*Context Sensitive Health Informatics: AI for Social Good*
*H. Ismaila et al. (Eds.)*


*doi:10.3233/SHTI250228*

# User-Centered Methods in Explainable AI Development for Hospital Clinical Decision Support: A Scoping Review

Bethany A VAN DORT[a,1], Thomas ENGELSMA[a,c], Stephanie MEDLOCK[b,c] and Linda DUSSELJEE-PEUTE[a]

[a] *Amsterdam UMC location University of Amsterdam, eHealth Living & Learning Lab Amsterdam, Department of Medical Informatics, Amsterdam, The Netherlands*

[b] *Amsterdam UMC location University of Amsterdam, Department of Medical Informatics, Amsterdam, The Netherlands*

[c] *Amsterdam Public Health Research Institute, Amsterdam, The Netherlands*

ORCiD ID: Bethany A Van Dort https://orcid.org/0000-0003-0052-6400, Thomas Engelsma https://orcid.org/0000-0001-8889-1669, Stephanie Medlock https://orcid.org/0000-0002-2679-8095, Linda Dusseljee-Peute https://orcid.org/0000-0002-9859-0912

**Abstract.** Explainable Artificial Intelligence (XAI) offers promising advancements in enhancing transparency and usability of AI-based Clinical Decision Support Systems (CDSS) in healthcare settings. These tools aim to improve clinical outcomes by assisting with diagnosis, treatment planning, and risk prediction. However, integrating XAI into clinical workflows requires effective involvement of healthcare professionals to ensure that the explanations provided by these tools are comprehensible, relevant, and actionable. This scoping review aimed to investigate how (potential) end users were involved in the design and development of XAI-based CDSS for hospitals. A systematic search of Medline, Embase, and Web of Science identified 11 studies meeting the inclusion criteria. Interviews and focus groups, mainly with physicians, were common, while some included nurses and developers. Four of the 11 studies engaged users across multiple stages, from pre-design to prototype testing, and specifically tested different explanation techniques with end-users. A quality assessment of papers found some studies had unclear recruitment strategies and insufficiently detailed analyses. Future work should engage end-users early in the design process, include health professionals with diverse experiences and backgrounds, and test explanation techniques to ensure appropriate methods that align with cognitive processes are chosen.

**Keywords.** Explainable artificial intelligence; Machine learning; Human computer interaction; Usability; Clinical decision support system.

[1] Corresponding Author: Bethany Van Dort, Amsterdam UMC location University of Amsterdam, Department of Medical Informatics, eHealth Living & Learning Lab Amsterdam, Meibergdreef 9, 1105AZ, Amsterdam, the Netherlands; E-mail: b.a.vandort@amsterdamumc.nl

## 1. Introduction

Clinical Decision Support Systems (CDSS) powered by artificial intelligence (AI) have the potential to support healthcare professionals with diagnosis, treatment planning, and risk assessment [1]. However, the complexity of these systems often makes their decision-making processes unclear to users, raising concerns about trust, usability, and accountability [2]. Explainable artificial intelligence (XAI) is a field aimed at making AI tools more transparent and understandable, by providing explanations that help clinicians understand and use AI-driven recommendations in their practice.[3]

A significant focus in XAI for healthcare is post-hoc explainability, where explainability methods are developed to understand the decisions of an AI model. These explanations aim to clarify how the system arrives at its outputs.[3] Yet, the effectiveness of these explanations largely depends on their relevance and comprehensibility to the end-users [3]. The unique workflows, time pressures, and interdisciplinary hospital settings require XAI tools to be accurate and seamlessly integrated into clinical routines, ensuring outputs are actionable without adding to healthcare professionals' cognitive load [4]. Previous studies have demonstrated that engaging users during design and development can help ensure that tools align with clinical workflows, meet the needs of diverse healthcare roles, and enhance the adoption of AI tools in practice [5, 6]. Only then, can the benefits of XAI CDSS be realised. This scoping review aimed to investigate how (potential) end users were involved in the design and development of XAI-based CDSS for hospitals. Specifically, we aimed to understand who is engaged, when in the design and development process, and what methods are used.

## 2. Methods

### *2.1. Search strategy*

Medline, Embase, and Web of Science were systematically searched to identify studies eliciting end-user perceptions in the development of XAI tools. The search strategy used a combination of text words and subject headings related to XAI, machine learning (ML), clinicians, and perspectives or user-centered design methods (e.g. survey, interview). The search was conducted on the 15$^{th}$ of July 2024.

We included studies reporting on the development of explainable AI/ML interventions for hospitals, where the participants were healthcare professionals. Reviews, commentaries, conference abstracts, and non-English papers were excluded. As the purpose of this review was to focus on the development phase, studies of XAI tools post implementation, and studies of AI prior to system development were excluded.

### *2.2. Study selection, data extraction and quality assessment*

References were imported from databases into Covidence for screening. Titles and abstracts were screened independently by two researchers (BV, TE) using the eligibility criteria. Full texts were then independently screened by two researchers (BV, TE) and disagreements were discussed until consensus was reached. Data was extracted by one researcher and checked by a second researcher to ensure accuracy. A quality assessment was conducted using the Critical Appraisal Skills Program (CASP) Checklist [7].

## 3. Results

Eleven studies were eligible for inclusion, after excluding 1168 studies through title and abstract screening, and 54 through full text screening. The goals of the XAI tools from the included studies are outlined in **Table 1**. All studies were published in the last 5 years.

**Table 1.** Included studies and the goal of the XAI tool in development

| Study | Year | Country | Purpose of XAI tool |
|---|---|---|---|
| [8] | 2020 | USA | Predict in-hospital mortality for paediatric ICU patients |
| [9] | 2023 | Switzerland | Predict onset of delayed cerebral ischemia in patients with aneurysmal subarachnoid haemorrhage |
| [10] | 2024 | USA | Predict postoperative complications |
| [11] | 2023 | USA | Predict risk of instability in ICU patients |
| [12] | 2023 | Norway | Predict polyp occurrence using colonoscopy images. |
| [13] | 2020 | China | Predict diagnosis and treatment outcome analysis |
| [14] | 2021 | Denmark | Predict ventricular tachycardia and ventricular fibrillation (VT/VF) in patients with cardiac implanted electronic devices |
| [15] | 2023 | Portugal | Predict EEG seizures in patients with epilepsy |
| [16] | 2022 | Germany | Predict risk of infection and graft loss within 90 days post kidney transplant |
| [17] | 2024 | Canada | Predict severity of COVID-19 |
| [18] | 2024 | Netherlands | Predict urgency of ICU admission for patients with sepsis |

EEG: Electroencephalogram, ICU: intensive care unit, XAI: Explainable artificial intelligence

The methods used to involve end-users, and the stages of development at which users were engaged, are summarised in **Table 2**. Interviews and/or focus groups were conducted in most studies. Participants were primarily physicians, however, some studies also sought perspectives of nurses[8-11] and developers.[9, 15].

**Table 2.** Participants and methods used to involve users at various stages of XAI tool development.

| Study | Participants | Requirements phase (Pre-design of XAI tool) | Development phase (Prototype testing) |
|---|---|---|---|
| [8] | *Physicians and nurses* | | *Focus groups and questionnaire (n=21) |
| [9] | *Physicians, nurses, and developers* | | Survey (n=95), focus group (n=6), interviews (n=11) and think aloud (n=7) |
| [10] | *Physicians and anaesthetic nurses* | Focus groups with card sorting activity (n=21) | Interviews (n=9), think aloud and survey (SUS) (n=10) |
| [11] | *Physicians, physician assistants, and nurses* | | Focus groups (n=23) |
| [12] | *Clinicians (primarily gastroenterologists)* | | *Survey (n=54) |
| [13] | *Senior physicians* | Interviews (n=2) | Scenario-based testing (n=2), interactive demonstration and interviews (n=7) |
| [14] | *Physicians (Cardiologists)* | Fieldwork observations and co-design workshops (n=2) | Feasibility survey, case walkthrough, and interviews (n=7) |
| [15] | *Clinicians and data scientists* | | *Interviews (n=10) |
| [16] | *Physicians* | | Interviews (n=14) |
| [17] | *Physicians* | Focus groups (n=7) | Informal feedback (n=2), simulated testing and interview (n=5) |
| [18] | *Physicians and Surgeons (5-10 years experience)* | | *Interviews (n=4) |

n= number of participants. * = also used to gather feedback on explanation techniques. XAI: Explainable artificial intelligence, SUS: System usability scale

Four studies [10, 13, 14, 17] involved end-users during the requirements phase (before designing or selecting the XAI model) and at multiple stages during development. Four studies [10, 11, 13, 17] also tested a low-fidelity prototype with participants before developing a high-fidelity prototype. Four studies [8, 12, 15, 18] specifically focused on testing different explanation techniques with end-users. For example, Barda et al. [8] created five displays with different explanation techniques and design features, which were presented to participants in focus groups. This helped identify user needs for improving understanding of explanations, such as expressing risk as a percentage, providing multiple plots, and showing directionality for trend-based features.

The quality assessment identified several issues in the reporting of the included studies. All papers lacked reflexivity, meaning they failed to examine the researcher's role on the research and potential bias. Several papers had unclear recruitment strategies [12, 13, 15, 17, 18], and insufficiently detailed qualitative data analyses [12, 13, 18]. The study design and methods used in the included studies were assessed as appropriate.

## 4. Discussion

This scoping review identified 11 studies highlighting the methods used to involve healthcare professionals in developing XAI-based CDSS for hospitals. Interviews and focus groups, primarily with physicians, were most common. Four studies engaged health professionals at multiple stages of design and development. A number of limitations were identified in some of the included studies, such as unclear recruitment strategies and data analysis, however, all data collection methods were reported sufficiently.

User centered XAI development has been explored in the broader literature but the processes identified have not been applied to XAI for the hospital context. For example, Chazette et al. [19] conducted a systematic review of recommended practices of XAI development across all industries. While many of the methods identified in our review aligned with Chazette et al. [19], only four studies reported involving healthcare professionals in the requirements phase. Further, several steps recommended by Chazette et al. [19] for the requirements phase (i.e. vision definition, stakeholder analysis, back-end analysis, and trade-off analysis) were not conducted in most of the studies reviewed. It should also be noted this review only examined studies reporting on XAI tool development, and there may be instances where user involvement occurred earlier but was unpublished or reported separately.

Four studies in this review reported testing different explanation techniques with health professionals. Previous XAI research has shown that users interpret and understand explanations differently, and testing explanations with end users is imperative [20]. The studies in this review used surveys, questionnaires, interviews, and/or focus groups to test explanations, which are valuable and widely used methods [19]. Additional methods such as A/B testing, scenario-based testing, observations, and development of mental models, have been successfully used in other contexts and could also be considered for testing XAI in the hospital context [19]. Involving a range of stakeholders is also essential. Two of the four papers involving end-users during the requirements phase included only two participants each. For example, Jin et al. [13] involved two senior clinicians to design a diagnostic and prognostic tool. Previous studies have found senior and junior doctors engage CDSS in different ways and have different approaches to clinical judgement [21, 22]. Therefore, involving a range of potential end-users is important for developing effective XAI tools.

## 5. Conclusion

This scoping review found studies engaged health professionals to varying degrees in the development of XAI CDSS, mainly using interviews and focus groups, and involvement in the development phase. Testing explanation techniques with end-users was limited. Future efforts should involve a diverse range of potential end-users early in the process and test explanation techniques to create effective XAI for clinical settings.

## References


[1] Giordano C, et al. Accessing artificial intelligence for clinical decision-making. Front Digit Health. 2021;3:645232.
[2] Hassan M, et al. Barriers to and Facilitators of Artificial Intelligence Adoption in Health Care: Scoping Review. JMIR Hum Factors. 2024;11:e48633.
[3] Loh HW, et al. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Comput Methods Programs Biomed. 2022;226:107161.
[4] Amann J, et al. To explain or not to explain? Artificial intelligence explainability in clinical decision support systems. PLOS Digital Health. 2022;1(2):e0000016.
[5] Owoyemi A, et al. End user experience of a widely used artificial intelligence based sepsis system. JAMIA Open. 2024;7(4).
[6] Panigutti C, et al. Co-design of human-centered, explainable AI for clinical decision support. ACM Transactions on Interactive Intelligent Systems. 2023;13(4):1-35.
[7] Critical Appraisal Skills Programme. CASP Qualitative Studies Checklist 2023 [Available from: https://casp-uk.net/casp-tools-checklists/qualitative-studies-checklist/.
[8] Barda AJ, et al. A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. BMC Med Inform Decis Mak. 2020;20(1):257.
[9] Bienefeld N, et al. Solving the explainable AI conundrum by bridging clinicians' needs and developers' goals. NPJ Dig Med. 2023;6(1):94.
[10] Fritz BA, et al. User-Centered Design of a Machine Learning Dashboard for Prediction of Postoperative Complications. Anesth Analg. 2024;138(4):804-13.
[11] Helman S, et al. Engaging Multidisciplinary Clinical Users in the Design of an Artificial Intelligence-Powered Graphical User Interface for Intensive Care Unit Instability Decision Support. Appl Clin Inform. 2023;14(4):789-802.
[12] Hicks S, et al. Visual explanations for polyp detection: How medical doctors assess intrinsic versus extrinsic explanations. PloS One. 2024;19(5):e0304069.
[13] Jin Z, et al. CarePre. ACM Transactions on Computing for Healthcare. 2020;1(1):3344258.
[14] Matthiesen S, et al. Clinician Preimplementation Perspectives of a Decision-Support Tool for the Prediction of Cardiac Arrhythmia Based on Machine Learning: Near-Live Feasibility and Qualitative Study. JMIR Hum Factors. 2021;8(4):e26964.
[15] Pinto MF, et al. The goal of explaining black boxes in EEG seizure prediction is not to explain models' decisions. Epilepsia Open. 2023;8(2):285-97.
[16] Samhammer D, et al. "Nothing works without the doctor:" Physicians' perception of clinical decision-making and artificial intelligence. Front Med. 2022;9(101648047):1016366.
[17] Shulha M, et al. Integrating Explainable Machine Learning in Clinical Decision Support Systems: Study Involving a Modified Design Thinking Approach. JMIR Form Res. 2024;8(101726394):e50475.
[18] Zilker S, et al. A machine learning framework for interpretable predictions in patient pathways: The case of predicting ICU admission for patients with symptoms of sepsis. Health Care Manag Sci. 2024(dl1, 9815649).
[19] Chazette L, et al., editors. How can we develop explainable systems? insights from a literature review and an interview study. Proceedings of the International Conference on Software and System Processes and International Conference on Global Software Engineering; 2022.
[20] Laato S, et al. How to explain AI systems to end users: a systematic literature review and research agenda. Internet Research. 2022;32(7):1-31.
[21] Baysari MT, et al. The influence of computerized decision support on prescribing during ward-rounds: are the decision-makers targeted? JAMIA. 2011;18(6):754-9.
[22] Nilsson MS, Pilhammar E. Professional approaches in clinical judgements among senior and junior doctors: implications for medical education. BMC Med Educ. 2009;9:1-9.