transparency for AI Algorithms - Behzad Benam - Medium. <https://medium.com/@bhbenam/how-much-transparency-for-ai-algorithms-c44837efc576>
- Benbunan-Fich, R., Desouza, K. C., & Andersen, K. V. (2020). IT-enabled innovation in the public sector: introduction to the special issue. *European Journal of Information Systems*, 29(4), 323–328. <https://doi.org/10.1080/0960085x.2020.1814989>
- Benedikt, L., Joshi, C., Nolan, L., Henstra-Hill, R., Shaw, L., & Hook, S. (2020). Human-in-the-loop ai in government: A case study. *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 488–497.
- Biran, O., & Cotton, C. (2017). Explanation and justification in machine learning: A survey. *IJCAI-17 workshop on explainable AI (XAI)*, 8(1), 8–13.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bryson, J., & Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. *Computer*, 50, 116–119. <https://doi.org/10.1109/MC.2017.154>
- Bullock, J. B. (2019). Artificial intelligence, discretion, and bureaucracy. *The American Review of Public Administration*, 49(7), 751–761.
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on Fairness, Accountability, and Transparency*, 77–91. <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>
- Busuioc, M. (2021). Accountable artificial intelligence: Holding algorithms to account. *Public Administration Review*, 81(5), 825–836. <https://doi.org/10.1111/puar.13293>
- Castleberry, A., & Nolen, A. (2018). Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in Pharmacy Teaching and Learning*, 10(6), 807–815. <https://doi.org/https://doi.org/10.1016/j.cptl.2018.03.019>
- Cheng, H.-F., Wang, R., Zhang, Z., O’Connell, F., Gray, T., Harper, F. M., & Zhu, H. (2019). Explaining decision-making algorithms through ui: Strategies to help non-expert stakeholders. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3290605.3300789>
- Chopard, K., & Przybylski, R. (2021). Case studies.
- Creswell, J. W., & Creswell, J. D. (2017, November 27). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications.
- Datatilsynet. (2022). *Nav - sluttrapport*. <https://www.datatilsynet.no/regelverk-og-verktoy/sandkasse-for-kunstig-intelligens/ferdige-prosjekter-og-rapporter/nav-sluttrapport/>
- De Fine Licht, K., & De Fine Licht, J. (2020). Artificial intelligence, transparency, and public decision-making. *AI & society*, 35(4), 917–926. <https://doi.org/10.1007/s00146-020-00960-w>
- DeBrusk, C. (2018). The risk of machine-learning bias (and how to prevent it). *MIT Sloan Management Review*.
- Dellermann, D., Calma, A., Lipusch, N., Weber, T., Weigel, S., & Ebel, P. (2019). The future of human-ai collaboration: A taxonomy of design knowledge for hybrid intelligence systems. <https://doi.org/10.24251/hicss.2019.034>

- Di Vaio, A., Hassan, R., & Alavoine, C. (2022). Data intelligence and analytics: A bibliometric analysis of human–Artificial intelligence in public sector decision-making effectiveness. *Technological Forecasting and Social Change*, 174, 121201. <https://doi.org/10.1016/j.techfore.2021.121201>
- Eggers, D., Fishman, T., & Kishnani, P. (2017). *Ai-augmented human services: Using cognitive technologies to transform program delivery*. Retrieved May 2, 2023, from [https://www2.deloitte.com/content/dam/insights/us/articles/4152\\_AI-human-services/4152\\_AI-human-services.pdf](https://www2.deloitte.com/content/dam/insights/us/articles/4152_AI-human-services/4152_AI-human-services.pdf)
- Ehsan, U., Liao, Q. V., Muller, M., Riedl, M. O., & Weisz, J. D. Expanding explainability: Towards social transparency in ai systems. In: 2021. <https://doi.org/10.1145/3411764.3445188>.
- Eisenhart, M. (1991). Conceptual frameworks for research circa 1991: Ideas from a cultural anthropologist; implications for mathematics education rese.
- Eriksson, C., & Olsen, K. (2022). A Case-Study on the Impact of Transparency in AI/Human Interaction for Case-Workers at the Norwegian Labour and Welfare Administration [Unpublished paper]. *Information Systems, University of Agder*.
- Eschenbach, V., & Warren, J. D. (2021). Transparency and the black box problem: Why we do not trust ai. *Philosophy & Technology*, 34(4), 1607–1622. <https://doi.org/10.1007/s13347-021-00477-0>
- Fails, J. A., & Olsen Jr, D. R. (2003). Interactive machine learning. *Proceedings of the 8th international conference on Intelligent user interfaces*, 39–45.
- Feuerriegel, S., Dolata, M., & Schwabe, G. (2020). Fair ai: Challenges and opportunities. *Business & Information Systems Engineering*, 62. <https://doi.org/10.1007/s12599-020-00650-3>
- Fineide, M. J., Hansen, G. V., & Haug, E. (2019). How does a new working method in the norwegian labour and welfare organization (nav) succeed in reducing sick leave rates? *International Journal of Integrated Care*, 19(4), 169. <https://doi.org/10.5334/ijic.s3169>
- Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, 14(3), 330–347. <https://doi.org/10.1145/230538.230561>
- Garcia, M. (2017). Racist in the machine: The disturbing implications of algorithmic bias. retrieved december 03, 2017.
- Gayialis, S., Papadopoulou, G., Ponis, S., Vassilakopoulou, P., & Tatsiopoulou, I. (2016). Integrating process modeling and simulation with benchmarking using a business process management system for local government. *International Journal of Computer Theory and Engineering*, 8(6), 482.
- GDPR. (2019). Recital 58 - The Principle of Transparency - General Data Protection Regulation (GDPR). <https://gdpr-info.eu/recitals/no-58/>
- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M. A., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. <https://doi.org/10.1109/dsaa.2018.00018>
- Green, G., & Johns, T. (2019). Exploring the relationship (and power dynamic) between researchers and public partners working together in applied health research teams. *Frontiers in Sociology*, 4. <https://doi.org/10.3389/fsoc.2019.00020>

- Gualdi, F., & Cordella, A. (2021). Artificial intelligence and decision-making: The question of accountability.
- Guenduez, A. A., Mettler, T., & Schedler, K. (2020). Technological frames in public administration: What do public managers think of big data? *Government Information Quarterly*, 37(1), 101406. <https://doi.org/https://doi.org/10.1016/j.giq.2019.101406>
- Guest, G., & McLellan, E. (2003). Distinguishing the trees from the forest: Applying cluster analysis to thematic qualitative data. *Field Methods*, 15(2), 186–201. <https://doi.org/10.1177/1525822x03015002005>
- Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, 61(4), 5–14.
- Halvorsen, K. (2008, January 1). *Å forske på samfunnet: En innføring i samfunnsvitenskapelig metode*.
- Haresamudram, K., Larsson, S., & Heintz, F. (2023). Three levels of ai transparency. *Computer*, 56(2), 93–100. <https://doi.org/10.1109/MC.2022.3213181>
- Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y. J., Van Den Hoven, J., Zicari, R. V., & Zwitter, A. (2017, February 25). *Will democracy survive big data and artificial intelligence?* Springer Nature. [https://doi.org/10.1007/978-3-319-90869-4\\_7](https://doi.org/10.1007/978-3-319-90869-4_7)
- Hellström, T., Dignum, V., & Bensch, S. (2020). Bias in machine learning—what is it good for? *arXiv preprint arXiv:2004.00686*.
- Horvitz, E. (1999). Principles of mixed-initiative user interfaces. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 159–166.
- IJtsma, M., Ma, L. M., Pritchett, A. R., & Feigh, K. M. (2019). Computational methodology for the allocation of work and interaction in human-robot teams. *Journal of Cognitive Engineering and Decision Making*, 13(4), 221–241. <https://doi.org/10.1177/1555343419869484>
- Ingrams, A., Kaufmann, W., & Jacobs, D. (2021). In ai we trust? citizen perceptions of ai in government decision making. *Policy & Internet*, 14(2), 390–409. <https://doi.org/10.1002/poi3.276>
- Israel, M., & Hay, I. (2006, January 1). *Research ethics for social scientists*. <https://doi.org/10.4135/9781849209779>
- Jacobsen, D. I. (2015, January 1). *Hvordan gjennomføre undersøkelser? Innføring i samfunnsvitenskapelig metode*.
- Jensen, M. V., & Lyngstad, C. P. (2019). Innspill til strategi for kunstig intelligens. [https://www.regjeringen.no/contentassets/0e36c85fcfe143a5b626c53cf292cb3b/strategi-kunstig-intelligens\\_innspill-fra-nav.pdf](https://www.regjeringen.no/contentassets/0e36c85fcfe143a5b626c53cf292cb3b/strategi-kunstig-intelligens_innspill-fra-nav.pdf)
- Kahneman, D. (2011, January 1). *Thinking, fast and slow*. <http://ci.nii.ac.jp/ncid/BB2184891X>
- Kempton, A. M., Parmiggiani, E., & Vassilakopoulou, P. (2023). Accountability in managing artificial intelligence: State of the art and a way forward for information systems research.
- Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. 2.
- Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2019). Discrimination in the Age of Algorithms. *Journal of Legal Analysis*, 10, 113–174. <https://doi.org/10.1093/jla/laz001>

- Lai, Y. (2021, January 5). *Human-ai collaboration in healthcare: A review and research agenda*. <https://scholarspace.manoa.hawaii.edu/items/48162f4c-8c99-47dc-8fb5-fe74e9cbb51d>
- Lakshmi, S., & Mohideen, M. A. (2013). Issues in reliability and validity of research. *International journal of management research and reviews*, 3(4), 2752.
- Larsson, S. (2019). The socio-legal relevance of artificial intelligence. *Droit et société*, N°103(3), 573. <https://doi.org/10.3917/drs1.103.0573>
- Larsson, S., & Heintz, F. (2020). Transparency in artificial intelligence. *Internet policy review*, 9(2). <https://doi.org/10.14763/2020.2.1469>
- Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, Transparent, and Accountable Algorithmic Decision-making Processes. *DSpace@MIT (Massachusetts Institute of Technology)*, 31(4), 611–627. <https://doi.org/10.1007/s13347-017-0279-x>
- Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of The ACM*, 61(10), 36–43. <https://doi.org/10.1145/3233231>
- Loonam, J., Eaves, S., Kumar, V., & Parry, G. (2017). Towards digital transformation: Lessons learned from traditional organisations. *Strategic Change*, 27. <https://doi.org/10.1002/jsc.2185>
- Lynn, T., Rosati, P., Conway, E., Curran, D., Fox, G., & O’Gorman, C. S. (2022, January 1). *Digital public services*. [https://doi.org/10.1007/978-3-030-91247-5\\_3](https://doi.org/10.1007/978-3-030-91247-5_3)
- Marathe, A. R., Schaefer, K. E., Evans, A. W., & Metcalfe, J. S. (2018). Bidirectional communication for effective human-agent teaming. In J. Y. Chen & G. Fragomeni (Eds.), *Virtual, augmented and mixed reality: Interaction, navigation, visualization, embodiment, and simulation* (pp. 338–350). Springer International Publishing.
- Mattu, S., Larson, J., Angwin, J., & Kirchner, L. (2020). How We Analyzed the COMPAS Recidivism Algorithm. <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>
- McCarthy, J. J., Minsky, M., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *Ai Magazine*, 27(4), 12. <https://doi.org/10.1609/aimag.v27i4.1904>
- McKinsey. (2022a). Forging the human–machine alliance. <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/forging-the-human-machine-alliance>
- McKinsey. (2022b). The state of AI in 2022 - and a half decade in review. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review>
- Mehr, H. (2017). Artificial intelligence for citizen services and government. *Cambridge, MA: Harvard Kennedy School, Ash Center for Democratic Governance And Innovation*. [https://ash.harvard.edu/files/ash/files/artificial\\_intelligence\\_for\\_citizen\\_services.pdf](https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf)
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6), 1–35. <https://doi.org/10.1145/3457607>



- Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. *Information Systems Management*, 39(1), 53–63. <https://doi.org/10.1080/10580530.2020.1849465>
- Middleton, S. E., Letouzé, E., Hossaini, A., & Chapman, A. (2022). Trust, regulation, and human-in-the-loop ai: Within the european region. *Communications of the ACM*, 65(4), 64–68.
- Mikalef, P., Fjørtoft, S. O., & Torvatn, H. Y. (2019, September 18). *Artificial intelligence in the public sector: A study of challenges and opportunities for norwegian municipalities*. Springer Science+Business Media. [https://doi.org/10.1007/978-3-030-29374-1\\_22](https://doi.org/10.1007/978-3-030-29374-1_22)
- Morakanyane, R., Grace, A., & O'Reilly, P. (2017). Conceptualizing digital transformation in business organizations: A systematic review of literature. <https://doi.org/10.18690/978-961-286-043-1.30>
- Moschella, D. (2022). AI Bias Is Correctable. Human Bias? Not So Much. <https://itif.org/publications/2022/04/25/ai-bias-correctable-human-bias-not-so-much/>
- NAV. (2021). NAV har fått ny FoU-plan - nav.no. <https://www.nav.no/no/nav-og-samfunn/kunnskap/fou-midler/nyheter/nav-har-fatt-ny-fou-plan>
- NAV. (2023a). Hva er NAV? - nav.no. <https://www.nav.no/hva-er-nav>
- NAV. (2023b, April 27). Samarbeidsavtaler med universitet og høyskoler - nav.no. <https://www.nav.no/no/nav-og-samfunn/kunnskap/fou-midler/samarbeid-med-universitet-og-hoyskoler>
- NAV. (2023c). 35 milliarder ekstra fra NAV - nav.no. <https://www.nav.no/no/nav-og-samfunn/kunnskap/analyser-fra-nav/nyheter/35-milliarder-ekstra-fra-nav>
- Neururer, M., Schlögl, S., Brinkschulte, L., & Groth, A. (2018). Perceptions on authenticity in chat bots. *Multimodal Technologies and Interaction*, 2(3). <https://doi.org/10.3390/mti2030060>
- Noor, K. B. M. (2008). Case study: A strategic research methodology. *American Journal of Applied Sciences*, 5(11), 1602–1604. <https://doi.org/10.3844/ajassp.2008.1602.1604>
- Norwegian Ministry of Local Government and Modernisation. (2020). *National strategy for artificial intelligence*. [https://www.regjeringen.no/contentassets/%201febbb2c4fd4b7d92c67dd353b6ae8/en-gb/pdfs/ki-strategi\\_en.pdf](https://www.regjeringen.no/contentassets/%201febbb2c4fd4b7d92c67dd353b6ae8/en-gb/pdfs/ki-strategi_en.pdf)
- Oates, B. J., Griffiths, M., & McLean, R. (2022). *Researching Information Systems and Computing*. Sage Publications Limited.
- Oecd ai principle - oecd.ai*. (2023). Retrieved April 24, 2023, from <https://oecd.ai/en/dashboards/ai-principles/P9>
- Olsen, J. P. (2014, May 1). *Accountability and ambiguity*. <https://doi.org/10.1093/oxfordhb/9780199641253.013.0013>
- O'Reilly-Shah, V. N., Gentry, K. R., Walters, A. M., Zivot, J., Anderson, C. T., & Tighe, P. J. (2020). Bias and ethical considerations in machine learning and the automation of perioperative risk assessment. *British Journal of Anaesthesia*, 125(6), 843–846.
- Osoba, O. A., & Welser, W. (2017). *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. RAND Corporation. <https://doi.org/10.7249/rr1744>

- Ostheimer, J., Chowdhury, S., & Iqbal, S. (2021). An alliance of humans and machines for machine learning: Hybrid intelligent systems and their design principles. *Technology in Society*, 66, 101647.
- Oxford-Dictionary. (2023). *Behaviour noun - definition, pictures, pronunciation and usage notes*. <https://www.oxfordlearnersdictionaries.com/definition/english/behaviour?q=behavior>
- Park, S. Y., Kuo, P.-Y., Barbarin, A., Kaziunas, E., Chow, A., Singh, K., Wilcox, L., & Lasecki, W. S. (2019). Identifying challenges and opportunities in human-ai collaboration in health-care. *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, 506–510. <https://doi.org/10.1145/3311957.3359433>
- Pawar, U., O’Shea, D., Rea, S., & O’Reilly, R. (2020). Explainable AI in Healthcare. *2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA)*. <https://doi.org/10.1109/cybersa49311.2020.9139655>
- Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020). Big data and ai – a transformational shift for government: So, what next for research? *Public Policy and Administration*, 35(1), 24–44. <https://doi.org/10.1177/0952076718780537>
- Punch, K. F. (2013, November 19). *Introduction to social research: Quantitative and qualitative approaches*. SAGE Publications Limited.
- Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and information technology*, 20(1), 5–14.
- Regjeringen. (2015, April). *Et nav med muligheter. bedre brukermøter, større handlingsrom og tettere på arbeidsmarkedet*. [https://www.regjeringen.no/globalassets/departementene/aid/dokumenter/2015/sluttrapport-ekspertgruppen-nav\\_9.4.15.pdf](https://www.regjeringen.no/globalassets/departementene/aid/dokumenter/2015/sluttrapport-ekspertgruppen-nav_9.4.15.pdf)
- Richards, T., Richards, L., et al. (1995). Using hierarchical categories in qualitative data analysis. *Computer-aided qualitative data analysis: Theory, methods, and practice*, 80–95.
- Riege, A. (2003). Validity and reliability tests in case study research: a literature review with “hands-on” applications for each research phase. *Qualitative Market Research: An International Journal*, 6(2), 75–86. <https://doi.org/10.1108/13522750310470055>
- Ringnes, I. F. (2018). *Infotrygd 40 år: Fortsatt lenge igjen til pensjonering*. <https://memu.no/artikler/infotrygd-40-ar-fortsatt-lenge-igjen-til-pensjonering/>
- Roberts, P., & Priest, H. (2006). Reliability and validity in research. *Nursing standard*, 20(44), 41–46.
- Ross, D. (2013). *Social Sustainability*. [https://doi.org/10.1007/978-3-642-28036-8\\_58](https://doi.org/10.1007/978-3-642-28036-8_58)
- Saunders, M. N. K., Thornhill, A., & Lewis, P. (2019, January 1). *Research methods for business students*.
- Schirner, G., Erdogmus, D., Chowdhury, K., & Padir, T. (2013). The future of human-in-the-loop cyber-physical systems. *Computer*, 46(1), 36–45.
- Schmager, S., Vassilakopoulou, P., Grøder, C., Parmiggiani, E., & Pappas, I. (2023). What do citizens think of ai adoption in public services? exploratory research on citizen attitudes through a social contract lens.

- Schramm, W. (1971). *Schramm, w. (1971, december). notes on case studies of instructional mediaprojects. working.* <https://www.scribd.com/document/402140881/Schramm-W-1971-December-Notes-on-Case-Studies-of-Instructional-Mediaprojects-Working>
- Setzu, M., Guidotti, R., Monreale, A., Turini, F., Pedreschi, D., & Giannotti, F. (2021). Glocalx - from local to global explanations of black box ai models. *Artificial Intelligence*, 294. <https://doi.org/10.1016/j.artint.2021.103457>
- Sheridan, T. B. (2012). Human supervisory control. *Handbook of human factors and ergonomics*, 990–1015.
- Shih, P. C. (2018). Beyond human-in-the-loop: Empowering end-users with transparent machine learning. *Human and machine learning: visible, explainable, trustworthy and transparent*, 37–54.
- Silberg, J., & Manyika, J. (2019). Notes from the ai frontier: Tackling bias in ai (and in humans). *McKinsey Global Institute*, 1(6).
- Sosialdepartementet, A. O. (2014). Brukernes møte med NAV. <https://www.regjeringen.no/no/dokumenter/Brukernes-mote-med-NAV/id2008441/>
- Stowers, K., Kasdaglis, N., Rupp, M., Chen, J., Barber, D., & Barnes, M. (2017). Insights into human-agent teaming: Intelligent agent transparency and uncertainty. In P. Savage-Knepshield & J. Chen (Eds.), *Advances in human factors in robots and unmanned systems* (pp. 149–160). Springer International Publishing.
- Tubella, A. A., Theodorou, A., Dignum, F., & Dignum, V. Governance by glass-box: Implementing transparent moral bounds for ai behaviour. In: *2019-August*. 2019, 5787–5793. <https://doi.org/10.24963/ijcai.2019/802>.
- UN. (2020). Take Action for the Sustainable Development Goals - United Nations Sustainable Development. <https://www.un.org/sustainabledevelopment/sustainable-development-goals/>
- Universitetet i Oslo. (2023). *Autotekst*. <https://autotekst.uio.no/>
- van Noordt, C., & Misuraca, G. (2022). Artificial intelligence for the public sector: Results of landscaping the use of ai in government across the european union. *Government Information Quarterly*, 39(3), 101714. <https://doi.org/https://doi.org/10.1016/j.giq.2022.101714>
- Vassilakopoulou, P., & Grisot, M. (2020). Effectual tactics in digital intrapreneurship: A process model. *The Journal of Strategic Information Systems*, 29, 101617. <https://doi.org/10.1016/j.jsis.2020.101617>
- Vassilakopoulou, P., Haug, A., Salvesen, L. M., & Pappas, I. O. (2023). Developing human/ai interactions for chat-based customer services: Lessons learned from the norwegian government. *European Journal of Information Systems*, 32(1), 10–22. <https://doi.org/10.1080/0960085X.2022.2096490>
- Vassilakopoulou, P., Parmiggiani, E., Shollo, A., & Grisot, M. (2022). Responsible ai: Concepts, critical perspectives and an information systems research agenda. *Scandinavian Journal of Information Systems*, 34(2), 3.

- Veale, M., Van Kleek, M., & Binns, R. (2018). Fairness and accountability design needs for algorithmic support in high-stakes public sector decision-making. <https://doi.org/10.1145/3173574.3174014>
- Velsberg, O., Westergren, U. H., & Jonsson, K. (2020). Exploring smartness in public sector innovation - creating smart public services with the Internet of Things. *European Journal of Information Systems*, 29(4), 350–368. <https://doi.org/10.1080/0960085x.2020.1761272>
- Ventures, M. (2019). *Mmc ventures*. <https://www.stateofai2019.com/summary/>
- Vollmer, N. (2022). Recital 58 EU General Data Protection Regulation (EU-GDPR). Privacy/Private according to plan. <https://www.privacy-regulation.eu/en/recital-58-GDPR.htm>
- Wang, D., Churchill, E. F., Maes, P., Fan, X., Shneiderman, B., Shi, Y., & Wang, Q. (2020). From Human-Human Collaboration to Human-AI Collaboration. <https://doi.org/10.1145/3334480.3381069>
- Wang, R., Harper, F. M., & Zhu, H. (2020). Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14.
- Webster, J., & Watson, R. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26. <https://doi.org/10.2307/4132319>
- Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S. M., Richardson, R., Schultz, J., Schwartz, O., et al. (2018). *Ai now report 2018*. AI Now Institute at New York University New York.
- Williams, M., & Moser, T. (2019). The art of coding and thematic exploration in qualitative research. *International Management Review*, 15(1), 45–. <https://www.questia.com/library/journal/1P4-2210886420/the-art-of-coding-and-thematic-exploration-in-qualitative>
- Wilson, C., & van der Velden, M. (2022). Sustainable ai: An integrated model to guide public sector decision-making. *Technology in Society*, 68, 101926. <https://doi.org/10.1016/j.techsoc.2022.101926>
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector — applications and challenges. *International Journal of Public Administration*, 42(7), 596–615. <https://doi.org/10.1080/01900692.2018.1498103>
- Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. <https://doi.org/10.1145/2601248.2601268>
- Woods, D. D. (1986). Cognitive technologies: The design of joint human-machine cognitive systems. *AI Magazine*, 6(4), 86–92.
- Wu, X., Xiao, L., Yixuan, S., Zhang, J., Ma, T., & He, L. (2021). A survey of human-in-the-loop for machine learning.
- Yin, K. (2013). Yin, r. k. (2009). case study research: Design and methods (4th ed.). thousand oaks, ca: Sage. *The Canadian Journal of Action Research*, 14(1), 69–71. <https://doi.org/10.33524/cjar.v14i1.73>
- Yin, R. K. (1994). Discovering the Future of the Case Study. *Method in Evaluation Research*. *Evaluation Practice*, 15(3), 283–290. <https://doi.org/10.1177/109821409401500309>

- Ylikoski, P., & Zahle, J. (2019). Case study research in the social sciences. *Studies in History and Philosophy of Science Part A*, 78, 1–4. <https://doi.org/https://doi.org/10.1016/j.shpsa.2019.10.003>
- Young, M. R., Bullock, J. B., & Lecy, J. D. (2019). Artificial discretion as a tool of governance: A framework for understanding the impact of artificial intelligence on public administration. *Perspectives on Public Management and Governance*. <https://doi.org/10.1093/ppmgov/gvz014>
- Zhu, H., Yu, B., Halfaker, A., & Terveen, L. (2018). Value-sensitive algorithm design: Method, case study, and lessons. *Proceedings of the ACM on human-computer interaction*, 2(CSCW), 1–23.

# Appendices

## A Consent form

### Vil du delta i forskningsprosjektet AI4Users?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å undersøke og designe for ansvarlig bruk av kunstig intelligens (KI). I dette skrevet gir vi deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

#### Formål

AI4Users-prosjektet tar et menneskesentrert perspektiv for utvikling av verktøy og designprinsipper for å hjelpe brukere å stole på KI-løsninger. AI4Users-prosjektet retter seg spesifikt mot ikke-eksperter og utvider dermed rekkevidden av forskningen på ansvarlig bruk av KI utover KI-eksperter og teknologer. Problemstillingene som blir tatt opp i prosjektet er knyttet til hvordan KI brukes til å løse arbeidsoppgaver, hvordan brukere kan forstå hvilken informasjon, evner og avgrensninger KI-løsninger har, og hvordan bruk av KI påvirker ansvarlighet og etterprøvbarehet innad i en organisasjon.

Opplysningene som oppgis til forskningsprosjektet vil kun bli brukt i forskningsøyemed. De anonymiserte funnene fra denne forskningen vil publiseres i designretningslinjer og forskningsartikler, og vil bli brukt som eksempler i undervisning.

#### Hvem er ansvarlig for forskningsprosjektet?

Prosjektet blir utført av et konsortium av samarbeidspartnere fra Universitetet i Agder (UIA), Universitetet i Oslo (UIO) og Norges teknisk-naturvitenskapelige universitet (NTNU). Polyxeni Vasilakopoulou (Professor ved Universitetet i Agder) er hovedansvarlig for prosjektet.

#### Hvorfor får du spørsmål om å delta?

Du har blitt spurt om å delta siden du er mellom 18 og 65 år og bor i Norge. Vi er interessert i innbyggernes syn på bruk av KI i offentlige tjenester. Det vil delta rundt 20-30 personer i denne brukerstudien.

#### Hva innebærer det for deg å delta?

Hvis du velger å delta, du skal bruke en digital prototype og svare på spørsmål knyttet til stola på KI-løsninger og forståelse av KI. Det vil ta deg ca. 30 minutter. Spørsmålene vil basere seg på hvordan du oppfatter bruken av KI i offentlige tjenester. Svarene blir tatt opp og transkribert for analyse, skjermen din er tatt opp og prototype interaksjonene dine blir analysert. Du som person vil anonymiseres, og ingen som leser rapporten vil kunne linke deg opp til svarene dine.

#### 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 vi oppbevarer og bruker dine opplysninger

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrevet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket. Navnet og kontaktopplysningene dine vil erstattes med en kode som lagres på egen navneliste adskilt fra øvrige data. Kun forskere tilknyttet prosjektet vil ha tilgang til opplysningene oppgitt under intervju og observasjon av interaksjon. Dette kan inkludere Ph.d.-studenter. Opplysningene vil bli lagret trygt på forskningsinstitusjonens servere og følge institusjonenes retningslinjer for sikker lagring.

#### Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Opplysningene anonymiseres når prosjektet avsluttes/oppgaven er godkjent, noe som etter planen er 30.06.2025.

#### Dine rettigheter

Så lenge 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 oss rett til å behandle personopplysninger om deg?**

Vi behandler opplysninger om deg basert på ditt samtykke.

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

#### **Hvor kan jeg finne ut mer?**

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Polyxeni Vasilakopoulou (prosjektleder, UIA)  
epost: [polyxenv@uia.no](mailto:polyxenv@uia.no)  
telefon: 41667639
- Miria Grisot (arbeidspakkeleder, UIO)  
epost: [miriag@ifi.uio.no](mailto:miriag@ifi.uio.no)  
telefon: 93421128
- Elena Parmiggiani (arbeidspakkeleder, NTNU)  
epost: [parmiggi@ntnu.no](mailto:parmiggi@ntnu.no)  
telefon: 73591464

Hvis du har spørsmål knyttet til NSD sin vurdering av prosjektet, kan du ta kontakt med:

- NSD – Norsk senter for forskningsdata AS på epost ([personvermtjenester@nsd.no](mailto:personvermtjenester@nsd.no)) eller på telefon: 55 58 21 17.

Med vennlig hilsen

Polyxeni Vasilakopoulou  
(Prosjektleder, Professor, Universitetet i Agder)

---

## **Samtykkeerklæring**

Jeg har mottatt og forstått informasjon om prosjektet AI4Users, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- å delta i brukerstudien

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

---

(Signert av prosjektdeltaker, dato)

## B Interview guide

### User Study 3 protocol - Caseworkers

#### Intro

Hello [PARTICIPANT\_NAME]

I really appreciate you taking time out of your day to participate in this session.

My name is [NAME\_RESEARCHER\_1] and I will moderate this session.

With me today is [NAME\_RESEARCHER\_2], and s/he will be taking notes and might also have one or two questions along the way.

Our research group is interested in the responsible implementation of technologies like AI and machine learning in the public sector. Our project takes a "Human-Centered AI" approach, which means we are focussing on the needs and perspectives of all the involved people, both citizens as well as employees of the organizations.

To give you a brief outlook on what we are going to do today.

- First, I will ask you some basic questions about who you are, and your experience working at NAV.
- Next, we will explore how such technologies could be used in the future within the public service, specifically in relation to your work. For that, I will give you a simple use case of such a technology, and then we will explore good and bad futures with this technology.

Is there anything you'd like to ask before we get going?

#### Start

- Share consent form
- Start recording





## Foresight

For this study, we are going to explore together possible futures in relation to AI in your workplace. We have this (fictitious) scenario related to work absences due to sick leave.

Based on the scenario, we would like you to think how different kinds of futures can look like. We would like you to think about these futures from different perspectives (roles):

### Scenario:

In our case, we assume there is a tool that can predict the probable duration of a citizen's sick leave. This prediction can help to avoid unnecessary dialog meetings in cases where a sick leave would end soon. The tool is part of the process when deciding if a dialog meeting has to take place. The outcome of the tool is a prediction of how long a citizen will probably still be out of work due to sickness

### Directions

#### Best case / worst case

- What could be the **best possible future** of such a system?
  
- What could be the **worst-case future** that you can imagine?

#### Probable futures

- What would you reasonably expect to happen with such a system?

#### Preferred futures

- What do you want to happen?
  
- How could we make that happen?

### Possible futures

- Could other things made be happen? And what could that be?

- What would produce those such futures?

### Discussion prompts (if needed)

- Think about the positive/negative effects
  - Employee perspective
    - Duration of work/workload
    - Objectivity / Impartiality
    - Documentation (public accountability)
  - For their role as a public servant
    - Social contract (individual vs. society)
- What types of output would you like to see?
  - Binary (sick leave will end before / beyond) vs. Exact days (beyond)
  - Which data is used? (representations examples)
  - Confidence level of the prediction? (representations examples)
  - "Feature importance"? (representations examples)
- Role of the case worker in the process?
  - Their role as being "in the loop"
  - Automated processes vs. human in the loop

## **C Systematic Literature Review**

Link to the full systematic literature review used in this thesis:

<https://docs.google.com/document/d/10Rlkqq5WBKDThYc1qwd56IDaQ6wlgSCQ7d14Vk9EYsM/edit?usp=sharing>

## D Interview guide: preliminary interviews

This interview guide is from the preliminary interviews we conducted (Eriksson & Olsen, 2022).

Interview guide
<p><b>Purpose of the interview</b> The purpose of this interview is to gather information on how employees at NAV <u>uses</u> AI, to get an understanding of how transparent AI is for them in their work environment</p>
<p><b>Declaration of confidentiality and interview usage</b> All of the participants will remain anonymous, and we will not collect any sensitive information about them. Potential recordings will be deleted within a month, once we are finished with the data analysis</p>
<p><b>Key components</b></p> <ul style="list-style-type: none"><li>- Present ourselves and the purpose of the interview</li><li>- Give the participants information on the confidentiality and duration of the interview</li><li>- Ask the participant if it is okay that we record the interview for analysis</li><li>- Present how the interview will be conducted</li><li>- Let the participant introduce themselves, and what job title they have at NAV</li><li>- <u>Conduct the interview</u> by asking questions and follow-up questions</li><li>- <u>Express gratitude</u> for their time and feedback</li></ul>
Questions for the interview
<ol style="list-style-type: none"><li>1. Do you have any questions about the interview before we start?</li><li>2. What is your job title?</li><li>3. Can you tell us <u>in general</u> what your work tasks are?</li><li>4. Are you familiar with <u>artificial intelligence</u>?</li><li>5. Have you used AI for your work?<ol style="list-style-type: none"><li>a. If you do not use AI for your work, can you think of any areas it could be applicable <u>for</u> your work?</li></ol></li><li>6. In what areas do you use AI in your work?</li><li>7. Can you explain any projects that <u>uses</u> AI in your company?</li><li>8. What challenges <u>do</u> AI solve in these projects?</li><li>9. When using AI, do you feel like you have a good understanding of what the system does?</li><li>10. Does the AI system give you any explanations as to what it does, why it does it, and how it ends up with the results that it gives you?</li><li>11. Do you feel like you can trust the results that <u>come from</u> using AI?</li><li>12. Are there any ethical challenges <u>by</u> using AI?</li><li>13. Are there any security risks <u>by</u> using AI?</li><li>14. Is there anything else you would like to add that is important to know before using AI?</li></ol>