10-14-2023

Addressing Uncertainty in AI Tool Development in Healthcare Through End-User Involvement

Ursula Sokolaj
Norwegian University of Science and Technology, ursula.sokolaj@ntnu.no

Casandra Grundstrom
Norwegian University of Science and Technology, casandra.a.grundstrom@ntnu.no

Aleisha Martul
Norwegian University of Science and Technology, martulalei22@gmail.com

ADDRESSING UNCERTAINTY IN AI TOOL DEVELOPMENT IN HEALTHCARE THROUGH END-USER INVOLVEMENT

Research paper

Sokolaj, Ursula, NTNU, Trondheim, Norway, ursula.sokolaj@ntnu.no

Grundstrom, Casandra, NTNU, Trondheim, Norway, casandra.a.grundstrom@ntnu.no

Martul, Aleisha, NTNU, Trondheim, Norway, martulalei22@gmail.com

Abstract

Artificial Intelligence (AI) development is an increasingly significant area of interest for Information Systems (IS) scholars, aiming to explore the socio-technical aspects of human-AI collaboration. This paper reports on a qualitative case study focused on a dementia-predicting AI tool currently under development. Through in-depth interviews with AI developers, we aim to gain insights into their expert practices and the uncertainties they encounter during the development process. Using a framework to outline the stages of AI tool development, we uncover four key uncertainties: feasibility, data, decision-making, and adoption. We examine how AI developers cope with these uncertainties through varying methods and present strategies for end-user involvement to reduce uncertainty when possible.

Keywords: AI development, uncertainty, end-user involvement, case study, healthcare.

1 Introduction

In the past decade, the innovative technologies of artificial intelligence (AI) and machine learning (ML) are becoming embedded into a variety of contexts, including healthcare (M. Braun et al., 2022). This adoption is incentivized by AI's potential to effectively transform various tasks (ibid) and meet increasing levels of case complexity and workload (Weber et al., 2022), therefore improving health services and quality of care (Lai et al., 2021). AI's advances in clinical decision-making and its capacity to learn allow humans and AI to work together mutually by leveraging and enhancing their respective strengths (Hemmer et al., 2022). As of today, in the healthcare domain, we talk about collaborative work between humans and AI rather than autonomous decision-making of the latter. Lai et al. (2021) define this human-AI collaboration "as an evolving, interactive process whereby two or more parties actively and reciprocally engage in joint activities aimed at achieving one or more shared goals" (p. 390), which in the context of clinical decision-making means "bilaterally sharing information and collaboratively forming decisions based on interactive knowledge exchange" (Hemmer et al., 2022, p. 2).

Despite the allure of interdependency, human-AI collaboration in clinical decision-making still needs to be improved. Impediments to the use of AI stem from imperfect algorithms, biases, and transparency or opacity, which demonstrate the uncertainty intrinsic to the black-box nature of these systems (Gagnon & De Regt, 2022; Lebovitz et al., 2022; Moltubakk Kempton & Vassilakopoulou, 2021). AI is commonly referred to as a computer program or intelligent system capable of mimicking human cognitive function (Asan & Choudhury, 2021). On the other hand, ML, essential in building these systems (Joshi, 2020), uses computational algorithms to make autonomous recommendations or decisions (Helm et al., 2020). AI tools are trained through ML to process, analyze, and learn from large data sets from various sources. With the ever-increasing digitalization of the healthcare sector in the form of electronic health records (EHRs) and services such as telehealth, the capacity for patient-generated data to train AI tools is most favourable. However, this data-driven aspect gives rise to several uncertainties related to the completeness, sufficiency and quality of the input data, including ground-truth labelling, algorithmic processing, or the accuracy of the algorithmic output and prediction (M. Braun et al., 2022; Dietz et al., 2021; Lebovitz et al., 2021).

Despite the great potential of collaborating with AI, decision-making in healthcare is more critical than in many other domains and can have fatal consequences if errors occur. In such cases, the medical experts remain primarily accountable nonetheless, which underscores the significance of the perceived trustworthiness of AI tools (Braun et al., 2022) and experts' ability to navigate situations of uncertainty (Lebovitz et al., 2022). A study by Weber et al. (2022) further emphasizes this relationship between the social and psychological factors—like trust or perception of technology—for a sustainable and successful adoption of and collaboration with AI tools. On the other hand, Hemmer et al. (2022) associate AI adoption with the presence or lack of user-centeredness during the design process, which determines the perceived validity by expert users, performance expectancy, and interferences with the workflow. Moreover, Bond et al. (2019) highlight the importance of placing humans at the center of AI solutions, as this can ensure that these systems will be adopted, usable, and ethical, which is closely linked to AI explainability, i.e., explanations of the tool's decision logic. The significance of this approach is also highlighted by Jacobs et al. (2021), who state that user involvement is crucial for contextual awareness, a lack of which can negatively impact how interpretable and useful these tools are, once embedded in a real-world scenario. It can, therefore, be argued that involving medical experts in the design and development process of AI tools could improve human-AI collaboration by mitigating uncertainty through transparency and usability through participation.

However, AI work has often prioritized increasing model accuracy rather than focusing on the needs of the intended users (Jacobs et al., 2021). Traditional design approaches face unique challenges for including end-users in AI tool development, such as explainability, but are poised to offer strong Human-AI interaction solutions (Bond, et al., 2019). While conventional obstacles to involving end-users in software development are known issues such as time demands, scalability, and contextual needs. In addition, although a growing number of publications focus on human-AI collaboration in the healthcare domain, including human-centered design and evaluation (Cai, Winter, et al., 2019; Jacobs et al., 2021; Lee et al., 2020), there is unexplored research in IS outlets for human-AI collaboration in clinical practices, such as prognosis and prevention, and disease-specific algorithms or tools (Lai et al., 2021). While IS researchers can contribute valuable insights to address the various challenges, the existing literature is primarily related to the technological use of AI and much less to its societal (or governmental) implications and recent advancements (Collins et al., 2021). The black-box of AI is a suitable metaphor to extend to the relatively unknown and understudied practices of AI developers (Dietz et al., 2021; Nascimento et al., 2019) and the differences from traditional software development practices, especially when working with data handling activities such as labelling (Amershi et al., 2019; Wan et al., 2020).

Due to such differences in AI development processes compared to traditional software development processes (Nascimento et al., 2019; Wan et al., 2020), novel uncertainties are emerging. Recent attention is being given to this phenomenon gap as "it is important to understand at which stage in the ML development process ML specific uncertainties occur and how they can be reduced and handled in order to steer a tremendous number of upcoming ML projects and ensure their success" (Dietz et al., 2021, p. 1). Consequently, investigating how AI experts cope with these emergent uncertainties in AI development projects requires particular attention (Cai, Reif, et al., 2019). Research that explores the developers' perspective (Grundstrom et al., 2023) and end-user collaboration (Hemmer et al., 2022) are moreover scarce in the extant literature, with early indications of phenomena to be of great interest to IS scholars (Dietz et al., 2021). To address the phenomenon above of interest for IS, we conduct an exploratory case study to answer the following research questions:

1. *What uncertainties do AI developers navigate?*
2. *How can end-user involvement help reduce uncertainty in the development of AI tools in healthcare?*

To address these questions, we first examine what uncertainties AI developers face during different AI development stages. Then we examine how end-users have been involved during different AI development stages to manage, mitigate, or cope with uncertainties of AI tools in a healthcare context. The paper is organized as follows. In section 2, we discuss the theoretical background for AI development processes used as the framework for this study. Section 3 presents information about the case study and research methodology. Section 4 presents the results, and section 5 discusses the contributions and implications of our findings.

2 Theoretical background

2.1 AI development

Research looking into the development practices and challenges of AI tools shows that there is currently no well-defined, standardized development process, possibly due to the ongoing evolution of the field (Nascimento et al., 2019). Several publications have proposed frameworks for the ML development process, each with a different number of stages, but all generally following a workflow familiar to traditional software development (Amershi et al., 2019; Wan et al., 2020). Based on a qualitative study with software developers working with ML projects, Nascimento et al. (2019) organized this workflow into a three-phase process—Elicitation, Development and Production—comprised of four stages, 1) Problem Understanding, 2) Data Handling, 3) Model Building, and 4) Model Monitoring. This model is visualized in Figure 1 and used as a framework in this article.

In the first stage, problem understanding, the team aims to collaboratively analyze the problem with the stakeholders, identify the requirements, define the scope and objectives of the project, and formulate a plan; hence, the 'Elicitation' phase. According to Wan et al. (2020), ML experts engage in dialogue with stakeholders to establish a realistic understanding of AI's capabilities and limitations, enabling them to manage expectations more effectively. The second phase, 'Development', consists of the data-oriented and model-oriented stages, 2 and 3, respectively. The Data Handling stage includes data acquisition, data labelling, data exploration, data structuring and feature engineering, tasks which aim to identify, collect, and process the dataset necessary to train the tool (Nascimento et al., 2019). In the last stage, model building, the developers choose, train, and tune the model. Preliminary tests evaluate the tool's performance using pre-defined evaluation measures. If the results are good, the model is deployed (ibid). Model monitoring is the last stage once the tool is applied to the production environment. Because in the 'Production' phase, the tool's training requires iterations as it learns with more and new data, continued performance monitoring is needed. This is done to identify errors or unexpected consequences and determine if changes over time invalidate the algorithm (Wan et al., 2021).

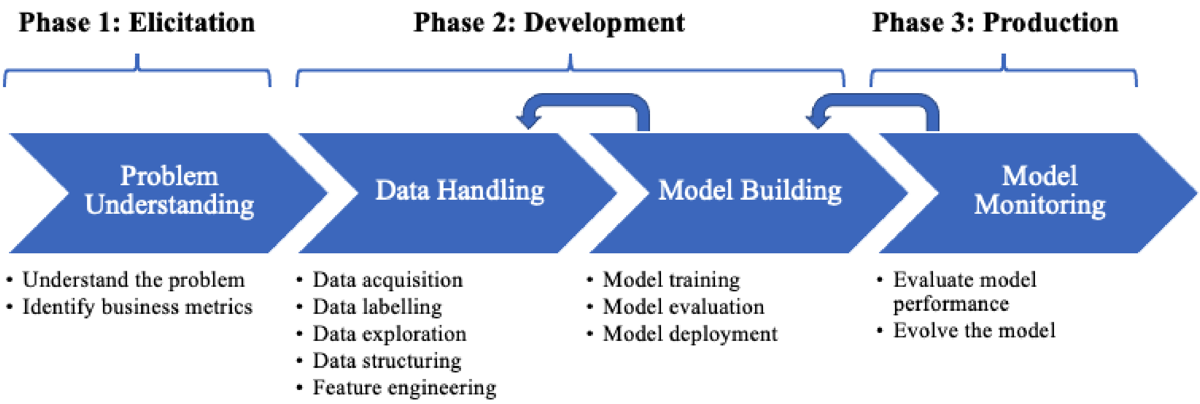

Figure 1. AI system development process stages were adopted from Nascimento et al. (2019) with three phases and four stages. Each stage includes the developer's tasks part of the stages. Solid arrows indicate iteration potential during the process.

The stages of the development and production phase are highly iterative, and feedback loops are executed continuously. Uncertainties may necessitate revisiting earlier stages, such as model building or data handling. Dietz et al. (2021) highlight that although the literature already gives an overview and description of the stages undertaken to develop AI tools, it does not elaborate on the uncertainties which arise at these respective steps. They address this gap in their study by integrating AI-specific uncertainties into the development framework and show how a collaborative approach with end-users in an organizational context can help address them. Our research takes a similar approach to fill this gap but with a context on AI developed in the healthcare domain and an emphasis on end-user involvement for uncertainty reduction strategies.

2.2 Uncertainty

Uncertainty emerges due to the complex nature of AI's decision-making process, making it challenging to explain the underlying logic and interpret the outcomes. Black-boxing the machine's decision-making process necessitates experts to cope with the uncertainty through various tactics (Grundstrom et al., 2023). In contrast, human decision-making is generally characterized to be traceable as it is more dynamic and is formed by a combination of intuitive and analytic reasoning styles, with cognitive limitations such as speed (Leyer et al., 2020). AI's ability to process and continuously learn from large datasets of different formats rapidly, all while delivering conclusions that resemble human execution, is intimidating (ibid). Considering that AI's self-learning is a continuous process, it can be said that AI development models are inherently dynamic and ever-changing and, therefore, never fully settled (Waardenburg & Huysman, 2022). As a result, it becomes challenging to draw the boundary between 'development' and 'use.' However, most of the extant AI research looks separately at either the process of creating the AI systems or the organizational impacts of their implementation (ibid). This stretches into classifying the actors that play a role in each as distinct entities that co-exist together but operate independently within their expertise. This distinction is heavily criticized as it hinders scholars from examining the socio-technical aspects of these processes, including but not limited to the context-specific designing of a tool or its impacts on its users and application environment (ibid).

In more recent research, a shift in this perspective is slowly taking place as scholars have started to recognize the importance of viewing the developers and users as co-creators of AI systems mutually dependent on one another (Lebovitz et al., 2021; van den Broek et al., 2021). This calls attention to the collaborative perspective of human-AI relationships in expert work practices, the opportunity for end-user collaboration, and a need for continual exchanges between technical and field-specific knowledge. Inevitably, emergent uncertainty between these dynamics of human-human and human-AI produces shifts in practice and collaboration, in addition to the inherent uncertainty to ML stochastic methods, which is more established in the field and can be classified by aleatoric (statistical) and epistemic (systematic) sources (Hüllermeier & Waegeman, 2021). Aleatoric uncertainty is inherent in ML due to the randomness in training data (such as flipping a coin) and is irreducible. In comparison, epistemic uncertainty refers to a lack of knowledge and can thus be reduced through additional data collection, expert knowledge, and model refinement (Chen & Geyer, 2022; Hüllermeier & Waegeman, 2021). Uncertainty, therefore, can be understood as the disparity between certainty and the knowledge available to the decision-maker at a given time (Nikolaidis et al., as cited by Dietz et al., 2021). In such instances, it is essential to identify strategies that can reduce or mitigate uncertainty, resulting in more successful AI tools. The choice between techniques that embrace or mitigate uncertainty depends on the nature of the uncertainty at hand. The former can be used when uncertainty is aleatoric, and coping strategies must be employed, as it is irreducible through additional data. Different techniques can be utilized to change the tolerances of the irreducibility, such as degrees of probabilistic outcomes in model variations. Epistemic, which refers to a gap between the current level of information and the threshold of complete details, where everything knowable is known, is reducible if the proper knowledge can be acquired. Collaborating with end-users offers a strategy to reduce epistemic uncertainty by facilitating the reduction of the knowledge gap.

3 Methodology

3.1 Case study

This qualitative research is the first phase of an exploratory case study to understand how experts handle uncertainty while developing AI tools. The exploratory case study is appropriate for studying our research question's “how” (Yin, 2003). As decision-making during development is a complex phenomenon entangled with experts' know-how and know-what knowledge, there are blurred boundaries between our area of interest and the context of the AI tool (Baxter & Jack, 2015). Furthermore, as the investigation of developers’ strategies for facing uncertainties part of the development process is relatively understudied for AI tools and has rich unknown phenomena of interest (Dietz et al., 2021), we see this research approach as highly appropriate. The case study occurred in a collaborative venture named “Aurora” (for anonymity). Aurora is based in northern Europe with a constellation of academic participants. It works exclusively on sustainable and trustworthy AI and represents a variety of perspectives, including the sociotechnical one. Our focus in Aurora was a specific and ongoing research project called “Cerebro,” an AI tool designed and developed for detecting the advancement of dementia. Cerebro was in the early stages of development when our study began in June 2022. The development is carried out centrally at Aurora, while clinicians' evaluation and user involvement are performed with international collaborators. Cerebro contains two components: the connector (sensor) and the detector (predictor). The connector examines whether the brain connectivity (how the different regions of the brain connect) exhibits risk factor signs of electrical signal degradation affected by dementia. Patterns in the electroencephalography (EEG) (a test with small electrodes attached to the scalp to record electrical activity) are analyzed to detect if subjects indicate some form of Mild Cognitive Impairment (MCI), which may develop into dementia in the future. The second component uses the data from the connector as an indicator for brain functionality and connectivity and, along with two other data inputs—cognitive scores and blood samples— assesses the risk for developing dementia and provides an output for clinicians in the form of a prediction with >95% accuracy. These together give the formulation for Cerebro to perform the prediction of a subject developing dementia, intending to enable clinical decision-making to intervene in a timely manner.

3.2 Data collection and analysis

This study is exploratory and therefore has several phases, but not all are included in this article. For this phase, nine in-depth interviews with eight different AI experts involved in developing the Cerebro research project were conducted; one interview was with the same expert. For this study, experts are characterized as AI developers who are part of the design and development of AI tools and may possess a variety of expertise necessary for application development (Piorkowski et al., 2021). Ethical approval was obtained through the appropriate governing board, and all participants provided informed consent to participate in the study.

Pseudonym*	**Expertise**	**Duration**	**Pseudonym***	**Expertise**	**Duration**
Ana	Statistics	56 mins + 35 mins	David	Machine learning	45 mins
Hari	Data science	56 mins	Omar	Machine learning	63 mins
Maria	Neuroimaging	45 mins	Luis	Machine learning	49 mins
Eric	Statistics	60 mins	Jorge	Computational science	56 mins
*Pseudonyms used throughout results to represent anonymized participants					

Table 1. Interview participant characteristics.

The in-depth interviews were semi-structured in design, with questions tailored around the decision-making practices of experts, involvement of clinicians, ways they cope with uncertainty, and their values before, during, or after AI tool testing and development phases. Details about the participants can be seen in Table 1. This qualitative data collection process has been conducted alongside the development of the Cerebro AI dementia tool, which involves a variety of experts in the Cerebro research group. During the first phase, developers working under the Cerebro research group were identified, contacted for participation, and interviewed over a period of 6 months from June-December 2022. This stage of the study only focuses on the developer's perspective, additional data collection from clinicians (as end-users) is ongoing in the form of semi-structured interviews and a questionnaire. The interviews were conducted digitally over Microsoft Teams or in-person for an average of 52 minutes in length, with video and written transcript recordings taken. Following the interviews, the transcripts were cleaned, finalized and the data was anonymized. Our interviews focused on gaining insights of participants' understanding of uncertainty, how they may quantify and work with uncertainty, as well as their decision-making processes and user involvement during AI tool development. This was based on the questions in the semi-structured interview guide; due to space considerations the interview guide will not be appended.

The data was analyzed following the six-phase thematic analysis process from Braun and Clarke (2006) with recursive inductive coding (Azungah, 2018). This is a back-and-forth process between data analysis and existing literature, with the aim of deriving themes by coding the raw data, as well as creating meaning and understanding the emerging concepts (ibid). This approach was selected to capture patterns of our phenomena of interest (uncertainty) throughout the phases of the system development process (Nascimento et al., 2019). An inductive approach to coding the data was chosen considering the exploratory nature of this case study, relatively limited research on types of uncertainties in the development process (Dietz et al., 2021), and capacity for theory and model building (Thomas, 2006). As the interviews were transcribed, they were entered into Dedoose (version 9.0.62 Mac) which is a computer-assisted qualitative data analysis software tool. Authors (US and AM) began familiarising themselves with the data, and early codes were assigned and collated to gain an initial understanding of the developers' ideas around uncertainty and decision-making in the development of Cerebro. As the iterative stages between data collection and analysis were carried out, different themes began emerging from the data, and new themes which became of interest were then discussed with participants in later interviews. Further subthemes were then generated based on emerging patterns of themes in the dataset, including aleatoric and epistemic uncertainty. Themes were reviewed and the four stages of the system development process were connected according to Nascimento et al., (2019): 1) Problem Understanding, 2) Data Handling, 3) Model Building, and 4) Model Monitoring to structure the phase of uncertainty and corresponding reducing actions. Continued definitions and names for each theme were iterated according to the overall narrative of uncertainty when developing an AI tool, resulting in main themes for uncertainty and corresponding actions. The authors revised literature to compare and inform theme descriptions to harmonize with relevant works such as Dietz et al. (2021). We present the results of our analysis in the format of the four system development phases, their corresponding uncertainties, and subsequent actions to reduce or cope with uncertainty.

4 Findings

This section presents the findings from the interviews with AI developers. The findings are arranged according to Nascimento et al.'s (2019) process model for AI development with three phases and four stages; see Figure 1. The participants had varying levels of experience and degrees of expertise relevant to working with the four stages of tool development. Therefore, they had unique perspectives on their uncertainties and coping strategies. The findings narratively account for the process of developing Cerebro and provide a comprehensive overview of these development stages and the uncertainties encountered, alongside a description of user involvement and strategies for uncertainty mitigation.

4.1 Elicitation

4.1.1 Stage 1: Problem Understanding

The very initial phase of developing an AI tool is elicitation. This phase refers to efforts to analyze, understand and define the problem at hand, determine the task and requirements of the AI solution, and understand the problem being solved. This stage lays the foundation and shapes the direction of the entire project. Once the problem being solved is understood, it becomes possible to identify what sort of approach, method, and data is needed. This also involves checking whether the required data is available and determining how it can be collected.

Uncertainty in this stage is related to the **feasibility** of the AI tool's value for the end-user or business and the capacity for whether the project can be executed to meet the end-user needs. A detailed analysis with critical reflection of all aspects of the proposed project must be informed by the problem understanding to ensure an effective and appropriate AI tool to address the case. This uncertainty is caused, on the one hand, by the nature and current state of AI technology itself. Despite significant advancements in the field, there are still limitations to what AI can do. Moreover, the development of an AI tool requires a deep understanding of the problem domain. If this understanding is limited, or the problem's complexity is very high, such as in healthcare, designing an efficient tool can be pretty challenging, and therefore its **feasibility** becomes uncertain.

"And for [Cerebro] as it is at the moment, at least in my view, it isn't really fixed, as this is going to be what it is. […] We don't quite know yet because the data hasn't been recorded yet, whether it's due to the data missing or due to us having made a mistake in our thought process, in our hypothesis." Maria

AI developers often need to gain the expert knowledge of healthcare professionals (or other domain-specific experts) using Cerebro to develop a more holistic understanding of the underlying problem. Moreover, they may need more intricate details that can be beneficial in understanding how certain models work with specific data or how a tool is applied in its end use. This is especially relevant in the medical field, where technicians need to gain expert knowledge utilized by medical professionals in their decision-making processes in practice. AI developers work to extend their knowledge into the relevant domain to varying degrees as a coping strategy for this.

"So I think the most important is to understand the biological or medical application to some extent. We usually try at the start to read a lot about the problem and what has been done there already. And then once we understand that, we talk a lot to the partners and maybe do some more reading." Hari

The **uncertainty of feasibility** further requires the acknowledgement of how significant the end-user is in the application, for business value but especially for meeting the needs of the experts who will be using the AI tool in practice, as well as for the implications of that use. Traditional IT tools which support work practices have differing consequences from AI tools due to their nature and real-world consequences. Cerebro will be supporting the decision-making of medical experts in the prognosis of a patient developing dementia. Therefore, this demands as much certainty as possible for patients and healthcare professionals for accuracy and usability. These implications are risky for the project and are considered by AI developers in the early phases. Aspects such as ethics and regulation impact the capacity for accessing relevant data, using AI tools in clinical practice, or decision-making outcomes.

" […] some AI that are embodied are not just running on the computer, they are running physical artifacts, controlling a robot or some device in a plant. And this is a risk at a different scale because it has a very large impact. So there is other risk that impacts [development] like regulations and laws […] there is extended risk when we actually then deploy and embody stuff in the real world and there are also impacts on other people or […] ethical implications of the decisions made and stuff." Luis

The AI developers indicate that early-stage user involvement can remove the disillusionment of building a successful AI tool as a strategy for reducing uncertainty and risk in the elicitation phase.

"If we just start right away to develop and to try out and create a prototype and we kind of forget at the moment about who's going to use it we might end up with something that is not usable, so I think it's very important in the applied to kind of involve early the users. I think that's kind of key because otherwise if you just sit on your own and you do the best technical implementation, then it's not usable for [many] reasons." Luis

Maria extends this strategy further, referring to communication between AI developers and end-users for the importance of creating a shared understanding of the Cerebro project's outcomes and realistic expectations in interdisciplinary environments and opportunities for involvement.

"Communication is one of the most important things, especially for these interdisciplinary projects. So, what we often do is we organize seminar workshops to give an overview of the basics of machine learning, and [the medical experts] also understand the limitations and the possibilities." Maria

User involvement from this initial stage of the AI tool development becomes crucial in reducing the **uncertainty of feasibility.**

4.2 Development

The next phase is the development, which includes two stages. The first stage is data handling with such tasks as the acquisition, labelling, and structuring of data for ML training. The second stage tasks the training, evaluation, and deployment of the model being built for the AI tool. These two stages are iterative as part of the development process.

4.2.1 Stage 2: Data handling

Data uncertainty at this stage of development is of extraordinary concern for AI developers. The variety of tasks necessary for collecting and preparing data for training the ML model is substantial. Several of the participants involved in the initial stages of AI tool development identified the significance of the data handling process in shaping the deliverable:

"This is one of the first steps we should take and the amount of data we collect. So, this is a very important process. After this process, then we have these data pre-processing phase. That means we get all the data we collect, we clean it up, we set-up the cleaning up process, which is also important because we have multiple partners." Hari

"We use a sentence [about data], [if you put garbage in then you get garbage out]. If you don't use efficient data, you cannot take any efficient response." Ana

For Cerebro developers, this manifested as uncertainty around the quality of the data, and the incompleteness of the data labels. As the data is used to train and test the ML models, quality data sets must be used, and where there are errors in the **data**, they should be understood so the model is more trustworthy:

"If you have a really poor data set to be able to make some sort of predictions, you will see it when you test it, you'll see that the model is performing in the pool, so then you will kind of directly get a feedback that this is not good." Eric

Several aspects of **data** collection and acquisition were identified by the participants connected to the quality of data, which is subjective. One of these aspects was the utilization of complete data sets with sufficient data and no missing values, which requires enormous efforts to clean and is very often not the case in open data sets. This is where the initial EEG training data comes from for Cerebro. Another aspect is ensuring that the **data** is representative of the population and unbiased.

"Uncertainty is everywhere, in data set, in model, in prediction and all parts include uncertainty now. If you don't take enough observation in your data set for training, there is uncertainty because of the underrepresented groups, for example. And in the real world you cannot be sure of all conditions." Ana

"It depends a bit on your use case, but usually we want the data set that is not biased...but also that represents reality because you can have a perfect data set, but then the clinical practice is not perfect. If we know OK, the reality is also biased, then we also have to work with biased data sets...But then the most important I think for us is the clinical usability at the end. We are not really interested in making models that cannot be used." David

Uncertainty in this stage is connected to something inherent in the **data** quality, like noisy **data** with missing values. Random elements that cannot be accounted for, much like the weather, represent the aleatoric type of uncertainty, which is irreducible and must be acknowledged and worked with. The participants kept the usability of the model in mind as the end goal and recognized there would often be bias or uncertainties in the **data** sets that could only be removed if the **data** was cleaned. Where uncertainty cannot be removed and instead must be acknowledged, AI developers expressed mitigation methods as a coping strategy.

"... usually one of the first things that I do is I try to remove noise from the data. The data I usually work on is not super clean or nice data. It's typically noisy. So, the typical step to, before I start even to kind of develop the [model], is [look at] what are the best ways to filter out noise. Depending on what data it is. [...] To remove uncertainty is like noise quantification, noise removal, filtering tricks, reprocessing. Basically, it's the answer to my uncertainty and mitigation." Luis

On the other hand, the AI developers also talked about epistemic uncertainty represented in data quality, which identifies a lack of knowledge and can therefore be reduced by collecting and cleaning more data. Part of this data collection and cleaning allows end-users and their medical expertise to be involved in relevant knowledge-building practices for data sets. The AI developers are engaged with this tacit knowledge transfer to data and clinical involvement in informing information gaps. However it is unclear to what degree of intensity the clinicians' expertise is utilized.

Cleaning out noise is generally done by the AI developers, but in some cases, it can also include the clinicians' input and feedback. The involvement of clinicians in uncertainty mitigation is most importantly beneficial due to their knowledge, both know-what and especially know-how, that medical experts, for instance doctors, have in their field. This background knowledge is necessary when developing the **data** sets for the ML model. If the AI developers are not experts in the field the tool is to be used in, they must learn necessary information through this collaboration. The experts can help the developers understand the data and the context in which this data can be used. In addition, they can check the quality of the data, identify the important features for the decision-making process, and how they are used to produce an outcome. Omar talked about the involvement of neuroscience domain experts during the Cerebro tool development:

"They help us with the design itself because we have to take into account features or characteristics that come from the neuroscience domain, because it does not help that we throw away all the knowledge from neuroscience." Omar

Participants also talked about uncertainty related to **data** structuring and the preparation part of the data handling stage. The data labelling causes explicitly this, and whether the medical experts did this or otherwise remains unclear. As it is used for training the ML model, knowing and trusting the **data** labels lends certainty to the overall outcome of the AI tool. Uncertainty in the labels is associated with the soft vs. hard ground truth, which is especially relevant in the medical context due to the know-how nature of situated, tacit knowledge.

"We have this principle of hard and soft ground truths. Let's say you have images of cancer, and you took biopsy of this, send it to a lab and then get the result and it exactly says this was cancer, so you have like a hard evidence that there was something. And soft ground truth is, if you would have a lot of images from polyps and you let three doctors annotate them to find the cancer, even if they might be correct a lot of the time, they might also make mistakes because they're just humans." David

In this scenario, user involvement when building data labels for domain-specific knowledge can create more **data uncertainty** for the AI developers, which is acknowledged as the human element of the development process. However, the AI developers proactively approach this **data** labelling uncertainty by asking the expert end-users to describe their levels of uncertainty in their own labels, which can later be accounted for in overall prediction estimations.

"Then you have more uncertainty where the doctors are doing annotations for example, and what we're trying to do there is at the same time as they're annotating, they think about how certain they are. Which we then can use in the models [...] But I think it basically comes from the human part where they don't really know what it is or maybe you don't have all the labels or the information in the data which also includes of course some uncertainty." David

This demonstrates an essential strategy for developers when reducing uncertainty through end-user estimation.

4.2.2 Stage 3: Model building

Once data is collected and cleaned, the developers must consider which method or algorithm is suitable for solving the identified problem. They then apply it to the data and look at the output to understand its accuracy and appropriateness. However, this stage is not a bounded one, especially in relation to the previous data-handling process. Typically, iterations between data handling and modelling are normal in the development process as a form of infinite regression. As the tool goes through development, AI developers simultaneously utilize layers of neural networks to contribute to the deep learning of the model, so it can continuously learn parameters from the data set. In addition, developers experiment with different sets of hyperparameters to fine-tune the output and explain data variations.

The **decision-making uncertainty** in this stage of development is mainly related to the algorithm's suitability and selection, which manifests in the model outcome. Uncertainty encountered in this stage stems first from choosing which algorithm is the best or most suitable to solve something:

"[…] uncertainty about what is the best model. For instance, there is also a kind of more qualitative uncertainty in these technical aspects. […] So, in that kind of uncertainty, then we try to bring down, so to say, to find the kind of quantity that enables us to take a decision." Jorge

Another form of **decision-making uncertainty** is related to the variability of the algorithm's performance across different times and settings, depending on dynamics in the end-user environment, but most importantly, stemming from the human factor of development:

"This is more like uncertainty within the algorithm that we use and there is uncertainty in our brain too. We might come up with two different answers on the same problem, in two different times, based on other aspects, noise, stress, the things that we have learned, so uncertainty is part of like the biological intelligence machinery as well. Not only the algorithm uncertainty as well." Luis

Maria and Eric also elaborated on uncertainty connected to how well the subject that the tool is intended to be used on fits within the parameters of the ML model and, ultimately, the sample data:

"You have your system and then you will later use it on new patients. So then usually in statistics when you think of quantifying uncertainty you assume that the new patients that arrive follow the same distribution as the data you use to train the model." Eric

One of the coping strategies used by AI developers to manage the uncertainty of algorithms and human factors is relying on intuition. The concept of intuition was associated with experience gained over time, which generally manifests in knowing which technique or algorithm has a high probability of being fit for a certain type of data:

"...there is of course intuition that comes out of experience, when you develop the tools, which technique would fit best for this kind of problem, which architecture can fit best? Because there is out there so many, many approaches for machine learning that really rely on intuition mostly because we know that for this type of data, this type of architecture worked well and then it should work probably." Omar

Two participants added that when different team members collaborate and their intuitive suggestions overlap, it becomes a sort of evidence which validates their decisions. This highlighted collaboration as a reduction strategy for **decision-making uncertainties:**

"With designing deep learning models, it's mainly that we agree on the model architecture, then we develop some kind of model architecture. We have these collaborative brainstorming sessions as well that's [to figure out] how we do it or which way we should do it? What components should we add or not? So there is a lot of collaboration in designing the model..." Hari

"...Sometimes it's better if we are in the team. Each one will make a suggestion. If the suggestion is the same from everybody, then this is strong evidence that we should start in that way." Jorge

Decision-making uncertainty towards outcomes can further be linked to the likelihood of the AI tool predictions being accurate and presenting explainable AI decision-making to end-users:

"... uncertainty in the output of the artificial intelligence algorithm that we are using... uncertainty would be then with how much certainty is this output the actual output that might be real and then there are ways of doing that […] to quantify some kind of an error margin on the output." Luis

Because the complexity of human decision-making may not be accurately mimicked by the machine, some developers identified the value in having users involved with the development of the tool and its decision-making processes to mitigate this problem. The participants recognized the inherent uncertainty in the nature of the clinicians' work and decision-making in practice. This should be considered in the tool to offload **decision-making uncertaint**y.

"No models are 100% correct. So that's why it's very important to have humans in the loop, especially the domain experts who are using it." Hari

Where tools are designed to assist clinicians in their clinical decision-making, alongside their know-how knowledge in the medical field, it's important that technicians can understand this uncertainty and reflect it in the tool development and outputs. As Cerebro is still in the development phase, limited evidence of the effect of user involvement at this stage was available but is planned iteratively throughout the project.

4.3 Production

4.3.1 Stage 4: Model Monitoring

The final phase of the process is production, which stages model monitoring through tasks of evaluation and evolving in the model. This takes place in iterations with the previous stage 3, model building.

Uncertainty during this phase is exhibited in the human-machine interactions, specifically clinicians' lack of know-how in how to use the AI tool in clinical practice due to little training or experience, willingness to engage with Cerebro, and perceived usefulness and safety for their patients. This was characterized as **adoption uncertainty** due to implications for the need to continuously evolve the AI tool through feedback loops, for further data acquisition to support ongoing ML training, and for end-users' willingness to incorporate the Cerebro solution into clinical practice. Either of these scenarios are significant barriers to sustainable AI learning and for the AI developers to monitor, evaluate, or evolve Cerebro.

Medical experts' involvement during the development of the tool could help mitigate this **adoption uncertainty** considering that knowledge exchange workshops are a planned part of the project process, as is expert involvement throughout the development lifecycle. Uncertainty, moreover, rises in relation to users' acceptance of the tool. Trustworthiness and explainability of the artifact were discussed as important factors that can play a key role. The explainability of AI tools was especially emphasized when uncertainty can only be minimized and not eliminated. Being able to estimate the degree of uncertainty of the tool makes it possible for the experts using it to understand the degree of accuracy of its outcome and therefore reduce the risk of making wrong decisions. David highlighted the importance of explaining the ML experts' decision-making process to the end-users, specifically in the medical domain:

"And we try to do that for each part in the pipeline in a way, because I think it's very important that they understand what we're doing. They might not have to understand the mathematics, but they have to understand we did this because of this and now we do this because of this." David

Efforts to increase the explainability of the dementia prediction tool are evident in the AI developers' practices. Since the health domain experts do not understand the AI experts' techniques, they should be provided with explanations of the decision-making process of the tool. These explanations can then lead to higher transparency and trust to support the reduction of adoption uncertainties of the artifact. *"Maybe if you can make it more transparent by design, if it's really opaque and you can explain it, then you increase the trustworthiness of the model. I use the opportunity to make my model more trustworthy* *so that I can tell the end-users that this is the probability of confidence or probability of happening that all that my model is saying yes or no." Hari*

However, the attitude for the acceptability of uncertainty also depends on the domain field in which the tool is being used, driving **adoption uncertainty**. Therefore, it is important to involve the end-users to determine this:

"[…] for some use cases you have very little acceptance for mistakes and for others you don't really care because everything that is better than random is acceptable in a way. And then of course, it depends also on the doctors, the final users, how much they would accept." David

According to Omar, uncertainty is a new topic in ML, at least how it is understood today. He observed that their development research team did not prioritize uncertainty until Cerebro's dementia prediction model in stage 3, but it should have been incorporated from the beginning. Since the tool is expected to go into the certification stage, which is heavily regulated, it is now important to properly prepare so all the testing should be done thoroughly. However, uncertainty at times can only be identified in "small scale, small instances", so the only solution is to embrace uncertainty as part of adopting an AI tool for decision-making. In such cases, it is important to have collaborative systems between humans and the machine, to reach better decisions and reduce uncertainty:

"... this goes hand in hand with the idea of the future of having collaborative system between AI and humans, because the person can know what is uncertain about the AI tool, in order to help. If the person is taking the decision at the end it will help to make a better decision, not to rely exactly on the machine learning tool when there is high uncertainty, or to correct the machine learning tool when it is highly uncertain about something..." Omar

Several AI developers proposed to assess the AI tool with the users to upgrade it according to their perspective needs and emergent requirements to cope with **adoption uncertainties**. The usage of quantitative criteria measurements as a strategy to evaluate the accuracy of the tool from the technical perspective is one possibility, a more qualitative approach can also be taken to evaluate and evolve the AI tool with the end-users. According to Jorge, two main things should be understood. First, whether the end-users consider the tool's outcome or rely on their own decision-making, and second, whether the reasons behind the expert decision-making are influenced by absent explanations or usability. This strategy was also echoed by Eric:

"The feedback we get from the medical experts affects how we develop the machine learning systems and how we try to modify and improve them…" Eric

5 Discussion

Our findings contribute to the emerging information about AI tool development uncertainty with theoretical and practical implications. Using the AI development process model from Nascimento et al. (2019) as a theoretical framework, we found four variations of uncertainty connected to AI tool development at different stages: feasibility, data, decision-making, and adoption uncertainties. We characterized coping strategies employed by AI developers when navigating uncertainty and examined how the end-users in all stages of development could mitigate uncertainty to varying degrees of reduction.

First, our work contributes to the literature on uncertainty in the design and development of AI tools. Our work directly contributes to the gap in developer perspectives when coping with uncertainty, as pointed out in a recent literature review (Grundstrom et al., 2023). In alignment with a similar study from a different domain, the four uncertainties we uncovered were harmonious with the eight uncertainties identified by Dietz et al., (2021) during comparable phases of the AI tool development process. More robust data collection and interview participation in the Dietz et al., (2021) article leads us to believe that missing uncertainties in our study are yet to be recognized. Our future research will follow a second phase with clinician involvement to further narrow the gaps of uncertainty in the Aurora case study from the end-user perspective, which we anticipate will enrichen the argumentation for involving end-users in AI tool development (Bond et al., 2019; Hemmer et al., 2022; Lai et al., 2021). Second, we further extend the theoretical importance of several key areas of AI research, such as explainability in healthcare for AI tools (Moltubakk Kempton & Vassilakopoulou, 2021). The explainability of AI tools was especially emphasized when uncertainty can only be minimized and not eliminated. Being able to estimate the degree of uncertainty of the tool makes it possible for the experts using it to understand the degree of accuracy of its outcome and therefore reduce the risk of making unacceptable predictions. Reducing the opacity of the black-box in the AI tool through end-user involvement was discussed as an important aspect of the implementation, especially for overcoming decision-making and adoption uncertainties. Ensuring engaged augmentation—embedding AI results in decision-making practices of experts (Lebovitz et al., 2022)—supports the value of end-user involvement for human-AI collaboration (Hemmer et al., 2022). According to user participation literature (Markus & Mao, 2004), active participation makes users perceive the system more relevant, leading to more optimistic attitudes and positively influencing system usage. The strategy for overcoming acceptance uncertainties in stage 4 complements this idea despite nearly 20 years having lapsed and differing development contexts. Moreover, involvement of users ensures that the end-user requirements are well-informed, and the resulting system will likely align with the business demand. It is widely acknowledged that user participation in development impacts attitudinal and behavioural project outcomes (He & King, 2008; Shen et al., 2013).

Third, our work has practical implications for AI developers working on projects that do not involve end-users early and throughout the AI development lifecycle. Reducing uncertainty when possible and managing it when necessary is highly impactful to the sustainability of the AI tool and creating a symbiotic human-AI collaboration to ensure ongoing data creation and learning. According to Dietz et al. (2021), uncertainties in developing conventional tools often stem from the unknown amount of necessary design work, or the accurate recognition of customer requirements, with strategies like requirements engineering or knowledge transfer between users and developers used to address them. In the case of conventional software, the strategies that directly reduce uncertainties and improve project performance are much more effective than the ones that merely mitigate their impact on development. On the other hand, when it comes to AI development projects, according to Dietz et al. (2021) and as apparent from our study, strategies that mitigate uncertainty impact are essential because, more often than not, uncertainty is inherent and not possible to be entirely reduced (Wan et al., 2020). Therefore, it is ultimately vital to embrace the differences between traditional IT implementation and AI practices, which emphasize a culture of acceptance and change as "the self-learning nature of AI systems calls for a sustained collaboration between developer and user, even when the tool is fully deployed in practice" (Waardenburg & Huysman, 2022, p. 6). We see this relationship as recursive between the developer and end-user in the data practices for labelling. Where AI developers were using medical experts to label their perceived uncertainty in the data, clinical experts expect AI developers to demonstrate their degree of certainty in the form of explainability in return. Future investigation into specific symbiotic relationships for data uncertainty reduction should be investigated in more detail.

Our study has limitations. Due to the nature of our exploratory case study, only emergent evidence of activities around the parts of the second phase and most of the third phase's uncertainty are available for analysis and presentation. The connecter component of Cerebro was delivered in March 2022, and additional data collection is ongoing. Therefore, further development of the results is needed to expand the theoretical contributions and practical implications. The number of participants in our study and triangulation through documentation will also be increased over time to ensure empirical data saturation and improve the case study validity (Yin, 2003). Even though uncertainty mitigation was a central part of the narrative, experts often observed that it is not necessarily an unwanted or negative feature. They recognized uncertainty mitigation is an inherent feature in many real-life and biological intelligence systems; therefore, removing it is not always the ultimate goal, as the AI model may fail to depict the system accurately if removed. Nevertheless, understanding uncertainty sources while developing the tool remains necessary, and future work to create taxonomies with mitigation strategies for strategic responses to uncertainty is undoubtedly valuable to the AI development process.

References

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software Engineering for Machine Learning: A Case Study. *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*, 291–300. https://doi.org/10.1109/ICSE-SEIP.2019.00042

Asan, O., & Choudhury, A. (2021). Research Trends in Artificial Intelligence Applications in Human Factors Health Care: Mapping Review. *JMIR Human Factors*, *8*(2), e28236. https://doi.org/10.2196/28236

Azungah, T. (2018). Qualitative research: Deductive and inductive approaches to data analysis. *Qualitative Research Journal*, *18*(4), 383–400. https://doi.org/10.1108/QRJ-D-18-00035

Baxter, P., & Jack, S. (2015). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2008.1573

Bond, R. R., Mulvenna, M. D., Wan, H., Finlay, D. D., Wong, A., Koene, A., Brisk, R., Boger, J., & Adel, T. (2019). Human Centered Artificial Intelligence: Weaving UX into Algorithmic Decision Making. *RoCHI*, 2–9.

Braun, M., Harnischmacher, C., Lechte, H., & Riquel, J. (2022). *Let's Get Physic(AI)l- Traonforming AI-Requirements of Healthcare into Design Principles*. ECIS.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cai, C. J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., Wattenberg, M., Viegas, F., Corrado, G. S., Stumpe, M. C., & Terry, M. (2019). Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3290605.3300234

Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 1–24. https://doi.org/10.1145/3359206

Chen, X., & Geyer, P. (2022). Machine assistance in energy-efficient building design: A predictive framework toward dynamic interaction with human decision-making under uncertainty. *Applied Energy*, *307*, 118240. https://doi.org/10.1016/j.apenergy.2021.118240

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, *60*, 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Dietz, J., Glaser, K., & Hoehle, H. (2021). *Uncertainty Reducing and Handling Strategies in ML Development Projects*.

Gagnon, E., & De Regt, A. (2022). How Humans and Machines Make Sense Together: Characteristics and Outcomes of a Case Survey Approach. *Academy of Management Proceedings*, *2022*(1), 11313. https://doi.org/10.5465/AMBPP.2022.11313abstract

Grundstrom, C., Mohanty, P., & Parmiggiani, E. (2023). *AI UNCERTAINTY IN EXPERT DECISION-MAKING: A QUALITATIVE EVIDENCE SYNTHESIS*.

He, J., & King, W. R. (2008). The Role of User Participation in Information Systems Development: Implications from a Meta-Analysis. *Journal of Management Information Systems*, *25*(1), 301–331. https://doi.org/10.2753/MIS0742-1222250111

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., Spitzer, A. I., & Ramkumar, P. N. (2020). Machine Learning and Artificial Intelligence: Definitions, Applications, and Future Directions. *Current Reviews in Musculoskeletal Medicine*, *13*(1), 69–76. https://doi.org/10.1007/s12178-020-09600-8

Hemmer, P., Schemmer, M., Riefle, L., Rosellen, N., Vössing, M., & Kühl, N. (2022). *Factors that influence the adoption of human-AI collaboration in clinical decision-making*. https://doi.org/10.48550/ARXIV.2204.09082

Hüllermeier, E., & Waegeman, W. (2021). Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, *110*(3), 457–506. https://doi.org/10.1007/s10994-021-05946-3

Jacobs, M., He, J., Pradier, M. F., Lam, B., Ahn, A. C., McCoy, T. H., Perlis, R. H., Doshi-Velez, F., & Gajos, K. Z. (2021). *Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens*. https://doi.org/10.48550/ARXIV.2102.00593

Joshi, A. V. (2020). Introduction to AI and ML. In A. V. Joshi, *Machine Learning and Artificial Intelligence* (pp. 3–7). Springer International Publishing. https://doi.org/10.1007/978-3-030-26622-6_1

Lai, Y., Kankanhalli, A., & Ong, D. (2021). *Human-AI Collaboration in Healthcare: A Review and Research Agenda*. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2021.046

Lebovitz, S., Levina, N., & Lifshitz-Assa, H. (2021). Is AI Ground Truth Really True? The Dangers of Training and Evaluating AI Tools Based on Experts' Know-What. *MIS Quarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/MISQ/2021/16564

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To Engage or Not to Engage with AI for Critical Judgments: How Professionals Deal with Opacity When Using AI for Medical Diagnosis. *Organization Science*, *33*(1), 126–148. https://doi.org/10.1287/orsc.2021.1549

Lee, M. H., Siewiorek, D. P., Smailagic, A., Bernardino, A., & Bermúdez i Badia, S. (2020). Co-Design and Evaluation of an Intelligent Decision Support System for Stroke Rehabilitation Assessment. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–27. https://doi.org/10.1145/3415227

Leyer, M., Oberlaender, A., Dootson, P., & Kowalkiewicz, M. (2020). *Decision-making with artificial intelligence: Towards a novel conceptualization of patterns*. Pacific Asia Conference on Information Systems (PACIS), Dubai, UAE.

Markus, M. L., & Mao, J.-Y. (2004). Participation in Development and Implementation—Updating An Old, Tired Concept for Today's IS Contexts. *Journal of the Association for Information Systems*, *5*(11), 514–544. https://doi.org/10.17705/1jais.00057

Moltubakk Kempton, A., & Vassilakopoulou, P. (2021). *Accountability, Transparency and Explainability in AI for Healthcare*. https://doi.org/10.18420/IHC2021_018

Nascimento, E. de S., Ahmed, I., Oliveira, E., Palheta, M. P., Steinmacher, I., & Conte, T. (2019). Understanding Development Process of Machine Learning Systems: Challenges and Solutions. *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM),* 1–6. https://doi.org/10.1109/ESEM.2019.8870157

Piorkowski, D., Park, S., Wang, A. Y., Wang, D., Muller, M., & Portnoy, F. (2021). How AI Developers Overcome Communication Challenges in a Multidisciplinary Team: A Case Study. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW1), 1–25. https://doi.org/10.1145/3449205

Shen, K. N., Khalifa, M., & Almulla, A. (2013). *When users are professionals: Effective User Participation for Information Systems Assimilation*. Thirty Fourth International Conference on Information Systems, Milan.

Thomas, D. R. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation*, *27*(2), 237–246. https://doi.org/10.1177/1098214005283748

van den Broek, E., Sergeeva, A., & Huysman Vrije, M. (2021). When the Machine Meets the Expert: An Ethnography of Developing AI for Hiring. *MIS Quarterly*, *45*(3), 1557–1580. https://doi.org/10.25300/MISQ/2021/16559

Waardenburg, L., & Huysman, M. (2022). From coexistence to co-creation: Blurring boundaries in the age of AI. *Information and Organization*, *32*(4), 100432. https://doi.org/10.1016/j.infoandorg.2022.100432

Wan, Z., Xia, X., Lo, D., & Murphy, G. C. (2020). How does Machine Learning Change Software Development Practices? *IEEE Transactions on Software Engineering*, 1–1. https://doi.org/10.1109/TSE.2019.2937083

Weber, S., Knop, M., & Niehaves, B. (2022). How Do Medical Professionals Perceive Artificial Intelligence: An Analysis of Reddit Data. *Wirtschaftsinformatik 2022 Proceedings*. Wirtschaftsinformatik.

Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed). Sage Publications.