

Design Guidelines for XAI in the Healthcare Domain

Iris Heerlien¹

¹University of Twente

Abstract

The workload in the healthcare sector is increasing, due to amongst other reasons the aging population. AI could be of help in this, by adding information from data to the knowledge and experience of healthcare professionals in decision-making processes. However, an AI algorithm without a user-centered explanation will not be able to be used in critical situations as the health professional needs to stay in control. This research focuses on creating user-centered guidelines for XAI representations in the selected domains, home (elderly) care and physiotherapy, using the Design Thinking method to include the users during the whole process.

Keywords

Design guidelines, XAI, physiotherapy, home (elderly) care

1. Context and motivation

Imagine a scenario where artificial intelligence (AI) lightens the workload on nurses and physiotherapists by combining data from (other) patients with their own knowledge and experience while keeping the health professional in control by explaining the reasoning behind the decisions of the AI, using Explainable AI (XAI). The workload in the healthcare sector is increasing, due to amongst other reasons the aging population. AI could be of help in this, by supporting the knowledge and experience of healthcare professionals in the decision-making processes with insights from data of e.g. other patients as this will eliminate the need of extensive dossier reading. However, the health professional needs to stay in control, making it insuperable to explain the AI algorithm in a way the professional understands. This research focuses on creating user-centered guidelines for XAI representations in the selected domains of home (elderly) care and physiotherapy, using the Design Thinking method to include the users during the whole process of creating the user-centered guidelines.

2. Key related work

Addressing personnel shortages is critical in the healthcare sector [1]. In this research, the decision was made to focus on two different healthcare domains to investigate the generic and domain-specific effects of XAI. The two selected domains are the home (elderly) care and physiotherapist domain.

Late-breaking work, Demos and Doctoral Consortium, colocated with The 2nd World Conference on eXplainable Artificial Intelligence: July 17–19, 2024, Valletta, Malta

✉ i.r.heerlien@utwente.nl (I. Heerlien)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

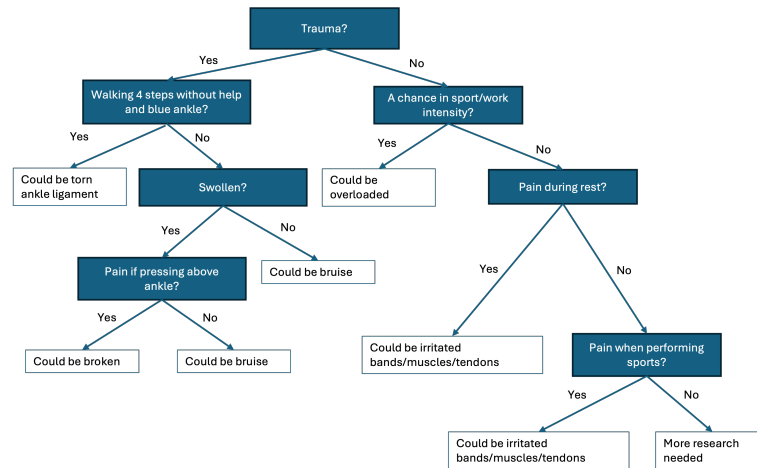


Figure 1: An example of a part of a decision tree supporting in the diagnosis of a painful ankle

Home care, suffering from an aging population [2], requires innovative solutions to meet the increasing demands to enhance efficiency and enabling individuals to live home for longer.

The domain of physiotherapy also suffers from an aging population and additionally the increase in chronically ill people [3]. Moreover, unfavorable employment conditions result in decreasing number of physiotherapists.

Using AI to extract information from data can be used to reduce the workload of health professionals. For instance, utilizing data from patient records eliminates the need for extensive dossier reading in home care. Besides, using patient-generated data [4], subsequently analyzed by physiotherapists, allows for less frequent in-person visits. Although this leads to less in-person contact, the time during visits can be spend more on personal contact as the data is already generated at home, increasing the effectiveness of the visits.

However, reliance on data-driven decisions in healthcare raises concerns. It is crucial to ensure healthcare professionals maintain control and responsibility, and the risks of AI system errors leading to patient harm is minimized [5]. Explainable AI (XAI) emerges as a solution, providing transparency in the decision-making process.

Various techniques, such as decision trees (Figure 1), have been proposed for explaining AI models. As model accuracy increases, the complexity of explainability also rises, presenting challenges in determining suitable representations [6] (Figure 2). To optimize XAI's utility, visualizations must align with the user's needs and knowledge.

The danger of over-reliance or neglecting explanations necessitates careful consideration of presentation formats. Designing effective XAI representations requires a user-centered approach to ensure healthcare professionals comprehend and utilize AI optimally. While prior research has explored XAI representations based on designers' experiences [8, 9], limited insight exists using direct user communication. The proposed research aims to extend the understanding of AI explanations. By adopting a user-centered approach, the research strives to create comprehensive guidelines for XAI representations applicable to the selected domains. This user-centric perspective ensures that the explanations generated align with the practical

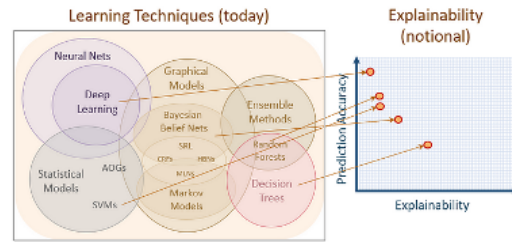


Figure 2: The relation between learning techniques and explainability [7]

needs and understanding of healthcare professionals, fostering responsible and informed use of AI in these critical domains.

3. Specific research questions, hypothesis and objectives

3.1. Objectives

This research focuses on creating guidelines for XAI representations based on user research to offer researchers and designers the tools to create XAI representations that fit the healthcare professionals in the selected domains. This will result in the possibility to use AI in these domains to support the health professional in a safe manner. By answering the research – and sub-questions below this goal will be met.

3.2. Research questions

RQ: How can XAI optimize the decision-making process of healthcare professionals in the selected domains?

SQ1: What XAI representation and design practices are used nowadays in the selected domains?

SQ2: What are the needs and wishes of the users to make the XAI representation guidelines fit their world of experience?

SQ3: What general and domain specific guidelines should XAI representations adhere to?

3.3. Hypothesis

The hypothesis is that XAI representations that follow the created user-centered guidelines will improve the acceptance of AI in the selected domains, which can eventually lower the workload of the healthcare professionals.

4. Research approach, methods, and rationale for testing the research hypothesis

The research will be performed in a design driven way, meaning that prototypes are created, tested, and improved based on the feedback by users. The methodology used in this research is

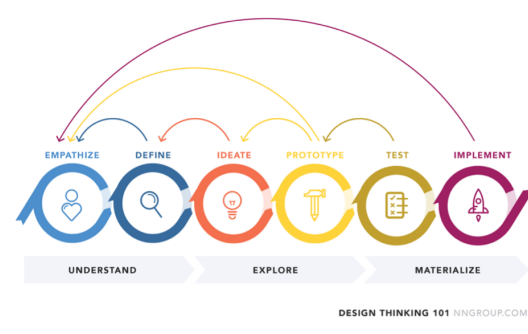


Figure 3: The Design Thinking method [11, 10]

the Design Thinking method as the goal is to maximize the usability of the XAI representations. This method shown in Figure 3, is an iterative process in which users are involved throughout the entire design process [10]. The first step is to empathize with the users and other stakeholders; who are they? How do they work nowadays? This step will also be used to empathize with the topic. The second step is to analyze the information received during the empathize phase. These first two steps focus on understanding the users and the problem. The next three steps focus on creating a solution based on these insights. The third step is to generate ideas which lead to prototypes to be tested and validated in the fourth and fifth step of this process. These prototypes are based on guidelines which are extracted from literature and user research. The sixth step, implementation, focuses on incorporating the final prototypes in the processes of the healthcare professionals to understand if using these prototypes adds value to the decision-making process of the professional and thereby reduce the workload.

The sections below explain in more detail how the steps will be undertaken.

4.1. Empathize and Define; understanding the problem

4.1.1. Step 1: systematic literature review XAI

The first step in this research is a systematic literature review [12] answering the question What XAI representations are used to communicate to the users of the algorithm?. A systematic literature review is the right method for this as it will give insight into what has already been done and how, to be able to build the proposed research on top of it and to relate it to existing knowledge.

The questions to be answered during this systematic literature review are as follows:

- What XAI representations are used nowadays?
- What XAI representations explicitly are not used nowadays and why?
- What are the existing XAI representation design practices?
- What are the pitfalls in creating XAI representations?

The answers to these questions will result in a broad understanding of how XAI is used until now and what general rules to consider when designing the representations during the proposed research. If the results are not specific enough to be used in the selected domains, a

systematic literature review focusing on the home (elderly) care and physiotherapy domain will be performed.

4.1.2. Step 2: Performing user research

During user research, research will be performed to understand the main challenge and to understand the visualizations used by the users. A part of this is understanding the needs and world of experience of the user regarding XAI and the capacity to understand data representations (SQ2). This knowledge is the basis of the guideline creation process to guarantee the XAI representations fits the user's needs.

The questions answered are as follows:

- What visualizations are used nowadays by the users?
- What is the opinion of the users about these visualizations?

Semi-structured interviews with health professionals will be held to answer these questions. This method is chosen as it leaves room for follow up questions to dive deeper and really understand how the participants think. This will lead to the goal of understanding which visualizations fit the participants best.

Participants As there is not a golden number how many interviews are needed as it depends e.g. on the diversity in the population, a good approach is to start small and add more interviews if necessary if saturation is not reached [13].

The participants in this research will be health professionals from the home (elderly) care domain as well as the physiotherapist domain using AI or having the possibility to use AI. Age will be used as a selection criterion as it is important that the chosen visualizations fit all health professionals from the selected domains. However, as it is expected that the participants from the same domain will use the same visualizations, the expectation is that 10 interviews spread over the different age groups is sufficient. If saturation is not reached, additional interviews will be held.

Preparation The participants will be asked to draw (or take pictures if possible, according to privacy regulations) over a period of one week the visualizations they encounter in their jobs. These visualizations will be collected and analyzed to remove doubles and to check if they adhere to the privacy regulations. Additionally, visualizations from relevant previous project of the research group the researcher is working for will be added to the collection. These steps will lead to a collection of visualizations from the selected domains which will be used during the interviews.

Interviews The collected visualizations will be discussed by asking questions about how they interpret the visualization, if they think the visualization is intuitive and what they think about the visualization. Follow up questions will be asked to understand better how the participants think and feel about the different kinds of visualizations.

Analysis The answers of the participants will be analyzed to create a complete overview of which visualizations fit them well and which don't together with the reasoning behind it. This information, combined with the information from the literature review will be used as input for guidelines to which the XAI visualizations should adhere to. These guidelines will be used to create the prototypes in the next steps explained below.

4.2. Step 3, 4, and 5: Ideate, prototype, test; creation and testing of guidelines (SQ3,4)

The next steps of this research is to create and validate guidelines for XAI based on the outcomes of the literature review and user research. Guidelines will be extracted and prototypes adhering to these guidelines will be created. The guidelines will be tested by performing at least 5 [14] thinking aloud usability tests [15]) in which representative users will be asked to join. Prototypes which adhere to these guidelines will be used during the tests. An interview afterwards in which highlights from this test will be discussed will give supplementary information about the quality of the representation. In case general guidelines could be created based on the outcomes of the user research, professionals from different health professions will be asked to test the guidelines as well. This will provide insights into whether the general guidelines can be used in other health professions as envisioned.

4.3. Step 6: Implement; incorporate the prototypes into the processes (RQ)

The prototypes adhering to the final guidelines will be included in the processes followed by the healthcare professionals from the selected domains to understand if and how it adds value and as a result reduce the workload. To measure the change in perceived workload, 10 - 20 participants [16] will be asked to fill in the NASA TLX [17] before and after using the prototype for a specific task and an interview will be held afterwards to discuss this. These results will be analyzed and the guidelines/prototypes will be improved, what could result in going back to different steps of the Design Thinking method.

4.4. Deliverables

The main result of this research will be a guidelines handbook stating the guidelines to design user-centered XAI representations. Besides, articles will be published contributing to knowledge and software prototypes will be created which will be used and improved during the ideate, prototype and validation phase.

5. Results and contributions to date

I started this PhD in January 2024. My first step is performing a systematic literature review to understand which XAI representations are used nowadays to communicate to users, for which I'm finishing up the protocol. This systematic literature review will result in the first contribution of this research.

6. Next steps and contribution to knowledge

6.1. Next steps

The next step is to finish and publish the named systematic literature review. If it seems necessary when the results of the current study are in, a literature review focusing on the selected healthcare domains will be the next step. Interviews with healthcare professionals

from these domains will follow after this. The information from the literature studies combined with the information from the interviews will result into prototypes which will be tested with the health professionals and improved based on these outcomes.

6.2. Contribution to knowledge

AI is used more and more nowadays, while the interpretability of AI can be lacking in critical situations, especially when people's health is concerned. There is not much known about how to adopt AI so that it incorporates principles of explainability to its users. Especially user-centered research into this domain needs attention. This research contributes by investigating in a user centered way what the guidelines are for explainable AI representations in the home care and physiotherapist domain, resulting in a standard way of designing explainable AI representations in these domains. This will help other researchers and developers to understand what their explainable AI representations should adhere to, guaranteeing these are interpretable by professionals in these domains.

Additionally, the methods in combination with the learnings from the different steps in this research can be used in follow-up research as well. User-centered research is crucial in countless research areas, making it possible to use the learnings from this research in many other research projects. The researcher of the proposed project could assist in writing research proposals and during the research itself. Also, the method and learnings as described in publications following from the proposed research can be used by researchers all over the world.

Moreover, the way AI can be used in the selected healthcare domains can be researched differently by using the assumption that the way of communicating is correct. This can elucidate other aspects of why AI models are used or not and how to improve this. This will result in the possibility of using AI in other processes than included in the proposed research as well.

Next to the healthcare domain, other critical domains such as the finance or legal domain, will benefit from using XAI in their processes. The method used in this research together with the learnings and results could be used as a starting point for research into user centered Explainable AI in these domains.

7. Acknowledgments

This work is (partially) supported by Regieorgaan SIA (part of the Dutch Research Council) in the context of the SPRONG DEMAND project¹.

References

- [1] W. e. S. Ministerie van Volksgezondheid, Personeelstekorten in de zorg, 2023. URL: <https://www.igj.nl/onderwerpen/personeelstekort>.
- [2] C. VGZ, Personeelstekort zorg, 2022. URL: <https://www.cooperatievgz.nl/cooperatie-vgz/zorg/personeelstekort-zorg>.

¹<https://demand.nl>

- [3] A. Amro, Fysiotherapie in beeld, 2023. URL: https://www.abnamro.nl/nl/media/Brochure_Brancheinformatie%20Fysiotherapie_2023-10_tcm16-153755.pdf.
- [4] P. José, W. Trees, d. B. Jolanda, F. Roland, Overzichtstudies - technologie in de zorg thuis: Nog een wereld te winnen!, 2013. URL: <https://www.nivel.nl/sites/default/files/bestanden/Rapport-Technologie-in-de-zorg-thuis.pdf>.
- [5] W. Nicholson Price 2, Risks and remedies for artificial intelligence in health care, 2019. URL: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrapes-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>.
- [6] M. Narayanan, E. Chen, J. He, B. Kim, S. Gershman, F. Doshi-Velez, How do humans understand explanations from machine learning systems? an evaluation of the human-interpretability of explanation, arXiv preprint arXiv:1802.00682 (2018).
- [7] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, J. Zhu, Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges, 2019, pp. 563–574. doi:10.1007/978-3-030-32236-6_51.
- [8] D. Saraswat, P. Bhattacharya, A. Verma, V. K. Prasad, S. Tanwar, G. Sharma, P. N. Bokoro, R. Sharma, Explainable ai for healthcare 5.0: Opportunities and challenges, IEEE Access 10 (2022) 84486–84517. doi:10.1109/ACCESS.2022.3197671.
- [9] O. Williams, Towards Human-Centred Explainable AI: A Systematic Literature Review, 2021. doi:10.13140/RG.2.2.27885.92645.
- [10] R. Van Der Wardt, De design thinking methode uitgelegd: in 5 fases naar innovatie, 2021. URL: <https://designthinkingworkshop.nl/design-thinking-methode/>.
- [11] G. Sarah, Design thinking 101, 2016. URL: <https://www.nngroup.com/articles/design-thinking/>.
- [12] B. Kitchenham, S. Charters, Guidelines for performing systematic literature reviews in software engineering 2 (2007).
- [13] R. Maria, How many participants for a ux interview?, 2021. URL: <https://www.nngroup.com/articles/interview-sample-size/>.
- [14] N. Jakob, Why you only need to test with 5 users, 2023. URL: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>.
- [15] J. Nielsen, Thinking aloud: The nr.1 usability tool, 2012. URL: <https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/>.
- [16] P. Laubheimer, Beyond the nps: measuring perceived usability with the sus, nasa-tlx, and the single ease question after tasks and usability tests, 2023. URL: <https://www.nngroup.com/articles/measuring-perceived-usability/>.
- [17] B. Gore, P. So, Nasa tlx: Task load index, 2020. URL: <https://humansystems.arc.nasa.gov/groups/tlx/>.