Proceedings Paper:
Hussain, Muhammad orcid.org/0009-0008-5116-3893, Iacovides, Jo orcid.org/0000-0001-9674-8440, Lawton, Tom et al. (8 more authors) (2024) Development and Translation of Human-AI Interaction Models into Working Prototypes for Clinical Decision-making. In: Vallgarda, Anna, Jonsson, Li, Fritsch, Jonas, Alaoui, Sarah Fdili and Le Dantec, Christopher A., (eds.) Proceedings of the 2024 ACM Designing Interactive Systems Conference, DIS 2024. 2024 ACM Designing Interactive Systems Conference, DIS 2024, 01-05 Jul 2024 Proceedings of the 2024 ACM Designing Interactive Systems Conference, DIS 2024 . Association for Computing Machinery, Inc , DNK , pp. 1607-1619.

https://doi.org/10.1145/3643834.3660697

Development and Translation of Human-AI Interaction Models into Working Prototypes for Clinical Decision-making

Muhammad Hussain*
Computer Science, University of York,
York, United Kingdom
muhammad.hussain@york.ac.uk

Ioanna Iacovides
Computer Science, University of York,
York, United Kingdom
jo.iacovides@york.ac.uk

Tom Lawton
Improvement Academy, Bradford Institute for Health Research,
Bradford, United Kingdom
tom.lawton@bthft.nhs.uk

Vishal Sharma
Improvement Academy, Bradford Institute for Health Research,
Bradford, United Kingdom
vishal.sharma@yhia.nhs.uk

Zoe Porter
Computer Science, University of York,
York, United Kingdom
zoe.porter@york.ac.uk

Alice Cunningham
Improvement Academy, Bradford Institute for Health Research,
Bradford, United Kingdom
alice.cunningham@yhia.nhs.uk

Ibrahim Habli
Computer Science, University of York,
York, United Kingdom
ibrahim.habli@york.ac.uk

Shireen Hickey
Improvement Academy, Bradford Institute for Health Research,
Bradford, United Kingdom
shireen.hickey@yhia.nhs.uk

Nee Ling Wong
Improvement Academy, Bradford Institute for Health Research,
Bradford, United Kingdom
Nee.Wong@bthft.nhs.uk

Yan Jia
Computer Science, University of York,
York, United Kingdom
yan.jia@york.ac.uk

Phillip Morgan
York Law School, University of York,
York, United Kingdom
phillip.morgan@york.ac.uk

ABSTRACT

In the standard interaction model of clinical decision support systems, the system makes a recommendation, and the clinician decides whether to act on it. However, this model can compromise the patient-centeredness of care and the level of clinician involvement. There is scope to develop alternative interaction models, but we need methods for exploring and comparing these to assess how they may impact clinical decision-making. Through collaborating with clinical, AI safety, and HCI experts, and patient representatives, we co-designed a number of alternative human-AI interaction models for clinical decision-making. We then translated these models into 'Wizard of Oz' prototypes, where we created clinical scenarios and designed user interfaces with different types of AI output. In this paper, we present alternative models of human-AI interaction and illustrate how we used a co-design approach to translate them into functional prototypes that can be tested with users to explore potential impacts on clinical decision-making.

DIS '24, July 01–05, 2024, IT University of Copenhagen, Denmark

ACM ISBN 979-8-4007-0583-0/24/07
https://doi.org/10.1145/3643834.3660697

CCS CONCEPTS

• **Human-centered computing**; • **Interaction design**; • **Interaction design process and methods**; • **Interface design prototyping**;

KEYWORDS

Human-AI interaction, healthcare, decision-making, prototyping, Wizard-of-OZ

ACM Reference Format:
Muhammad Hussain*, Ioanna Iacovides, Tom Lawton, Vishal Sharma, Zoe Porter, Alice Cunningham, Ibrahim Habli, Shireen Hickey, Nee Ling Wong, Yan Jia, and Phillip Morgan. 2024. Development and Translation of Human-AI Interaction Models into Working Prototypes for Clinical Decision-making. In *Designing Interactive Systems Conference (DIS '24), July 01–05, 2024, IT University of Copenhagen, Denmark*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3643834.3660697

1 INTRODUCTION

Artificial Intelligence (AI)-based solutions have become prevalent in several societal sectors over the past ten years. For instance, recent developments in machine learning are speeding up the integration of collaborative human-AI decision-making tools, particularly in safety-critical domains like healthcare [10, 37]. In clinical practice, Clinical decision support systems (CDSS) are electronic or non-electronic systems designed to support clinicians directly during decision-making, in which various characteristics of an individual patient are used to generate patient-specific considerations and recommendations that are then presented to clinicians [38]. AI-based

CDSS are the tools aimed to assist clinicians with diagnosis and treatment decisions by employing AI models trained on data from patients relevant to the specific use case, in contrast to traditional CDSS that match patient characteristics to an established knowledge base [10, 51, 69]. CDSSs assist clinicians through different modes, including alert systems, monitoring systems, recommendation systems, and prediction systems [40]. This paper is focused on recommendation CDSS systems. Research is being conducted on improving the recommendation systems for drug prescriptions [42, 45], laboratory tests [34], and polypharmacy management [52], as well as for diseases such as diabetes [3], and rare diseases [16]. Despite the research occurring to develop and improve different AI algorithms and explanations for decision support systems [9, 41, 62, 81], several issues related to the human-computer interaction (HCI) domain are currently contributing to the low adoption of CDSSs. These include a lack of patient-centred human-AI (HAI) interaction models [7], a lack of human-centred design for AI systems [78], issues around trust and reliance [9], levels of autonomy and clinician concerns about liability[43].

Recently, there has been also a push to provide guidance for designing AI systems. Big tech companies like Google [60], IBM [32], Microsoft [13], Amazon [76], and Facebook [25] have released white papers with advice on creating AI systems and conversational AI. To create systems that are fit for purpose, a clear understanding of people, AI systems, and their interaction in a particular environment is required. Human-centred AI (HCAI) can provide an adequate framework and mindset to achieve this goal. HCI professionals should explore how people and AI work together, understand human-AI collaboration, and consider users' perspectives when developing AI systems [29, 53].

The standard human-AI interaction model within CDSSs involves the AI system making a recommendation, and the human clinician deciding whether to act on it—a kind of 'sense check' on the machine [43]. However, there are concerns with this model. First, many AI systems have little room to incorporate patient beliefs and values [7], potentially compromising patient-centeredness in clinical practice. These issues raise ethical concerns related to patient autonomy [31, 48] and contradict the NHS's ethos of supporting patients to be actively involved in their care [56]. Secondly, the approach may not be good for clinicians, potentially disenfranchising them from the decision-making process and turning them into mere safeguards on the AI system. Such issues can contribute to a culture of treating clinicians using AI as 'liability sinks,' potentially holding them legally responsible for outcomes without providing them with sufficient control or understanding of the AI system [43]. Thirdly, the standard model may not be beneficial for the whole system, including healthcare providers and regulators, as trust and responsibility issues could contribute to the low adoption of AI in healthcare.

To overcome these issues, different HAI models for CDSSs need to be explored and then tried out with users to understand their potential impact on decision-making. While user experience (UX) methodology and design prototyping literature do not always specify how to test AI-based experiences [20, 49], one cost-effective way is through utilising the 'Wizard of Oz' method. The Wizard of Oz method is ideal for testing AI experiences as researchers can simulate a system's computations and gather feedback early in the development process, without requiring a fully functional system. To overcome challenges in developing prototypes for human-AI interaction [82], it is crucial to ensure close collaboration between various stakeholders by adopting a co-design approach.

In this research, we have developed a number of potential HAI interaction models for CDSSs to investigate their impact on clinical decision-making by collaborating with Clinical experts, AI safety experts, HCI experts and a patient involvement panel. A co-design approach was used to translate the developed HAI models into 'Wizard of Oz' prototypes and to develop a set of clinical scenarios within Diabetes and Obstetrics cases. In this paper, we present the HAI interaction models and discuss our approach to translating them into working prototypes involving an Obstetric scenario that could then be tested with end users to examine how the different models may influence clinical decision-making. We contribute to discussions about how to co-design human-centred CDSS systems and how we can explore the impact of different human-AI configurations on users in the early stages of development.

2 RELATED WORK

2.1 Clinical Decision Support Systems

In recent years, artificial intelligence techniques like neural networks [28] and knowledge graphs [74] have contributed to the growth of AI technologies to enhance CDSSs performance and accuracy [59]. These new AI methods, along with access to extensive clinical datasets, are changing the way CDSSs work, moving from rule-based approaches to a more data-driven approach [17, 70]. It is expected that these new generations of CDSSs utilising AI technologies may become more prevalent than traditional methods for generating clinical decision support rules [50]. Some of them have already shown capabilities similar to domain experts in various clinical decision-making tasks, such as cancer diagnosis in histopathology [14], detection of diabetic retinopathy [24], or assessment of X-ray scans for conditions like pneumonia [68].

AI technology has brought many benefits to people's lives and work, but its development is often technology-centred rather than user-centred [64, 65]. This approach has led to failures in many AI systems [79], resulting in more than 2000 AI accidents reported by the AI incident database, such as autonomous cars hitting pedestrians, Alexa recommending a dangerous TikTok challenge to a ten-year-old girl, or trading algorithms causing market crashes [75]. As AI-based decision-making systems become more prevalent in industries and government services, skewed decisions based on data biases can directly impact people's lives, potentially causing harm. Sectors like healthcare, which heavily rely on accurate recommendations, are particularly vulnerable to such incidents.

Combining humans and AI in collaborative decision-making is thought to enhance decision quality in healthcare [26]. However, recent studies reveal that trust is a significant concern, which can lead to the acceptance of incorrect recommendations (due to too much trust) or the rejection of correct ones (due to too little) [35]. Explainability helps humans understand AI recommendations, promoting trust and control [30]. It assists users in deciding when to accept or reject AI recommendation and is crucial for establishing accountability [15, 80]. Previous studies have considered different types of explanations for AI-based CDSSs. These include Local explanations [27], Counterfactual explanations [67], Example-based explanations [11], and Global explanations [77]. Local explanations justify the model's reasoning by focusing on specific decisions that have been made in relation to a particular user. Counterfactual explanations, adopt a 'what-if' approach by providing alternative outputs related to different input values. Example-based explanations support AI decisions by providing real-world examples from the dataset that share similar characteristics with the input data. Finally, Global explanations provide a holistic understanding of the model's logic by revealing the relative importance of different features in the decision-making process. Example-based and Counterfactual explanations were considered understandable but time-consuming for clinicians, who often lack the time to study them in detail and may contain biases. Conversely, Local and Global explanations required additional design considerations and interactive approaches to make them usable for end-users. The user interfaces for these explanations were also not easily understandable, where information was presented through tables and graphs [54]. There is a need to explore alternative explanations for AI recommendations that are easy for clinicians to understand and tailored to their decision-making needs. In addition, it will be important to explore how different types of clinicians affect their level of trust in the system.

Many advanced AI-based CDSS systems such as recommender systems are still in the developmental phase. Due to the limited scale of deployment, uncertainties persist regarding how such systems are perceived and utilised by their intended users, such as clinicians. Additionally, there is a need to further understand the existing barriers and challenges in their deployment, to determine optimal levels of involvement with respect to clinicians, patients, and AI in decision-making, and to identify effective models of human-AI interaction. Gaining a thorough understanding of these aspects will assist the designers and developers of AI-CDSS in comprehending and addressing potential issues in integrating AI-CDSS into clinical practice. Our study aims to contribute to this knowledge gap by proposing potential human-AI interaction models and translating them into working prototypes for further testing.

2.2 Challenges in the implementation of AI-CDSS

In order to explore different forms of human-AI interaction, it is necessary to consider both challenges and opportunities for enabling human-centred AI that is fit for purpose. Seven main issues for human interaction with AI are highlighted within the HCI community [78]. Firstly, the potential for unexpected machine behaviour and biased outcomes, which may evolve as the machine learns [61]. Secondly, the limitations of machine intelligence, which cannot fully replicate advanced human cognitive capabilities [84]. Thus, there is a need to find ways to integrate human role into AI to ensure human-controlled decisions [83]. Thirdly, the autonomous nature of machines, which may handle some operational situations not fully anticipated [57]. Fourthly, the ongoing debate surrounding human-AI collaboration, addressing how to work with AI as a teammate and establish a collaborative relationship [12, 57]. Fifthly, the explainability of AI output, as AI systems may exhibit a 'black box' effect, obscuring the reasoning behind their decisions and causing users to question when to trust AI [53]. Sixthly, the design of user interfaces for AI systems, ensuring they are natural and adaptive to human capabilities [1]. Lastly, ethical considerations such as privacy, fairness, decision-making authority, and responsibility [43, 85].

AI-based decision support systems in healthcare face initial challenges that require thorough exploration. Some crucial aspects include understanding their integration into existing healthcare structures, considering user differences, addressing potential biases, identifying the contribution of both human and AI actors, integrating human factors considerations, as well as highlighting the importance of physical patients and their data representation [73]. Given that most AI systems in healthcare function as complex interventions designed for clinical decision support rather than autonomous agents, the interactions between the AI systems, their users, and the implementation environments play a crucial role in determining the overall potential effectiveness of these AI interventions. Therefore, moving AI systems from achieving mathematical performance to practical clinical utility requires a carefully phased implementation and evaluation approach. This approach should consider the complexities involved in the collaboration between two distinct forms of intelligence, going beyond measures of effectiveness alone [66]. Despite indications that some AI-based algorithms now have the same accuracy to human experts in preclinical in silico studies [2], there is limited high-quality evidence demonstrating improved clinician performance or patient outcomes in clinical studies [23, 71]. Reasons suggested for this gap include a lack of necessary expertise for translating a tool into practice, insufficient funding available for translation, a general underappreciation of clinical research as a translation mechanism, and, more specifically, a disregard for the potential value of the early stages of clinical evaluation and the analysis of human factors [47, 72].

Within the domain of AI research in healthcare, there is a lack of information about the development process of human-AI interaction models and their translation into prototypes, leaving readers curious about how models are created, refined, and implemented, though clinicians have some involvement but it's not clear how they were involved in the process [7]. One possible method for exploring how to translate different models is the Wizard of Oz technique, as it allows for quickly producing different interfaces without having to develop fully functional AI technologies. By using this method, user feedback can be gathered at an earlier stage, helping to identify the most promising models and functionalities to pursue. For example, in automotive research, this approach has been used to understand user preferences for future in-vehicle interfaces in automated vehicles [19], and to develop prototypes for on-road evaluation of futuristic human-machine interfaces (HMIs) [22]. However, it's worth noting that such practices are less common in healthcare research, where the complexities of patient care and regulatory considerations may pose unique challenges.

As HCI professionals, we navigate not only the conventional 'interaction' between humans and machines but also new forms of human-machine relationships that may not exist yet. Considerable research has explored human-machine teaming, indicating that humans and AI systems may be more effective when operating collaboratively as a unified unit rather than as individual entities

[5, 6, 18]. It has been argued that the more intelligent the AI system, the greater the need for collaborative capabilities [36]. There are discussions about whether AI systems can effectively work as teammates with humans [63]. People should not have to adjust to non-human 'teammates'; instead, designers should develop technology to act as a cooperative team player (or a highly effective tool) alongside humans [1, 63]. To preserve the autonomy of human actors, we believe that a shared design goal in human-AI interaction should be focused on.

There is a lack of human factors and HCI studies that investigate how best to support communication and collaboration within shared human-AI decision-making systems in healthcare. To find the most suitable interaction model we need to explore different human-AI configurations. Yet, questions remain about the different ways in which AI technologies can be implemented into clinical practice, the optimum level of involvement for both human and AI actors, and how any recommendations or explanations should be presented to clinicians. Thus, there is a need to develop different models of human-AI interaction, and also to test these models through translating them into prototypes that can be used with clinicians to explore how they influence clinical decision-making.

In this paper, we aim to address this gap by proposing different human-AI interaction models that vary according to the level of clinician, AI, and patient involvement. To enable the translation of these human-AI interaction models into actual systems, we collaborated with clinical experts and adopted a Wizard of Oz approach to create interactive prototypes based on those models. The design prototypes will not only assist developers and organisations in exploring different ways to implement AI solutions in healthcare but will also facilitate the study of clinical decision-making and the different factors that may influence human-AI interaction.

3 CO-DESIGN METHOD

Our project aims to evaluate a range of models for Human-AI Interaction (HMI) for clinical decision support, to explore their impact on shared decision-making. The project team includes experts with diverse backgrounds: 2 from HCI, 2 specialising in AI safety, 3 from clinical fields, 1 focused on AI ethics, 1 in Law, and 2 in Psychology. The team members have a track record of working together on projects exploring the safety and accountability of AI systems in healthcare, making use of a broad cross-disciplinary base that considers clinical as well as ethical, legal, engineering, human-computer interaction, and safety angles. To create working prototypes, we used the 'Wizard of Oz' prototyping method, where participants are made to believe that they are acting with a real system, but instead, the experimenter acts as the 'wizard', a proxy for the system behind the scenes [39].

We have divided the process into three phases. In Phase 1, we developed a set of alternative Human-AI Interaction (HAI) models and initiated considerations for potential cases and scenarios. Moving to Phase 2, we created six clinical scenarios related to Diabetes and Obstetrics. that were to be integrated into each model. In Phase 3, we translated the models and scenarios into working prototypes, designing an interface resembling a real Electronic Patient Record (EPR) system. Participants from both within the project team and outside of it contributed throughout all three phases, the specifics of

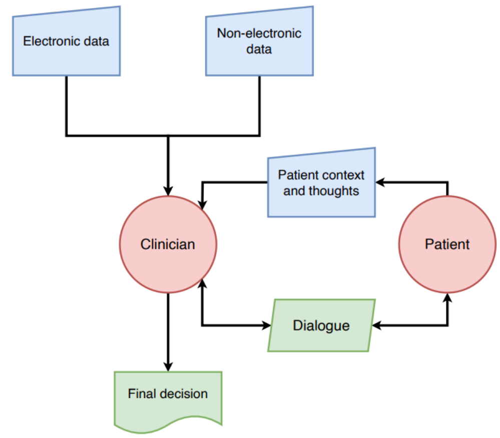


Figure 1: Model - 1: Traditional model (No-AI).

which are detailed in Table 1. The details of each phase are detailed below,

3.1 Phase 1: HAI interaction models

In this phase, different HAI interactive models were developed through a focus group workshop led by clinical experts. Various topics concerning AI decision support systems were considered, including objectives, target users, primary goals, tasks, functionality, and guidelines for AI systems. The workshop particularly emphasised guidance issued by the National Health Service (NHS) in England regarding decision-support systems [4]. The team started with Model 1- the traditional (no-AI) model (Figure 1), which describes the current patient-clinician interaction model, without any involvement of AI. Model 2 is the standard recommendation model that reflects the NHS guidance [4], emphasising that the final decision should be made by a healthcare professional (Figure 2). However, the standard recommendation model could potentially adversely affect clinicians who are facing numerous cognitive and practical challenges while monitoring automation such as facing a binary choice of either accepting the AI recommendation or ignoring it to revert to a traditional (no-AI) approach.

As illustrated in Figure 2, the model involves the AI system making a recommendation, which the human clinician reviews and decides whether to act upon. As explained earlier, this setup may limit the incorporation of patient beliefs and values, potentially compromising patient autonomy and clinician involvement in decision-making. One of the potential consequences of this is that clinicians are likely to become legally responsible for outcomes but may not have adequate understanding or control of the AI system [43].

Considering these challenges, several alternative models were developed by the team for further discussion which aimed to explore potential solutions that could incorporate both clinical and patient perspectives in AI decision-making. The goal was to develop models which were clinically realistic, achievable, actionable, and based on the NHS guidance towards AI [4]. In some of these models, the AI's output may not necessarily be a decision or recommendation; instead, it presents the data that underlies a potential

Table 1: Participants and their contribution in each phase

Phases	Project Team	External contributors
Phase 1 (Development of HAI models)	Whole project team, led by clinical experts to develop HAI models	
Phase 2 (Development of Clinical scenarios)	3 Clinical experts 2 HCI experts Reviewed by the project team	10 Clinicians - further scenario refinement (6 for Diabetes, 4 for Obstetrics) 5 Member Patient Panel - provided feedback on scenarios and how patients might respond in each scenario
Phase 3 (Development of prototypes)	2 HCI experts 3 Clinical experts Reviewed by the project team	4 Clinicians – validated prototypes

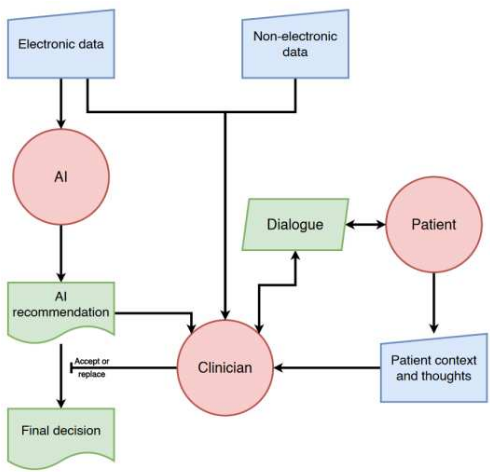


Figure 2: Model - 2: Recommendation-only model.

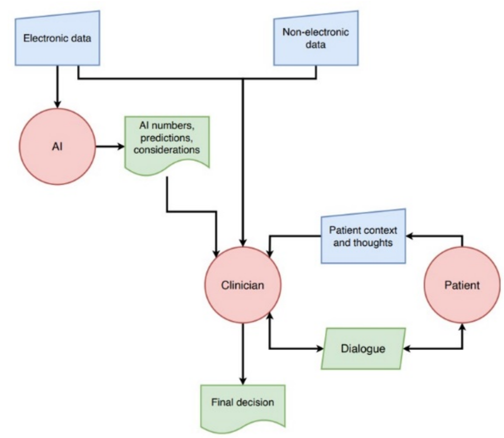


Figure 3: Model - 3: Underlying data model.

recommendation so the clinician can review this. In addition to Model 1 (Traditional non-AI model) and 2 (Recommendation only model), four alternative models were created, each varying in terms of how the clinician, patient and the AI system interact with each other when making a decision. The alternative models are the Underlying Data model (Figure 3), Recommendation Plus model (Figure 4), Patient-AI model (Figure 5), and Patient-AI-Clinician model (Figure 6). A detailed description of each model is provided below.

- Model - 1: Traditional model (No-AI)

Model-1(Figure 1) is a traditional non-AI approach, where the clinician independently makes decisions by reviewing both electronic and non-electronic data. The clinician engages in a dialogue with the patient to understand the patient's context and thoughts before making a decision about treatment.

- Model - 2: Recommendation-only model.

Model 2 (Figure 2) is the current standard recommendation model for clinical decision support systems, where the AI makes a specific treatment recommendation based on electronic data. The clinician has access to the same electronic data, as well as non-electronic data that usually comes from a dialogue with the patient providing more information about their thoughts and context. The clinician has to choose whether they will accept the AI recommendation or replace it with another form of treatment.

- Model - 3: Underlying data model.

In this model (Figure 3), the AI doesn't provide a specific treatment recommendation to the clinician. Instead, it offers a summary of relevant underlying data in the form of numbers, predictions, and other risk-related considerations specific to the patient. The clinician then makes the final treatment decision considering the AI output, non-electronic data and the patient's context and thoughts that are elicited through a dialogue.

- Model - 4: Recommendation Plus model.

Model 4 (Figure 4) combines features from both Model 2 and 3, where the AI provides a specific treatment recommendation in combination with the numbers, predictions, and other risk-related considerations that underlie that recommendation. Again, the clinician makes the final decision considering the AI output, non-electronic data and the patient's context and thoughts that are elicited through a dialogue.

- Model - 5: Patient-AI model.

In this model (Figure 5), an extra feature is introduced to the AI, allowing it to incorporate patient values and preferences. To

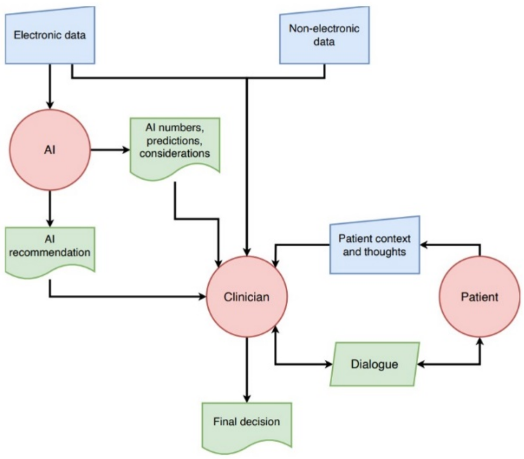


Figure 4: Model - 4: Recommendation Plus model.

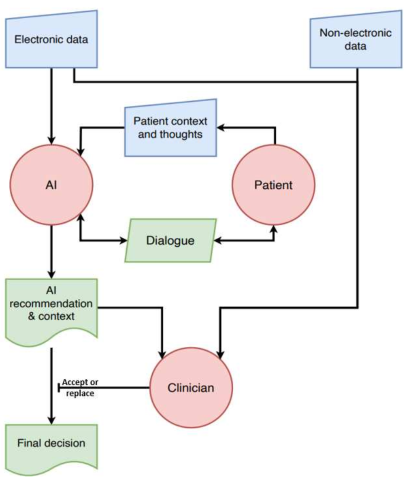


Figure 5: Model - 5: Patient-AI model.

improve efficiency, the patient engages in a dialogue with the AI but not the clinicians. The AI then provides a specific treatment recommendation, as well as the underlying numbers, predictions, and other risk-related considerations specific to the patient. The clinician then decides whether to accept or reject the AI recommendation after viewing the recommendation, the underlying data, and the AI dialogue with the patient. To simulate conversations between AI and patients for Model 5 and 6, we utilised roleplay, a method commonly employed in HCI [46].

- Model - 6: Patient-AI-Clinician model.

This model (Figure 6) is similar to model 5, but the difference lies in how the chat and context are handled, as the clinician can now also have a direct dialogue with the patient. Again, they are also provided with the recommendation, the underlying data, and the AI dialogue with the patient before they make a decision about treatment. In both models, patients engage in a conversation with the AI, but they differ in terms of the conversation's purpose, with one

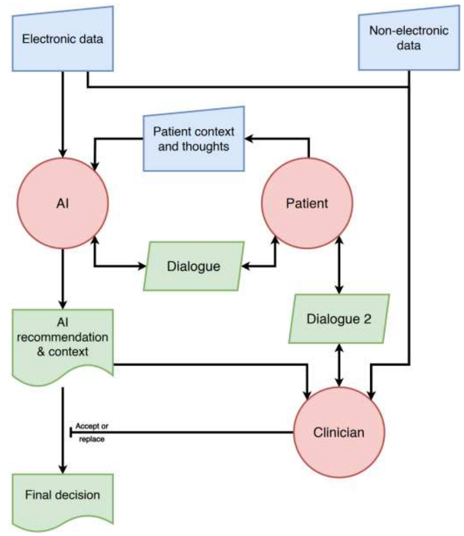


Figure 6: Model - 6: Patient-AI-Clinician model.

guiding towards a decision (Model 5) and the other solely providing information and autonomously generating a recommendation without informing the patient through the chat (Model 6).

3.2 Phase 2: Clinical Scenarios

In order to progress towards translating the model into interfaces that could be used with clinicians, we also needed to decide on potential use cases. Six scenarios were simulated, three for Diabetes, and three for Obstetrics. In consultation with the wider project team, three clinicians developed the scenarios based on their clinical knowledge and experience. To ensure the scenarios were realistic, a focus group was held with five patient and public involvement (PPI) panel members. The focus group participants, based on their experience, provided feedback on the possible thoughts, feelings, and questions a patient in each scenario might experience. Additionally, the group provided their view on how the AI should provide output, and who they felt was responsible for the final decision. This feedback was used to refine the scenarios and also provided useful information for later stages in the project when the prototypes would be used by real clinicians and patient actors.

The scenarios included patient background information, information from previous consultations, medical history, medication details, test results, and observations, suitable treatment options, AI treatment responses, patient preferences, potential treatment consequences, and transcripts of conversations between the patient and the AI system (for Model 5 and 6). The scenarios were all loosely based on actual cases, with modifications made to ensure patient confidentiality. To further enhance the scenarios, a final validation stage involved presenting them to ten additional clinicians—six specialising in Diabetes and four in Obstetrics. This step aimed to ensure the scenarios felt real and natural and were suitable to be implemented in the prototyping phase.

The AI responses for each model and scenario were also simulated by the same clinicians from the team who developed the scenarios. These responses considered both National guidance and

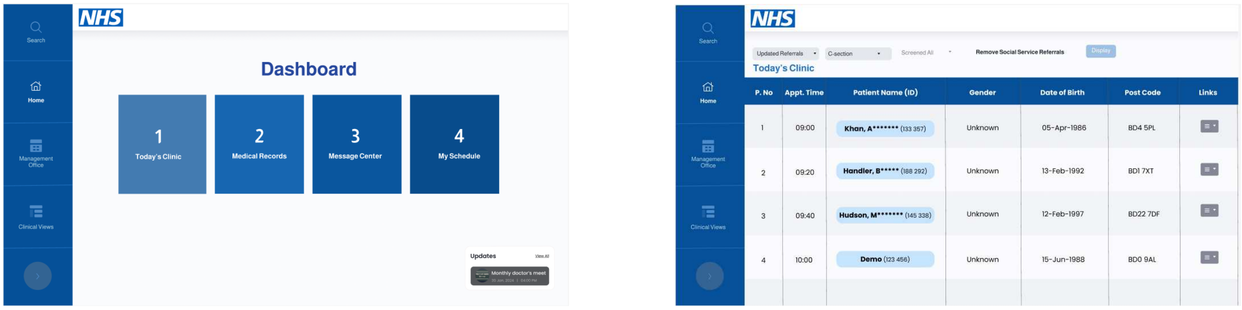


Figure 7: Dashboard and patient list interface of the prototype designed with Figma.

clinical experience. The patient-AI chat was prepared by clinicians in the style of ChatGPT. For the Diabetes scenarios, the scenarios focused on whether a patient would need to start taking insulin or not, while the Obstetrics scenarios, focused on whether a patient should undergo a caesarean section or not. In addition, for each domain, we developed three types of scenarios: (1) with a clear decision where a clinician was likely to agree with the AI output; (2) with a clear decision, but the AI recommendation would differ from what a clinician would likely decide; and (3) a less clear-cut decision, where either course of action would generally be considered reasonable. Through these variations, we aim to explore how clinicians would respond to different AI outputs under a range of circumstances.

3.3 Phase 3: Prototype development

In this phase, we translated the above models and scenarios into working prototypes by using the Wizard of Oz method. The Wizard of Oz approach strikes a balance between fidelity and feasibility, leading to robust and user-centric prototypes. We chose this approach for prototype development due to its ability to simulate realistic user interactions that would allow us to gather insights into valuable clinician perspectives and explore their reactions to different implementations of AI much earlier on in the development process. This approach allows us to consider different human-AI configurations before requiring fully functioning AI systems. While some concerns have been raised about the user of Wizard of Oz e.g., within human-robot interaction relating to the need for deception [55], we argue that some level of deception is acceptable in order to gain realistic responses from users when a fully functioning system has not been developed, and as long as participants are fully debriefed afterwards.

The goal was to design an interface that closely resembled real-life EPR systems used in the UK, with an additional AI plug-in for models 2-6. The HCI experts examined several EPR interfaces and consulted clinical experts in the project for their input on design and system requirements. Starting with sketching, the initial design underwent multiple revisions based on feedback from clinicians in project workshops. Once the design was agreed upon, the HCI experts crafted an interactive prototype using Figma [21], refining it several times based on feedback from the clinicians involved in the project.

The final design of the interactive prototype is provided below, featuring an illustration from an Obstetrics scenario. In Figure 7, the Dashboard exhibits a familiar interface for clinicians. They will navigate to the 'Today's Clinic' section to proceed to the next page, presenting a list of four patients. In this example, the list comprises Obstetrics patients according to the three scenarios and a demo scenario.

Clicking on the patient's name (in this case, 'Khan') directed them to the next page (Figure 8 - left), featuring patient information and a visit summary. The patient information section contains details such as Name, Allergies, Age, Ethnicity, Gender, Patient ID, Date of Birth, and other relevant information. The visit summary section is categorised into four expandable sections: 'Previous Consultation,' 'Medical History,' 'Results and Observations,' and 'Medication.' By clicking the 'Expand' button, a pop-up overlay appears, revealing the entire section, as illustrated in Figure 8 - right (Detailed summary of Khan). On the right side of the Visit Summary, there is a button for an AI-based Decision Support System (Shared CAIRE), which produces AI-based advice for each scenario.

For the prototype based on Model 1, no AI tool was present on the interface as there was no involvement of AI in decision-making. The purpose of this prototype was to allow for comparison with the other prototypes which did involve AI. For the remaining five prototypes, different outputs were generated according to the underlying HAI interaction model format, discussed above. These are static outputs, developed by clinicians within the project team (as described in phase 2), which are presented to the users as information to consider as part of their decision-making process. The AI output for patient Khan according to each prototype is given below (Figure 9, 10, 11, 12, 13).

Model 2 is a standard recommendation model where AI makes a specific recommendation. For this prototype, a recommendation is provided, after the clinician clicks on the AI tool - Shared CAIRE. Thus, the user interface is designed to only show the recommended treatment to the clinician based on electronic data (Figure 9). In this scenario, the recommended treatment from AI is 'Vaginal Birth after Caesarean' without any further details.

For the Model 3 prototype (Figure 10), the interface is designed to provide a summary of underlying data in the form of numbers, predictions, and other risk-related considerations to the clinician. This data is what would have been used as the basis of a recommendation, but in this prototype, a specific recommendation is not

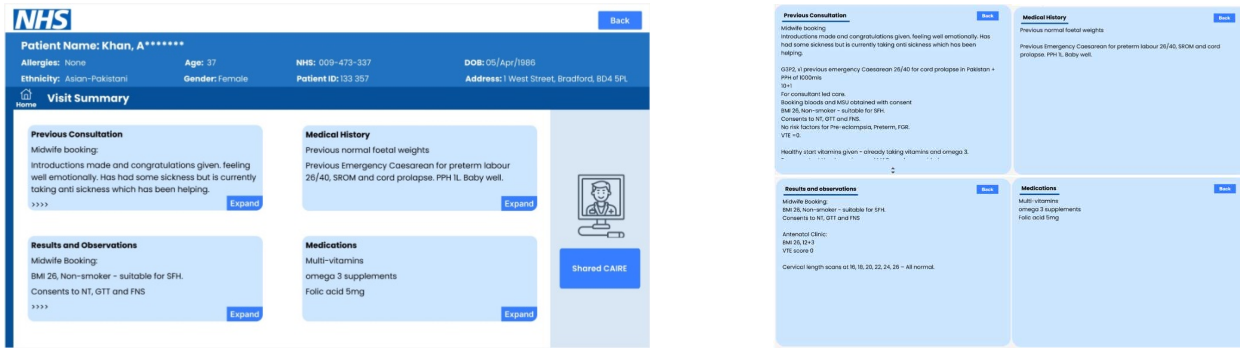


Figure 8: User interface for patient information, visit summary and AI decision support tool (Left). The popup screens show simulated insights of the patient for each expandable button in the visit summary (Right).

Figure 9: Simulated AI output according to Model 2 (Recommendation-only model)

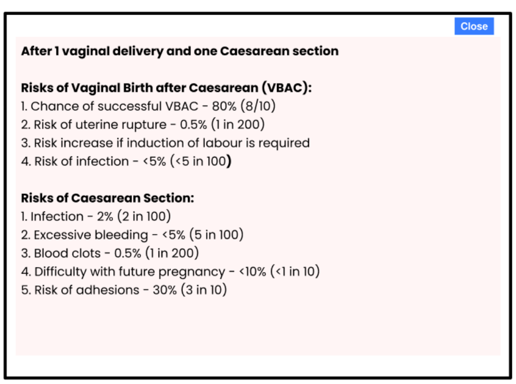


Figure 10: Simulated AI output according to Model-3 (Underlying data model)

provided. For the same scenario, it shows the important factors to consider along with the risks of vaginal birth after caesarean and risks of caesarean section in the form of numbers and percentages.

In the translation of Model 4, the interface is designed to show the output of both the Model 2 and 3 prototypes. For the same

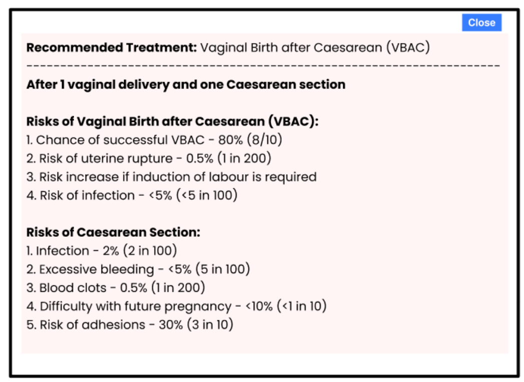
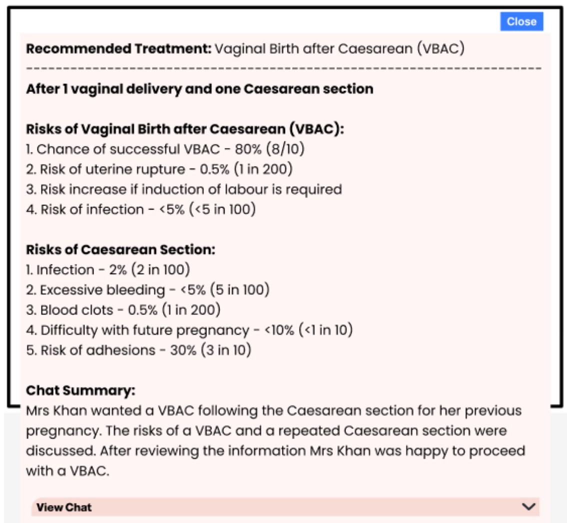


Figure 11: Simulated AI output according to Model-4 (Recommendation plus model)

patient scenario, it displays the recommended treatment, as well as the underlying factors to consider, which in this case are risks of vaginal birth after caesarean, and risks of caesarean section based on electronic data.

For the Model 5 and Model 6 prototypes, an extra feature is introduced to the AI, allowing it to incorporate patient values and preferences. We decided to adopt a ChatGPT-style dialogue between AI and patient. In both models, patients engage in a conversation with the AI, but they differ in terms of the conversation's purpose, with one guiding towards a decision like less-trained staff members such as Physician Associates (Model 5) and the other solely providing information like a medical student (Model 6). Based on the chat, the AI now has additional data to process related to patient preferences and values. The AI creates a summary of the chat to provide decision context for the clinician with the option to view the full chat.

For the prototype based on Model 5, the interface includes recommended treatment, factors to consider, risks of vaginal birth after caesarean, and risks of caesarean section based on electronic data, as well as a chat summary based on the patient's chat with AI, with the option to view the full chat. There is no direct interaction between patient and clinician, making the nature of the



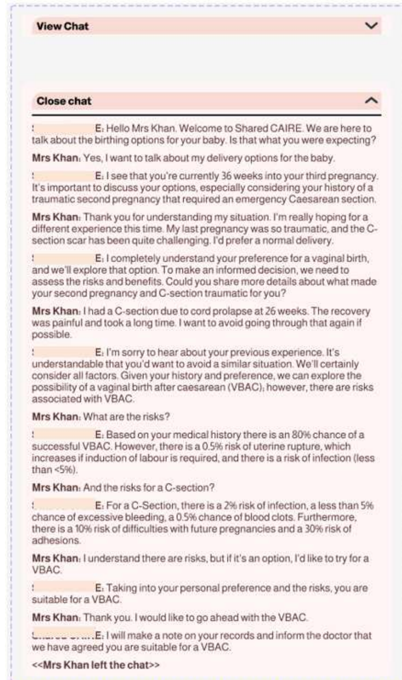


Figure 12: Simulated AI output according to model-5 (Patient-AI model) with an option to view full chat. Simulated full chat between AI and patient (right).

chat more decision-focused. As part of the conversation, the AI and patient agree on a suitable treatment based on the information provided by the AI and the patient's preferences. The AI creates a summary of the conversation detailing the agreed next steps, which is made available to the clinician to accept or replace the AI recommendation.

For the Model 6 prototype, the interface is similar to the Model 5 version but with an additional physical dialogue between clinician and patient to confirm the conversation and gather additional insights. Here, the nature of the chat between AI and patient before meeting the clinician is also different from the previous prototype, focusing on obtaining more information and insights regarding patient thoughts without agreeing on any particular treatment. The AI again creates a summary of the conversation with the option to view the full chat on the interface. This is made available to the clinician, who can accept or replace the recommendation with the option to have a dialogue with the patient.

During the development of the prototype, feedback was also obtained from four clinicians outside the project team about the layout, what information should be presented, where it should be presented, and how. The clinicians were from the following specialties, Palliative Care, Emergency Medicine, General Practice, and Oncology. The feedback was obtained from clinicians one at a time, where they were encouraged to think aloud whilst using the interface. The feedback suggested that they appreciated the consolidation of all information on a single interface, and understood how to use the systems, indicating the interface was suitable for use in clinical settings. The final version of the interface was agreed upon by the project team due to its simplicity, ease of use, and accessibility to relevant information from the patient record.

4 DISCUSSION

Artificial Intelligence is growing in both public and private healthcare worldwide. Its integration into healthcare, particularly in the context of decision support systems, poses a number of challenges including concerns from clinicians and patients. In this research, we have presented various potential human-AI interaction models for clinical decision-making and how we have translated them into working prototypes that can be used to explore the impact of different human-AI configurations on decision-making in healthcare.

The interaction model without AI involves clinicians making decisions based on electronic data and dialogue with patients, without AI involvement. However, as CDSSs advance to include AI-based recommendations to aid clinician decision-making, a standard model for human-AI interaction (Figure 2) appears to have been adopted. The standard model involves feeding electronic data into an algorithm, a machine-learned model, which generates a recommendation for a human clinician. Nevertheless, this model lacks patient-centeredness, raising concerns among clinicians about its usefulness. Additionally, the explainability of recommendations from such a model is problematic due to the black-box nature of AI [69]. Recognising the importance of exploring alternative human-AI interaction models alongside the need for accurate AI algorithms is crucial to ensuring that AI systems can be trusted and accepted by clinicians and patients. A crucial question in HCI is the focus of design for interaction (or mutual collaboration); specifically, who

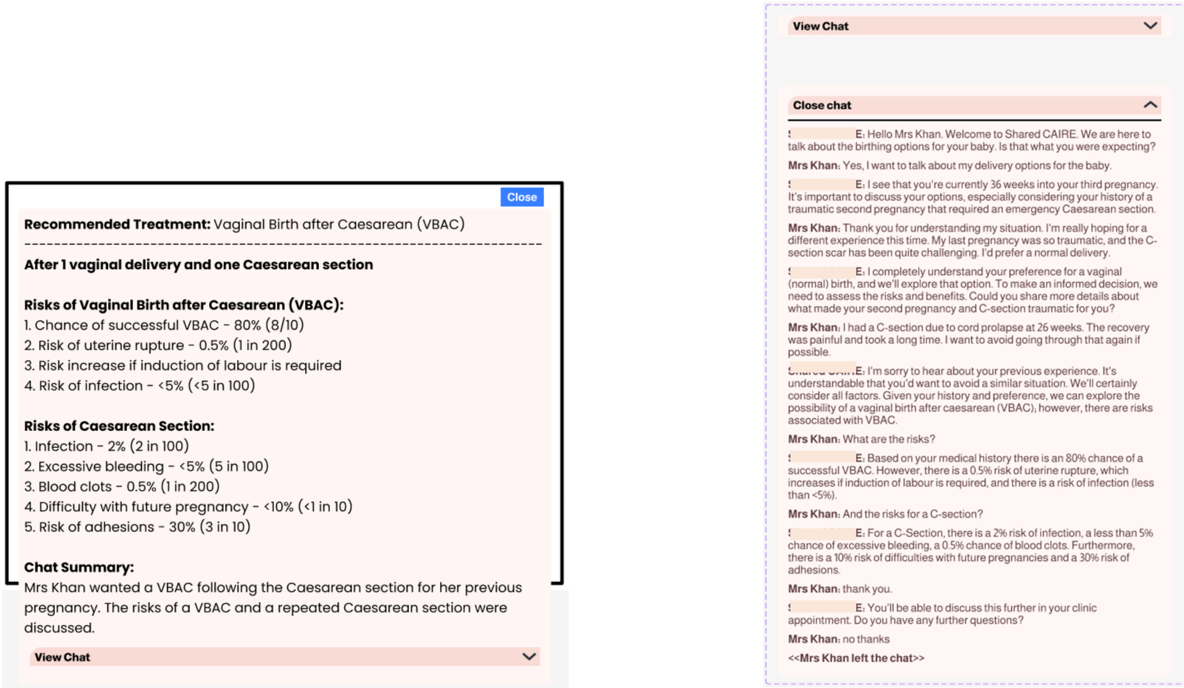


Figure 13: Simulated AI output according to model-6 (Patient-AI-Clinician model) with an option to view full chat. Simulated full chat between AI and patient (right).

holds the ultimate decision-making authority? It is essential to explore how we can guarantee that in complex decision-making systems, humans maintain a pivotal role in decision-making, as opposed to AI systems [33]. In this study, we propose four alternative human-AI interaction models, each differing in terms of recommendation style, incorporation of patient values and preferences, level of clinician and patient involvement, and explainability.

Previous research has proposed exploring interaction pathways that integrate patient values and attitudes about risk into AI recommendation systems for breast cancer screening [7]. However, there is a lack of clarity regarding the design process for such interaction models, including who was involved in the process and how to translate them into working prototypes or real systems. While there has been some previous research exploring interaction models [7], there is generally a lack of clarity about how to then implement these models into practice. In our approach, we describe how we developed human-AI interaction models in collaboration with stakeholders such as clinicians and then translated these models into prototypes. Adopting a Wizard of Oz approach, we could explore different implementations of AI without the risk of introducing them as actual systems in real-life workplaces.

Several methods can be employed to incorporate information about a patient's values, such as utilising a risk-profiling questionnaire to explore their attitudes toward various outcomes, a common practice in finance [58]. Another way could be the risk assessment questionnaire, but these closed-ended questionnaires may limit patients from openly expressing their thoughts and preferences. With the introduction of ChatGPT, the concept of employing a chat or dialogue format emerged as an innovative way to integrate patient values into AI decision-making which will give the patient an option to write openly about their thoughts and preferences. This approach can assist HCI, AI, and clinical researchers in testing them in clinical settings before progressing to the implementation stage of AI-based decision-making. There is also a lack of clarity regarding the clinical insights generated by AI that are deemed useful and actionable for clinicians within the human-AI interaction models. In this research, we used a conceptual prototype for AI-generated insights in Obstetrics tailored to each human-AI interaction model through collaboration with clinicians specialising in Obstetrics. The AI output was based on what AI can generate and can be helpful for clinicians to achieve their goal of making a suitable decision which is also one of the definitions of explainability for such systems. We have also tried to address transparency when integrating patient feelings and preferences into AI, by giving an option to view full chat to the clinicians on the interface.

In the prototype for Model 2, treatment recommendations were made to clinicians based solely on electronic data, lacking additional information. The prototype for Model 3 was designed to provide clinicians with underlying data in numerical form and other risk-related considerations to aid in decision-making. This model aimed to give clinicians more autonomy in their decision-making process without feeling pressured into a specific decision. The interface for Model 4, on the other hand, involved AI making specific treatment recommendations along with numerical risk considerations. Although less complex, Model 4 provided clinicians with more information and recommendations, allowing for the clinician a chat interaction with patients to incorporate their thoughts and preferences. However, a common drawback of Model 2-4 is that AI outputs are solely based on electronic data without considering patient preferences. The explanation of AI-specific treatment recommendation is justified through the numbers and risk-related considerations which is different from the justification made through previously used explanations including Local explanations [27], Example-based [11], counterfactual [67] and Global explanations [77].

With Model 5 and 6, we aim to reflect the two advanced ways that AI might be used in the future in healthcare where the AI directly communicates with the patient before making a recommendation. Neither are necessarily right or wrong, but they have a different emphasis. Model 5 focuses on enhancing efficiency, where the clinician has potential to see more patients, as the AI has spoken to the patient directly and makes a recommendation that is also based on this data. Clinicians are simply checking the AI output, speeding up the process, though in a small proportion of cases they may need to see the patient themselves. This is similar to care models currently being tested with less-trained members of staff such as Physician Associates. Conversely, Model 6 focuses on enhancing effectiveness, with clinicians still doing what they're best at through directly interacting with the patient- but the AI has also been able to engage with the patient directly and use this information to make a recommendation. After speaking to the AI, the patient is better informed and has all the time they need to discuss things before having a standard consultation with the doctor. The doctor also should benefit from the AI managing to bring out useful bits of information. This process has parallels with the way a medical student might operate in a clinic, presenting the patient to the doctor before the doctor has a consultation - but being unable to make the final decision themselves. In practice, the patient may like the extended contact (or they may see it as a waste of time prior to the 'real' consultation!). Explanations for AI recommendations considering AI patient preferences are expected to increase clinician's trust.

4.1 Future Work

To unpack the effects of these different models, our future plan involves qualitatively evaluating these with clinicians and patient actors within role-play simulations to gain insights from a clinical perspective, and also with respect to potential legal concerns. We have recently concluded 21 evaluation sessions with clinicians in Diabetes and Obstetrics (who each engaged with three different combinations of the scenarios and prototypes described in this paper). Clinicians were observed interacting with the prototypes and patient actors, before being interviewed afterwards (and debriefed). Using Thematic Analysis [8], we will delve into participant responses, to explore clinicians' thoughts and concerns regarding AI-based decision-support tools within the different scenarios. We aim to understand how different AI outputs and human-AI interaction models influence their decision-making processes, contributing to an in-depth understanding of human-AI interaction in healthcare. While we will again engage with a patient panel to share our findings, further research will be needed to more explicitly consider the patient perspective within the context of AI decision support. Such research will also be particularly important for prototypes based on models such as 5 and 6, which require the patient to directly engage with AI as part of the consultation process.

5 CONCLUSION

We want to understand how best to support clinicians when interacting with decision-support tools in healthcare. However, there are questions about the different ways in which human-AI interaction can occur, including the role of AI in the decision-making process, and with respect to how to present information to clinicians. So, we need to develop ways of exploring these different implementations. In this research, we developed several potential human-AI interaction models through collaboration with clinical experts, AI safety experts, patients, and HCI experts. The challenge of translating these models into working prototypes for further testing was also addressed in this paper via co-designing clinical scenarios and the user interface using a Wizard of Oz method for two cases: Diabetes and Obstetrics. We explained the method for developing interaction models for such AI systems and how to translate them into working prototypes by involving end users. With this approach, different prototypes can be developed to facilitate the exploration and application of various decision-support systems in healthcare.

ACKNOWLEDGMENTS

This work was supported by The MPS Foundation Grant Programme (2022-0000000206). The MPS Foundation was established to undertake research, analysis, education and training to enable healthcare professionals to provide better care for their patients and improve their own well-being. To achieve this, it supports and funds research across the world that will make a difference and can be applied in the workplace. We are grateful to Dr Uriana Boye for her comments on the paper. The work was also supported by the Engineering and Physical Sciences Research Council (EP/W011239/1).

REFERENCES

[1] 5 ways to help robots work together with people: 2018. http://theconversation.com/5-ways-to-help-robots-work-together-with-people-101419. Accessed: 2024-01-29.

[2] A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis - The Lancet Digital Health: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(19)30123-2/fulltext. Accessed: 2024-01-30.

[3] Alcalá-Rmz, V. *et al.* 2021. Identification of People with Diabetes Treatment through Lipids Profile Using Machine Learning Algorithms. *Healthcare.* 9, 4 (Apr. 2021), 422. DOI:https://doi.org/10.3390/healthcare9040422.

[4] Artificial Intelligence: 2023. https://transform.england.nhs.uk/information-governance/guidance/artificial-intelligence/. Accessed: 2024-01-15.

[5] Bansal, G. *et al.* 2019. Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing.* 7, (Oct. 2019), 2–11. DOI:https://doi.org/10.1609/hcomp.v7i1.5285.

[6] Bansal, G. *et al.* 2019. Updates in human-AI teams: understanding and addressing the performance/compatibility tradeoff. *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence* (Honolulu, Hawaii, USA, Jan. 2019), 2429–2437.

[7] Birch, J. *et al.* 2022. Clinical decisions using AI must consider patient values. *Nature Medicine.* 28, 2 (Feb. 2022), 229–232. DOI:https://doi.org/10.1038/s41591-021-01624-y.

[8] Braun, V. and Clarke, V. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health.* 11, 4 (Aug. 2019), 589–597. DOI:https://doi.org/10.1080/2159676X.2019.1628806.

[9] Bussone, A. *et al.* 2015. The Role of Explanations on Trust and Reliance in Clinical Decision Support Systems. *2015 International Conference on Healthcare Informatics* (Oct. 2015), 160–169.

[10] Cai, C.J. *et al.* 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* (New York, NY, USA, May 2019), 1–14.

[11] Cai, C.J. *et al.* 2019. The effects of example-based explanations in a machine learning interface. *Proceedings of the 24th International Conference on Intelligent User Interfaces* (New York, NY, USA, Mar. 2019), 258–262.

[12] Christopher Brill, J. *et al.* 2018. Navigating the Advent of Human-Machine Teaming. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* 62, 1 (Sep. 2018), 455–459. DOI:https://doi.org/10.1177/1541931218621104.

[13] Conversational user experience: 2022. https://learn.microsoft.com/en-us/azure/bot-service/bot-service-design-principles?view$=$azure-bot-service-4.0. Accessed: 2024-01-08.

[14] Coudray, N. *et al.* 2018. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nature Medicine.* 24, 10 (Oct. 2018), 1559–1567. DOI:https://doi.org/10.1038/s41591-018-0177-5.

[15] Dazeley, R. *et al.* 2021. Levels of explainable artificial intelligence for human-aligned conversational explanations. *Artificial Intelligence.* 299, (Oct. 2021), 103525. DOI:https://doi.org/10.1016/j.artint.2021.103525.

[16] Decherchi, S. *et al.* 2021. Opportunities and Challenges for Machine Learning in Rare Diseases. *Frontiers in Medicine.* 8, (2021).

[17] Deep neural network improves fracture detection by clinicians | PNAS: https://www.pnas.org/doi/abs/10.1073/pnas.1806905115. Accessed: 2024-01-30.

[18] Demir, M. *et al.* 2018. A conceptual model of team dynamical behaviors and performance in human-autonomy teaming. *Cognitive Systems Research.* 52, (Dec. 2018), 497–507. DOI:https://doi.org/10.1016/j.cogsys.2018.07.029.

[19] Detjen, H. *et al.* 2020. A Wizard of Oz Field Study to Understand Non-Driving-Related Activities, Trust, and Acceptance of Automated Vehicles. *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (New York, NY, USA, Sep. 2020), 19–29.

[20] Dove, G. *et al.* 2017. UX Design Innovation: Challenges for Working with Machine Learning as a Design Material. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2017), 278–288.

[21] Figma: 2024. https://www.figma.com. Accessed: 2024-01-29.

[22] Flohr, L.A. *et al.* 2023. Prototyping Autonomous Vehicle Windshields with AR and Real-Time Object Detection Visualization: An On-Road Wizard-of-Oz Study. *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh PA USA, Jul. 2023), 2123–2137.

[23] Freeman, K. *et al.* 2021. Use of artificial intelligence for image analysis in breast cancer screening programmes: systematic review of test accuracy. *BMJ.* 374, (Sep. 2021), n1872. DOI:https://doi.org/10.1136/bmj.n1872.

[24] Gargeya, R. and Leng, T. 2017. Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology.* 124, 7 (Jul. 2017), 962–969. DOI:https://doi.org/10.1016/j.ophtha.2017.02.008.

[25] General Best Practices - Messenger Platform - Documentation: https://developers.facebook.com/docs/messenger-platform/introduction/general-best-practices/. Accessed: 2024-01-08.

[26] Green, B. and Chen, Y. 2019. The Principles and Limits of Algorithm-in-the-Loop Decision Making. *Proceedings of the ACM on Human-Computer Interaction.* 3, CSCW (Nov. 2019), 50:1-50:24. DOI:https://doi.org/10.1145/3359152.

[27] Guidotti, R. *et al.* 2018. Local Rule-Based Explanations of Black Box Decision Systems. arXiv.

[28] Hansen, L.K. and Salamon, P. 1990. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* 12, 10 (Oct. 1990), 993–1001. DOI:https://doi.org/10.1109/34.58871.

[29] Hoffman, R.R. *et al.* 2018. Explaining Explanation For "Explainable Ai." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting.* 62, 1 (Sep. 2018), 197–201. DOI:https://doi.org/10.1177/1541931218621047.

[30] Holzinger, A. 2021. The Next Frontier: AI We Can Really Trust. *Machine Learning and Principles and Practice of Knowledge Discovery in Databases* (Cham, 2021), 427–440.

[31] Human Factors in Healthcare AI: https://ergonomics.org.uk/resource/human-factors-in-healthcare-ai.html. Accessed: 2024-01-08.

[32] IBM Design for AI: https://www.ibm.com/design/ai/www.ibm.com/design/ai. Accessed: 2024-01-08.

[33] Inkpen, K. *et al.* 2019. Where is the Human? Bridging the Gap Between AI and HCI. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2019), 1–9.

[34] Islam, M.M. *et al.* 2021. Deep into Laboratory: An Artificial Intelligence Approach to Recommend Laboratory Tests. *Diagnostics.* 11, 6 (Jun. 2021), 990. DOI:https://doi.org/10.3390/diagnostics11060990.

[35] Jacobs, M. *et al.* 2021. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Translational Psychiatry.* 11, 1 (Feb. 2021), 1–9. DOI:https://doi.org/10.1038/s41398-021-01224-x.

[36] Johnson, M. and Vera, A. 2019. No AI Is an Island: The Case for Teaming Intelligence. *AI Magazine.* 40, 1 (Mar. 2019), 16–28. DOI:https://doi.org/10.1609/aimag.v40i1.2842.

[37] Jones, C. *et al.* 2023. Artificial intelligence and clinical decision support: clinicians' perspectives on trust, trustworthiness, and liability. *Medical Law Review.* 31, 4 (Nov. 2023), 501–520. DOI:https://doi.org/10.1093/medlaw/fwad013.

[38] Kawamoto, K. *et al.* 2005. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ.* 330, 7494 (Mar. 2005), 765. DOI:https://doi.org/10.1136/bmj.38398.500764.8F.

[39] Kelley, J.F. 1983. An empirical methodology for writing user-friendly natural language computer applications. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Dec. 1983), 193–196.

[40] Khalfallah, H.B. *et al.* 2023. Decision support systems in healthcare: systematic review, meta-analysis and prediction, with example of COVID-19. *AIMS Bioengineering.* 10, 1 (2023), 27–52. DOI:https://doi.org/10.3934/bioeng.2023004.

[41] Kim, J.T. 2018. Application of Machine and Deep Learning Algorithms in Intelligent Clinical Decision Support Systems in Healthcare. *Journal of Health & Medical Informatics.* 09, 05 (2018). DOI:https://doi.org/10.4172/2157-7420.1000321.

[42] Lau, L. *et al.* 2021. Comparison of Clinical Importance of Drug Interactions Identified by Hospital Pharmacists and a Local Clinical Decision Support System. *The Canadian Journal of Hospital Pharmacy.* 74, 3 (2021), 203–210. DOI:https://doi.org/10.4212/cjhp.v74i3.3147.

[43] Lawton, T. *et al.* 2024. Clinicians Risk Becoming "Liability Sinks" for Artificial Intelligence. *Future Healthcare Journal.* (Feb. 2024), 100007. DOI:https://doi.org/10.1016/j.fhj.2024.100007.

[45] Madar, R. *et al.* 2021. A Web Interface for Antibiotic Prescription Recommendations in Primary Care: User-Centered Design Approach. *Journal of Medical Internet Research.* 23, 6 (Jun. 2021), e25741. DOI:https://doi.org/10.2196/25741.

[46] Matthews, M. *et al.* 2014. Taking part: role-play in the design of therapeutic systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2014), 643–652.

[47] McCradden, M.D. *et al.* 2020. Clinical research underlies ethical integration of healthcare artificial intelligence. *Nature Medicine.* 26, 9 (Sep. 2020), 1325–1326. DOI:https://doi.org/10.1038/s41591-020-1035-9.

[48] McDougall, R.J. 2018. Computer knows best? The need for value-flexibility in medical AI. *Journal of Medical Ethics.* (Nov. 2018). DOI:https://doi.org/10.1136/medethics-2018-105118.

[49] McElroy, K. 2016. *Prototyping for Designers: Developing the Best Digital and Physical Products.* O'Reilly Media, Inc.

[50] Middleton, B. *et al.* 2016. Clinical Decision Support: a 25 Year Retrospective and a 25 Year Vision. *Yearbook of Medical Informatics.* 25, S 1 (2016), S103–S116. DOI:https://doi.org/10.15265/IYS-2016-s034.

[51] Moja, L. *et al.* 2014. Effectiveness of Computerized Decision Support Systems Linked to Electronic Health Records: A Systematic Review and Meta-Analysis. *American Journal of Public Health.* 104, 12 (Dec. 2014), e12–e22. DOI:https://doi.org/10.2105/AJPH.2014.302164.

[52] Mouazer, A. *et al.* 2022. Decision-support systems for managing polypharmacy in the elderly: A scoping review. *Journal of Biomedical Informatics.* 130, (Jun. 2022), 104074. DOI:https://doi.org/10.1016/j.jbi.2022.104074.

[53] Mueller, S.T. *et al.* 2019. Explanation in Human-AI Systems: A Literature Meta-Review, Synopsis of Key Ideas and Publications, and Bibliography for Explainable AI. arXiv.

[54] Naiseh, M. *et al.* 2023. How the different explanation classes impact trust calibration: The case of clinical decision support systems. *International Journal of Human-Computer Studies.* 169, (Jan. 2023), 102941. DOI:https://doi.org/10.1016/j.ijhcs.2022.102941.

[55] Nasir, J. *et al.* 2022. Questioning Wizard of Oz: Effects of Revealing the Wizard behind the Robot. *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (Aug. 2022), 1385–1392.

[56] NHS England»Involving people in their own care: https://www.england.nhs.uk/ourwork/patient-participation/. Accessed: 2024-01-24.

[57] O'Neill, T. *et al.* 2022. Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature. *Human Factors.* 64, 5 (Aug. 2022), 904–938. DOI:https://doi.org/10.1177/0018720820960865.

[58] Pan, C.H. and Statman, M. 2012. Questionnaires of Risk Tolerance, Regret, Overconfidence, and Other Investor Propensities.

[59] Patel, V.L. *et al.* 2009. The coming of age of artificial intelligence in medicine. *Artificial Intelligence in Medicine.* 46, 1 (May 2009), 5–17. DOI:https://doi.org/10.1016/j.artmed.2008.07.017.

[60] People + AI Guidebook: https://pair.withgoogle.com/guidebook. Accessed: 2024-01-08.

[61] Rahwan, I. *et al.* 2019. Machine behaviour. *Nature.* 568, 7753 (Apr. 2019), 477–486. DOI:https://doi.org/10.1038/s41586-019-1138-y.

[62] Schoonderwoerd, T.A.J. *et al.* 2021. Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. *International Journal of Human-Computer Studies.* 154, (Oct. 2021), 102684. DOI:https://doi.org/10.1016/j.ijhcs.2021.102684.

[63] Shneiderman, B. 2020. Design Lessons From AI's Two Grand Goals: Human Emulation and Useful Applications. *IEEE Transactions on Technology and Society.*

1, 2 (Jun. 2020), 73–82. DOI:https://doi.org/10.1109/TTS.2020.2992669.

[64] Shneiderman, B. 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction*. 36, 6 (Apr. 2020), 495–504. DOI:https://doi.org/10.1080/10447318.2020.1741118.

[65] Shneiderman, B. 2020. Human-Centered Artificial Intelligence: Three Fresh Ideas. *AIS Transactions on Human-Computer Interaction*. 12, 3 (Sep. 2020), 109–124. DOI:https://doi.org/10.17705/1thci.00131.

[66] Skivington, K. *et al.* 2021. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *BMJ*. 374, (Sep. 2021), n2061. DOI:https://doi.org/10.1136/bmj.n2061.

[67] Sokol, K. and Flach, P. 2019. Counterfactual explanations of machine learning predictions: 2019 AAAI Workshop on Artificial Intelligence Safety, SafeAI 2019. *Proceedings of the AAAI Workshop on Artificial Intelligence Safety 2019*. 2301, (Jan. 2019).

[68] Stephen, O. *et al.* 2019. An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. *Journal of Healthcare Engineering*. 2019, (Mar. 2019), e4180949. DOI:https://doi.org/10.1155/2019/4180949.

[69] the Precise4Q consortium *et al.* 2020. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*. 20, 1 (Dec. 2020), 310. DOI:https://doi.org/10.1186/s12911-020-01332-6.

[70] Tian, F. *et al.* 2019. What Can Gestures Tell? Detecting Motor Impairment in Early Parkinson's from Common Touch Gestural Interactions. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2019), 1–14.

[71] Vasey, B. *et al.* 2021. Association of Clinician Diagnostic Performance With Machine Learning–Based Decision Support Systems: A Systematic Review. *JAMA Network Open*. 4, 3 (Mar. 2021), e211276. DOI:https://doi.org/10.1001/jamanetworkopen.2021.1276.

[72] Vasey, B. *et al.* 2021. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nature Medicine*. 27, 2 (Feb. 2021), 186–187. DOI:https://doi.org/10.1038/s41591-021-01229-5.

[73] Vasey, B. *et al.* 2022. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature Medicine*. 28, 5 (May 2022), 924–933. DOI:https://doi.org/10.1038/s41591-022-01772-9.

[74] Wang, Z. *et al.* 2014. Knowledge Graph Embedding by Translating on Hyperplanes. *Proceedings of the AAAI Conference on Artificial Intelligence*. 28, 1 (Jun. 2014). DOI:https://doi.org/10.1609/aaai.v28i1.8870.

[75] Welcome to the Artificial Intelligence Incident Database: https://incidentdatabase.ai/. Accessed: 2024-02-06.

[76] What Is Conversational AI? - Alexa Skills Kit Official Site: https://developer.amazon.com/en-US/alexa/alexa-skills-kit/conversational-ai.html. Accessed: 2024-01-08.

[77] Wu, W. *et al.* 2020. Towards Global Explanations of Convolutional Neural Networks With Concept Attribution. (2020), 8652–8661.

[78] Xu, W. *et al.* 2023. Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI. *International Journal of Human–Computer Interaction*. 39, 3 (Feb. 2023), 494–518. DOI:https://doi.org/10.1080/10447318.2022.2041900.

[79] Yampolskiy, R.V. 2018. Predicting future AI failures from historic examples. *foresight*. 21, 1 (Jan. 2018), 138–152. DOI:https://doi.org/10.1108/FS-04-2018-0034.

[80] Yang, F. *et al.* 2020. How do visual explanations foster end users' appropriate trust in machine learning? *Proceedings of the 25th International Conference on Intelligent User Interfaces* (New York, NY, USA, Mar. 2020), 189–201.

[81] Yang, Q. *et al.* 2023. Harnessing Biomedical Literature to Calibrate Clinicians' Trust in AI Decision Support Systems. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg Germany, Apr. 2023), 1–14.

[82] Yang, Q. *et al.* 2020. Re-examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, Apr. 2020), 1–13.

[83] Zanzotto, F.M. 2019. Viewpoint: Human-in-the-loop Artificial Intelligence. *Journal of Artificial Intelligence Research*. 64, (Feb. 2019), 243–252. DOI:https://doi.org/10.1613/jair.1.11345.

[84] Zheng, N. *et al.* 2017. Hybrid-augmented intelligence: collaboration and cognition. *Frontiers of Information Technology & Electronic Engineering*. 18, 2 (Feb. 2017), 153–179. DOI:https://doi.org/10.1631/FITEE.1700053.

[85] Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems Version 2.