An Explorative Study on Enabling Clinicians to Express Ethical Values in the Development of a Machine Learning Clinical Decision Support System

Ilse, P., Faber

Department of Industrial Design, University of Technology Eindhoven, i.p.faber@student.tue.nl

Machine Learning (ML) has great potential in the domain of healthcare, yet integration into clinical practice is a challenge. A naturally important part of implementing medical ML are the ethical considerations that accompany the system. Drawing upon human-centered AI and value sensitive design, this explorative research investigates how clinicians can be enabled to express their ethical values regarding a ML Clinical Decision Support System. A toolkit with accompanying workshop was designed that creates mutual understanding of the ML envisioned implementation context between designer and clinician in a physical manner, creating a canvas to place ethical values into. The workshop was researched with clinicians, using a case study model predicting cardiac risk. Results show benefits of creating mutual understanding of the ML implementation and the toolkit's use of value cards. Recommendations for future work include examining adaptations to the toolkit, and the effects of mapping a (different) set of values.

CCS CONCEPTS • •Human-centered computing~Human computer interaction (HCI)~Empirical studies in HCI

Additional Keywords and Phrases: Human-centered AI, Human-AI Interaction, Clinical Decision Support Systems, Machine Learning in Clinical Context, Participatory Design, Value Sensitive Design, Ethical Values in AI.

1 INTRODUCTION

Exceeding human performance in diverse domains, recent interest in artificial intelligence (AI) and machine learning (ML) has increased [15]. ML is an application of AI, defined as a program that utilizes training data to algorithmically learn to predict, decide or perform tasks [3]. ML models, and thus AI, has great potential in the domain of healthcare by being able to aid in diagnosis [21,40], treatment selection [64] and prognosis or risk prediction [7,29,62]. Furthermore, AI could presumptively improve patient and staff experience and lower healthcare costs [6]. However, the field is predominantly focused on improving ML performance rather than its implementation, while integration into practice and adoption by users is a prominent challenge [6]. Example bottlenecks include the explainability and transparency of increasingly complex systems [24] closely tied with what the users' need to trust the system [6,7,29], and technical considerations such as model bias or discrimination [10].

These bottlenecks demonstrate how ethical challenges and human factors are inherent to implementing machine learning in healthcare [10,55]. Doing so from the outset of the development on could prevent unnecessary risks [55,57]. This stance is reflected in value sensitive design (VSD), addressing human values in new technologies [11,16], and human-centered AI (HCAI), linking ethics and practice by placing human needs and capabilities at the core of AI [23,51,52,67]. Despite these approaches and the field's many ethical value overviews, e.g. [14,22,27,37], due to the complexity and cross-disciplinary nature of the emergent field of healthcare machine learning, the current identification of ethical concerns is generally ad hoc, reactive, and fragmented [9], without reinforcement incentive [22]. Besides, literature states ethics of AI is currently often discussed on abstract level, calling for a situation-specific approach and translation of values to technical implications, while simultaneously advocating a social perspective [1,22], arguing for stakeholder engagement [1,9]. Similarly, critique on VSD calls for more participatory stakeholder consultation in value reflection and implementation [11]. Indeed, benchmarking (see section 2.3) identified a current underrepresentation of ethical AI tools that combine the end user with a comprehensive set of ethical considerations. From a human-centered and value-based perspective, focusing on early development and implementation phases, this research explores: *How can we enable clinicians to express their ethical values regarding a Machine Learning Clinical Decision Support System, using a toolkit that promotes mutual understanding of the implementation context?*

To perform this research, value sensitive design and participatory design were used to create a contextually fitting toolkit and supplementary workshop. The toolkit aims to establish a mutual understanding of the implementation context between clinician and facilitator, simultaneously creating a physical canvas to map the clinician's ethical values to, elicited by a card set of ethical AI values [27]. Using a case study ML model on cardiovascular risk developed at the Leiden University Medical Centre to explore the workshop's possible use in practice, the workshop was evaluated on usability and utility with five clinicians. The outcomes of this explorative and qualitative evaluation lead to insights and recommendations on the design and the clinical HCAI field in general. In this way, the research aims to create a starting point for those involved in clinical ML development to take a more human-centered stance.

2 RELATED WORK

2.1 Ethics in machine learning and AI

Consideration of ethical values in machine learning is crucial for its implementation [10], resulting in a rise of publications on the topic. However, these papers rarely define the terms "ethics" and "values". On overarching level a distinction is made between ethics of ML creation and use by humans, and the technological "machine ethics" [38,41]. Synthesizing the

first with definitions by the Oxford dictionary and VSD [16,68,69]1, this research defines ethical values of ML as: *people's beliefs and principles about what is morally right and important, in life and concerning a ML concept or model.*

Within found literature, ethical values of clinical end-users often surface implicitly in research with a broader scope, such as ML case study usability testing [8] and research on clinician case-study ML needs [7], or clinical stakeholder attitudes towards ML [50]. In contrast, when explicitly mentioning on ethical values, research is often zoomed in, e.g. stakeholder views on trust and collaboration [26] or on explainable AI [17]. These value-specific studies also extend beyond user level, e.g. research on explainability and "black box" medicine [4,45], accuracy versus explainability [33] and trust [6]. So, although generally featuring in literature, research combining clinician-involvement and ethical values in a comprehensive or practical manner [28] seems to be underrepresented.

On this overarching, comprehensive level are emerging overviews of ethical principles or guidelines, although there is not one universal set [22]. For example, the evaluations by Hagendorff [2020] Jobin et al. [2019] and Fjeld et al. [2020], show similar values such as 'privacy', 'transparency' and 'fairness', yet diverge on 'sustainability' or 'solidarity'. Nonetheless, definitions of diverging terms often overlap, while simultaneously matching principles can have different nuances. E.g. 'transparency' and 'explainability' differ in being separated or grouped. Found meta-analyses specific to ethical AI values in healthcare, for example [19,28,37,42,53], overlap greatly with the general AI values. Only Morley et al. distinguish ethical patterns rather than values [2019]. Literature on HCAI features similar principles such as transparent-, explainable-, ethical-, fair-, trustworthy-, responsible- and sustainable AI [23].

Due to its resemblance with research on ethical values in clinical ML and strong recognition in the research community, the overview provided by Jobin et al. [2019] was chosen as foundation of ethical values in this research.

2.2 Human-centered and value-based approaches to AI

2.2.1 HCAI and participatory design

Human-centered AI exhibits a human-centric approach, placing the user at center to avoid individual to societal ethical issues [23]. It strives for AI that amplifies human capability, being usable and based on users' needs, for example through UX design methods and participatory design [23,51,52,67]. In participatory design the end user is involved in the design process as co-designer, aiming to integrate existing skills and tacit knowledge in the process [13], and values to surface [25]. Participatory design can be executed through generative tools, utilizing hands-on toolkits to allow non-designers to express their context-specific needs and aspirations [48,49].

2.2.2 Value sensitive design

Value sensitive design (VSD) is a well-reviewed approach that proactively incorporates human values, with a focus on morality and ethics, in designing new technologies, primarily applied in the field of human-computer interaction [11,36]. VSD is based on a theoretical foundation, a "interactional stance" that entails the interrelatedness of the technology, its stakeholders human and technological values [11,16,66]. As result, VSD iteratively integrates conceptual, empirical and technological investigations [16]. The conceptual investigations comprise an analytical, theoretical or philosophical contextual analysis, identifying (in)direct stakeholders and values. Empirical investigations investigate these notions in practice, the human context, potentially through qualitative and quantitative methods stemming from social sciences.

[1] In the Oxford dictionary, ethical is defined as "morally correct or acceptable" [73], and values as "beliefs about what is right and wrong and what is important in life" [74]. Quite similar, VSD defines values as: "what is important to people in their lives, with a focus on ethics and morality." [19].

Technological investigations focus on the innovation itself [16,66]. Within the theoretical foundation, VSD established a set of methods that allow analysis of the context in a value-based manner. In section 3.1 “Workshop design”, the methods used in this research will be explained.

Umbrello and van de Poel [2021] propose a method of using VSD to design AI for social good through contextual analysis, value identification, translating values into design requirements, and prototyping [56]. The first steps, aligning with conceptual and empirical investigation, are utilized in this research, approached in a participatory manner.

2.3 A benchmark of ethical AI tools

To understand the current landscape of tools for user-centered and ethical AI, a benchmark was performed. Visualizing their aim, the tools were mapped on their targeted development stage and audience (figure 1). Appendix A1 contains a more extensive analysis. As seen in the figure, benchmarked tools predominantly focus on the internal team of designers, developers, (project) managers or organizations. Although they strive for developer awareness of the AI’s value [70,71], ethical values [46,72–77] [34,78,79], or taking on a user-perspective [47,70,74,80,81], they rarely actually involve the user in this, or do so in a broad manner [81,82]. No tools were found that specifically aim for designer-user collaboration in the ethical value creation and consideration in a comprehensive manner during development, on which this study will focus.

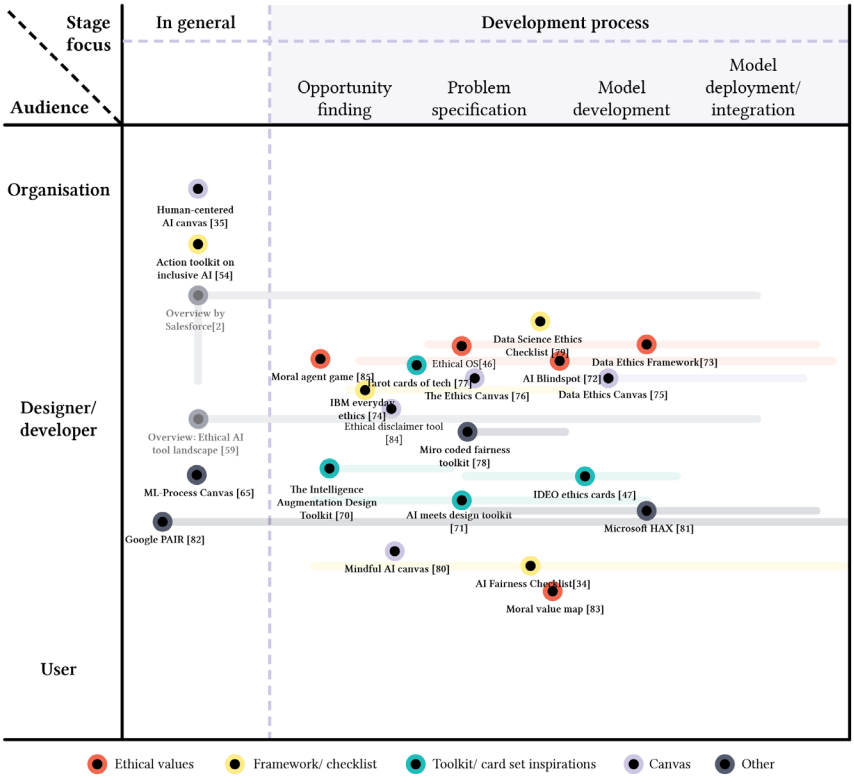


Figure 1: Benchmark of existing tools for ethical and user-centered AI. Dots represent the placement of the tool, a line is added if the tool is suited in multiple places. [2,34,35,46,47,54,54,59,65,70–78,80–85]

2.4 The research's case study

In accordance with the introduction, this research uses a ML model under development at the LUMC by van Os et al. [unpublished] as case study. As current cardiovascular risk prediction models focus on older adults and overlook gender differences, a prediction model was created for the risk of first cardiovascular events in men and women aged 30-49. The model extracts its data from electronic health records (EHR), potentially functioning as population screening. STIZON EHR data from 542,147 patients was used for data-driven risk predictor selection and training the prediction model [44].

3 METHODOLOGY

The research's two-folded process, following the double diamond structure, is visualized in appendix A2 [12]. First, converging and diverging actions of research scoping and designing the probe were intertwined with literature research (section 3.1), showing VSD conceptual investigation. Subsequently, empirically investigating, the workshop and toolkit were qualitatively evaluated (section 3.2). Expert interviews, the pilot- and final research were approved by the Ethical Review Board of the researcher's institute and conducted under informed consent (appendix A3).

3.1 Workshop design

3.1.1 Design process

Initial scoping within the current landscape of ethics in ML was performed through literature review, benchmarking and informal consultations on ML development with a MSc student and PhD candidate in the field. After scoping, semi-structured interviews with experts were conducted to validate and specify the research focus and workshop ideas. Insights from consulting a Philips PhD candidate reflected found literature: although working in a user-centered manner, specific ethical values such as "transparency" are arbitrarily validated in sprints, yet a comprehensive manner of validating ethical values is lacking. Furthermore, a program leader Impuls Hartzorg at NVVC elaborated on the complex cardiac care path. Besides, due to the research focus on clinician involvement in ML development rather than the usual evaluation, they advised maximizing the concreteness of the context of the ML to enable discussion. Interviewing PhD candidate Fan Li, working on combining ethical value cards and a Smart Service Blueprint Scape [31], validated the research scope of an ethical toolkit including the context, aimed at initial design phases.

The HCAI-, VSD- and interview-based designed probe was iteratively prepared for evaluation, validating it with the developers of the LUMC model, a pilot research conducted with three MSc Industrial Design students (appendix A4: pilot protocol and photos), the Philips PhD-er and two Philips graduation students design and data science.

3.1.2 The research probe

To foster a synthesis of HCAI and VSD, participatory design through generative techniques and VSD methods was applied to design the research probe: a workshop and toolkit. The workshop has three phases, see figure 2, with clinicians as participants and in this study facilitated by the researcher. In envisioned use, the ML developer or designer engaging in model implementation would facilitate. In phase 1 and 2 the toolkit is used to physicalize the implementation context, serving as canvas for phase 3 in which ethical values are mapped. The corresponding protocol can be found in appendix A5. Using a generative approach of describing experience through the artefact, aims at uncovering the tacit knowledge and needs of the clinician [58]. Its physicality, somewhat ambiguous modular design and whiteboard texture foster this.

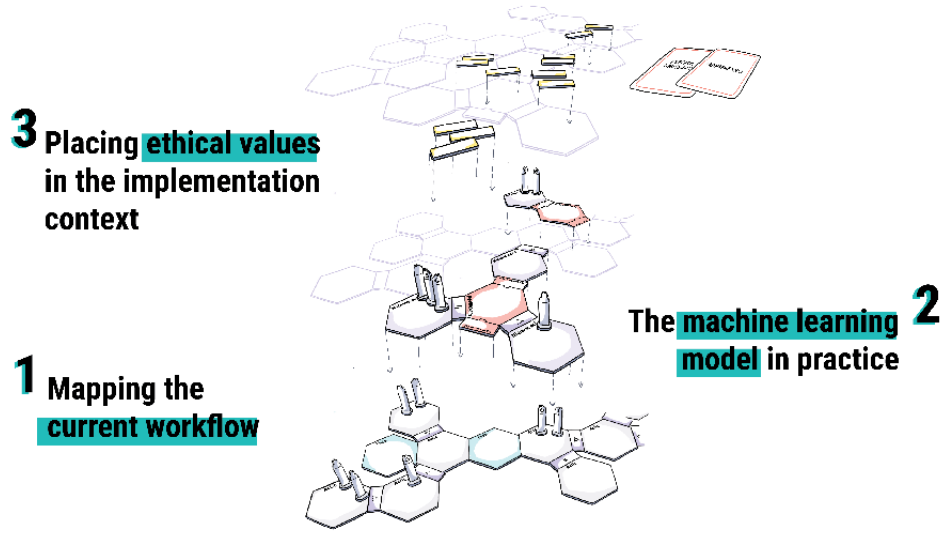


Figure 2: An overview of the workshop's phases.

Phase 1

In phase 1, the current clinician workflow is mapped through modular and physical journey [18] (see a mapping replication in figure 3). Inspired by timeline creation, a generative contextmapping technique, this phase aims to sensitize the clinician to the topic [58]. Besides, the step aims to create a mutual understanding of the context between clinician and facilitator, and to create a reference for following workshop phases. For time efficiency, a basic journey was used as starting point (appendix A6), based on the interview with the NVVC employee and consulting a general practice assistant.

Based on feedback from the pilot and the Philips PhD-er, the journey was made more concrete and versatile by creating a distinction between phase and action tiles and adding connectors to the previously one-dimensional journey (see appendix A4.2). Furthermore, while normally adopted in solitude, the VSD method of stakeholder tokens was added to the journey to concretely map out roles and relations [63].

Phase 2

In phase 2, a general introduction to machine learning is given, followed by an explanation of the case study model and its training method. The clinician and designer map the model in the workflow, adhering to VSD in making its scenario of use concrete by formulating and placing new actions and involved stakeholders [16,66] (see figure 4).

Although the use of a value scenario instead of a general explanation of the model was considered, this was anticipated to be steering, so a joint creation of the scenario was opted for [61,66]. Furthermore, while originally the model implementation was mapped by mapping the model tile into the workflow (appendix A4.2), after the pilot the envisioned implementation was made more concrete by formulating "new action" tiles in the workflow. A model diagram (appendix A5.1) and LUMC model risk predictor tables served as backup explanations.

Phase 3

In phase 3, utilizing ethical value cards as inspiration along with a value-oriented semi-structured interview [66], the designer and clinician map values in the workflow (see figure 5), aiming to translate these into design requirements [56]. The value cards are based on the literature review on ethical values in AI by Jobin et al. [2019]. Sustainability was eliminated from the set to reduce overwhelming, due to estimated contextual irrelevance. See appendix A7 for the final value cards.

While originally the tool included tiles for the clinicians' to note their value interpretation, this was eliminated after the pilot to reduce time and complexity. Furthermore, the value cards were revised to become more accessible, placing the term on the front and its simplified explanation and illustration on the back (see appendix A8 the original cards). Examples were removed and kept as facilitator backup information (see appendix A7.2).

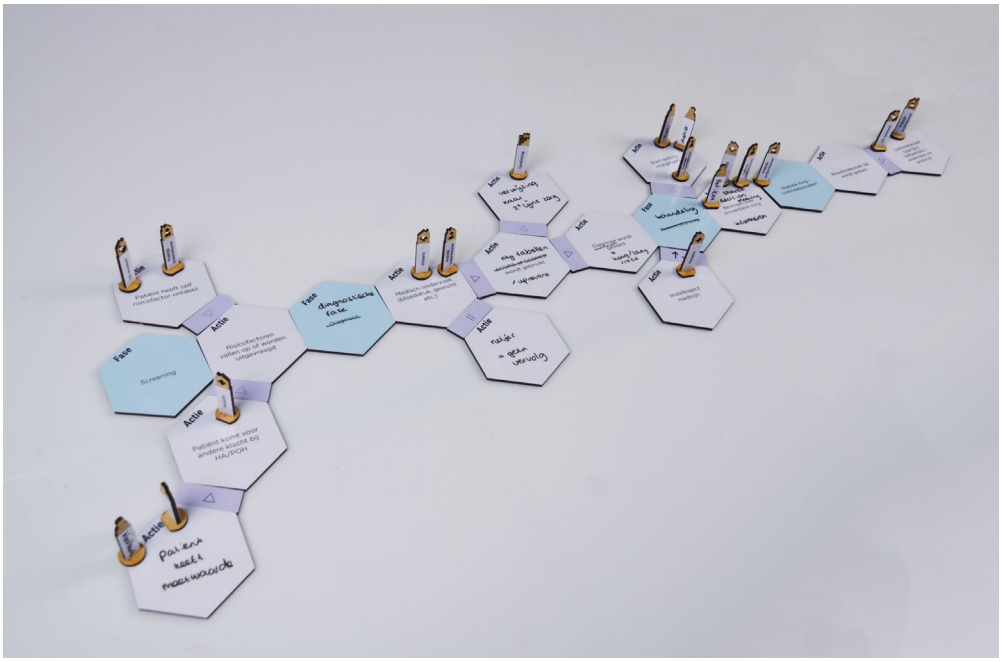


Figure 3: A possible layout of the tool after phase 1, recreated based on mappings during the research.

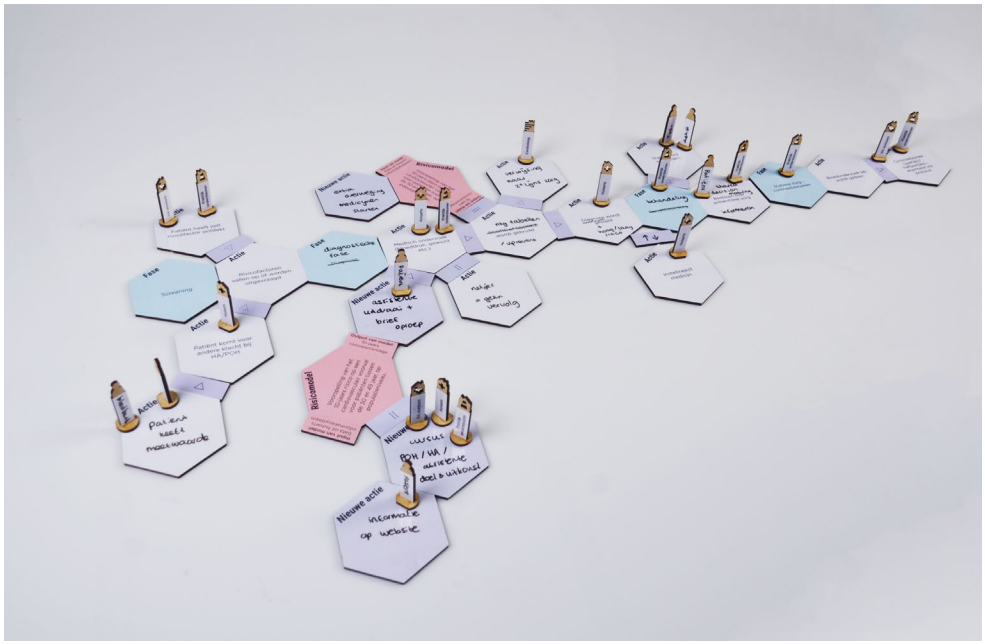


Figure 4: A possible layout of the tool after phase 2, recreated based on mappings during the research.

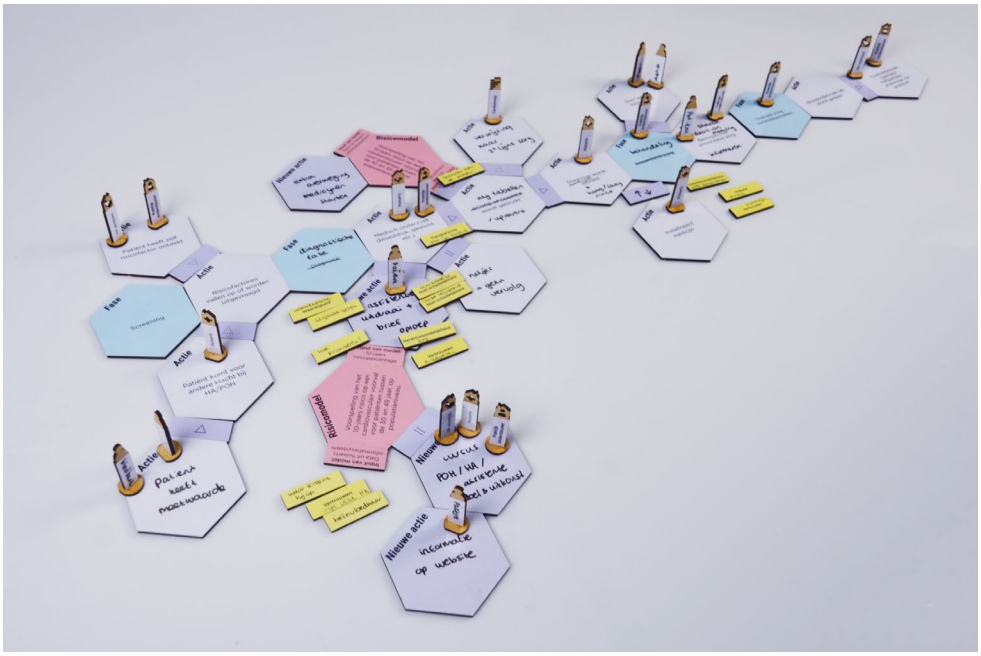


Figure 5: A possible layout of the tool after phase 3, recreated based on mappings during the research.

3.2 Evaluation of the workshop

3.2.1 Procedure

The case study model by the LUMC is intended to be implemented in general practice, where cardiac primary preventative care is generally performed by general practitioners (GP) and general practice assistants (GPA). Therefore the research was conducted with two GPs and three GPA's, a number proven effective in usability research [43]. A protocol change on the explained model accuracy and a consultation validating the general journey before testing makes the first GPA test (p0) a pilot. The workshop including evaluation lasted between 1 hour and 45 minutes to 2 hours and 45 minutes, see the timeline in appendix A5.

3.2.2 Analysis

Analysis was based on toolkit evaluation research. The tested workshop and its objective of value expression inhibits empirical research through an artefact [60]. Therefore, evaluation a combination of holistic observations and data gathering through conducting the workshop, followed by a semi-structured interview on the user experience was chosen [60]. This two-folded research approach allowed investigation of the workshop usability and utility, to ensure the tool is conceptually clear, easy to use, and valuable [20,30].

Workshop and evaluation transcripts were theoretically and inductively thematically analyzed [5]. Theoretically, staying close to the research objective, overall themes resembled usage and utility of the workshop phases. Inductively, codes and additional themes were extracted to utilize the rich data. Final themes and codes can be found in appendix A10. Video recordings and photographs were used for analyzing interactions with the toolkit.

4 RESULTS

Results will be discussed adhering globally to the thematic structure of the toolkit's usage and utility. P1 and p2 were GP's, p0, p3 and p4 GPA's. Again, p0 was a pilot due to accuracy numbers being mentioned. Because of the small sample size and broad workshop coverage extending far beyond accuracy, their insights are included in results yet always indicated with p0. The created toolkit mappings are displayed per test in appendix A9. A more elaborate version of the results can be found in appendix A12.

4.1 Mapping the implementation context

A general pattern in the usage of phase 1 is clinician initiating questions or elaborations on the tiles of the basic journey, leading to summarizing facilitator suggestions to which participants agreed or provided alternatives. Stakeholder relations, actions and facilitator validations often inspired each other. Furthermore, the physical journey reminded both participant and facilitator to return to previously mentioned actions (all), or aided participants in revising the journey's completeness (p0,p1,p4)[2]. The overarching 'Phase' tiles were confused with actions (all), 'Experience' tiles and connectors were mostly used by the facilitator (p1,p2,p3,p4). In the end evaluation phase 1 was often mentioned as fun, specificity giving new insights into existing work patterns (p1,p3,p4).

As for the usage of phase 2, while some immediately envisioned the model as the developers' intended population screening (p0,p2), other participants placed it here later, after assumptions, questions and possible misunderstandings by the participant surfaced throughout the phase. Only p1 deviated by suggesting governmental screening, meeting

[2] Participants will be referred to as p1, p2, with p0 being the pilot.

requirements of consent and accuracy. All participants added an additional implementation placement during care delivery. Participants often hesitated mapping the envisioned implementation, asking for validation, not initiating new actions or formulating these based on previous workflows, and mentioning difficulties envisioning the future in the end evaluation. Evaluating, p3 indicated ongoing uncertainty on her model understanding, p0 and p4 brought up satisfaction with the end overview.

4.2 Expressing ethical values

The mapping of the model (phase 2) and its values (phase 3) turned out to inspire each other. In phase 2 ethical issues (p1,p2) or other considerations (all) came up reacting to the model and making implementation and actions specific. E.g. p0 on the call-in patient letter: *"if you properly explain, 'research has showed that...'... Hm that's difficult, it becomes ethical. Because people with lower SES... You can't put that in a letter."*[3]. Vice versa, implementation was sometimes altered after values surfaced in phase 3.

As for cards usage, participants differed in not using the back-side explanation (p2) or using it to understand values (p1,p0), throughout their value selection (p3,p4) or when reviewing chosen values (p0). The explanations did not always provide clarity *"This is too difficult. […] It is way too much information."* (p1). Specifically 'beneficence' was vague (p0,p2,p3,p4). Interaction-wise, most participants mapped cards in the journey or sorted cards. P4 laid them out statically and p3 stacked discussed ones.

Mentioned values or model requirements did not always explicitly belong to the card values, so the facilitator noted them on an empty value tile, or suggested an interpretation. Participants associated values with different implications, so improvising, the facilitator made notes on the small space on the value tiles (see appendix A9). Furthermore, participants commented on Jobin et al.'s terms overlapping (p0,p3) or not grasping meaning, e.g. 'trust' not covering 'trustworthiness' (p1). As for placement, the facilitator often took initiative in suggesting or actively asking. If not translating to concrete model implications, participants expressed placement difficulties due to terms being general (p0,p2,p3,p4), unconscious, or a natural part of healthcare (p3,p4). P4: *"Yeah… When do people start to, for example, trust you?"*.

Overall, participants mentioned the value cards providing guidance: *If you have to think of them yourself I probably would not have mentioned them." (*p2), and specificity *"You make the aspects more concrete. […] What element does it influence?" (*p1). P2 did mention more values could be added beforehand or by participants, reasoning: "*This also limits people right? If you only provide a few.".* P0 mentioned the cards inspired, but was glad not all had to be mapped due to placement difficulties.

4.2.1 Expressed values

See appendix A12.2.2 for elaborate results on mentioned values. When using the value cards, values mentioned as essential, but not previously mentioned in phase 2 surfaced. Values often naturally translated into model expectations, questions or requirements. As seen in the overview in appendix A11, it stands out that most values, apart from beneficence, were mentioned in all tests, although sometimes differing in interpretation. Remarkably, 'transparency', 'privacy' and 'trust' were mentioned earlier in the conversation than others.

While overlapping, values and specifically mentioned prerequisites also differed between participants. For example, only p1 and p2 were concerned with the model's "value", interpreted as accuracy, and "patient permission" determined p1's model placement and p4's actions, but was not mentioned by others.

[3] Participant quotes are translated from Dutch, so nuances might differ from the original.

4.3 Evaluating workshop utility

Evaluating workshop utility, participants mentioned ethics being important for ML to fit clinical practice (p0, p1,p4): *"What are the ethical aspects and where does that fit in daily practice, why does a risk model need to be used, and by who? […] That seems very valuable to me."* (P1). Besides, participants mentioned the workshop potentially surfacing value patterns between participants (p0,p2). The GP's were most positive about the ML implementation (p1) and ethics discussion (p1,p2). Other participants indicated the ethical values were the most difficult part of the workshop (p0,p3,p4). P0 reasoned to personally find ethics a difficult topic, p4 struggled with the general nature of the values, and p3 mentioned difficulties with the conceptual stage of the model, needing to remind themselves of the workshop objective of the ethics relating to the model. The workshop phases were mentioned to surface new insights throughout (p0,p2), following each other up logically (p3) and required to build onto each other: p2: *"I need this [the workflow] to be able to think of it, where should this come [ML]? You build onto it".* Lastly, physicality was seen as conversation starter (p1,p3) and helped in thinking (p2). However, p2 did mention interactions by the facilitator made her think less actively about tile placement. The workshop was experienced as intensive but no information was redundant (p0,p2,p3).

5 DISCUSSION AND FUTURE WORK

This section discusses the interpretation of results, adhering to the research question: *How can we enable clinicians to express their ethical values regarding a Machine Learning Clinical Decision Support System, using a toolkit that promotes mutual understanding of the implementation context?* Insights and suggestions for future work cover the probe itself and abstractions to the field of clinical HCAI. The section concludes with limitations of the study. A more elaborate discussion is found in appendix A13.

5.1 Creating a mutual understanding of the ML implementation context

For the facilitator, the interactive mapping of the current workflow and stakeholders created context of the clinicians' needs and expectations of the model. Inversely, the concrete envisioned implementation mapping, interchanged with questions or clarifications, showed the facilitator the participant's vision of the model and iteratively increased clinicians knowledge. Abstracting this insight to clinician involvement in ML, establishing a common ground between developer and clinician is recommended due to its potential in creating both a shared mental foundation of knowledge, as well as a physical foundation to discuss specifics of the ML.

5.2 Starting a conversation on ethical values

While the discussion on ethical values was the main goal of phase 3 of the workshop, unexpectedly values organically surfaced in phase 2, showing how concrete implementation already triggers and enables clinician value elicitation. Furthermore, the value cards proved effective: besides their value being mentioned in the evaluation, it stood out they sparked conversation both on values previously mentioned as well as new insights, and often resulted in model prerequisites. Whereas phase 2 values – apart from p2 - focused more on the practical requirements and effects of the tool, in phase 3 more dimensions of the tool were discussed. See the elaborated discussion on mentioned values in appendix A13.2.1. However, due to all users selecting almost all values, and p3 and p4 seemingly feeling expected to, it could be argued that providing a set of values can steer or bias participants, also a point of discussion in VSD [11,16]. Concluding, both making implementation concrete (phase 2), as well as providing an ethical framework (phase 3) elicits value discussion, yet a recommendation for future work is to further investigate the balance of unguided elicitation and (different selections of) outlined values in consulting clinical end users.

Content-wise, mentioned values overlapped, yet unexpectedly some major prerequisites differed between participants. Because of this, a suggestion for future end-user involvement is exploring the VSD "value dams and flows" method, implementing model prerequisites throughout evaluation with different participants [66]. Another suggestion would be to explore the research's workshop in group sessions, to encourage stakeholder discussion.

Furthermore, unexpectedly, the cards term-side, was used predominantly compared to the backside explanation, and examples were not discussed. Although causes could be card (un)clarity, the workshop also did not provide space – cognitive or time-wise - to explore the values in-depth. This links to another insight: although the first two workshop phases created a contextual canvas, they diverged workshop focus from the values. A future suggestion is to conduct the conversation on ML ethics in multiple sessions, separating the envisioned ML implementation from a session treating the ethical values in-depth. Besides, an iteration reviewing the card content could make their exploration threshold lower.

5.3 Physicality and modularity

The physicality of the tool made all discussed matters – workflows, the use of the ML and values – very concrete in an time-efficient manner. Through this concreteness, physicality unexpectedly contributed to eliciting ethical values. Furthermore, evaluation comments and interactions proved physicality and modularity as conversation catalyst. Practical workshop suggestions would be to iterate on the 'phase' tiles, include writing space on the value tiles and include more ambiguous tiles to capture unexpected workshop input.

It did stand out that the facilitator interacted more with the tool than the participant. While time efficient, and the mutual summarizing and adapting of content feeling like a joint mapping, clinician non-interaction was also commented on to cause less active thought on placement. To align more with participatory design, conducting the workshop in multiple sessions would create space for more interaction by the participant.

5.4 Multi-stakeholder perspectives

Amongst the research's limited sample size GPs seemed more comfortable discussing ethics than GPAs, both seen in workshop interactions as in the end evaluation. Evaluation comments linked this to different background knowledge (p4), personal preference (p0,p3) and unclarity of the workshop objective (p3). One of the resulting recommendations for clinical HCAI is to thoughtfully adapt background knowledge and questions to the audience, and making sure the objective of clinician involvement is communicated concretely. Furthermore, the effect of group setting involvement on participant confidence levels is a suggested exploration.

Lastly, the created tool is designed for, but not limited to the healthcare setting. Therefore, future developers in different contexts can explore the use of (separate elements of) the toolkit with the end users or stakeholders of their model.

5.5 Limitations of the study

A limitation of the study is that the facilitator did not have prior experience with ML development and was not part of the LUMC model development team. Mitigation of this affecting protocol explanations was attempted through coach revision, developer consultation and pilot testing. However, researcher background comprised (technical) details on the LUMC model and sometimes answers to clinician questions, possibly influencing participants' understanding of model capabilities and limitations. An example is the faulty and eventually eliminated pilot accuracy. However, researcher novelty did add a first-person perspective to the research, fitting the participants introduction to ML. Besides, the conceptual LUMC model description was perhaps more realistic in the research focus on the initial development stage.

Related to the researchers novelty, yet caused by a time-wise scoping restriction, the workshop's translation into practice was not incorporated in this research. Workshop value did come forward in expert interviews, indicating positivity towards more concrete workflow mapping and the novelty of involving end-users comprehensively in ethical values. However, future work could examine both the developers' ability to facilitate the workshop, as well as its utility in the translation of results to the developer team and to the technical implications for the ML. Examining utility for developers would also provide future guidelines on how explicitly values should conform with value sets, and whether specifying exact placement of values, which participants experienced difficulties with, is needed.

The use of a set of general AI values was chosen for its comprehensiveness and simultaneous ambiguity, offering overview yet inspiration. After the research, a literature review on ethical values specific to the healthcare context has been published [32]. This overview greatly overlaps with Jobin er al.'s [2019], though discrepancies show. Besides, this overview does contain sustainability, showing that eliminating the term from this research was a misjudgment. Building onto the previous suggestion of exploring the costs and benefits of a (different) set of values, future work could explore using ethical healthcare AI values.

Lastly, a general research limitation is the low sample size, potentially making individual differences between participants more apparent.

6 CONCLUSION

A toolkit and supplementary workshop were designed to explore how clinicians can be enabled to express their ethical values regarding a Machine Learning Clinical Decision Support System. This design toolkit embedded VSD and participatory design to promote mutual understanding of the implementation context and direct the conversation towards ethical values, aligning the ML development process with HCAI. The toolkit featured 1) mapping the current workflow for to sensitize facilitator and clinician and to serve as canvas for step 2) mapping the envisioned implementation workflow, and 3) mapping ethical values guided by value cards. Evaluating the toolkit usage and utility with GPs and GPAs leads to several contributions to the field of ethical HCAI. Firstly, the benefits of creating a joint understanding surfaced, as concretely mapping the envisioned implementation educated the facilitator and both uncovered model unclarities for the clinician, as well as elicited initial values. Physicality allowed workflow and value mapping to be intertwined throughout the phases, iteratively creating overview. The usage of ethical value cards proved effective in sparking value discussion, leading to model expectations and prerequisites. However, the balance between inspiration versus bias through the usage of a predetermined value set is a discussion point to be researched. Other recommendations for future work includes investigating whether the workshop and its results could be adopted by the ML design- and development team and exploring the workshop in group setting or multiple sessions. In this way, the toolkit potentially contributes to a more ethical value-oriented, human-centered AI development process.

ACKNOWLEDGMENTS

First of all, I would like to greatly thank Sara Colombo and Loes van Renswouw for coaching me throughout the project, and Lu Yuan, Indre Kalinauskaite and Nick Ruijs for their expert input. Thanks to all participating in the pilot- and final research. Thank you to María Villabos for welcoming me on the LUMC team, and Margot Rakers and Hine van Os for the explanations on the development of the DECIDE model and for your time and input while developing and evaluating my workshop. Lastly, a huge thanks to Roosmarijn Ovaa, Sabeth Diks, Alicia Poels, Jolanda Faber and Britt Stegers for help content-wise, and always being available for me to bounce ideas off of.

REFERENCES

[1] Kars Alfrink, Ianus Keller, Gerd Kortuem, and Neelke Doorn. 2022. Contestable AI by Design: Towards a Framework. *Minds and Machines*. https://doi.org/10.1007/s11023-022-09611-z

[2] Kathy Baxter. 2019. Ethical AI frameworks, tool kits, principles, and certifications - Oh my! *Salesforce AI Research*. Retrieved January 13, 2023 from https://blog.salesforceairesearch.com/frameworks-tool-kits-principles-and-oaths-oh-my/

[3] Andrew L. Beam and Isaac S. Kohane. 2018. Big Data and Machine Learning in Health Care. *JAMA* 319, 13: 1317. https://doi.org/10.1001/jama.2017.18391

[4] Jens Christian Bjerring and Jacob Busch. 2021. Artificial Intelligence and Patient-Centered Decision-Making. *Philosophy & Technology* 34, 2: 349–371. https://doi.org/10.1007/s13347-019-00391-6

[5] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2: 77–101. https://doi.org/10.1191/1478088706qp063oa

[6] Jacob T Browne, Saskia Bakker, Bin Yu, Jeroen Raijmakers, Jon Pluyter, Nick Ruijs, and Eva Deckers. 2022. Towards a Framework for Trust in Clinical AI: Expanding the Unit of Analysis. 10.

[7] Carrie J. Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2019. "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW: 104:1-104:24. https://doi.org/10.1145/3359206

[8] Carmen Carroll, Phil Marsden, Pat Soden, Emma Naylor, John New, and Tim Dornan. 2002. Involving users in the design and usability evaluation of a clinical decision support system. *Computer Methods and Programs in Biomedicine* 69, 2: 123–135. https://doi.org/10.1016/S0169-2607(02)00036-6

[9] Danton S. Char, Michael D. Abràmoff, and Chris Feudtner. 2020. Identifying Ethical Considerations for Machine Learning Healthcare Applications. *The American Journal of Bioethics* 20, 11: 7–17. https://doi.org/10.1080/15265161.2020.1819469

[10] Danton S. Char, Nigam H. Shah, and David Magnus. 2018. Implementing Machine Learning in Health Care — Addressing Ethical Challenges. *The New England journal of medicine* 378, 11: 981–983. https://doi.org/10.1056/NEJMp1714229

[11] Janet Davis and Lisa Nathan. 2015. Value Sensitive Design: Applications, Adaptations, and Critiques. *Handbook of Ethics, Values, and Technological Design: Sources, Theory, Values and Application Domains*: 11–40. https://doi.org/10.1007/978-94-007-6970-0_3

[12] Design Council. What is the framework for innovation? Design Council's evolved Double Diamond. *Design Council*. Retrieved May 10, 2021 from https://www.designcouncil.org.uk/news-opinion/what-framework-innovation-design-councils-evolved-double-diamond

[13] Pelle Ehn. 2008. Participation in Design Things. 101. https://doi.org/10.1145/1795234.1795248

[14] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. 2020. Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. https://doi.org/10.2139/ssrn.3518482

[15] Alexander L. Fogel and Joseph C. Kvedar. 2018. Artificial intelligence powers digital medicine. *npj Digital Medicine* 1, 1: 1–4. https://doi.org/10.1038/s41746-017-0012-2

[16] Batya Friedman and David Hendry. 2019. Value Sensitive Design: Shaping Technology with Moral Imagination. *MIT Press*. Retrieved December 29, 2022 from https://mitpress.mit.edu/9780262039536/value-sensitive-design/

[17] Julie Gerlings, Millie Søndergaard Jensen, and Arisa Shollo. 2022. Explainable AI, But Explainable to Whom? An Exploratory Case Study of xAI in Healthcare. In *Handbook of Artificial Intelligence in Healthcare: Vol 2: Practicalities and Prospects*, Chee-Peng Lim, Yen-Wei Chen, Ashlesha Vaidya, Charu Mahorkar and Lakhmi C. Jain (eds.). Springer International Publishing, Cham, 169–198. https://doi.org/10.1007/978-3-030-83620-7_7

[18] Sarah Gibbons. 2018. Journey Mapping 101. *Nielsen Norman Group*. Retrieved May 11, 2021 from https://www.nngroup.com/articles/journey-mapping-101/

[19] Magali Goirand, Elizabeth Austin, and Robyn Clay-Williams. 2021. Implementing Ethics in Healthcare AI-Based Applications: A Scoping Review. *Science and Engineering Ethics* 27, 5: 61. https://doi.org/10.1007/s11948-021-00336-3

[20] Saul Greenberg and Bill Buxton. 2008. Usability evaluation considered harmful (some of the time). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), 111–120. https://doi.org/10.1145/1357054.1357074

[21] Robert A. Greenes. 2014. *Clinical Decision Support: The Road to Broad Adoption*. Academic Press.

[22] Thilo Hagendorff. 2020. The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines* 30, 1: 99–120. https://doi.org/10.1007/s11023-020-09517-8

[23] Maria Hartikainen, Kaisa Väänänen, Anu Lehtiö, Saara Ala-Luopa, and Thomas Olsson. 2022. Human-Centered AI Design in Reality: A Study of Developer Companies' Practices. *Nordic Human-Computer Interaction Conference*. Retrieved October 19, 2022 from https://dl.acm.org/doi/fullHtml/10.1145/3546155.3546677

[24] Andreas Holzinger. 2018. From Machine Learning to Explainable AI. In *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, 55–66. https://doi.org/10.1109/DISA.2018.8490530

[25] Ole Sejer Iversen, Kim Halskov, and Tuck Wah Leong. 2010. Rekindling values in participatory design. In *Proceedings of the 11th Biennial Participatory Design Conference on - PDC '10*, 91. https://doi.org/10.1145/1900441.1900455

[26] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3411764.3445385

[27] Anna Jobin, Marcello Ienca, and Effy Vayena. 2019. Artificial Intelligence: the global landscape of ethics guidelines. 42.

[28] Golnar Karimian, Elena Petelos, and Silvia M. A. A. Evers. 2022. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. *AI and Ethics*. https://doi.org/10.1007/s43681-021-00131-7

[29] Sujata Khedkar, Vignesh Subramanian, Gayatri Shinde, and Priyanka Gandhi. 2019. Explainable AI in Healthcare. https://doi.org/10.2139/ssrn.3367686

[30] David Ledo, Steven Houben, Jo Vermeulen, Nicolai Marquardt, Lora Oehlberg, and Saul Greenberg. 2018. Evaluation Strategies for HCI Toolkit Research. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI '18), 1–17. https://doi.org/10.1145/3173574.3173610

[31] Fan Li and Yuan Lu. 2021. Engaging End Users in an AI-Enabled Smart Service Design - The Application of the Smart Service Blueprint Scape (SSBS) Framework. *Proceedings of the Design Society* 1: 1363–1372. https://doi.org/10.1017/pds.2021.136

[32] Fan Li, Nick Ruijs, and Yuan Lu. 2023. Ethics & AI: A Systematic Review on Ethical Concerns and Related Strategies for Designing with AI in Healthcare. *AI* 4, 1: 28–53. https://doi.org/10.3390/ai4010003

[33] Alex John London. 2019. Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Center Report* 49, 1: 15–21. https://doi.org/10.1002/hast.973

[34] Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (CHI '20), 1–14. https://doi.org/10.1145/3313831.3376445

[35] Albéric Maillet. 2019. Introducing the Human-Centered AI Canvas. *Medium*. Retrieved January 3, 2023 from https://medium.com/@albmllt/introducing-the-human-centered-ai-canvas-a4c9d2fc127e

[36] Noëmi Manders-Huits. 2011. What Values in Design? The Challenge of Incorporating Moral Values into Design. *Science and Engineering Ethics* 17, 2: 271–287. https://doi.org/10.1007/s11948-010-9198-2

[37] Stuart McLennan, Hannes Kahrass, Susanne Wieschowski, Daniel Strech, and Holger Langhof. 2018. The spectrum of ethical issues in a Learning Health Care System: a systematic qualitative review. *International Journal for Quality in Health Care* 30, 3: 161–168. https://doi.org/10.1093/intqhc/mzy005

[38] J.H. Moor. 2006. The Nature, Importance, and Difficulty of Machine Ethics. *IEEE Intelligent Systems* 21, 4: 18–21. https://doi.org/10.1109/MIS.2006.80

[39] Jessica Morley, Caio Machado, Christopher Burr, Josh Cowls, Mariarosaria Taddeo, and Luciano Floridi. 2019. The Debate on the Ethics of AI in Health Care: A Reconstruction and Critical Review. https://doi.org/10.2139/ssrn.3486518

[40] Clara Mosquera-Lopez, Sos Agaian, Alejandro Velez-Hoyos, and Ian Thompson. 2015. Computer-Aided Prostate Cancer Diagnosis From Digitized Histopathology: A Review on Texture-Based Systems. *IEEE Reviews in Biomedical Engineering* 8: 98–113. https://doi.org/10.1109/RBME.2014.2340401

[41] Vincent C. Müller. 2020. Ethics of Artificial Intelligence and Robotics. *Stanford Encyclopedia of Philosophy*. Retrieved December 31, 2022 from https://plato.stanford.edu/entries/ethics-ai/?fbclid=IwAR3zBI5BYERCGCdEBZhAvLHExNJhPUJA9SYkvwteRUdmXBgB3ILfUk6y81o

[42] Kathleen Murphy, Erica Di Ruggiero, Ross Upshur, Donald J. Willison, Neha Malhotra, Jia Ce Cai, Nakul Malhotra, Vincci Lui, and Jennifer Gibson. 2021. Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Medical Ethics* 22, 1: 14. https://doi.org/10.1186/s12910-021-00577-8

[43] Jakob Nielsen. 2000. Why You Only Need to Test with 5 Users. *Nielsen Norman Group*. Retrieved January 4, 2023 from https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

[44] Hendrikus J.A. van Os, Jos P. Kanning, Tobias N. Bonten, Margot M. Rakers, Hein Putter, Mattijs E. Numans, Ynte M. Ruigrok, Rolf H.H. Groenwold, and Marieke J.H. Wermer. Cardiovascular risk prediction in men and women under 50 years using routine care data, unpublished.

[45] Jeremy Petch, Shuang Di, and Walter Nelson. 2022. Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology. *Canadian Journal of Cardiology* 38, 2: 204–213. https://doi.org/10.1016/j.cjca.2021.09.004

[46] Goncalo Praca. Ethical OS Toolkit - a guide to anticipating the future impact of today's technology. *RRI tools*. Retrieved January 3, 2023 from https://rri-tools.eu/-/ethical-os-toolkit

[47] Ovetta Sampson and Michael Chapman. AI Needs an Ethical Compass. This Tool Can Help. *IDEO*. Retrieved January 3, 2023 from https://www.ideo.com/blog/ai-needs-an-ethical-compass-this-tool-can-help

[48] E. B.-N. Sanders. 2000. Generative Tools for Co-designing. In *Collaborative Design*, Stephen A. R. Scrivener, Linden J. Ball and Andrée Woodcock (eds.). Springer London, London, 3–12. https://doi.org/10.1007/978-1-4471-0779-8_1

[49] L Sanders and P.J. Stappers. 2012. *Convivial toolbox: Generative Research for the Front End of Design.* BIS Publishers, Amsterdam.

[50] Ian A Scott, Stacy M Carter, and Enrico Coiera. 2021. Exploring stakeholder attitudes towards AI in clinical practice. *BMJ Health & Care Informatics* 28, 1: e100450. https://doi.org/10.1136/bmjhci-2021-100450

[51] Ben Shneiderman. 2020. Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. *ACM Transactions on Interactive Intelligent Systems* 10, 4: 26:1-26:31. https://doi.org/10.1145/3419764

[52] Ben Shneiderman. 2022. *Human-Centered AI*. Oxford University Press.

[53] Marianne W. M. C. Six Dijkstra, Egbert Siebrand, Steven Dorrestijn, Etto L. Salomons, Michiel F. Reneman, Frits G. J. Oosterveld, Remko Soer, Douglas P. Gross, and Hendrik J. Bieleman. 2020. Ethical Considerations of Using Machine Learning for Decision Support in Occupational Health: An Example Involving Periodic Workers' Health Assessments. *Journal of Occupational Rehabilitation* 30, 3: 343–353. https://doi.org/10.1007/s10926-020-09895-x

[54] Ela Stopford Sackville, Anusha Raghav, and Sophie Lambin. Action Toolkit On Inclusive AI.

[55] Mark Sujan, Dominic Furniss, Kath Grundy, Howard Grundy, David Nelson, Matthew Elliott, Sean White, Ibrahim Habli, and Nick Reynolds. 2019. Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ health & care informatics* 26, 1: e100081. https://doi.org/10.1136/bmjhci-2019-100081

[56] Steven Umbrello and Ibo van de Poel. 2021. Mapping value sensitive design onto AI for social good principles. *Ai and Ethics* 1, 3: 283–296. https://doi.org/10.1007/s43681-021-00038-3

[57] Effy Vayena, Alessandro Blasimme, and I. Glenn Cohen. 2018. Machine learning in medicine: Addressing ethical challenges. *PLOS Medicine* 15, 11: e1002689. https://doi.org/10.1371/journal.pmed.1002689

[58] Froukje Sleeswijk Visser, Pieter Jan Stappers, Remko van der Lugt, and Elizabeth B.-N. Sanders. 2005. Contextmapping: experiences from practice. *CoDesign* 1, 2: 119–149. https://doi.org/10.1080/15710880500135987

[59] Edwin Wenink. 2021. Ethical AI Tool Landscape. *edwinwenink.github.io*. Retrieved January 13, 2023 from https://edwinwenink.github.io/ai-ethics-tool-landscape/

[60] Jacob O. Wobbrock and Julie A. Kientz. 2016. Research contributions in human-computer interaction. *Interactions* 23, 3: 38–44. https://doi.org/10.1145/2907069

[61] Jill Woelfer, Amy Iverson, David Hendry, Batya Friedman, and Brian Gill. 2011. Improving the safety of homeless young people with mobile phones: Values, form and function. 1707–1716. https://doi.org/10.1145/1978942.1979191

[62] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19), 1–11. https://doi.org/10.1145/3290605.3300468

[63] Daisy Yoo. 2021. Stakeholder Tokens: a constructive method for value sensitive design stakeholder analysis. *Ethics and Information Technology* 23, 1: 63–67. https://doi.org/10.1007/s10676-018-9474-4

[64] A. Zamora, F. Fernández de Bobadilla, C. Carrion, G. Vázquez, G. Paluzie, R. Elosua, M. Vilaseca, A. Martín-Urda, A. Rivera, N. Plana, L. Masana, VALIDA Study Group, and Network of Lipid Units of Catalonia (XULA). 2013. Pilot study to validate a computer-based clinical decision support system for dyslipidemia treatment (HTE-DLP). *Atherosclerosis* 231, 2: 401–404. https://doi.org/10.1016/j.atherosclerosis.2013.09.029

[65] Zhibin Zhou, Qing Gong, Zheting Qi, and Lingyun Sun. 2019. ML-Process Canvas: A Design Tool to Support the UX Design of Machine Learning-Empowered Products. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI EA '19), 1–6. https://doi.org/10.1145/3290607.3312859

[66] 2020. About VSD. *VSD Lab*. Retrieved December 29, 2022 from https://vsdesign.org/vsd/

[67] 2021. What is human-centered AI? *IBM Research Blog*. Retrieved December 29, 2022 from https://research.ibm.com/blog/what-is-human-centered-ai

[68] Ethical - adjective. *Oxford Learner's Dictionary*. Retrieved December 31, 2022 from https://www.oxfordlearnersdictionaries.com/definition/english/ethical?q=ethical

[69] Value - noun. *Oxford learner's dictionaries*. Retrieved December 31, 2022 from https://www.oxfordlearnersdictionaries.com/definition/english/value_1?q=values

[70] The Intelligence Augmentation Design Toolkit. *Futurice*. Retrieved January 3, 2023 from https://futurice.com/ia-design-kit

[71] AI Meets Design Toolkit. *AIxDesign*. Retrieved January 3, 2023 from https://www.aixdesign.co/toolkit

[72] AI Blindspot: A Discovery Process for preventing, detecting, and mitigating bias in AI systems. *AI Blindspot*. Retrieved January 3, 2023 from http://aiblindspot.media.mit.edu/

[73] Data Ethics Framework. *GOV.UK*. Retrieved January 3, 2023 from https://www.gov.uk/government/publications/data-ethics-framework

[74] Everyday ethics for AI. *IBM Design for AI*. Retrieved January 3, 2023 from https://www.ibm.com/design/ai/ethics/everyday-ethics/www.ibm.com/design/ai/ethics/everyday-ethics

[75] The Data Ethics Canvas. *The ODI*. Retrieved January 3, 2023 from https://www.theodi.org/article/the-data-ethics-canvas-2021/

[76] The Ethics Canvas. *Online ethics canvas*. Retrieved January 3, 2023 from https://www.ethicscanvas.org/

[77] The Tarot Cards of Tech. *Artefact*. Retrieved January 3, 2023 from https://www.artefactgroup.com/case-studies/the-tarot-cards-of-tech/

[78] Coded Fairness Toolkit template. *Miro - miroverse*. Retrieved January 3, 2023 from https://miro.com/miroverse/coded-fairness-toolkit/

[79] About — Deon. *Deon*. Retrieved January 3, 2023 from https://deon.drivendata.org/#data-science-ethics-checklist

[80] Introducing The Mindful AI Canvas. *Linkedin*. Retrieved January 3, 2023 from https://www.linkedin.com/pulse/introducing-mindful-ai-canvas-mike-edmonds/

[81] Microsoft HAX Toolkit. *Microsoft*. Retrieved January 3, 2023 from https://www.microsoft.com/en-us/haxtoolkit/

[82] People + AI Research (PAIR). *Google research*. Retrieved from https://research.google/teams/brain/pair/

[83] Ethics for Designers — Moral value map. *Ethics for Designers*. Retrieved January 3, 2023 from https://www.ethicsfordesigners.com/tools-1/moral-value-map

[84] Ethics for Designers — Ethical Disclaimer tool. *Ethics for Designers*. Retrieved January 3, 2023 from https://www.ethicsfordesigners.com/ethical-disclaimer

[85] Ethics for Designers — Moral Agent game. *Ethics for Designers*. Retrieved January 3, 2023 from https://www.ethicsfordesigners.com/moral-agent

A APPENDICES

A0 Reflection
A1 Extensive benchmark text
A2 Process visualization
A3 Ethical Review Board Approvals and consent forms
A4 Pilot
A5 Research protocol
    A5.1 LUMC model diagram
A6 Starting journey
A7 Final value cards
A8 Value cards iteration 1
A9 Results: photos per participant
A10 Themes and codes
A11 Expressed values tables
A12 Extensive results
    A12.2.1 Expressed values
A12 Extensive discussion
    A13.2.1 Expressed values

A0 REFLECTION

The project I carried out over the past semester allowed both a gain in practical skills as well as understanding of the overarching design research process. Together, this facilitated development within my professional identity and vision on the design field and society.

A0.1 The expertise areas

A0.1.1 User and society (US), creativity and aesthetics (CE) and technology and realization (TR)

Due to my interest in creative facilitation and design tools, a longstanding goal was creating a design toolkit myself. An insight in doing so is how much goes into designing a workshop from scratch, carefully considering the audience, materials, and envisioned results to reach the objective and ensure relevance in the grander research context (US and CE). Subsequently, the realization of the toolkit has to fit all these dimensions as well (CE and TR). Although I could use my background experience in grounded and empathic approaches, it was new to select and merge general approaches or theoretical background into a tool (US & CE). These insights together have validated the part of my identity that thrives on acquiring new perspectives through human-centered creative methods, making me certain I want to expand this expertise in future projects.

Furthermore, relating to CE, it was interesting to notice the required perspective shift in communicating the research results to different audiences. I gained presenting skills in adapting my explanation to Demoday visitors with no background in research, machine learning or healthcare, compared to presenting it to the LUMC consortium.

A0.1.2 Math, data and Computing (MDC) and technology and realization (TR)

The field of AI and ML was completely novel to me. Besides understanding ML in a general fashion, I believe the focus on ethics also allowed me to gain a very nuanced understanding, gaining insights on its capabilities, but also its limitations and current clinical implementation bottlenecks. These practical MDC and TR insights translate to a more versatile profile as designer.

Furthermore, iterating on themes and combining inductive and theoretical coding within explorative, qualitative research was a novel thematic analysis method to me. It taught me a manner of gathering rich and grounded insights while also teaching me the need for a constant awareness of the research scope. This is a strong addition to the more general analysis methods I usually consult.

A0.1.3 Business and entrepreneurship (BE)

Although for feasibility the research was scoped towards user involvement rather than the broader stakeholder landscape, expert and stakeholder consultation were used to validate research relevance. Input from Philips employees and LUMC developers showed potential value of the tool in real-world practices, teaching me manners of creating entrepreneurial value through research.

Moreover, it was very educational to see research in a professional setting through the LUMC consortium, seeing its involved disciplines, structure, and organization of user- and stakeholder involvement.

A0.2 Design research processes and PIV

The project allowed me to tackle my PDP goal of strengthening my research skills and understanding. The first half of the project, my challenge with literature research and scoping allowed many learnings. I learned that I do not lack skills in literature research, but in self-confidence, possibly tied to my recently diagnosed ADD. This makes it hard for me to look

at literature on the right level of abstraction. This realization, combined with acquiring new structuring skills, makes me more confident in future work. Scoping-wise, the project made me realize I want to approach problems in a thorough manner, causing a broad and explorative scope. While this strategic perspective is a strength of mine in design projects, I learned in research the focus is more on in-depth exploration of a single element of a context. This awareness will allow me to be more realistic in considering time and manpower during future scoping. The second stage of the research taught me to constantly link the research protocol, probe and analysis to the research question.

Based on this research understanding, I wanted to use the project to steer my identity, assessing my interest in research. I learned that although I am analytical, the combination of my strategic interest and human-centered interest makes me value being inspired by- and applying insights directly into a real-world context. While explorative and qualitative contextual- and user research is where my strength lies in design, I realized the more framed nature of research brings out my self-doubt, a weakness.

During this project, I adopted an approach more steered by literature and intuition than usual user- and expert consultation. Although I see the benefit of basing decisions on a broader frame of reference and progressing without experts, less reflection from practice led to a personally weaker experienced relevance and enhanced insecurities about my newness to the topic. Based on these insights, I want to enhance the part of my identity and vision that favors a method-based, human-centered approach by also relying more on designerly instinct and literature.

A PDP goal was to develop my vision on emergent technologies and my role as a designer in this. Although I am still hesitant, the project has taught me that at least in AI, the field is aware of the need for human-centered approaches. Although this intrigues me and motivates me to strive for normalizing and creating more ethical and human-centered approaches and methods within emerging technology, I also witnessed the technological side is daunting to me. Therefore, social design and HCD with a focus on health will remain my focus, yet I am open to AI-based projects if they surface in my future work.

While conducting individual research with all included scoping and protocol insecurities, and designing in a novel field to me, was educational, combined with personal circumstances and ongoing battle with perfectionism this was a challenging semester. This has taught me to prioritize health by setting boundaries and being more accepting of making mistakes, which I hope to express in next projects.

To conclude, acquiring skills and insights on the expertise areas and design research processes has allowed me to become a more skilled designer. Additionally, combining new approaches with utilizing my expertise, has strengthened my identity and vision.

A1 EXTENSIVE BENCHMARK TEXT

To understand the current landscape of work for value or user-centered and ethical AI, a benchmark was performed. To grasp the aim of the tools, they were mapped on the development stage and audience they target, see figure 1. The circle is where the tool gravitates towards, yet tools often focus on multiple development stages, hence their lower opacity "tails" indicating where they feature. It stands out that current tools focus mostly on the internal team: the designer or developer, policymakers or (project) managers. Most bottom toolkits, aiming to act in line with an approach based on creating human value yet not actively involving users, include goals of enabling designers to create value through ML or AI (intelligence augmentation design toolkit [10] and AI meets design toolkit [11]), advocating a user-centered- (IDEO ethics cards [3], Futurice IA design kit [10], HAX Microsoft [12], IBM everyday ethics for AI [13], the mindful AI canvas [14]), or even participatory perspective (Google PAIR [15]). Generally less at the bottom, specifically aimed at the developer in their context are tools raising awareness on ethical concerns or values (ethical OS [7], AI blindspot [16], IBM everyday ethics [13], the ethics canvas [17], data ethics framework [18], data ethics canvas [19], tarot cards of tech [20]) or applying ethics through checklists or frameworks (coded fairness toolkit [21], the AI fairness checklist [5], deon [22]). Several mediums provide an overview of ethical tools, such as [8] and [1]. Again, these envelop internal tools - for managers, developers, policymakers etcetera, to conform with (overviews of) ethical values. However, no concrete tools were found that specifically aim for designer-user collaboration in ethical value creation and consideration in a comprehensive manner. Only the moral value map is stated to be fit to use with stakeholders, yet no theoretical substantiation or proof of use has been found on this tool [23]. Besides, found tools rarely focus on the specific implementation context.

A2 PROCESS VISUALIZATION

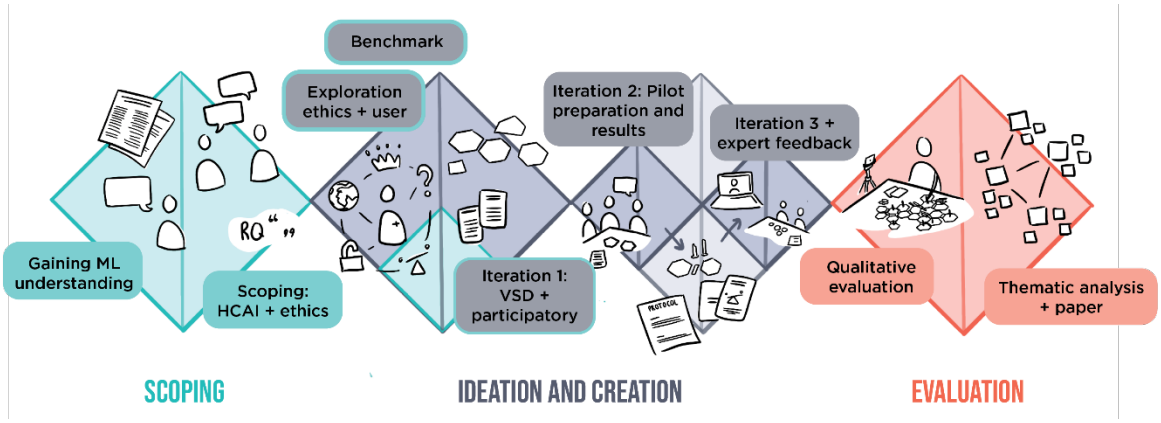


Figure A2.1 Process visualization

Figure A.2.1 displays a visual representation of the design research process. As seen in the colors, overall distinctions between scoping, ideation and creation and evaluation were conducted, using diverging and converging actions to gather information and subsequently remain relevantly focused. The green color depicts the literature on ML and AI, its stakeholders and ethics that was consulted throughout the first two phases, the second phase more focused on translating this into a probe.

A3 ETHICAL REVIEW BOARD APPROVALS AND SIGNED CONSENT FORMS

A3.1 ERB form expert interviews

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form Education

(Version 17.07.2020)

This Ethical Review Form should be completed for every research study that involves human participants or personally identifiable data. The form should be submitted and approved by your supervisor before potential participants are approached to take part in the research study.

Part 1: General Study Information		
1	Student name and email	I.P. Faber, i.p.faber@student.tue.nl
2	Supervisor name and email	Sara Colombo, s.colombo@tue.nl
3	Degree Program	Industrial Design
4	Bachelor/master	Master
5	Bachelor/master end project?	No
6	Course name and code	DPM120 Project 2 design research
7	Project title	Clinician values in a CDSS for heart failure
8	Research location	Microsoft teams or at university
9	Research period (start/end date)	Between 10/10/2022 and 12/01/2023
10	[If Applicable] Proposal already approved by (external) Ethical Review Board: Add name, date of approval, and contact details of the ERB	
11	Research question	What considerations are important in designing a tool that identifies clinician values in the design process of a ML Clinical Decision Support System (for cardiac risk assessment)?
12	Description of the research method	The overall research this study is part of is a workshop design for clinicians, that creates a workflow with possible uses of a ML model and maps their ethical values onto this. In order to understand the design context and prepare this research probe, expertise from practice will be used. For this, professionals from the ML (in healthcare), ethics or cardiac field will be consulted in 30-60 min semi-structured interviews.
13	Description of the research population, in- and exclusion criteria	A population professionals, working or researching in the field of ML/AI or health/cardiology.
14	Number of participants	Around 5 participants.
15	Explain why the research is socially important.	The technological advancements in the use of AI and ML in healthcare offer various opportunities for diagnosis and risk assessment (Greenes, 2014; Mosquera-Lopez et al., 2015; Zamora et al., 2013; Cai et al., 2019; Khedkar et al., 2019; Yang et al., 2019). However, the AI field is very much focussed on developing models that perform well, while integration into practice and adoption by users is a challenge (Browne et al., n.d.). To bring the tools into practice, proper

1

		consideration of ethical issues is vital (Jobin et al., n.d.; Karimian et al., 2022) This research is part of the DECIDE-VerA project in collaboration with LUMC, Leiden Law school, TU Eindhoven and Hogeschool Rotterdam. The DECIDE-VerA project features a AI-Clinical Decision Support System (AI-CDSS) that is built to aid in risk analysis for cardiovascular patients. The project focuses on how the design process of such a ML system can be altered to involve the ethical values of all stakeholders, specifically to eventually improve the shared decision making of doctors and patients. This research presents a starting point of this research, by exploring a manner in which the ethical values important to the general practitioners (GP's) can be discovered. The goal of this research is to validate a workshop format. This is designed in a twofold of (1) sensitizing the GP by going through a general mapping of the current workflow and plotting the use of the ML tool on this workflow, and (2) prioritizing and matching different ethical values of AI in health to this workflow.
16	Describe the way participants will be recruited	The participants will be found and contacted through personal and professional networks.
17	Provide a brief statement of the risks you expect for the participants or others involved in the research and explain. Take into consideration any personal data you may gather and privacy issues.	There are no risks involved for the professionals. The questions will be about their experience from practice and their view and input on the project. No personal data will be shared outside the project team. Information provided in the session will be used to understand the research space and to possibly steer research directions and to communicate or report to colleagues and other researchers at the TU/e. Audio recordings will be made to process information more efficiently, these will be transcribed as soon as possible after the conversation, and deleted directly after. In the meantime, recordings will be stored on SurfDrive, a password-protected server trusted by the TU/e. Data is anonymised and stored on SurfDrive, a password-protected server trusted by the TU/e.

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

	Part 2: Checklist for Minimal Risk		
		Yes	No
1	Does the study have a medical scientific research question or claim (see definition below) *Medical/scientific research is research which is carried out with the aim of finding answers to a question in the field of illness and health (etiology, pathogenesis, signs/symptoms, diagnosis, prevention, outcome or treatment of illness), by systematically collecting and analysing data. The research is carried out with the intention of contributing to medical knowledge which can also be applied to populations outside of the direct research population.'*	☐ If yes or maybe: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with question 2
2	Does the study involve human material (such as surgery waste material derived from non-commercial organizations such as hospitals)?	☐ If yes or maybe: This is only allowed if your supervisor has consulted with the medical coordinator. Continue with question 3	☒ If no: Continue with question 3
3	Will the participants give their explicit consent – on a voluntary basis – either digitally or on paper? Or have they given consent in the past for the purpose of education or for re-use in line with the current research question?	☒ If yes: Continue with question 4	☐ If no: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval
4	Will the study involve discussion or collection of personal data? (e.g. name, address, phone number, email address, IP address, BSN number, location data) or will the study collect and store videos, pictures, or other identifiable data of human subjects?	☐ If yes: The handling, storing and de-identification of the personal data should be discussed with your supervisor. Continue with question 5 if you met all requirements for handling personal data (see …)	☒ If no: Continue with question 5

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

		Yes	No
5	Does the study involve participants who are particularly vulnerable or unable to give informed consent? (e.g. children, people with learning difficulties, patients, people receiving counselling, people living in care or nursing homes, people recruited through self-help groups)?	☐	☒
		If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 6
6	May the research procedure cause harm or discomfort to the participant in any way? (e.g. causing pain or more than mild discomfort, stress, or anxiety)	☐	☒
		If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 7
7	Will the participants receive any compensation for their participation? Such as a coupon or a chance to win a prize?	☐	☒
		If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 8 or 10, depending on the type of study (see red text below)
The following questions 8-9 are for *observational* research (e.g. (semi-)structured interviews; focus groups; (participatory) observations). If your research is *experimental*, then skip questions 8-9 and continue with question 10			
8	Will it be necessary for participants to take part in the study without their knowledge and consent at the time? (e.g. covert observation of people)?	☐	☒
		If yes: This is only allowed when observing behavior in public space. If so, continue with question 9. If you observe people in non-public space without their consent, your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 9
9	Will participants be asked to discuss or report sexual experiences, religion, alcohol or drug use, or suicidal thoughts, or other topics that are highly personal or intimate?	☐	☒
		If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with part 3

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

The following questions 10-13 are for *experimental* research (e.g. measurements on yourself or another person; testing a prototype/device; influencing behavior through manipulation (e.g. light or temperature). If your research is *observational*, then skip questions 10-13 and continue with part 3			
		Yes	No
10	Is the study invasive (i.e. it affects the body such as puncturing the skin; taking blood or other body material (such as DNA) from the participant)?	☐	☐
		If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 11
11	Does the device have a medical purpose sucs as diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease or injury?	☐	☐
		If yes or maybe: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 12
12	Will the experiment involve the use of physical devices that are 'CE' certified for unintended use (meaning you will use existing CE certified devices for other things than they were originally intended for?	☐	☐
		If yes: This is only allowed if they are completely harmless. They should have a harmless voltage of <5V and hazardous waste (fumes/gas/substances) should not be released. You should discuss with your supervisor whether you need to have the device tested for safety	If no: Continue with question 13
13	Will the experiment involve the use of physical devices that are not 'CE' certified?	☐	☐
		If yes: This is only allowed if they are completely harmless. They should have a harmless voltage of <5V and hazardous waste (fumes/gas/substances) should not be released. You should discuss with your supervisor whether you need to have the device tested for safety	If no: Continue with part 3

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

Part 3: Enclosures and Signature		
1	Enclosures (tick if applicable): ☒ Informed consent form (link to template); ☐ The survey the participants need to complete, or a description of other measurements (such as interview questions or a description of the prototype); ☐ Text used to find participants (such as brochures, flyers, etc); ☐ Approval other research ethics committee;	Informed consent form Experts
2	I hereby declare that I have completed this form truthfully Signature(s) of the student(s) Date	

Discuss this form with your supervisor. If any of the boxes your ticked in Part 2 suggest that your supervisor should submit your study to the ERB for ethical approval, try to change your research design in such a way that your supervisor can approve it instead. If this is not possible, ask your supervisor to submit the proposal to the ERB. It will take two to five weeks before you receive a decision from the ERB.

Part 4: Review by supervisor		Yes	No
1	Does the data storage adhere to all requirements of responsible data management (link toevoegen)?	☒	☐
		If yes: Continue with question 2	If no: Discuss with your student the necessary steps to adhere to the requirements
2	Does the research proposal adhere to all requirements for automatic approval?	☒	☐
		If yes: Please skip the questions 3-6 and sign the form	If no: Discuss with your student if any alterations can be made in order to adhere to the requirements for automatic approval. If you decide that the study cannot adhere to the requirements, then you as a supervisor need to submit the proposal to the ERB. Please answer the following additional questions (3-6)

6

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

Additional questions for ERB approval		
3	Elaborate on the topics from part 2 that do not allow for automatic approval. Describe how you safeguard any potential risk for the research participant for each topic.	
4	Describe and justify the number of participants you need for this research, taking into account the risks and benefits	
5	Explain if your data are completely anonymous, or whether they will be de-identified (pseudonymized or anonymized) and if so, explain how	
6	Who will have access to the data?	

Part 5: Signature by supervisor		
	I hereby declare that I have completed this form truthfully Signature of the supervisor Date 31/10/2022	

A3.2 Consent form expert interviews

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Consent form

This document offers information on participation in an expert interview for a research project that is being carried out within the *Inclusive design and thoughtful technology* project group within the faculty of Industrial Design at the Technical University of Eindhoven. The student that carries out this session is **Ilse Faber**, who can be contacted via i.p.faber@student.tue.nl. Please read this document carefully to understand the procedure and to give consent to participating voluntarily.

The goal of the expert interview is to learn about your experiences from practice or research, within the themes either heart failure, or of Machine Learning (ML) in a healthcare context and ethical issues regarding ML (in healthcare).
The procedure of the research entails that the researcher will ask questions about this topic and perhaps show visualizations or protocols to receive feedback on, based on your expertise.
The interview will last approximately 30 minutes to 1 hour.

Right to refrain: Your participation in the research is completely voluntary. You are free to refrain from participating without any specific reason, and can withdraw participation at any moment during the research. These decisions will not have negative consequences.
Risks: The research does not contain risks or negative effects. In case unpublished/ confidential information is discussed, this must be disclosed and will not be used in the project.

Confidentiality

Confidential: No personal data will be shared outside the project team. Information provided in the session will be used to understand the research space and to possibly steer research directions and to communicate or report to colleagues and other researchers at the TU/e.
Audio and photos: Audio recordings will be made to process information more efficiently, these will be transcribed as soon as possible after the conversation, and deleted directly after. In the meantime, recordings will be stored on SurfDrive, a password-protected server trusted by the TU/e.
Data storage: Data is anonymised and stored on SurfDrive, a password-protected server trusted by the TU/e. Data will be deleted 1 month after the research (March 2023).
If you are not satisfied with the way the privacy of your data is handled, you can file a complaint at the Chief Information & Security Officer, the Privacy & Security Officer and/or the Data Protection Officer of the Technical University of Eindhoven via privacy@tue.nl, or by contacting the Dutch Data Protection Authority.

More information

To receive more information about the research and the research group, you can contact researcher Ilse Faber (i.p.faber@student.tue.nl) or squad leader and project coach Sara Colombo (s.colombo@tue.nl).

Certificate of consent

I, (name)..... have read and understood this consent form, and have had the possibility to ask questions. I agree to participate voluntarily with the research of Ilse Faber that is part of the *Inclusive design and thoughtful technology* project group of the faculty Industrial Design at the Technical University Eindhoven.

_____ _____

Autograph participant: Date:

A3.3 ERB form pilot study

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form Education

(Version 17.07.2020)

This Ethical Review Form should be completed for every research study that involves human participants or personally identifiable data. The form should be submitted and approved by your supervisor before potential participants are approached to take part in the research study.

Part 1: General Study Information

1	Student name and email	I.P. Faber, i.p.faber@student.tue.nl
2	Supervisor name and email	Sara Colombo, s.colombo@tue.nl
3	Degree Program	Industrial Design
4	Bachelor/master	Master
5	Bachelor/master end project?	No
6	Course name and code	DPM120 Project 2 design research
7	Project title	Clinician values in a CDSS for heart failure
8	Research location	On location at general practitioners, university or at home
9	Research period (start/end date)	Between 31/10/2022 and 30/11/2023
10	[If Applicable] Proposal already approved by (external) Ethical Review Board: Add name, date of approval, and contact details of the ERB	
11	Research question	A pilot of the research on "What considerations are important in designing a tool that identifies clinician values in the design process of a ML Clinical Decision Support System (for cardiac risk assessment)?".
12	Description of the research method	This study is part of a bigger research. In the eventual research, a workshop will be tested with clinicians. This study aims to pilot this tool and add to it, in a session with two or three designers. The study will take place through a semi-structured interview using a physical research probe. For the pilot, a general version of the GP's workflow will be mapped using the research probe in advance. During the study, a scenario on a future ML tool for in this workflow will be discussed, and its possible use will be mapped in the workflow. Lastly, general descriptions of ethical AI values based on the paper by Jobin et al. (2019) will be shown to the participant to discuss how these values play a role in their envisioned use of the ML system. After being part of the session in a similar manner as the GPs will be, an evaluation will be held to assess how the designers experienced the session and its materials. Lastly, the designers will be asked to participate in a short codesign activity, on their inputs on the tool, especially the value examples and the "Role of ML" part of the tool.

1

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

13	Description of the research population, in- and exclusion criteria	Industrial Design master students and PhD'ers.
14	Number of participants	2 or 3 participants
15	Explain why the research is socially important.	The technological advancements in the use of AI and ML in healthcare offer various opportunities for diagnosis and risk assessment (Cai et al., 2019; Greenes, 2014; Khedkar et al., n.d.; Mosquera-Lopez et al., 2015; Yang et al., 2019; Zamora et al., 2013). However, the AI field is very much focused on developing models that perform well, while integration into practice and adoption by users is a challenge(Browne et al., n.d.). To bring the tools into practice, proper consideration of ethical issues is vital (Jobin et al., 2019; Karimian et al., 2022). This research is part of the DECIDE-VerA project in collaboration with LUMC, Leiden Law school, TU Eindhoven and Hogeschool Rotterdam. The DECIDE-VerA project features a AI-Clinical Decision Support System (AI-CDSS) that is built to aid in risk analysis for cardiovascular patients. The project focuses on how the design process of such a ML system can be altered to involve the ethical values of all stakeholders, specifically to eventually improve the shared decision making of doctors and patients. The research presents a starting point of this research, by exploring a manner in which the ethical values important to the general practitioners (GP's) can be discovered. The goal of this research is to validate a workshop format. This is designed in a twofold of (1) sensitizing the GP by going through a general mapping of the current workflow and plotting the use of the ML tool on this workflow, and (2) prioritizing and matching different ethical values of AI based on the work by Jobin to this workflow (Jobin et al., 2019). **This pilot ensures that the eventual research involving GPs will be more efficient and better catered to their knowledge and creativity.**
16	Describe the way participants will be recruited	The participants will be found and contacted through personal networks.
17	Provide a brief statement of the risks you expect for the participants or others involved in the research and explain. Take into consideration any personal data you may gather and privacy issues.	The research does not contain direct risks or negative effects, yet cardiac issues or cardiac failure could be a sensitive topic. In this case, participants can refrain from participating. The personal data of the participants and photos and recordings of the interviews will be encrypted and stored on SurfDrive, a password protected environment trusted by the TUE. Recordings will be transcribed and deleted directly after transcription. Photos will not show participants' faces or any recognizable trait, so they will be completely anonymous.

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

	Part 2: Checklist for Minimal Risk		
		Yes	No
1	Does the study have a medical scientific research question or claim (see definition below) *Medical/scientific research is research which is carried out with the aim of finding answers to a question in the field of illness and health (etiology, pathogenesis, signs/symptoms, diagnosis, prevention, outcome or treatment of illness), by systematically collecting and analysing data. The research is carried out with the intention of contributing to medical knowledge which can also be applied to populations outside of the direct research population.'*	☐ If yes or maybe: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with question 2
2	Does the study involve human material (such as surgery waste material derived from non-commercial organizations such as hospitals)?	☐ If yes or maybe: This is only allowed if your supervisor has consulted with the medical coordinator. Continue with question 3	☒ If no: Continue with question 3
3	Will the participants give their explicit consent – on a voluntary basis – either digitally or on paper? Or have they given consent in the past for the purpose of education or for re-use in line with the current research question?	☒ If yes: Continue with question 4	☐ If no: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval
4	Will the study involve discussion or collection of personal data? (e.g. name, address, phone number, email address, IP address, BSN number, location data) or will the study collect and store videos, pictures, or other identifiable data of human subjects?	☐ If yes: The handling, storing and de-identification of the personal data should be discussed with your supervisor. Continue with question 5 if you met all requirements for handling personal data (see …)	☒ If no: Continue with question 5

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

		Yes	No
5	Does the study involve participants who are particularly vulnerable or unable to give informed consent? (e.g. children, people with learning difficulties, patients, people receiving counselling, people living in care or nursing homes, people recruited through self-help groups)?	☐ If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with question 6
6	May the research procedure cause harm or discomfort to the participant in any way? (e.g. causing pain or more than mild discomfort, stress, or anxiety)	☐ If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with question 7
7	Will the participants receive any compensation for their participation? Such as a coupon or a chance to win a prize?	☐ If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with question 8 or 10, depending on the type of study (see red text below)
	The following questions 8-9 are for *observational* research (e.g. (semi-)structured interviews; focus groups; (participatory) observations). If your research is *experimental*, then skip questions 8-9 and continue with question 10		
8	Will it be necessary for participants to take part in the study without their knowledge and consent at the time? (e.g. covert observation of people)?	☐ If yes: This is only allowed when observing behavior in public space. If so, continue with question 9. If you observe people in non-public space without their consent, your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with question 9
9	Will participants be asked to discuss or report sexual experiences, religion, alcohol or drug use, or suicidal thoughts, or other topics that are highly personal or intimate?	☐ If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with part 3

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

The following questions 10-13 are for *experimental* research (e.g. measurements on yourself or another person; testing a prototype/device; influencing behavior through manipulation (e.g. light or temperature). If your research is *observational*, then skip questions 10-13 and continue with part 3			
		Yes	No
10	Is the study invasive (i.e. it affects the body such as puncturing the skin; taking blood or other body material (such as DNA) from the participant)?	☐	☐
		If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 11
11	Does the device have a medical purpose sucs as diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease or injury?	☐	☐
		If yes or maybe: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 12
12	Will the experiment involve the use of physical devices that are 'CE' certified for unintended use (meaning you will use existing CE certified devices for other things than they were originally intended for?	☐	☐
		If yes: This is only allowed if they are completely harmless. They should have a harmless voltage of <5V and hazardous waste (fumes/gas/substances) should not be released. You should discuss with your supervisor whether you need to have the device tested for safety	If no: Continue with question 13
13	Will the experiment involve the use of physical devices that are not 'CE' certified?	☐	☐
		If yes: This is only allowed if they are completely harmless. They should have a harmless voltage of <5V and hazardous waste (fumes/gas/substances) should not be released. You should discuss with your supervisor whether you need to have the device tested for safety	If no: Continue with part 3

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

Part 3: Enclosures and Signature		
1	Enclosures (tick if applicable): ☐ Informed consent form (link to template); ☐ The survey the participants need to complete, or a description of other measurements (such as interview questions or a description of the prototype); ☐ Text used to find participants (such as brochures, flyers, etc); ☐ Approval other research ethics committee;	Consent form pilot study Pilot research protocol Ilse Faber M1.2
2	I hereby declare that I have completed this form truthfully Signature(s) of the student(s) Date	

Discuss this form with your supervisor. If any of the boxes your ticked in Part 2 suggest that your supervisor should submit your study to the ERB for ethical approval, try to change your research design in such a way that your supervisor can approve it instead. If this is not possible, ask your supervisor to submit the proposal to the ERB. It will take two to five weeks before you receive a decision from the ERB.

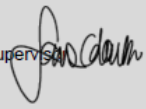
Part 4: Review by supervisor		Yes	No
1	Does the data storage adhere to all requirements of responsible data management (link toevoegen)?	☒	☐
		If yes: Continue with question 2	If no: Discuss with your student the necessary steps to adhere to the requirements
2	Does the research proposal adhere to all requirements for automatic approval?	☒	☐
		If yes: Please skip the questions 3-6 and sign the form	If no: Discuss with your student if any alterations can be made in order to adhere to the requirements for automatic approval. If you decide that the study cannot adhere to the requirements, then you as a supervisor need to submit the proposal to the ERB. Please answer the following additional questions (3-6)

6

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

Additional questions for ERB approval		
3	Elaborate on the topics from part 2 that do not allow for automatic approval. Describe how you safeguard any potential risk for the research participant for each topic.	
4	Describe and justify the number of participants you need for this research, taking into account the risks and benefits	
5	Explain if your data are completely anonymous, or whether they will be de-identified (pseudonymized or anonymized) and if so, explain how	
6	Who will have access to the data?	

Part 5: Signature by supervisor		
	I hereby declare that I have completed this form truthfully Signature of the supervisor  Date 31/10/22	

A3.4 Consent form pilot study

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Consent form

This document offers information on participation in a pilot study for a research project that is being carried out within the *Inclusive design and thoughtful technology* project group within the faculty of Industrial Design at the Technical University of Eindhoven. The student that carries out this session is **Ilse Faber**, who can be contacted via i.p.faber@student.tue.nl. Please read this document carefully to understand the procedure and to give consent to participating voluntarily.

The goal of the study is to pilot a workshop that will be carried out with general practitioners, and to gain creative input on the workshop's materials.
The procedure: The study consists of the workshop itself, a short evaluation and a brainstorm. The workshop itself entails mapping possible uses of a ML model in healthcare on a clinicians journey, and discussing how different ethical values play a role in this. The evaluation is about the contents of the session, but also about possible improvements in deliveries. Lastly, two brainstorm prompts on the workshop materials will be used to brainstorm in the group. The study will last approximately 1 hour and 15 minutes.

Right to refrain: Your participation in the research is completely voluntary. You are free to refrain from participating without any specific reason, and can withdraw participation at any moment during the research. These decisions will not have negative consequences.
Risks: The research does not contain direct risks or negative effects, yet cardiac issues or cardiac failure could be a sensitive topic. In this case, participants can refrain from participating.

Confidentiality

Confidential: No personal data will be shared outside the project team. Information provided in the session will be used to improve the workshop materials. Overall, abstracted or anonymised insights can be communicated or reported to colleagues and other researchers at the TU/e in published materials.
Audio and photos: Audio recordings will be made to process information more efficiently, these will be deleted after analysis of session results.
Data storage: Data is anonymised and stored on SurfDrive, a password-protected server trusted by the TU/e. Recordings will be transcribed and deleted directly as soon as possible after the research. Any photos taken will be without persons on them or anonymized in publications, originals will be deleted after the research.

If you are not satisfied with the way the privacy of your data is handled, you can file a complaint at the Chief Information & Security Officer, the Privacy & Security Officer and/or the Data Protection Officer of the Technical University of Eindhoven via privacy@tue.nl, or by contacting the Dutch Data Protection Authority.

More information

To receive more information about the research and the research group, you can contact researcher Ilse Faber (i.p.faber@student.tue.nl) or squad leader and project coach Sara Colombo (s.colombo@tue.nl).

Certificate of consent

I, (name)..... have read and understood this consent form, and have had the possibility to ask questions. I agree to participate voluntarily with the research of Ilse Faber that is part of the *Inclusive design and thoughtful technology* project group of the faculty Industrial Design at the Technical University Eindhoven.

Autograph participant: Date:

A3.5 ERB form final research

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form Education

(Version 17.07.2020)

This Ethical Review Form should be completed for every research study that involves human participants or personally identifiable data. The form should be submitted and approved by your supervisor before potential participants are approached to take part in the research study.

Part 1: General Study Information

1	Student name and email	I.P. Faber, i.p.faber@student.tue.nl
2	Supervisor name and email	Sara Colombo, s.colombo@tue.nl
3	Degree Program	Industrial Design
4	Bachelor/master	Master
5	Bachelor/master end project?	No
6	Course name and code	DPM120 Project 2 design research
7	Project title	Clinician values in a CDSS for cardiac risk assesment
8	Research location	On location at general practitioners, university or at home
9	Research period (start/end date)	Between 18/11/2022 and 12/01/2023
10	[If Applicable] Proposal already approved by (external) Ethical Review Board: Add name, date of approval, and contact details of the ERB	
11	Research question	What considerations are important in designing a tool that identifies clinician values in the design process of a ML Clinical Decision Support System (for cardiac risk assessment)?
12	Description of the research method	The study will take place through a semi-structured interview using a physical research probe. During the study, a tool loosely based on user journey mapping will be used to map the current heart failure workflow of a general practitioner (Gibbons, 2018). Afterwards, A scenario on a future ML tool for in this workflow will be discussed, and its possible use will be mapped in the workflow. Lastly, general descriptions of ethical AI values based on the paper by Jobin et al. will be shown to the participant to discuss how these values play a role in their envisioned use of the ML system (Jobin et al., 2019). At the end of the session, a short evaluation will be held to assess how the GP experienced the session and its materials. Additionally, when in scope, the workshop application and its results will be discussed with designers with experience in designing clinical ML systems.
13	Description of the research population, in- and exclusion criteria	For the main research: General practitioners and general practitioner assistants. Inclusion criteria are that they are Dutch. For the additional assessment: Industrial design FMP students or graduates who have worked on clinical ML development/design projects.

1

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

14	Number of participants	Around 3 to 6 clinicians. Around 2 to 4 designers.
15	Explain why the research is socially important.	The technological advancements in the use of AI and ML in healthcare offer various opportunities for diagnosis and risk assessment (Cai et al., 2019; Greenes, 2014; Khedkar et al., n.d.; Mosquera-Lopez et al., 2015; Yang et al., 2019; Zamora et al., 2013). However, the AI field is very much focussed on developing models that perform well, while integration into practice and adoption by users is a challenge(Browne et al., n.d.). To bring the tools into practice, proper consideration of ethical issues is vital (Jobin et al., 2019; Karimian et al., 2022). This research is part of the DECIDE-VerA project in collaboration with LUMC, Leiden Law school, TU Eindhoven and Hogeschool Rotterdam. The DECIDE-VerA project features a AI-Clinical Decision Support System (AI-CDSS) that is built to aid in primary risk analysis for cardiovascular patients. The project focuses on how the design process of such a ML system can be altered to involve the ethical values of all stakeholders, specifically to eventually improve the shared decision making of doctors and patients. This research presents a starting point of this research, by exploring a manner in which the ethical values important to the general practitioners (GP's) can be discovered. The goal of this research is to validate a workshop format. This is designed in a twofold of (1) sensitizing the GP by going through a general mapping of the current workflow and plotting the use of the ML tool on this workflow, and (2) prioritizing and matching different ethical values of AI based on the work by Jobin to this workflow (Jobin et al., 2019). In the research described in this ERB form, the workshop will be tested with GPs and GPAs, and optionally evaluated with clinical ML designers.
16	Describe the way participants will be recruited	The participants will be found and contacted through personal networks, and an attempt will be made to recruit participants via professional networks such as the Nederlandse Vereniging voor Cardiologie, Hartvolgers, Landelijke Huisartsen Vereniging and Nederlands Huisartsen Genootschap. ML designers will be recruited through personal network.
17	Provide a brief statement of the risks you expect for the participants or others involved in the research and explain. Take into consideration any personal data you may gather and privacy issues.	GPs and GPAs will sign "Dutch informed consent GP and GPA". Designers will sign "Informed consent form Experts" The contact information of the participant will be encrypted in a password protected file. During the workshop, audio recordings will be made. Besides, video recordings will be made to know in hindsight which places in the journey the participant is

		referring to. The video will be set up with the journey in frame, yet possibly sometimes the participant will be in frame when placing objects. Video and audio recordings will be stored on SurfDrive, a password protected environment trusted by the TUE. Audio recordings will be deleted directly after transcription as soon as possible after the session, top-down video recordings will be deleted after analysis. Audio will be removed from the videos. To show the outcomes to the designers, pictures will be made of the created journey. The participant will not be in these pictures or made anonymous as soon as possible after the session. Anonymous versions will be used in publications. Non-anonymous originals will be deleted. During the designer interviews, audio recordings will be made that will be stored in SurfDrive and deleted after transcription, which is done as soon as possible after the session.

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

	Part 2: Checklist for Minimal Risk		
		Yes	No
1	Does the study have a medical scientific research question or claim (see definition below) *Medical/scientific research is research which is carried out with the aim of finding answers to a question in the field of illness and health (etiology, pathogenesis, signs/symptoms, diagnosis, prevention, outcome or treatment of illness), by systematically collecting and analysing data. The research is carried out with the intention of contributing to medical knowledge which can also be applied to populations outside of the direct research population.'*	☐	☒
		If yes or maybe: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 2
2	Does the study involve human material (such as surgery waste material derived from non-commercial organizations such as hospitals)?	☐	☒
		If yes or maybe: This is only allowed if your supervisor has consulted with the medical coordinator. Continue with question 3	If no: Continue with question 3
3	Will the participants give their explicit consent – on a voluntary basis – either digitally or on paper? Or have they given consent in the past for the purpose of education or for re-use in line with the current research question?	☒	☐
		If yes: Continue with question 4	If no: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval
4	Will the study involve discussion or collection of personal data? (e.g. name, address, phone number, email address, IP address, BSN number, location data) or will the study collect and store videos, pictures, or other identifiable data of human subjects?	☒	☒
		If yes: The handling, storing and de-identification of the personal data should be discussed with your supervisor. Continue with question 5 if you met all requirements for handling personal data (see …)	If no: Continue with question 5

4

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

		Yes	No
5	Does the study involve participants who are particularly vulnerable or unable to give informed consent? (e.g. children, people with learning difficulties, patients, people receiving counselling, people living in care or nursing homes, people recruited through self-help groups)?	☐ If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with question 6
6	May the research procedure cause harm or discomfort to the participant in any way? (e.g. causing pain or more than mild discomfort, stress, or anxiety)	☐ If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with question 7
7	Will the participants receive any compensation for their participation? Such as a coupon or a chance to win a prize?	☐ If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with question 8 or 10, depending on the type of study (see red text below)
The following questions 8-9 are for *observational* research (e.g. (semi-)structured interviews; focus groups; (participatory) observations). If your research is *experimental*, then skip questions 8-9 and continue with question 10			
8	Will it be necessary for participants to take part in the study without their knowledge and consent at the time? (e.g. covert observation of people)?	☐ If yes: This is only allowed when observing behavior in public space. If so, continue with question 9. If you observe people in non-public space without their consent, your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with question 9
9	Will participants be asked to discuss or report sexual experiences, religion, alcohol or drug use, or suicidal thoughts, or other topics that are highly personal or intimate?	☐ If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	☒ If no: Continue with part 3

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

The following questions 10-13 are for *experimental* research (e.g. measurements on yourself or another person; testing a prototype/device; influencing behavior through manipulation (e.g. light or temperature). If your research is *observational*, then skip questions 10-13 and continue with part 3			
		Yes	**No**
10	Is the study invasive (i.e. it affects the body such as puncturing the skin; taking blood or other body material (such as DNA) from the participant)?	☐	☐
		If yes: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 11
11	Does the device have a medical purpose sucs as diagnosis, prevention, monitoring, prediction, prognosis, treatment or alleviation of disease or injury?	☐	☐
		If yes or maybe: Your supervisor should submit the study to the ERB. You cannot get automatic ethical approval	If no: Continue with question 12
12	Will the experiment involve the use of physical devices that are 'CE' certified for unintended use (meaning you will use existing CE certified devices for other things than they were originally intended for?	☐	☐
		If yes: This is only allowed if they are completely harmless. They should have a harmless voltage of <5V and hazardous waste (fumes/gas/substances) should not be released. You should discuss with your supervisor whether you need to have the device tested for safety	If no: Continue with question 13
13	Will the experiment involve the use of physical devices that are not 'CE' certified?	☐	☐
		If yes: This is only allowed if they are completely harmless. They should have a harmless voltage of <5V and hazardous waste (fumes/gas/substances) should not be released. You should discuss with your supervisor whether you need to have the device tested for safety	If no: Continue with part 3

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

Part 3: Enclosures and Signature		
1	Enclosures (tick if applicable): ☒ Informed consent form (link to template); ☒ The survey the participants need to complete, or a description of other measurements (such as interview questions or a description of the prototype); ☒ Text used to find participants (such as brochures, flyers, etc); ☐ Approval other research ethics committee;	Consent form GP and GPA Dutch consent form GP and GPA Informed consent form Experts Research protocol Ilse Faber M1.2
2	I hereby declare that I have completed this form truthfully Signature(s) of the student(s) Date	21/11/2022 

Discuss this form with your supervisor. If any of the boxes your ticked in Part 2 suggest that your supervisor should submit your study to the ERB for ethical approval, try to change your research design in such a way that your supervisor can approve it instead. If this is not possible, ask your supervisor to submit the proposal to the ERB. It will take two to five weeks before you receive a decision from the ERB.

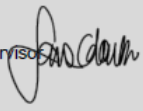
Part 4: Review by supervisor		Yes	No
1	Does the data storage adhere to all requirements of responsible data management (link toevoegen)?	☒	☐
		If yes: Continue with question 2	If no: Discuss with your student the necessary steps to adhere to the requirements
2	Does the research proposal adhere to all requirements for automatic approval?	☒	☐
		If yes: Please skip the questions 3-6 and sign the form	If no: Discuss with your student if any alterations can be made in order to adhere to the requirements for automatic approval. If you decide that the study cannot adhere to the requirements, then you as a supervisor need to submit the proposal to the ERB. Please answer the following additional questions (3-6)

7

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Ethical Review Form

Additional questions for ERB approval		
3	Elaborate on the topics from part 2 that do not allow for automatic approval. Describe how you safeguard any potential risk for the research participant for each topic.	Video recordings will be taken during the research with clinicians. This will be done in a top-view manner so the participant is not in frame. Regardless, identifiable traits might be in frame at some moments, these frames will not be used to discuss results with other designers. Videos will be stored on SurfDrive without audio attached to it. The recordings will be deleted after analysis is finished.
4	Describe and justify the number of participants you need for this research, taking into account the risks and benefits	
5	Explain if your data are completely anonymous, or whether they will be de-identified (pseudonymized or anonymized) and if so, explain how	
6	Who will have access to the data?	

Part 5: Signature by supervisor		
	I hereby declare that I have completed this form truthfully Signature of the supervisor  Date 22/11/22	

A3.6a Consent form GP and GPA Dutch

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Toestemmingsformulier

Dit document geeft informatie over deelname in een onderzoek wat wordt uitgevoerd binnen de faculteit Industrial Design van de Technische Universiteit van Eindhoven. Het onderzoek wordt uitgevoerd door masterstudent **Ilse Faber**, met wie contact opgenomen kan worden via i.p.faber@student.tue.nl. Lees dit document alstublieft zorgvuldig door om de procedure van het onderzoek te begrijpen en om toestemming te geven voor het vrijwillig deelnemen in het onderzoek.

Het doel van het onderzoek is om de effectiviteit van een tool te testen, die clinici betrekt bij het ontwerp en de implementatie van klinische data-driven tools. Middels een workshop zal de tool getest worden. De tool heeft als doel om de behoeften en waarden van huisartsen en praktijkondersteuners in kaart te brengen binnen het ontwerpen van een data-driven tool voor cardiovasculaire risico voorspelling.
De procedure: het onderzoek bestaat uit de workshop zelf, en een korte evaluatie hierna. De workshop zelf heeft drie delen: het in kaart brengen van de huidige manier van werken rondom een patiënt die wordt gescreend en preventief behandeld voor cardiovasculair risico, in kaart brengen hoe een data-driven interventie hier in kan bijdragen, en het bespreken van de ethische waarden hierin een rol spelen. De evaluatie achteraf gaat over hoe de deelnemer de sessie heeft ervaren en hoe zij hun betrokkenheid bij het design process van zo'n tool zien. Het onderzoek duurt maximaal 1 uur en 30 minuten.

Recht op afbraak: Deelname aan het onderzoek is volledig vrijwillig. U bent vrij om uw deelname af te breken op elk moment, zonder het aangeven van een reden hiervoor. Deze beslissing heeft geen negatieve gevolgen.
Risico's: Het research heeft geen directe risico's of negatieve effecten.

Vertrouwelijkheid

Vertrouwelijkheid: Er worden geen persoonlijke gegevens gedeeld buiten het projectteam, uw contactgegevens worden opgeslagen in een versleuteld bestand. De informatie die tijdens de sessie wordt verstrekt, wordt gebruikt om de effectiviteit van de workshop en de tool te beoordelen. Geanonimiseerde inzichten worden in gepubliceerde materialen gerapporteerd.
Verzamelde gegevens en opslag: Er zullen audio- en video-opnames worden gemaakt om informatie efficiënter te verwerken en de interacties met de tool te analyseren. Gegevens worden geanonimiseerd en opgeslagen op SurfDrive, een met een wachtwoord beveiligde server die door de TU/e wordt vertrouwd. Audio-opnamen worden zo snel mogelijk na de sessie getranscribeerd en vervolgens verwijderd. Top-view video-opnames worden na analyse zo snel mogelijk verwijderd, en opgeslagen zonder audio. Van het invullen van de tool en de ingevulde tool worden top-down foto's gemaakt om de resultaten te analyseren en met ontwerpers te bespreken. Alle gemaakte foto's worden geanonimiseerd bewaard en mogelijk gepubliceerd, originelen worden verwijderd.

Indien u niet tevreden bent over de wijze waarop met de privacy van uw gegevens wordt omgegaan, kunt u een klacht indienen bij de Chief Information & Security Officer, de Privacy & Security Officer en/of de Functionaris Gegevensbescherming van de Technische Universiteit Eindhoven via privacy@tue.nl, of door contact op te nemen met de Autoriteit Persoonsgegevens.

Meer informatie

Voor meer informatie over het onderzoek en de onderzoeksgroep kun je contact opnemen met onderzoeker Ilse Faber (i.p.faber@student.tue.nl) of squadleider en projectcoach Sara Colombo (s.colombo@tue.nl).

Geving van toestemming

Kies je toestemmingen:

- ☐ Ik geef **wel** toestemming om audio opnames te maken tijdens de sessie, deze worden zo snel mogelijk na de sessie getranscribeerd en hierna verwijderd.
- ☐ Ik geef **wel** toestemming om video opnames te maken tijdens de sessie, deze worden na analyse verwijderd.

Kies één van onderstaande opties:

- ☐ Ik geef **wel** toestemming om tijdens de sessie gemaakte foto- en video opnames van mij **geanonimiseerd** te gebruiken in het evalueren van de tool met andere designers **en** ik geef toestemming om genomen foto's van mij **geanonimiseerd** in (online) publicaties gerelateerd aan dit project te gebruiken.
- ☐ Ik geef **wel** toestemming om tijdens de sessie gemaakte foto- en video opnames van mij **geanonimiseerd** te gebruiken in het evalueren van de tool met andere designers **en** ik geef **geen** toestemming om genomen foto's van mij **geanonimiseerd** in (online) publicaties gerelateerd aan dit project te gebruiken.
- ☐ Ik geef **geen** toestemming om tijdens de sessie gemaakte foto- en video opnames van mij **geanonimiseerd** te gebruiken in het evalueren van de tool met andere designers **en** ik geef **wel** toestemming om genomen foto's van mij **geanonimiseerd** in (online) publicaties gerelateerd aan dit project te gebruiken.
- ☐ Ik geef **geen** toestemming om tijdens de sessie gemaakte foto- en video opnames van mij **geanonimiseerd** te gebruiken in het evalueren van de tool met andere designers **en** ik geef **geen** toestemming om genomen foto's van mij **geanonimiseerd** in (online) publicaties gerelateerd aan dit project te gebruiken.

Ik, (naam)..... heb dit toestemmingsformulier gelezen en begrepen, en de mogelijkheid gehad om vragen te stellen. Ik ga ermee akkoord vrijwillig deel te nemen aan het onderzoek van Ilse Faber binnen de faculteit Industrieel Ontwerpen aan de Technische Universiteit Eindhoven.

Handtekening deelnemer: Datum:

A3.6b Consent form GP and GPA English

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

Consent form

This document offers information on participation in a study for a research project that is being carried out within the faculty of Industrial Design at the Technical University of Eindhoven. The student that carries out this session is **Ilse Faber**, who can be contacted via i.p.faber@student.tue.nl. Please read this document carefully to understand the procedure and to give consent to participating voluntarily.

The goal of the study is to assess the effectiveness of a tool for involving clinicians in the design and implementation of clinical, data-driven tools. Through a workshop, the tool will be tested. The tool aims to gather needs and values of general practitioners and general practitioners assistants within the design of a data driven tool for cardiac risk assessment.
The procedure: The study consists of the workshop itself, and a short evaluation afterwards. The workshop itself entails mapping the current workflow surrounding a patient that is screened and treated on cardiac risk, the possible use of a data driven tool helping in this screening, and discussing how different ethical values play a role in this role. The evaluation is about how the participant experienced the session and their involvement in the design process of the tool. The study will last at maximum 1 hour and 30 minutes.

Right to refrain: Your participation in the research is completely voluntary. You are free to refrain from participating without any specific reason, and can withdraw participation at any moment during the research. These decisions will not have negative consequences.
Risks: The research does not contain direct risks or negative effects.

Confidentiality

Confidential: No personal data will be shared outside the project team, your contact details will be stored in an encrypted file. Information provided in the session will be used to assess the effectiveness of the workshop and tool. Anonymised insights will be reported in published materials.
Data gathering and storage: Audio and video recordings will be made to process information more efficiently and enable analysis of the interactions with the tool. Data is anonymised and stored on SurfDrive, a password-protected server trusted by the TU/e. Audio recordings will be transcribed and subsequently deleted as soon as possible after the session. Top-view video recordings of the tool will be used for analysis, audio will be removed from the video file. Top-view photos will be taken of interactions with the tool and the filled in tool, to analyze results and discuss with other designers. Any photos taken will be stored and possibly published anonymously, originals will be deleted.

If you are not satisfied with the way the privacy of your data is handled, you can file a complaint at the Chief Information & Security Officer, the Privacy & Security Officer and/or the Data Protection Officer of the Technical University of Eindhoven via privacy@tue.nl, or by contacting the Dutch Data Protection Authority.

More information

To receive more information about the research and the research group, you can contact researcher Ilse Faber (i.p.faber@student.tue.nl) or squad leader and project coach Sara Colombo (s.colombo@tue.nl).

Certificate of consent

Choose your permissions:

- ☐ I **do** give permission to make audio recordings during the session, which are transcribed and deleted directly after.
- ☐ I **do** give permission to make video recordings during the session, which are deleted after analysis.

Choose one of the options below:

- ☐ I **do** give permission to use photo and video recordings of the session in an **anonymised manner** in evaluating the tool with other designers **and** I give permission to use photos taken during the session in an **anonymised manner** in (online) publications related to the project.
- ☐ I **do** give permission to use photo and video recordings of the session in an **anonymised manner** in evaluating the tool with other designers **and I do not** give permission to use photos taken during the session in an **anonymised manner** in (online) publications related to the project.
- ☐ I **do not** give permission to use photo and video recordings of the session in an **anonymised manner** in evaluating the tool with other designers **and I do** give permission to use photos taken during the session in an **anonymised manner** in (online) publications related to the project.
- ☐ I **do not** give permission to use photo and video recordings of the session in an **anonymised manner** in evaluating the tool with other designers **and I do not** give permission to use photos taken during the session in an **anonymised manner** in (online) publications related to the project.

I, (name)…….. have read and understood this consent form, and have had the possibility to ask questions. I agree to participate voluntarily with the research of Ilse Faber within the faculty Industrial Design at the Technical University Eindhoven.

Autograph participant: Date:

A4 PILOT PROTOCOL AND PHOTOS

A4.1 Pilot protocol

Goal of the research

This research is part of the DECIDE-VerA project in collaboration with LUMC, Leiden Law school, TU Eindhoven and Hogeschool Rotterdam. The DECIDE-VerA project features a AI-Clinical Decision Support System (AI-CDSS) that is built to aid in risk analysis for cardiovascular patients. The project focuses on how the design process of such a ML system can be altered to involve the ethical values of all stakeholders, specifically to eventually improve the shared decision making of doctors and patients.

The overall research this study is part of, presents a starting point of this research, by exploring a manner in which the ethical values important to the general practitioners (GP's) can be discovered. The goal of this research is to validate a workshop format. This is designed in a twofold of (1) sensitizing the GP by going through a general mapping of the current workflow and plotting the use of the ML tool on this workflow, and (2) prioritizing and matching different ethical values of AI in health to this workflow.

The overall **research question:** Which considerations are important in designing a workshop that identifies clinician ethical values in the early stages of the design process of a ML Clinical Decision Support System (for cardiac risk assessment)?

This study is a pilot for this research. It tests the workshop elements with design students who either have a background in machine learning and AI, or who do not. This will allow examination of whether the workshop materials are low-threshold and can spark conversation in general.

Sub-questions:

- Does a physical workflow enable participants to map the potential role of a ML model in the workflow?
- Are value cards a way to aid in discussing the ethical AI values that matter to participants?
- Can value cards on ethical AI values spark discussion on how the ML system should be designed?

Target population and recruitment strategy

The target audience of this workshop are master or PhD students of Industrial Design, either with background knowledge of AI or not. The students are recruited through the researchers personal network.

Procedure:

Participants' introduction to the study

The participants will be introduced to the overall research and main setup of the study beforehand.

Measures

The session with designers will consist of a mockup of the relevant sections of the GP session (figure 1), followed by a short co-design on the material and an evaluation. In advance, the participants will be asked to sign the informed consent form (appendix 1).

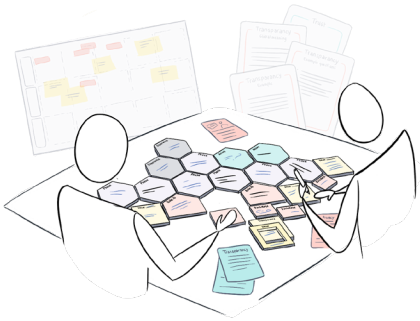


Figure 1: representation of the workshop.

Part 1: Workshop

See “GP Research protocol Ilse Faber M1.2” for the regular workshop setup. Throughout the workshop, workflow, ML and value tokens will be used (see figure 2 and appendix 2). The initial part, workflow mapping, will be skipped by the designers due to lack of medical knowledge. An overall journey with general stages will be mapped in advance.

Afterwards, the session will advance as usual: the ML model will be explained through the use of a scenario (Martin & Hanington, 2012). Not having spoken to the developer yet, this scenario will be based on an overall impression of the system from the project briefing by Sara Columbo (see interview guide). Afterwards, the designer will map possible uses of this model with the tokens, or add to these with empty ones.

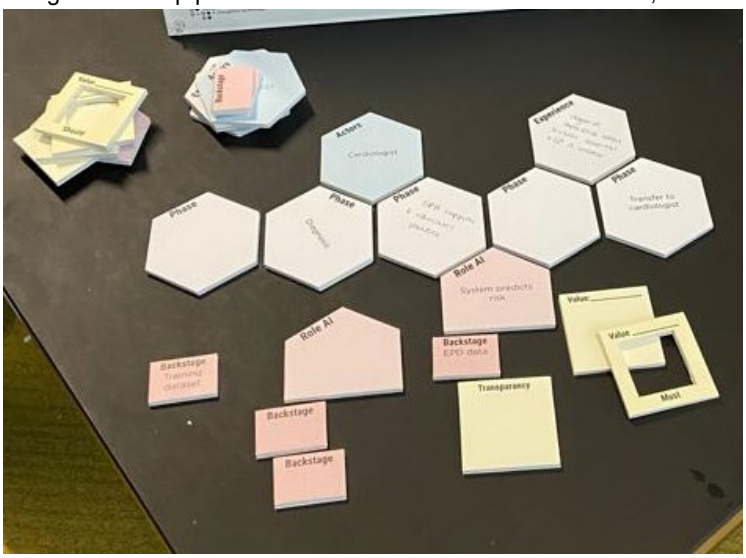


Figure 2: demonstration of the physical workshop elements and a possible way these can be filled in

Part 2: ethical value mapping

To understand general ethical values of AI, the work of Jobin et al., (2019), who formulated 11 main ethical AI values based on a literature review, was studied. The work of Jobin et al. (2019) was summarized in ethical value cards. When needed for clarification or to provide a more specific health-related view on a value, other explanations from literature specifically on ethical AI values in the domain of health were consulted. These are to be used in a card sorting (Martin & Hanington, 2012) way and to map onto the tokens. See an example in figure 3 and the complete set in appendix 3.

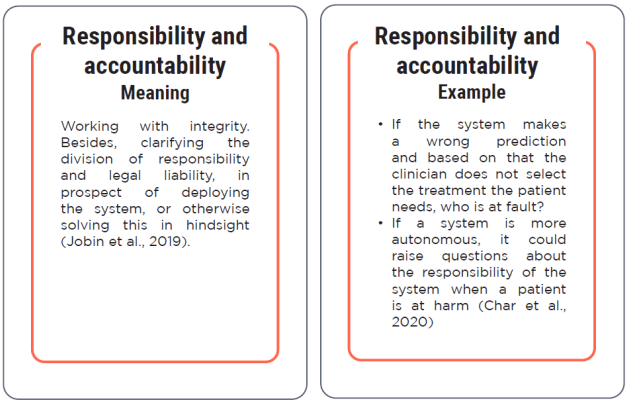


Figure 3: example card from the value set.

Part 3: evaluation

After the workshop element, the participants will be briefly questioned on how they experienced the workshop,

how it fit their expectations and what recommendations they have to improve the workshop.

Part 4: cocreation

In a cocreation setting using (visual) brainstorming, the participants will be asked to think of 1) more examples that go with the ethical values, 2) specific ways in which the model can be used, 3) more creative/ user friendly manners of executing the value cards.

Workshop overview

The complete workshop with its timing will therefore look like figure 4:

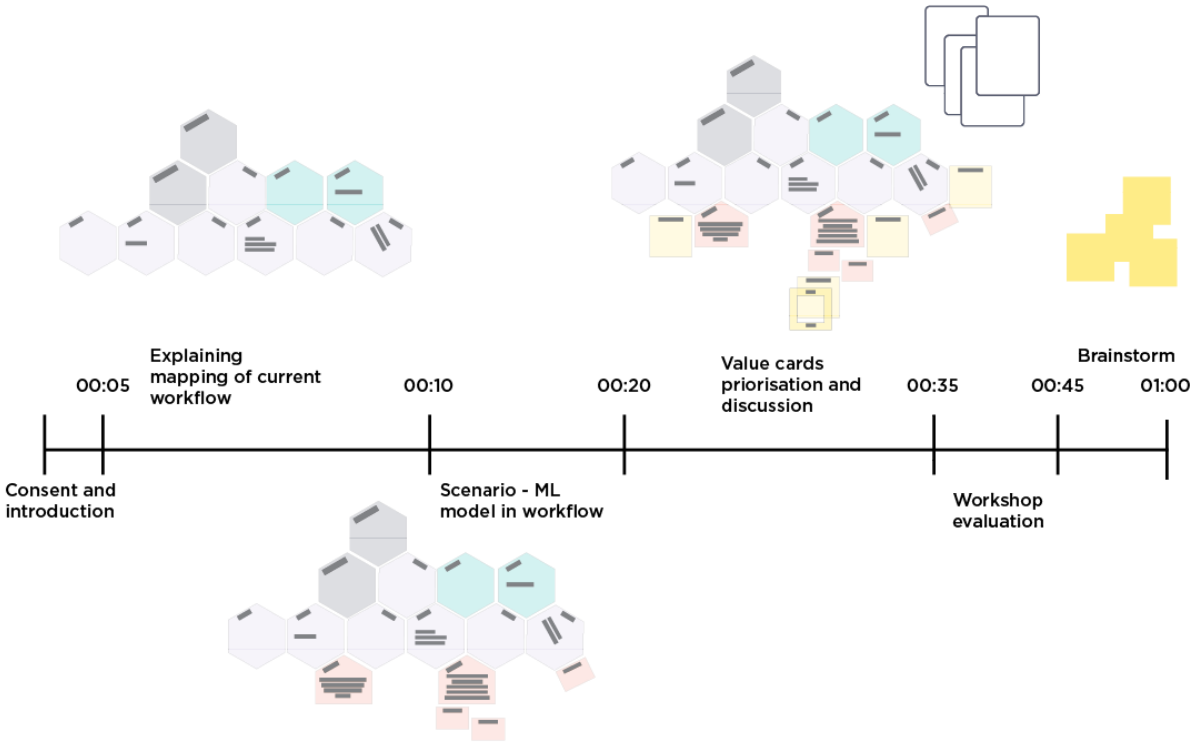


Figure 4: workshop timeline

Analysis of results

Improvement suggestions and brainstorm results will be taken up into the protocol of the GP research.

Plan

Workshop: 30 min

1. Sensitizing **(5 min)**: current workflow is prefilled. Do you think any phases or important people are missing? (Making clear this is not for content input but to get familiar with the journey)
    - Assumptions about model for now:
        - It is about cardiac risk prediction for patients who are not yet known as heart patients. Two options: complaints or not.
        - Workflow from arrival at GP (since not yet heart patient) to treatment selection and monitoring (is based on risk after all)
2. Explanation of ML and case **(10 min + mapping)**
    a. Is this clear?
    b. Are you missing information?
3. Role of the ML
    a. Again, making clear I know they do not have medical background. About seeing how they reason.
    b. Where in the workflow could the model be of use?
        i. In what way? writing blank cards
    c. Data in and outputs
        i. See obstructions?
        ii. How to use output?
        iii. Output handy at other moments?
4. Ethical values **(15 min)** : what do you think is important in mapping such a model?
    a. Cards clear?
    b. Which stand out in importance?
    c. How do they feature in the way the model is now placed in the workflow?
    d. Asking per value
        i. How important? Why? How to solve issues?

Feedback on workshop (15 min) - or embedded in different workshop elements?

- General remarks?
- Workflow mapping:
    - do-able?
    - Do you think it should be in more detail to work (more of a question to developers I guess)
    - would general personas help?
- ML in workflow: do-able? level of explanation?
- Ethical values: understood?

Brainstorm

1. How to make journey set-up more fun?
2. How to make cards more fun, creative, easier to integrate, easier to focus on, easier to sort through?
3. ML Model: specific ways in which it can be implemented?

Script

Introductie

- Uitleg project (5 min - 15:35)
    - Ik kan achteraf nog meer vertellen over de grotere context van mijn project, jullie weten het deels misschien ook al, maar om niet te beïnvloeden zal ik het kleinschalig uitleggen.
    - Research project: welke ethische waarden een rol spelen in het ontwikkelen van een ML tool binnen hartfalen. Zijn die losse onderdelen misschien nog vaag nu, maar hier gaan we vandaag dieper in duiken.
    - Deels gaan we door de workshop, dan een stukje evaluatie daarop, en een korte brainstorm op een van de workshop elementen.
        - Gaat ook om of het gesprek aan te wakkeren is, hoe de timing klopt. Niet perse om echt de inhoud, dus geen zorgen over foute opmerkingen/vragen.
    - Consent form gezien en getekend?

Workshop gedeelte

- Workshop gedeelte:
    - Gaan eerst kijken naar de workflow, dan hoe het model hierin past, vervolgens naar de ethische waarden die hierin een rol spelen.
    - sensitising
- **Journey (5 min - 15:40)**
    - Het model van het project helpt uiteindelijk bij **het screenen van patienten die de komende 10 jaar mogelijk hartproplemen onder de 50 die hier nog niet eerder last van hadden.**
    - In de sessie wil ik daarom de workflow van de huisarts in de praktijk in kaart brengen **- eerste contact tot stabiele status van patient**
    - Snap dat jullie de medische achtergrond niet hebben, maar ik heb het zo veel mogelijk voorbereid
        - Normaal zou ik aan de HA vragen naar wanneer patient bij HA komt, welke stappen onderneemt, wat tot stabiele toestand, gezamelijk besluiten maken, wie betrokken zijn.
    - *Ga er even doorheen, kijk of het sense maakt, of het is wat je zou verwachten.*
        - *Nog extra kaartjes, mist er nog iets wat je logisch zou vinden?*
        - *Toevoegingen?*
        - *Dingen die niet logisch zijn?*
        - *Zijn er nog knelpunten die jullie voor je zien, positieve en negatieve ervaringen?*

- **Rol van ML model (10 min - 15:50)**
    - Dan gaan we daar nu het ML model aan toevoegen.
        - Heeft men weleens gehoord van machine learning?
            - Zo ja, wat weet men hierover?
            - Zo niet: machine learning kan heel behulpzaam zijn in medische context, met diagnosticeren of beslissingen. *machine learning werkt eigenlijk volgens een soort uitgebreide beslisboom of bepaalde statistische*

netwerken. Een model wordt getraind met een trainings dataset, waarmee het model zichzelf leert met welke verbanden een bepaalde uitkomst tot stand komt. Uiteindelijk stop je daadwerkelijke data in het model en kan deze voorspellingen over gevallen uit de praktijk doen. Zo kan een ML bijvoorbeeld helpen met diagnoses stellen vanuit scans. Een vorm van ML is een CDSS: deze helpt zorgprofessionals keuzes maken met zijn voorspellingen.

- *Dit klinkt natuurlijk heel ideaal, maar er zijn nog stappen te nemen voor we daar zijn: wanneer de training dataset niet alle patientgroepen omvat of fouten heeft, kan het model bepaalde vooroordelen of risico's hebben: bijvoorbeeld als een bepaalde etnische groep of een gender niet of minder deel is geweest van het trainen van het model, en deze hebben wel een andere relatie tussen hun risicogedrag en prognose, kan het model hier niet mee omgaan en geeft het verkeerde uitkomsten zonder dat dit merkbaar is. Of wanneer een model verkeerde variabelen met elkaar verbindt, die niet echt een relatie hebben. In de uitvoering van het model is daarom de rol die het model inneemt in vergelijking met de expertise van de arts belangrijk (transparant, interpreteerbaar, uitkomt icm eigen kennis (Vokinger et al., 2021). Daarnaast is het belangrijk dat een model goed bij het zorgpad van de patiënt en dokter past, door begrijpelijk en te vertrouwen te zijn, en het gesprek en de zorg te bevorderen.*

- Dan heb ik nu een nieuwe laag kaarten om eraan toe te voegen.
  - ***EERSTE model SCORE2***
  - ***Scenario: er wordt op het moment een tool ontwikkeld die op basis van de informatie in het elektronisch patiënten dossier een voorspelling kan maken van het risico op hartproblemen voor patiënten onder de 50 jaar, die niet eerder iets met hun hart hebben gehad.***
    - ***Data kaarten: de data van het EPD gaat hier dus in, en de risicocalculatie komt eruit.***
- *Wat is je eerste reactie op deze nieuwe innovatie?*
  - *Is dit positief of negatief?*
  - *Waarom?*
- *Zou je dit voor je zien in clinische context?*
  - *Waarom wel/ niet?*
- *Als we de workflow erbij halen*
  - *Wanneer zou je het voor je zien?*
  - *Hoe past het in de workflow?*
  - *Hoe zou je het voor je zien?*
    - *In elke interactie moment, wat voor informatie heb je nodig?*
    - *Via welke systemen zou het kunnen werken?*
- *Bij welke beslissingen zou het kunnen helpen?*
  - *voorspellen?*
  - *feedback?*
  - *keuzes?*
- *Zie je een rol voor deze innovatie binnen je interactie met een patiënt?*
  - *Hoe zou dit eruit kunnen zien?*

        - *Hoe denken jullie dat een patiënt hierop zou reageren?*
            - *Gezamelijk of achter de schermen?*

        - *Hoe zou dit ontwikkeld moeten worden om het voor jou/ in de praktijk te laten werken?*
            - *Zijn er bepaalde kansen die je voor je ziet?*
            - *Zijn er risico's die je voor je ziet?*
            - *Wat is voor jou de belangrijkste overweging?*
        - ***Data laag***
- **Values (15 min - 16:05)**
    - Om te zorgen dat het model goed in de workflow past en de verschillende mensen die hierin een rol spelen respecteert, zijn ethische waarden belangrijk om te onderzoeken. Ik heb een aantal waarden uit de literatuur verzameld om jullie hiermee op weg te helpen.
    - *(met kaarten)*
        - *Kijk ze vooral even door, wat is je eerste reactie?*
            - *Begrijpt u wat de kaarten inhouden?*
        - Uitleg kaartjes waarden: op het mooment dat je een waarde belangrijk vind op een moment in de journey, kan je hem hierop schrijven. Vervolgens schrijf je eronder hoe dit een rol zou moeten spelen. Bijvoorbeeld privacy, kan geuit worden in de manier dat de patient schriftelijk toestemming moet geven voor het gebruik van zijn data in de workflow.
            - *Per interactiemoment in de workflow, welke waarden zijn belangrijk?*
                - *Op welke manier?*
                - *Hoe uit zich dit?*
        - *Zijn er bepaalde die eruit springen qua relevantie?*
        - *Zitten er tussen die niet belangrijk zijn?*
    - ***(met voorbeelden erbij)***

Evaluatie (15 min-16:20)

- Algemeen:
    - Wat vond je van de workshop?
        - Wat vond je het leukst?
        - Wat het minst leuk/ interessant?
- Journey
    - Is deze denk je (makkelijk) in te vullen?
        - Hoe kan ik dit nog duidelijker maken?
- Role ML en data
    - Hoe vond je het mappen van het model op de workflow?
    - Wat vond je van het nadenken over de data laag?
    - Was dit te doen? Te begrijpen, te breed of te specifiek?
    - data laag?
    - vooral score2
- Waarden
    - Waren de waarden te begrijpen met de kaarten?
    - Wat vond je ervan deze toe te passen op de workflow?
    - Hoe kan ik dit onderdeel verbeteren?
    - kaarten eruit?

- Algemeen 2.0
    - Wat vond je van de hoeveelheid informatie die werd gegeven tijdens de workshop?
    - Stel zo'n machine learning tool wordt echt ontwikkeld, op welke manier zou jij de betrokkenheid van huisartsen voor je zien bij het ontwerpproces daarvan?
    - Wat kan ik het beste nog veranderen voor ik de workshop aan huisartsen voorleg?

Brainstormvraag (optioneel) (15 min - 16:35)

- In welke vorm/ manier kunnen de waarden nog beter in de workshop geintegreerd worden? (nu kaarten misschien overwhelming)
    - Zijn er manieren waarop je meer creativiteit voor je ziet in de sessie (bijvoorbeeld in het gebruik van de kaarten)?

- Hoe kan de tool het gesprek over values in de journey integreren nog beter faciliteren?
- **Hoe kan de tool meer voor zich spreken (het gesprek faciliteren), minder gesprek vanuit mij nodig?**

Afronding (5 min)

Additionele value vragen

- Transparantie
    - Wat voor informatie heb je nodig om dit systeem te gebruiken/ begrijpen?
    - Moet je kennis hebben van de ontwikkeling van het systeem om deze te willen/kunnen gebruiken?
    - Is het belangrijk te weten hoe precies het systeem tot een beslissing/ diagnose is gekomen?
        - Wat wil je hieraan precies weten?
    - Hoeveel moet een patient weten om het systeem te gebruiken?
- Rechtvaardigheid, fairness (eerlijkheid/ redelijkheid), gelijkheid
    - Kan het systeem voor alle patienten gebruikt worden?
        - Waarom wel/ niet?
        - Wie wordt buitengesloten, is dat kwalijk?
    - Kan het model op een manier discrimineren tussen patienten?
- Niet-schadelijkheid
    - Zijn er vormen van schadelijkheid waar je je zorgen om maakt?
    - Wie krijgen de uitkomsten te weten? Wat kunnen zij hiermee?
- Verantwoordelijkheid
    - Wie is er verantwoordelijk bij een beslissing van het model?
    - Hoe beinvloedt dit het contact met de patient?
- Privacy
    - Wat voor problemen zouden er zich op kunnen doen tijdens het gebruik van het systeem?
    - Is het iets waar een patiënt bijvoorbeeld schriftelijke toestemming voor zou moeten geven?

- Weldadigheid
    - Op welke manieren kan het model goed doen?
        - Hoe kan dit tot stand komen?
- Vrijheid en autonomie
    - Hoe blijft jouw autonomie als huisarts in stand tijdens het gebruik van het systeem?
    - Hoe wordt de autonomie van de patient gerespecteerd tijdens het gebruik van het systeem?
- Vertrouwen
    - Zou je het systeem vertrouwen?
    - Wat is er voor jou nodig om het systeem te vertrouwen?
    - Hoe denk je dat een goede balans tussen te veel en te weinig vertrouwen tot stand komt?
- Waardigheid
    - Wat is er nodig zodat het systeem mensen in hun waarde laat?
    - Zie je moeilijkheden bij bepaalde patientgroepen voor je?
- Solidariteit
    - Denk je dat het model effect kan hebben op sociale interacties, professioneel of met patienten? Goed of slecht, hoe kan dit in goede banen geleid worden?

A4.2 Pilot photos

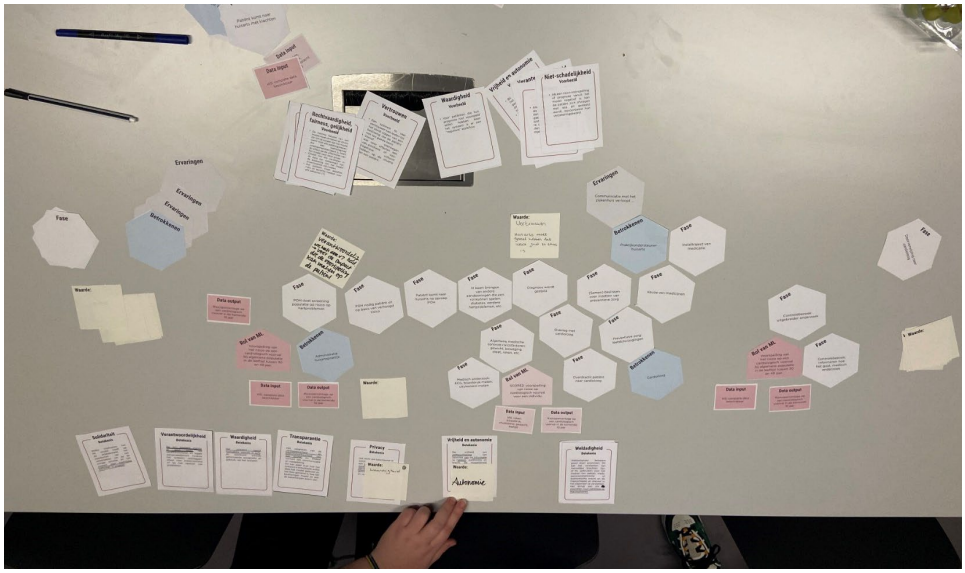


Figure A4.2.1 The mapping resulting from the pilot.

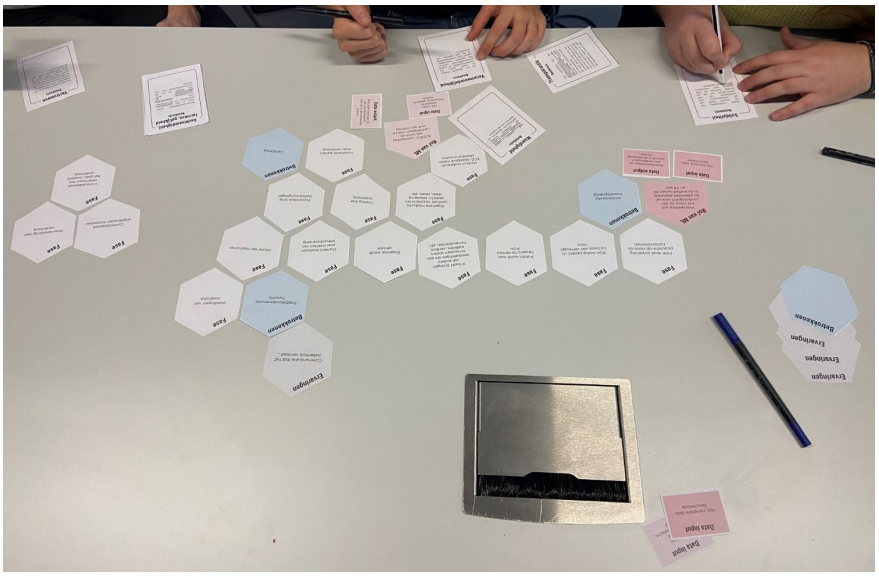


Figure A4.2.2 The pilot participants interacting with the probe. In the photo, participants are crossing out what information they felt was redundant on the ethical value cards.

Purpose and goal of research

Machine Learning (ML) has great potential in the healthcare sector by being able to aid in diagnosis (Greenes, 2014; Mosquera-Lopez et al., 2015), treatment selection (Zamora et al., 2013) and prognosis or risk prediction (Cai et al., 2019; Khedkar et al., 2019; Yang et al., 2019). Besides, utilizing ML could presumptively improve patient and staff experience and lower healthcare costs (Browne et al., n.d.). However, the AI field is very much focussed on developing models that perform well, while integration into practice and adoption by users is a challenge (Browne et al., n.d.). To bring the tools into practice, proper consideration of ethical issues is vital (Jobin et al., 2019; Karimian et al., 2022)

This research is part of the DECIDE-VerA project in collaboration with LUMC, Leiden Law school, TU Eindhoven and Hogeschool Rotterdam. The DECIDE-VerA project features a AI-Clinical Decision Support System (AI-CDSS) that is built to aid in risk analysis for cardiovascular patients. The project focuses on how the design process of such a ML system can be altered to involve the ethical values of all stakeholders, specifically to eventually improve the shared decision making of doctors and patients.

This research presents a starting point of this research, by exploring a manner in which the ethical values important to the general practitioners (GP's) can be discovered. The goal of this research is to validate a workshop format. This is designed in a twofold of (1) sensitizing the GP by going through a general mapping of the current workflow and plotting the use of the ML tool on this workflow, and (2) prioritizing and matching different ethical values of AI in health to this workflow.

Research question: Which considerations are important in designing a workshop that identifies clinician ethical values in the early stages of the design process of a ML Clinical Decision Support System (for cardiac risk assessment)?

Sub-questions:

- What does the current workflow of the clinician look like?
    - Where can de ML CDSS aid in this workflow?
- Can a physical tool create overview over the current workflow of the clinician?

- Can a physical tool create overview in the role of a ML model in the workflow?
- Are value cards a way to aid the clinician in discussing the ethical AI values that matter to them?
- Can value cards on ethical AI values spark discussion on how the ML system should be designed?

Target population and recruitment strategy

The target population of the study are General Practitioners (GPs) and General Practitioners Assistants (GPAs). The inclusion criteria is that they work in the Netherlands, to ensure a consistency in workflow. Participants will be contacted via personal networks, through a recruitment message (see appendix 1).

Procedure:

Participants' introduction to the study

The GP's and GPA's will be given general information about the session via the recruitment message. The informed consent form will also be sent via email beforehand (appendix 2), so participants know what to expect.

Measures

The session with GP's will consist of the workshop itself, and a short evaluation afterwards. The workshop itself has three parts: workflow, ML model and value mapping (figure 1). In advance, the GPs will be asked to sign the informed consent form (appendix 2).

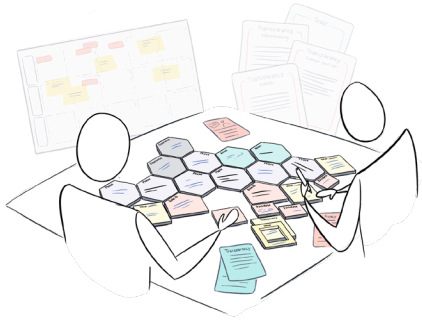


Figure 1: representation of the workshop.

Part 1: workflow mapping

Through a semi-structured interview, the participants will be guided through the tool.

The mapping of the workflow is intended to serve as a sensitizing assignment, getting the GPs into the context and triggering latent needs and thoughts (Sanders & Stappers, 2012). The basic shape of a user journey map is used (*Journey Mapping 101*, n.d.), of which layers are selected based on their relevance. Due to the complex, multi-stakeholder environment and the interest in how the workflow can be improved, this selection is “Phases” and “Actions”, “Actors” and “Experiences”.

Furthermore, the workflow gives a reference point to map the possible use of the ML model. In order to add the role of the ML model to the journey, the essence of a service blueprint was used as inspiration (*Service Blueprints*, n.d.). The execution shape of the ML model fulfills the “frontstage technology” layer. The data represents the support processes of the model. These terms are not used during the workshop, but were used to think of the right questions to ask.

The ML model will be explained to the GP through the use of a scenario (Martin & Hanington, 2012). This scenario will be based on the model’s description by the developer. Afterwards, the clinician is able to map possible uses of this model through pre-filled design tokens, or add to these with empty ones. The tokens will be designed in cooperation with the developers and Industrial Design students, and an expert from the NVVC and a GPA are consulted to understand the basic workflow.

The journey and model “blueprint” are represented in a physical manner, allowing the participant to shift upon adding the model and later add the values in a dynamic way.

The tool will be partially filled in beforehand, based on literature, expert interviews and interviewing the ML model project owner. See figure 2 for a representation of the tool:

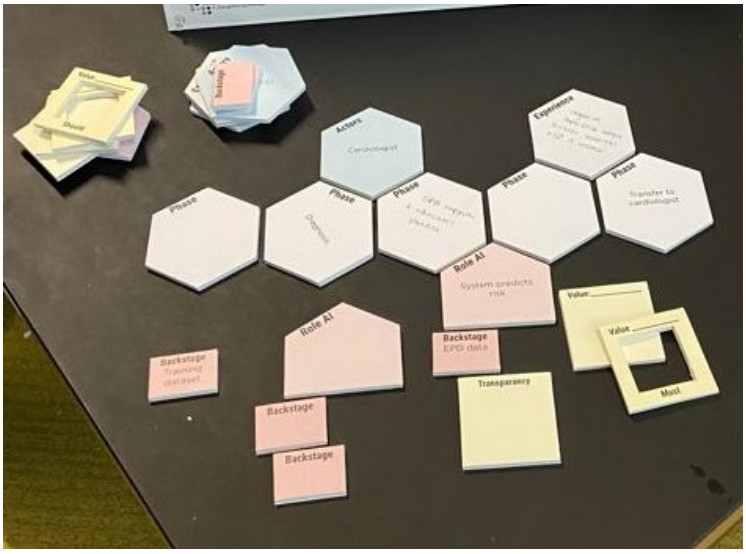


Figure 2: demonstration of the physical workshop elements and a possible way these can be filled in

Part 2: ethical value mapping

To understand general ethical values of AI, the work of Jobin et al., (2019), who formulated 11 main ethical AI values based on a literature review, was studied. The work of Jobin et al. (2019) was summarized in ethical value cards. when needed for clarification or to provide a more specific health-related view on a value, other explanations from literature specifically on ethical AI values in the domain of health were consulted. These are to be used in a card sorting (Martin & Hanington, 2012) way and to map onto the tokens. See an example in figure 3.

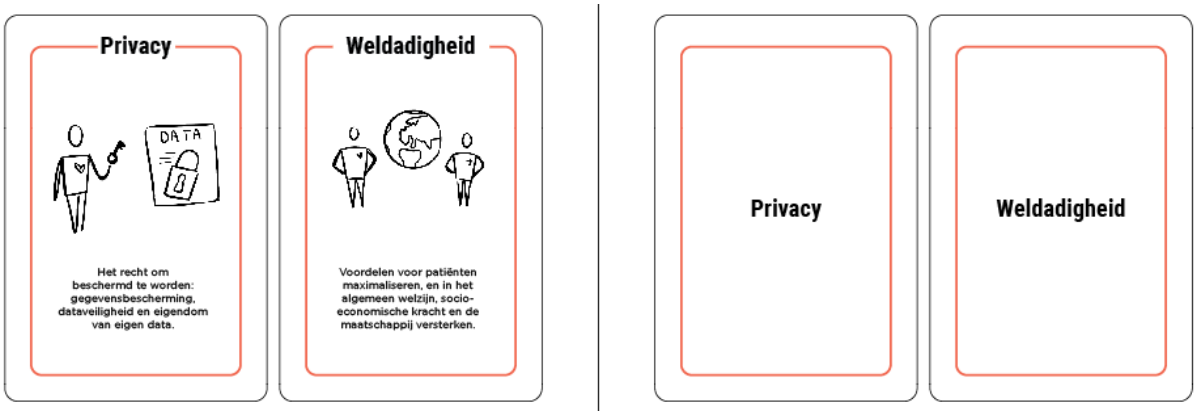


Figure 3: example card from the value set.

Part 3: Evaluation

After the workshop element, the participants will be briefly questioned on how they experienced the workshop, how it fit their expectations and how they see these technological design processes in the future.

Workshop overview

The complete workshop with its timing will therefore look like figure 4:

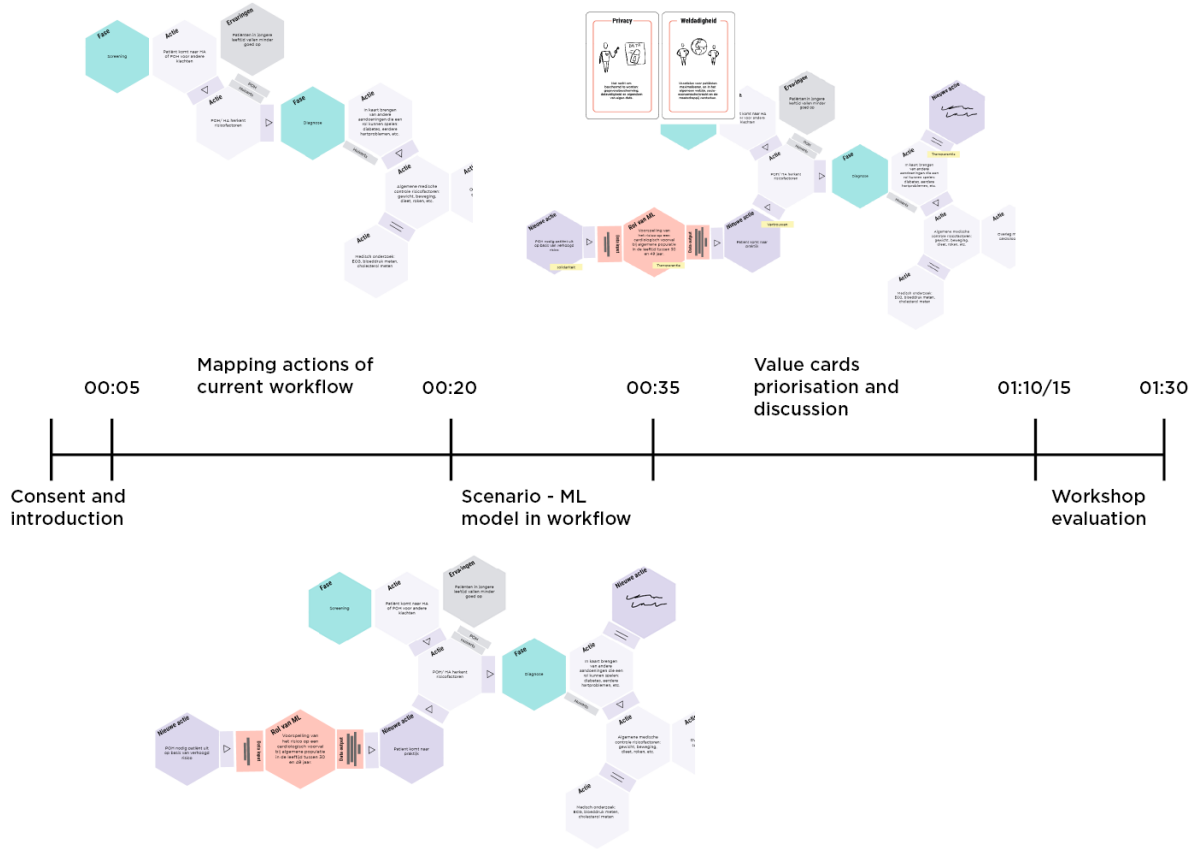


Figure 4: workshop timeline

Analysis of results

Observations will be noted after the workshop. The audio of the session itself will be transcribed, after which an inductive thematic analysis will be performed to uncover patterns in the systems usage and utility.
Video analysis will be used to understand what parts of the journey the participant is referring to, and added to the transcription between brackets.

If in scope, in subsequent evaluation, the sessions resulting filled out tool will be shown to designers experienced in ML development, to evaluate the added value of the results in the design process.

Semi-structured workshop interview guide (see Appendix 3 for english version)

- telefoon en ipad opgeladen, horloge om

- Basis flow neerleggen
    - andere kaarten neerleggen (stakeholders overeind, kaarten per stapel)
- Camera opzetten (telefoon)
- Audio opzetten (ipad)
- consent form
- protocol en value uitleg bij de hand

1. Inleiding project

- Ilse, bachelor ID met interesse in ontwerpen voor de eindgebruiker in gezondheidszorg- eindscriptie bachelor over verbeteren cardio revalidatie.
- Nu bezig met het research project van mijn master. Hiervoor ben ik aangesloten bij een onderzoek van het LUMC, over de ethische waarden binnen clinical decision ML systems (data-gedreven innovaties) binnen gezondheidszorg.
    - Ik kijk naar hoe de menselijke waarden of afwegingen van de huisarts en praktijkondersteuner betrokken kunnen worden bij het ontwerpprocess van dit soort vernieuwingen binnen de gezondheidszorg. Dit doe ik met een case study van een tool voor cardiovasculaire risicovoorspellingen.
    - Ik heb een workshop ontworpen, hiermee kunnen uiteindelijk ontwerpers in gesprek gaan met HA en POHs en andere clinici over de implementatie van hun systeem en de ethiek hieromheen.
    - Uiteindelijk draagt het onderzoek eraan bij dat in de toekomst de behoeften van de clinici beter in een technisch process kunnen worden betrokken. En voor u kan het interessant zijn over dit soort nieuwe ML tools te horen, veel potentie in de toekomstige gezondheidszorg.
- Vandaag gaan we samen de tool gebruiken, hij heeft 3 onderdelen, we gaan eerst de huidige manier van werken rondom een cardiovasculaire patiënt in kaart te brengen, kijken naar hoe de ontworpen interventie hierin een rol kan spelen, en welke ethische waarden hierin belangrijk zijn te overwegen.
    - Tijdens de sessie zou ik dus de ontwerper zijn die uiteindelijk de tool voor in de praktijk gaat ontwerpen.

- Er zijn geen foute antwoorden of als u iets niet weet is dit niet erg: het is een oriënterend onderzoek op of deze tool werkt, en daarom zijn ook de vragen die u stelt of dingen die niet duidelijk zijn relevant.
    - Denk vooral hardop tijdens elk onderdeel.
- Dan heb ik een consent form, waarin u toestemming geeft dat ik de antwoorden die u geeft mag verwerken, en dat ik audio en video opnamen mag maken van het gesprek om later uit te werken. (verwijderen en anoniem)

2. Algemene vragen

- Achtergrond: hoe lang HA/POH, waar?
- Gebruikt u op het moment tools om gezondheidsrisico voorspellingen te doen? (Binnen cardio, daarbuiten?)
- Hoe kijkt u over het algemeen naar nieuwe technische ontwikkelingen in de zorg?
    - Hoe krijgt u hiermee te maken?
    - Bent u hier tevreden mee?
- Hoe kijk je naar ethische waarden binnen de gezondheidszorg?
    - Wat voor rol zouden deze moeten spelen in de praktijk?
    - En bij het ontwikkelen van nieuwe innovaties?

3. Workflow

- De innovatie die is ontworpen speelt een rol in het behandeltraject van een cardiovasculaire patiënt.
    - In het algemeen: wat voor rol speelt u voor deze patiënten?

- *Voor de eerste stap van de tool zou ik graag **in ongeveer 15 minuten** samen het zorgpad voor een cardiovasculaire patiënt visueel te maken. Zo kan ik uw werk begrijpen en hebben we een fysiek iets om de volgende stappen aan te relateren.*
- *Het gaat om het zorgpad van de primaire preventie van een cardiovasculaire patiënt, dus zonder event. Op basis van input van experts heb ik een eerste opzet van dit zorgpad gemaakt, die ziet u hier.*
- *Het is het zorgpad wat de HA en POH en mogelijk andere experts uitvoeren, gedurende het behandelen van een primaire preventie CV patient.*
    - *Tegels voor fasen, activiteiten/acties, en connectors voor ertussen.*

        - *Met deze kan je aangeven wie er bij het process betrokken zijn en wanneer.*
        - *Ook zijn er tegels voor bepaalde positieve en negatieve ervaringen, die we erna toevoegen.*
- Wat ik tot nu toe begreep zijn er een aantal globale fasen: de screening van de patient, de diagnose, preventieve zorg en een stabiele fase.
    - Klopt dit hoog over naar uw ervaring?
- En als we kijken naar de acties van de verschillende fasen
    - Kloppen deze?
    - Wat mist er?
    - **Inzoomend, als we kijken naar de screening**
        - Wanneer komt een cardiovasculaire patiënt het eerste bij de huisartspraktijk terecht, in het geval van primaire preventieve zorg (dus geen event)?
        - Welke acties worden ondernomen?
        - Wie heeft er met de patient te maken?
    - **Diagnose**
        - Welke stappen worden er ondernomen om de patiënt te onderzoeken en diagnosticeren?
            - Werken jullie met u-prevent?
            - Werken jullie met SCORE?
        - Wie zijn hierbij betrokken? Hoe?
        - Speelt hierin het gezamenlijk besluiten maken?
            - Op welke manier?
    - **Preventieve zorg**
        - Welke acties/ stappen worden doorlopen?
        - Wie is hierbij betrokken?
        - Speelt hierin het gezamenlijk besluiten maken?
            - Op welke manier?
    - **Stabiele controles**
        - Welke acties/ stappen worden doorlopen?
        - Wie is hierbij betrokken?
- **Stakeholders**
    - VSD: Wie zijn belangrijke mensen, groepen of gemeenschappen die betrokken zijn?
        - Waar in de workflow worden ze betrokken? Met wie hebben ze te maken, hoe?

    - VSD: directe betrokkenen van workflow hebben we nu, zijn er ook nog indirecte betrokkenen?
  - Wie worden er nog meer betrokken in het zorgpad? Missen er nog mensen/ partijen?
  - VSD: Wat zijn de relaties tussen de betrokkenen?

- **Ervaringen**
  - Zijn er bepaalde knelpunten waar u tegenaan loopt in het zorgpad?
    - Zijn er dingen die wel heel goed gaan?

- ***Maak foto van neergelegde flow***

4. ML Scenario

- *De komende 15 minuten wil ik nemen om te kijken of de innovatie die door het LUMC wordt ontwikkeld waarde kan toevoegen in de zorgpad die we net hebben opgesteld.*
- Heeft men wel eens gehoord van machine learning?
  - Zo ja, wat weet men hierover?
  - Zo niet: *machine learning werkt eigenlijk volgens een soort uitgebreide beslisboom of bepaalde statistische netwerken. Een model wordt getraind met een trainings dataset, waarmee het model zichzelf leert met welke verbanden een bepaalde uitkomst tot stand komt. Uiteindelijk stop je daadwerkelijke data in het model en kan deze voorspellingen over gevallen uit de praktijk doen. ((Zo kan een ML bijvoorbeeld helpen met diagnoses stellen vanuit scans, wanneer hij is getraind op fotos van tumoren kan hij deze vormen uiteindelijk herkennen.)) Zo kan het model in dit geval kan cardiovasculaire risicofactoren gebruiken voor een voorspelling. Daar vertel ik zo meer over*
  - *Dit klinkt natuurlijk heel ideaal, maar er kunnen ook nadelen aan een model zitten: wanneer de training dataset geen goede representatie is van de daadwerkelijke data, kan het model bepaalde vooroordelen of risico's hebben, het model overgeneraliseert dan. Als bijvoorbeeld een gender niet of minder deel is geweest van het trainen van het model, en deze hebben wel een andere relatie tussen hun risicogedrag en prognose, kan het model hier niet mee omgaan en geeft het verkeerde uitkomsten. Daarnaast heeft het model niet door dat het fout zit.*

    - *Naast het goed trainen van het model is het goed implementeren van het model belangrijk: hoe het past in de manier van werken en het bevorderen van de zorg voor de patiënt. Daarom wil ik samen gaan kijken hoe jij naar het ontwikkelde model kijkt.* *(transparant, interpreteerbaar, uitkomt icm eigen kennis (Vokinger et al., 2021)).*

- *Jullie kennen nu de tabellen van SCORE: bloeddruk, cholesterol, gender, roken, gewicht, leeftijd. -- niet bekend, tabel laten zien*
- ***Scenario: er wordt op het moment een ML model ontwikkeld die op basis van de informatie in het elektronisch patiënten dossier/ HIS de 10jaars risico op een cardiovasculair event in leeftijd 30-49 kan voorspellen. In plaats van dat SCORE dit doet op gemeten medische data, doet dit model het op basis van alle bekende data uit het HIS.***
- ***Het is een statistisch model, een soort beslisboom. Deze is getraind met data van 542,147 (540 duizend) patiënten, waarvan 51% vrouw zijn, afkomstig van STIZON, een database van reguliere zorg die elektronisch patiëntendossiers ontvangt. Met de uiteindelijke cardiovasculaire events in 10 jaar van 80% van de dossiers zijn risicofactoren geïdentificeerd en gebruikt om het model te trainen deze te gebruiken voor een risico voorspelling. Vervolgens is het model gevalideerd met de overige 20% van de data.***
- ~~***Het model is 95% accuraat, wat betekent dat in 5% van de gevallen hij fout zit: een risico voorspelt ook al is dat er niet, of geen risico voorspelt ook al is dat er wel.***~~
- ***Ook al is SCORE2 betrouwbaar bij diagnose van een individu, biedt dit model mogelijkheden voor een initiele screening van patienten op populatieniveau, op basis van HIS bekende data. Dit kan bijvoorbeeld data zijn over: anticonceptiepil, reuma medicatie, diabetes, depressie etc.***
- ***Dit hoeft niet los ingevuld te worden, dit is de al bestaande data.***
    - ***(Extra info als nodig: initiele screening, nog niet in de molen want kan zonder medische data. Huidig idee is helemaal hier vooraan. Zie je daar waarde? Waar anders?)***
    - ***Bij geen interesse/ moeite uitleggen waar de developers het voor zich zien: bij de POH een heads up in het systeem bij hoog risico, die patiënten opbellen voor medische check.***

- Wat is je eerste reactie op deze nieuwe innovatie?
    - Is dit positief of negatief?

        - Waarom?
- Zou je dit voor je zien in de praktijk?
    - Waarom wel/ niet?

    - ***De bestaande medische risicofactoren zijn meer valide, het model heeft niet perse een hogere accuraatheid (want vanuit HIS), maar het voordeel is dat je al een indicatie hebt wanneer je nog niet de medische waarden hebt.***

- *Als we de workflow erbij halen* Zou graag het model in de journey mappen, elke keer dat je er mee zou interacteren.
    - Zou het systeem in je huidige workflow passen? Waar? Hoe?-- (mogelijk extra uitleg/sturing)
        - Op welke momenten van de journey heeft dit nog meer effect?
        - (Overleg met patiënt, overleg met andere medici?)
    - Hoe zou je het gebruiken? Welke acties komen erbij?
        - Ervoor - input systeem
            - Denk je dat het systeem kan werken met de beschikbare data?
            - Hoe komt deze data in het HIS?
                - Zijn hier extra acties voor nodig?
        - Erna- output systeem
            - Tijdens interactie, wat voor informatie zou je willen van het systeem?
                - Waarmee kan het helpen? Op welke manier?
            - Wat kan je met uitkomst/ wanneer zou je er iets mee kunnen? Wat zet het in gang?

    - **Shared decision making/ interaction patient:**
        - Zie je een rol voor deze innovatie binnen je interactie met een patiënt?
            - Hoe zou dit eruit kunnen zien?

            - Hoe zou een patiënt hierop reageren?
                - Gezamenlijk of achter de schermen?
                - Zou het systeem een rol spelen in gezamenlijk besluiten maken?

        - Hoe?
- **Stakeholders**
    - Wie krijgen te maken met het systeem?
    - Zijn er nieuwe mensen bij betrokken? Op wie heeft het systeem op welk moment effect?
    - VSD: Wie geeft er nog meer om deze kwestie en waarom? Wordt er nog iemand buiten gelaten?
    - VSD: wat zijn de relaties tussen de betrokkenen?
        - Op welke manier ga jij om met…
            - De huisarts, cardioloog, assistente, verzekeraar, patient?
        - Op welke manier gaan ze met elkaar om?

    - **Zie je op een andere plek dan de initiële risico voorspelling en screening waarde voor het systeem?**
        - **Hoe zou dit eruit zien?**

- Hoe zou het systeem ontwikkeld moeten worden om het voor jou te laten werken?
    - Zijn er bepaalde kansen die je voor je ziet?

- VSD wat voor effect zou dit hebben nederland breed?
    - / verder in de toekomst?

- ***Maak foto van neergelegde flow***

5. Values

- *Laatste onderdeel van de workshop is om te kijken welke ethische overwegingen voor jou belangrijk zijn bij het implementeren van zo'n systeem in de workflow.*
- *Hier wil ik 30 minuten voor nemen.*

- Allereerst wil ik het als open vraag stellen: zijn er bepaalde ethische overwegingen die bij je opkomen nu we aan het bespreken zijn hoe het model in praktijk zou kunnen werken?
    - Zijn er risico's die je voor je ziet?
    - Wat is voor jou de belangrijkste over/afweging voordat je het systeem zou kunnen gebruiken?

    - *Vanuit literatuur heb ik een overzicht van waarden die een rol spelen bij het ontwikkelen van nieuwe data-gedreven innovaties. Het is relevant te kijken welke van deze waarden in deze context een rol spelen en op welke manier. Ik ben benieuwd hoe jij naar de waarden kijkt.*
        - *er staat een uitleg op de achterkant, ik kan voorbeelden geven als u wil.*

- Kijk ze vooral even door, wat is uw eerste reactie?
    - Begrijpt u wat de kaarten inhouden?
- Zijn er bepaalde die eruit springen qua relevantie?
    - Welke?
    - Waarom?
- Per interactiemoment in de workflow, welke waarden zijn belangrijk?
    - Op welke manier?
    - Hoe uit zich dit?
    - Waarom?
- Spelen er nog waarden op andere momenten in de workflow een rol?
- Zitten er tussen die niet belangrijk zijn?

- Zijn er dingen die je vind dat de POH, HA of patient sowieso zou moeten weten van het systeem voordat deze geïmplementeerd wordt? (over de mogelijkheden/ limitaties, nauwkeurigheid)

- ***Maak foto van neergelegde flow***

6. Eindreflectie - einde workshopgedeelte

- Wat vond je van de workshop?
    - Welk aspect ben je het positiefst over?
    - Over wat het negatiefst/ zou je veranderen?
    - Wat vond je van het gebruiken van de tool?
- Aan het begin vroeg ik naar de rol die ethische waarden zouden moeten hebben in nieuwe innovaties binnen de gezondheidszorg, is er iets veranderd hierin?
    - Waarom wel/ niet?
- Hoe vond je het mappen van de workflow?
    - Was dit moeilijk, makkelijk?
    - Hoe was het om de betrokkenen in kaart te brengen?

    - Ben je eerder op zo'n manier met je werkproces bezig geweest?
- Hoe vond je het horen over het ML Model?
    - Hoe vond je het mappen van het model op de workflow?
        - Was dit moeilijk, makkelijk?
- **Wat vond je van het derde gedeelte, over het in kaart brengen van de ethische waarden?**
    - Wat vond je ervan de waarden een plek te geven in de workflow?
        - Moeilijk, makkelijk?
            - zijn er dingen die dit makkelijker/ moeilijker maakten?
    - Wat vond je van het gebruik van de ethics kaarten?
        - wat was hier positief aan?
        - wat zou je hieraan verbeteren?
        - Hielpen de kaarten bij het mappen van de waarden op de workflow?
- Wat vond je van de driedeling tussen workflow - ML - values? (Logisch, omslachtig, nuttig, verwarrend)
- Wat vond je van de hoeveelheid informatie die werd gegeven tijdens de workshop?
    - Heb je het idee dat je de benodigde kennis en skills had om de tool te gebruiken? Hoe merkte je dat?
- **Zie je het nut in van de workshop? Waar zit dat nut?**
    - Vond je bepaalde onderdelen nuttiger dan andere?
- **Wat voor rol zouden medici moeten hebben in het ontwikkelen van dit soort nieuwe innovaties? (Zou jij hierin betrokken willen worden, waarom wel/niet, op welke manier, begin vs testen?)**

References

Browne, J. T., Bakker, S., Yu, B., Raijmakers, J., Pluyter, J., Ruijs, N., & Deckers, E. (n.d.). *Towards a Framework for Trust in Clinical AI: Expanding the Unit of Analysis*. 10.

Cai, C. J., Winter, S., Steiner, D., Wilcox, L., & Terry, M. (2019). "Hello AI": Uncovering the Onboarding Needs of Medical Practitioners for Human-AI Collaborative Decision-Making. *Proceedings of the ACM on Human-Computer Interaction*,

3(CSCW), 104:1-104:24. https://doi.org/10.1145/3359206

Greenes, R. A. (2014). *Clinical Decision Support: The Road to Broad Adoption*. Academic Press.

Jobin, A., Ienca, M., & Vayena, E. (2019). *Artificial Intelligence: The global landscape of ethics guidelines*. 42.

Journey Mapping 101. (n.d.). Nielsen Norman Group. Retrieved October 24, 2022, from https://www.nngroup.com/articles/journey-mapping-101/

Karimian, G., Petelos, E., & Evers, S. M. A. A. (2022). The ethical issues of the application of artificial intelligence in healthcare: A systematic scoping review. *AI and Ethics*. https://doi.org/10.1007/s43681-021-00131-7

Khedkar, S., Subramanian, V., Shinde, G., & Gandhi, P. (2019). *Explainable AI in Healthcare* (SSRN Scholarly Paper No. 3367686). https://doi.org/10.2139/ssrn.3367686

Martin, B., & Hanington, B. (2012). *Universal Methods of Design*. Rockport Publishers Inc. https://www.bol.com/nl/nl/f/universal-methods-of-design/39682951/

Mosquera-Lopez, C., Agaian, S., Velez-Hoyos, A., & Thompson, I. (2015). Computer-Aided Prostate Cancer Diagnosis From Digitized Histopathology: A Review on Texture-Based Systems. *IEEE Reviews in Biomedical Engineering*, *8*, 98–113. https://doi.org/10.1109/RBME.2014.2340401

Sanders, L., & Stappers, P. J. (2012). *Convivial toolbox: Generative Research for the Front End of Design.* BIS Publishers.

Service Blueprints: Definition. (n.d.). Nielsen Norman Group. Retrieved October 24, 2022,

from https://www.nngroup.com/articles/service-blueprints-definition/

Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S. (2021). Mitigating bias in machine learning for medicine. *Communications Medicine*, *1*(1), Article 1. https://doi.org/10.1038/s43856-021-00028-w

Yang, Q., Steinfeld, A., & Zimmerman, J. (2019). Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–11. https://doi.org/10.1145/3290605.3300468

Zamora, A., Fernández de Bobadilla, F., Carrion, C., Vázquez, G., Paluzie, G., Elosua, R., Vilaseca, M., Martín-Urda, A., Rivera, A., Plana, N., Masana, L., VALIDA Study Group, & Network of Lipid Units of Catalonia (XULA). (2013). Pilot study to validate a computer-based clinical decision support system for dyslipidemia treatment (HTE-DLP). *Atherosclerosis*, *231*(2), 401–404. https://doi.org/10.1016/j.atherosclerosis.2013.09.029

Appendix 1 - recruitment message

Mijn naam is Ilse, ik doe de master Industrieel Ontwerpen aan de TU Eindhoven. Ik ben op het moment bezig met een research project, waarvoor ik uw expertise als huisarts goed kan gebruiken! In mijn onderzoek werk ik mee aan een project over hoe ethische waarden van een machine learning tool (die helpt in de behandeling van hartpatiënten) in kaart kunnen worden gebracht. Ik ben hierin specifiek geïnteresseerd in hoe de ervaring en mening van huisartsen het beste in het designproces kan worden betrokken.

Voor het onderzoek wil ik een workshop-format testen. Hierin wil ik samen met huisartsen de huidige workflow omtrent een patiënt met hartfalen in kaart brengen, en het gesprek aangaan over wat hierin verbeterd kan worden en hoe de machine learning tool hierin waarde kan toevoegen.

Ik zou heel graag de workshop met u doen, om te kijken of het kan helpen uw waarden in kaart te brengen. Dit duurt een uur tot anderhalf uur, waarvoor ik langs zou kunnen komen in de praktijk of thuis. U hoeft niks voor te bereiden, ik vul met u samen de workshop tools in. Voor u kan het interessant zijn te horen wat voor mogelijke toepassingen machine learning kan hebben in de zorg, en natuurlijk helpt u mij enorm met mijn onderzoek!

Het onderzoek speelt zich af in de week van 21 november. Mocht u het leuk vinden om mee te werken kunnen we een moment zoeken dat u uitkomt.

Ik hoor graag of u geïnteresseerd bent, via i.p.faber@student.tue.nl of 0617892089.

Appendix 3 - English interview questions

Introduction

- Ilse, did ID bachelor with an interest in human-centered design in healthcare - bachelor's thesis on improving cardio rehabilitation.
- Currently working on my master's research project. I am part of a study by the LUMC about the ethical values within clinical decision ML systems (data-driven innovations) within healthcare.

    - I look at how the human values and considerations of the end user, so the general practitioner and practice nurse can be involved in the design process of these types of new innovations. I do this with a case study tool on cardiovascular risk prediction.
    - I designed a workshop for designers to conduct with clinicians, to talk about the implementation of their system and the ethics surrounding it.
    - Ultimately, the research contributes to the needs and values of the clinician being included in the technical development and implementation process of new innovations from the start on. And it may be interesting for you to hear about these new ML tools, potential in the future of healthcare.
- Today we will use the tool together. It has three parts. First we will map the current way of working surrounding a cardiovascular patient, to look at how the designed intervention can play a role in this, and which ethical values are important to consider in this.
    - So hypothetically I would be the designer that is developing this tool.
- There are no wrong answers or if you don't know something, don't worry: it's an exploratory study into how this tool works, so the questions you ask about things that aren't clear are also relevant.
    - Think aloud etc.
- Then I have a consent form, in which you give permission that I process your answers and input, and that I may make audio recordings of the conversation for later elaboration (will be deleted and anonymous).

- **General questions**

- Background: for how long HA, where?
- Do you currently use tools to make health risk predictions? (within CVD, outside?)
- How do you generally view new technical developments in healthcare?
    - How do you come into contact with new technical developments/ innovations?
    - Are you satisfied with this?
- How do you view ethical values within healthcare?
    - What role should they play in practice?
    - And when developing new innovations?

Workflow

- *The innovation that has been designed plays a role in the treatment process of a cardiovascular patient.*
- Very generally speaking: where do you play a role for these patients?
    - What role is this?

- *For the first step of the tool, I would like to visualize the care path for a cardiovascular patient together in about 15 minutes. This way I can understand your work and we have a physical mapping to relate the next steps to.*
- *It concerns the care path of the primary prevention of a cardiovascular patient (no event). Based on input from earlier interviews, I have made a first draft of this care path, which you can see here.*
- *It is the path of a patient, seen through the eyes of the healthcare professionals.*
    - *Tiles for phases, activities/actions, and connectors between them.*
    - *With tokens you can indicate who is involved in the process and when.*
    - *There are also tiles for certain positive and negative experiences, which we can add after.*
- From what I understood so far there are a number of global phases: patient screening, diagnosis, preventive care and a stable phase.
    - Does this match your experiences?
- And if we look at the actions of the different stages
    - Are these correct?
    - What's missing?

- ***Zooming in, looking at the screening***
    - When does a cardiovascular patient first come to the general practice, in the case of primary preventive care (i.e. no event)?
    - Who is involved with the patient?
    - What actions are taken?
- ***Diagnosis***
    - What steps are taken to examine and diagnose the patient?
        - Do you work with u-prevent?
        - Do you work with SCORE?
    - Who are involved?
    - What are next steps?
    - Does joint decision-making play a role in this?
        - How?
- ***Preventive care***

    - Which actions/steps are followed?
    - Who is involved?
    - Does joint decision-making play a role in this?
        - How?
- ***Stable check-ups***
    - Which actions/steps are followed?
    - Who is involved?

- **Stakeholders**
    - VSD: Who are key people, groups or communities involved?
        - Where in the workflow are they involved? Who do they deal with, how?
        - VSD: we now have direct stakeholders of workflow, are there also indirect stakeholders?
    - Who else is involved in the care pathway? Are there still people / parties missing?
    - VSD: What are the relationships between those involved?

- **Experiences**
    - Are there any bottlenecks you encounter in the care pathway?
    - Are there things that are going really well?

- **Make picture of flow**

<u>Role ML model</u>

- *I want to take the next 15 minutes to map whether the innovation designed by the LUMC can add value to this workflow.*
- *Have you ever heard of machine learning?*
    - *If so, what do you know about this?*
    - *If not: machine learning actually works according to a kind of extensive decision tree or certain statistical networks. A model is trained with a training dataset, with which the model learns itself through which variables a certain outcome is achieved. Ultimately, you put actual data into the model and it can make predictions about real-life cases. ((For example, an ML can help with diagnosis from scans, when trained on photos of tumors, it can eventually recognize these shapes)). In this case, it can recognize*

cardiovascular risk factors to make a prediction. I will tell more about this later.

- *This sounds very ideal, of course, but there can also be disadvantages to a model: when the training dataset is not a good representation of the actual data, the model can have certain biases or risks, the model then overgeneralizes. If, for example, a gender has not been part of the training of the model, or has been less so, and they do have a different relationship between their risk behavior and prognosis, the model cannot deal with this and gives wrong results. Besides, the model does not know it is wrong.*
- *In addition to properly training the model, proper implementation of the model is important: what role the model plays in relation to the expertise of the doctor and how it fits into the way of working and promoting patient care. Therefore, I want to see how you view the developed model. (transparent, interpretable, based on own knowledge (Vokinger et al., 2021)).*

- Explaining the model:
  - *SCORE as comparison: you now know the tables of SCORE: blood pressure, cholesterol, gender, smoking, weight, age.*
  - ***Scenario: An innovation is currently being developed that can predict the 10-year risk of a cardiovascular event in 30-50 y/o based on the information in the electronic patient record/HIS. Instead of SCORE doing this on measured medical data, this model uses all known data in HIS.***
  - ***It is a statistical model, a kind of decision tree. This has been trained with data from 542,147 (540 thousand) patients, 51% of whom are women, from STIZON, a database of regular care that receives electronic patient records. With the final cardiovascular events in 10 years of 80% of the files, risk factors were identified and used to train the model to use them for risk prediction. The model was then validated with the remaining 20% of the data.***
  - ~~***The model is 95% accurate, which means that in 5% of the cases it is wrong: it predicts a risk even if there is none, or predicts no risk even if there is.***~~
  - ***Although SCORE2 is reliable in diagnosing an individual, this model offers possibilities for an initial screening of patients at the population level, based on known HIS data. For example, this can be***

data about: contraceptive pill, rheumatism medication, diabetes, depression, etc.

- ***This does not have to be filled in separately, this is the already known data.***
    - ***(Additional info if necessary: initial screening, not yet in the mill because it can be done without medical data. Current idea is all the way here at the front. Do you see value there? Where else?)***
    - ***If there is no interest/difficulty, explain where the developers envision it: at the POH a heads up in the system at high risk, which patients call for a medical check.***

- What is your first reaction to this new innovation?
    - Is this positive or negative?
        - Why?
    - Would you envision this in practice?
        - Why/why not?

- **The existing medical risk factors are more valid, the model does not necessarily have a higher accuracy (because from HIS), but the advantage is that you already have an indication when you do not yet have the medical values.**

- *If we bring in the workflow* Would like to map the model into the journey, every time you interact with it.
    - Would the system fit into your current workflow? Where? How?-- (possible additional explanation/guidance)
        - At which moments of the journey does this have even more effect?
        - (Consultation with patient, consultation with other physicians?)
    - How would you use it? What actions are included?
        - Before - input system
            - Do you think the system can work with the available data?
            - How does this data get into the HIS?
                - Are additional actions required for this?
        - After-output system
            - During interaction, what kind of information would you like from the system?
                - What can it help with? How?

        - What can you do with the outcome/when could you do something with it? What new actions does it trigger?

    - **Shared decision/ interaction patient?**
        - Do you see a role for this innovation in your interaction with a patient?
            - What could this look like?

        - How would a patient react to this?
            - Would you want to use it together or behind the scenes?
            - Would the system play a role in joint decision making?
                - How?

- **Stakeholders**
- Who will be affected by the system?
- Are there new people involved? Who does the system affect at what time?
- VSD: Who else cares about this issue and why? Is anyone else being left out?
- VSD: what are the relationships between those involved?
    - How do you deal with…
        - The general practitioner, cardiologist, assistant, insurer, patient?
    - How do they interact with each other?

- **Do you see added value of the system in another part of the journey, besides the initial screening?**
    - **What would that look like?**

- How should this be developed to make it work for you?
    - Are there certain opportunities that you envision?

- VSD What effect would this have for the Netherlands as a whole?
    - / further in the future?

- **Make picture flow**

Ethical values

- *The last part of the workshop is to see which ethical considerations are important to you when implementing such a system.*
- *I want to take 30 minutes for this.*

- First of all, I want to ask it as an open question: Are there any ethical considerations that come to mind as we discuss how the model might work in practice?
    - Are there any risks you see ahead of you?
    - What are the most important considerations for you before you could use the system?

    - *From literature I have an overview of values that play a role in the development of new data-driven innovations. It is relevant to see which of these values play a role in this context and in what way. I'm curious how you look at the values.*
        - there is an explanation on the back, I can give examples if you want.

- Have a look through them, what is your first reaction?
    - Do you understand what the cards mean?
- Are there certain ones that stand out in terms of relevance?
    - Which?
    - Why?
- Per interaction moment in the workflow, which values are important?
    - In what way? How does that manifest itself?
    - Which?
    - Why?
    - (Ask further. Use additional questions sheet for examples and questions)
- Are there any that are not important?

- Are there things that you think the POH, HA or patient should know about the system before it is implemented? (about the possibilities/limitations, accuracy)

Post-workshop end reflection

- What did you think of the workshop?
    - What did you like the most?

    - What was the least fun/interesting?
    - What did you think of using the tool?
- At the beginning I asked about the role that ethical values should have in new innovations in healthcare, has anything changed in this?
    - Why/why not?
- How did you like the workflow mapping?
    - Was this hard, easy?
    - What was it like to identify stakeholders?
    - Have you looked at your work process in such a way before?
- What did you think of hearing about the ML Model?
    - How did you like the mapping of the model to the workflow?
        - Easy, hard?
- What was it like thinking of the ethical values?
    - What did you think of giving the values a place in the workflow?
        - Difficult/ easy?
            - Are there things that made this easier/harder?
    - What did you think of using the ethics cards?
        - what was positive about this?
        - what would you improve on this?
        - Did the maps help map the values to the workflow?
    - What did you think of the tripartite division between workflow - ML - values? (Logical, cumbersome, useful, confusing)
    - What did you think of the amount of information given during the workshop?
        - Do you feel that you had the necessary knowledge and skills to use the tool? How did you notice that?
- Do you see the purpose of the workshop? What is that?
    - Did you find certain parts more useful than others?
- What role should physicians play in developing these kinds of new innovations? (Would you like to be involved, why/why not, how, start vs testing?)

A5.1 LUMC model diagram

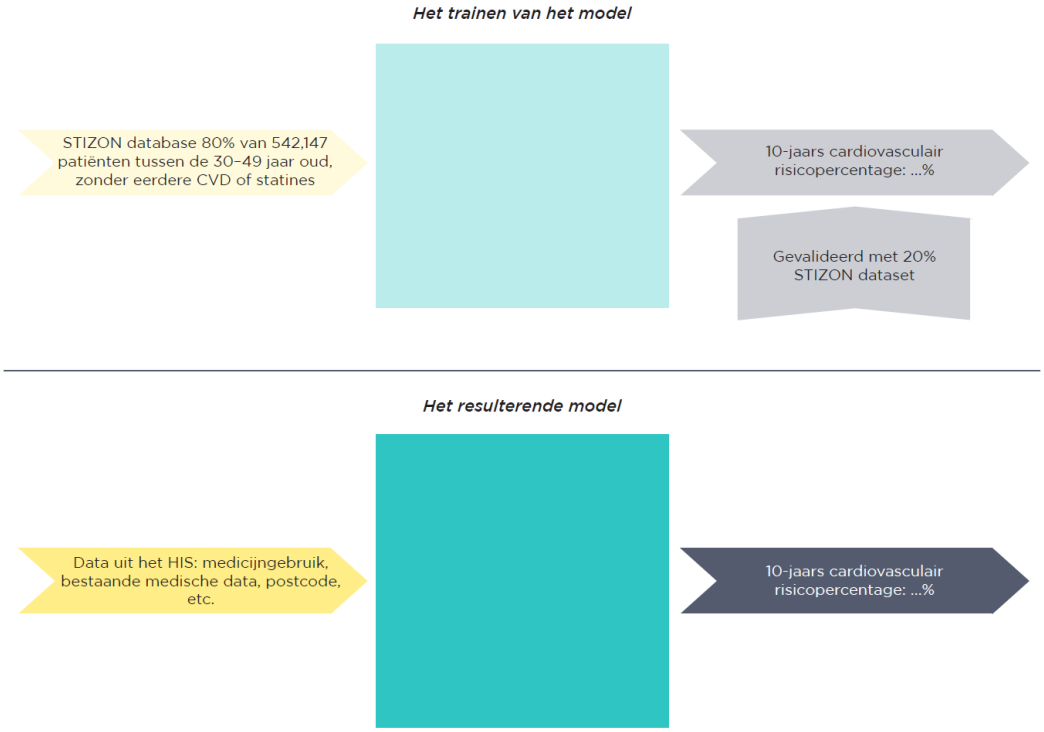


Figure A5.1.1: A visual diagram of the data flow in the training of the LUMC model and the eventual deployment, possibly used to clarify explanations during workshop phase 2 and 3.

A6 BASIC FORM JOURNEY

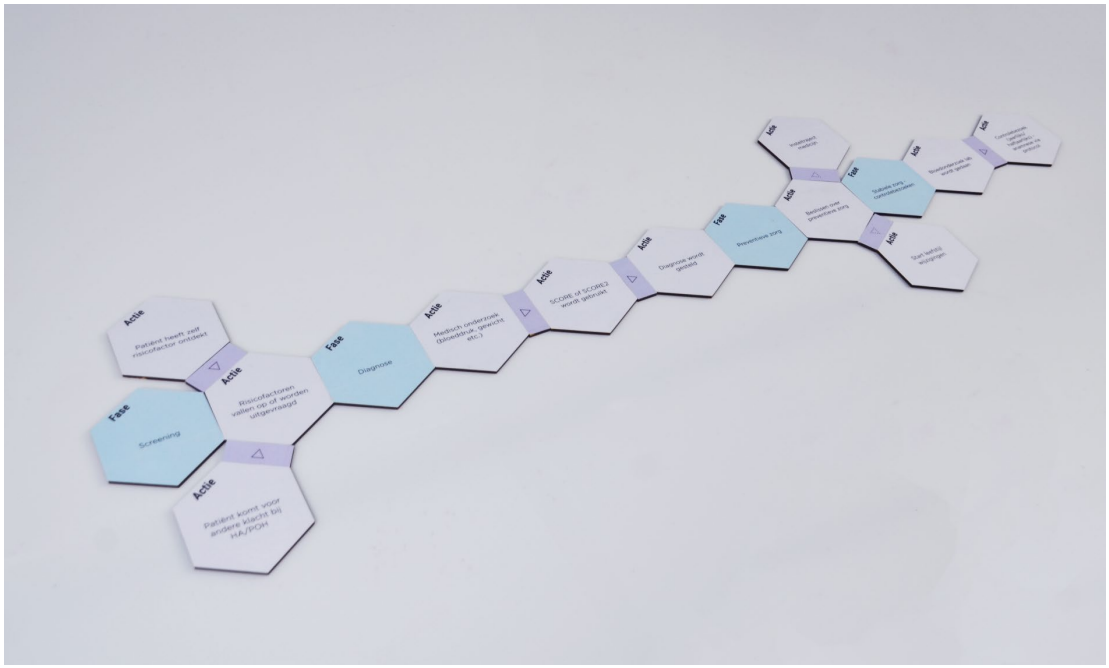
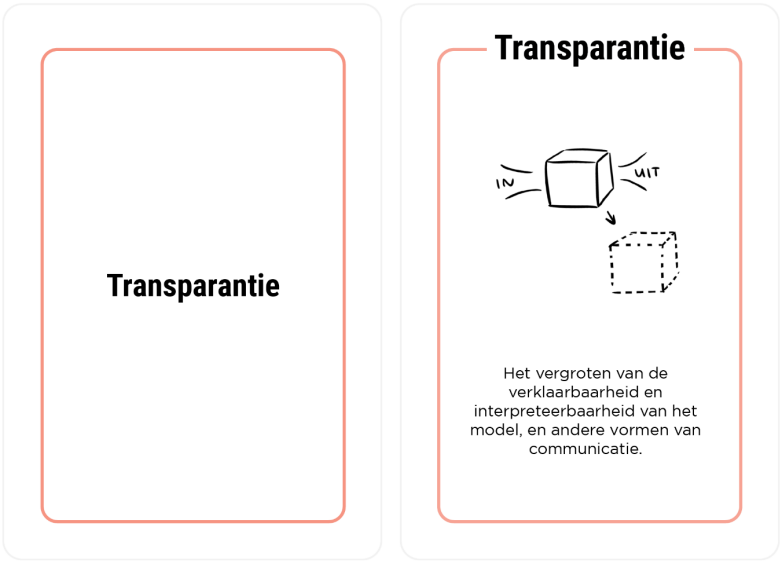


Figure A6.1 The starting journey that was presented to the participants in phase 1 of the workshop.

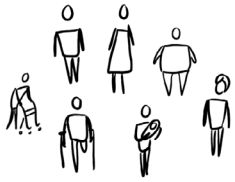
A7 FINAL VALUE CARDS

A7.1 Dutch final value cards



Rechtvaardigheid, fairness, gelijkheid

Rechtvaardigheid, fairness, gelijkheid



Het voorkomen van- of omgaan met vooroordelen en discriminatie door het model. Gelijke, redelijke en inclusieve behandeling van patiënten.

Verantwoordelijkheid en aansprakelijkheid

Verantwoordelijkheid en aansprakelijkheid

Integer te werk gaan, de verantwoordelijk en aansprakelijkheid rondom het systeem vastleggen.

Privacy

Privacy

Het recht om beschermd te worden: gegevensbescherming, dataveiligheid en eigendom van eigen data.

Vrijheid en autonomie

Vrijheid en autonomie

Zelfbeschikking, autonomie en kracht hebben. Kennis van de tool, en de mogelijkheid deze wel of niet te gebruiken.

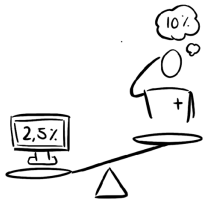
Weldadigheid

Weldadigheid

Voordelen voor patiënten maximaliseren, en in het algemeen welzijn, socio-economische kracht en de maatschappij versterken.

Vertrouwen

Vertrouwen



Het genoeg, te veel of te weinig vertrouwen van de tool, beïnvloed door educatie, transparantie en integriteit van het systeem en verantwoordelijkheid, privacy en rechtvaardigheid.

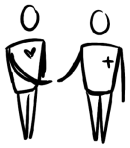
Solidariteit

Solidariteit

Effecten op de arbeidsmarkt, de sociale cohesie en het respect voor kwetsbare personen of groepen.

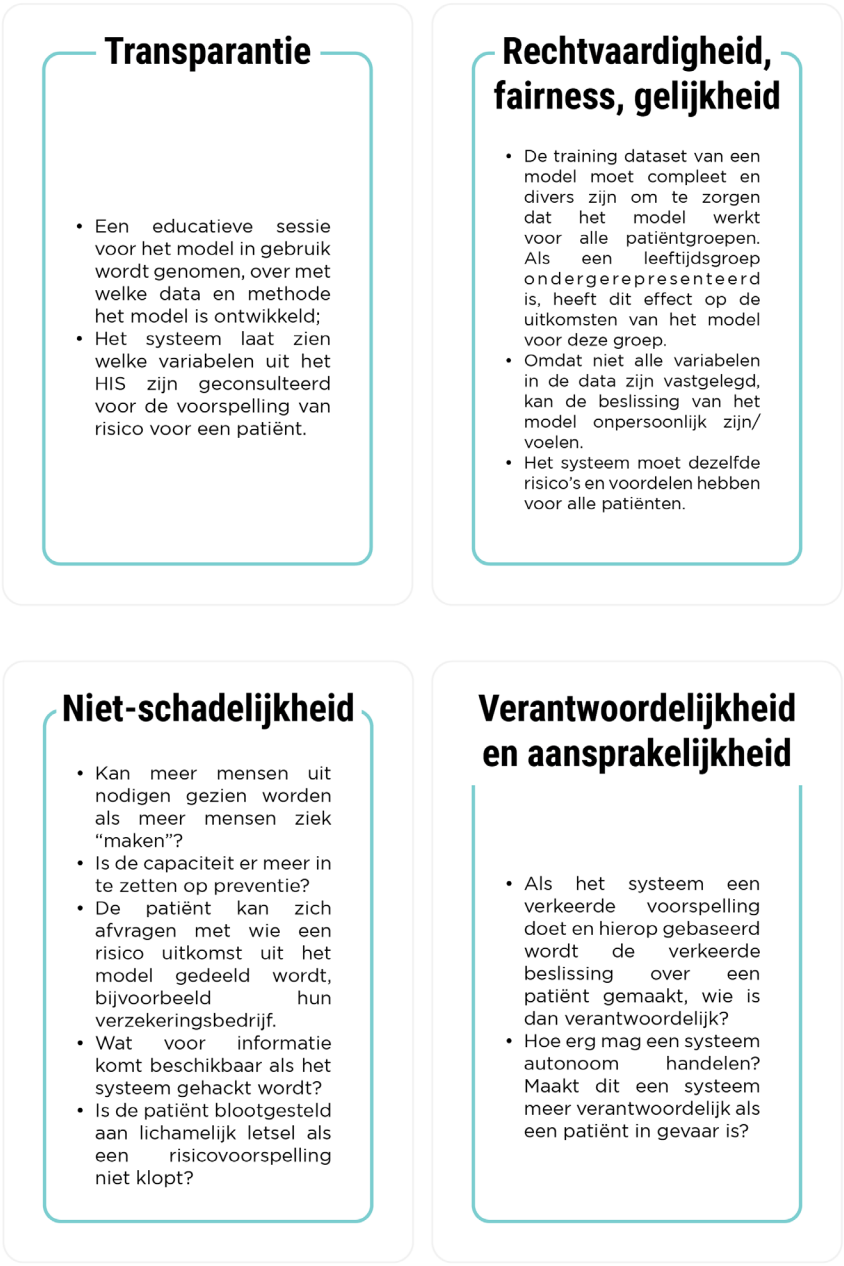
Waardigheid

Waardigheid



Het systeem neemt menselijke waarde in acht.

A7.2 Dutch value explanations for facilitator



Privacy

- Bij een voorspelling op basis van bekende data, krijg je als patiënt zijnde “opeens” een diagnose of risico te horen. Hoe verloopt deze ervaring?
- Moet een patiënt toestemming geven om in het systeem meegenomen te worden? Op welke manier?

Weldadigheid

- Wat voor voordelen/ effect heeft een model wat medische voorspellingen doet op populatieniveau? En op individueel niveau?
- Om het systeem te gebruiken om de situatie van individuele patiënten ten goede te doen, moet de arts het systeem en zijn outputs begrijpen.

Vrijheid en autonomie

- Wordt elke patient meegenomen in het model? Wat als ze dit niet willen?
- Bij elke invoeging in het EPD dat het model gaat gebruiken moet de patiënt getekend toestemming geven.
- Het ML model zal, naast de patiënt en arts, een rol vervullen in het gezamenlijk besluiten maken over het behandelen van een gezondheidsprobleem.

Vertrouwen

- Een systeem te veel vertrouwen kan leiden tot het blind volgen van een beslissing die de patiënt niet ten goede komt.
- Wat voor effecten zouden een vals-positieve of vals-negatieve uitkomst hebben op de situatie van een patiënt?

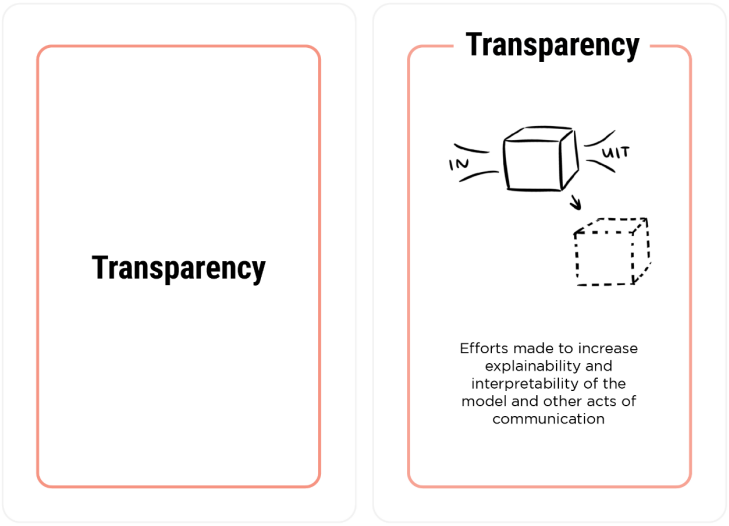
Waardigheid

- Wordt de patiënt nog steeds als mens/ individu gezien/ behandeld?
- Voor patiënten die hun prognose niet voorspeld willen hebben door het systeem is er een “reguliere” workflow.

Solidariteit

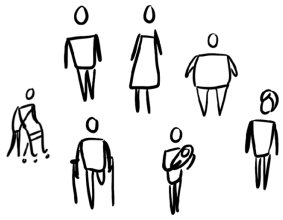
- The ML model als derde beslissingsmaker kan effect hebben op de interactie tussen medisch personeel die samen verantwoordelijk hebben over een patiënt.
- Je identificeert meer mensen - wat is de bemensbaarheid? Kan de POH wel meer mensen zien?

A7.3 English final value cards



Justice & fairness and equity

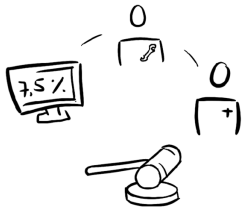
Justice & fairness and equity



Prevention or mitigation of bias and discrimination. Equal, inclusive and fair treatment of patients

Responsibility and accountability

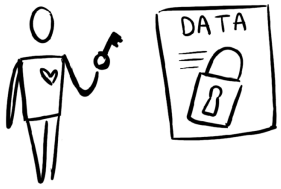
Responsibility and accountability



Working with integrity. Clarifying the division of responsibility and legal liability

Privacy

Privacy



The right of being protected. Privacy relates to data protection, data security and data ownership.

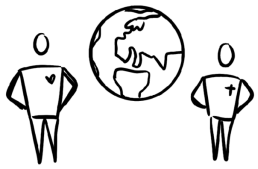
Freedom & autonomy

Freedom & autonomy

Self-determination, autonomy and empowerment. To be able to refrain from the use of the technology.

Beneficence

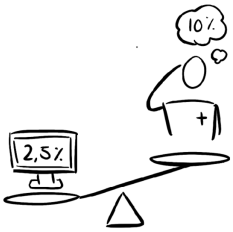
Beneficence



Using ML to maximize patient benefits, and in general, create human well-being, socio-economic strength, and benefit society.

Trust

Trust



The right amount of trust, over- or undertrust, influenced by transparency, education, system's integrity, responsibility, privacy and fairness

Solidarity

Solidarity

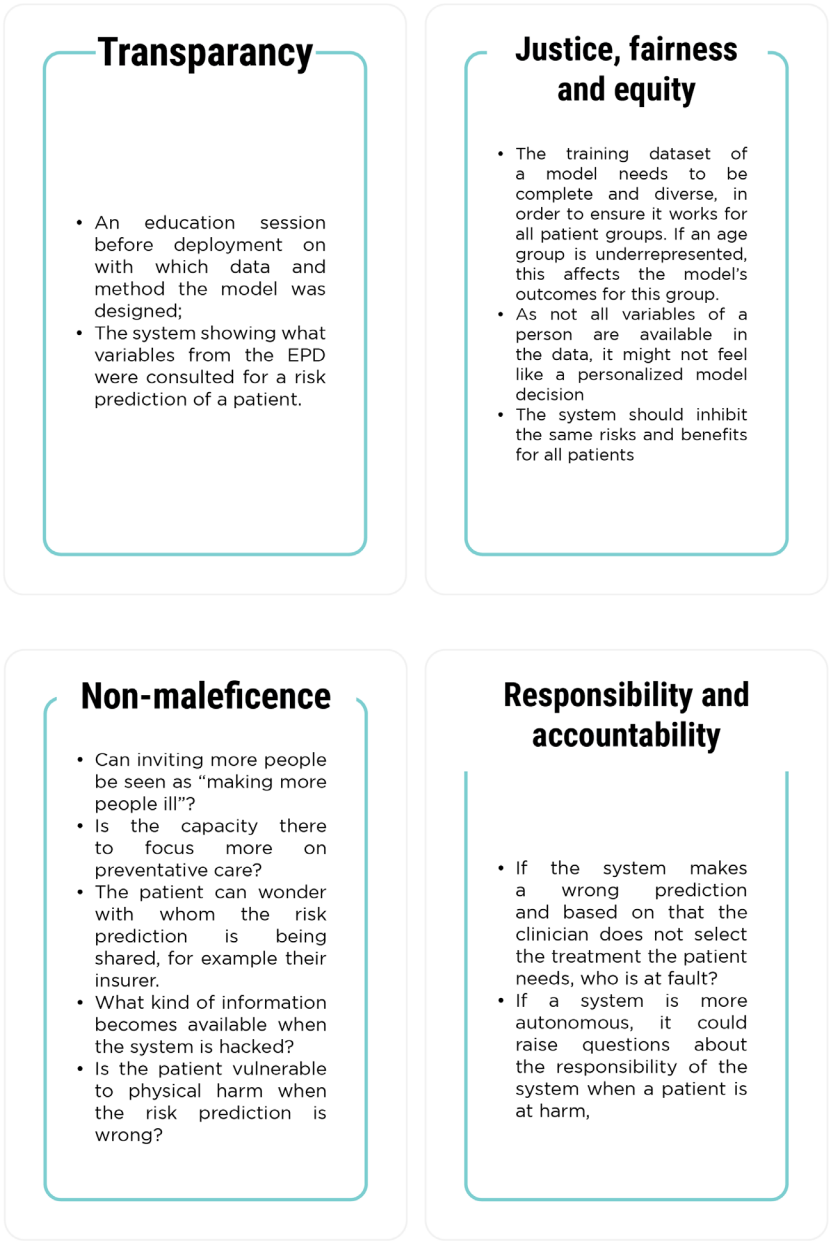
What implications the ML has for the labor market, social cohesion or the respect of vulnerable persons or groups.

Dignity

Dignity

The system takes human value and rights into account.

A7.4 English value explanation for facilitator



Privacy

- With a prediction based on known data, a patient "suddenly" hears a diagnosis or risk. How does this experience go?
- Does a patient have to give permission to be included in the system? How?

Beneficence

- What are the benefits/ effects of a model that makes medical predictions at the population level? And on an individual level?
- To use the system to benefit the situation of individual patients, the clinician must understand the system and its outputs.

Freedom & autonomy

- Is every patient processed by the model? What if they do not want to be?
- With all entries submitted into the EPD that the system will use, the patient gives signed consent.
- Next to the patient and doctor, the ML model will be another "actor" in the shared decision making on treatment of a health issue.

Trust

- Trusting a system too much could lead to blindly following a decision that might not fit the individual patient.
- What would be the effects of a false positive or false negative outcome?

Dignity

- Is the patient still being seen and treated as human/ individual?
- For patients who do not want to have their prognosis predicted by the system, a “regular” workflow is in place.

Solidarity

- The ML model as third decision maker might have an effect on the way clinicians interact when having a dual responsibility over a patient.
- You identify more people, what is the work capacity? Can the GPA see more people?

A8 EXAMPLE VALUE CARDS ITERATION 1

A9 RESULTS: PHOTOS PER PARTICIPANT

Figure A9.1 Participant 0 results phase 1

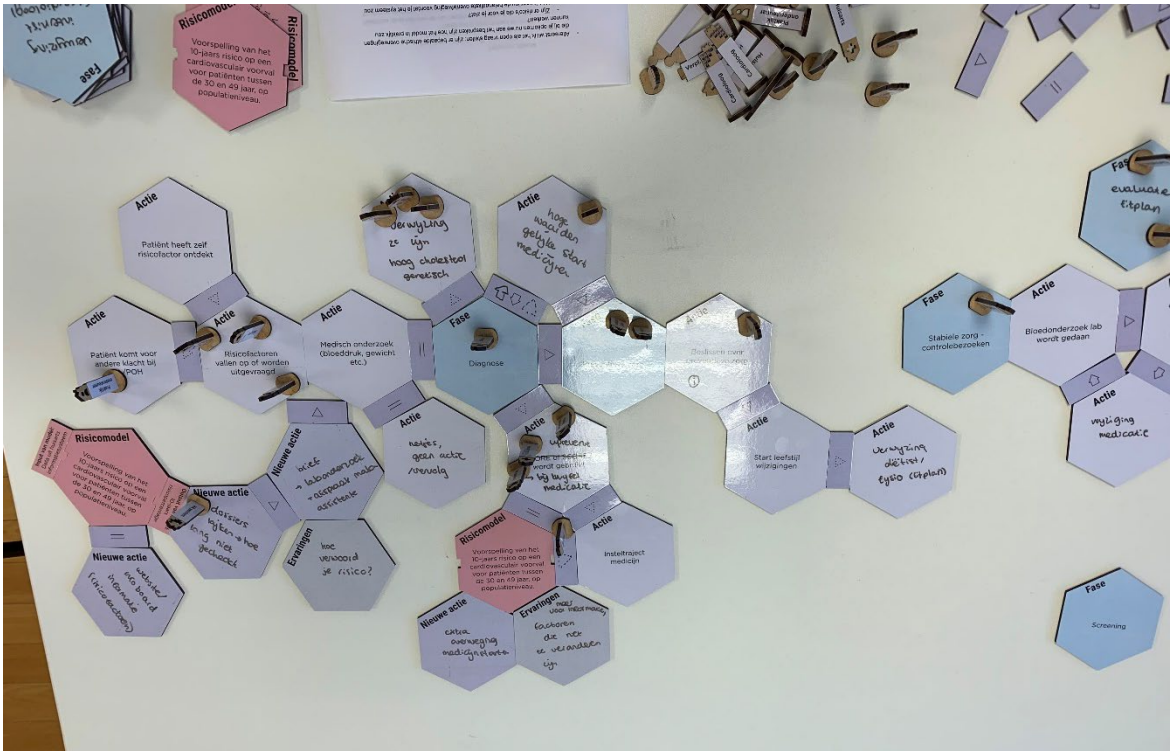


Figure A9.2 Participant 0 results phase 2

Figure A9.3 Participant 0 results phase 3

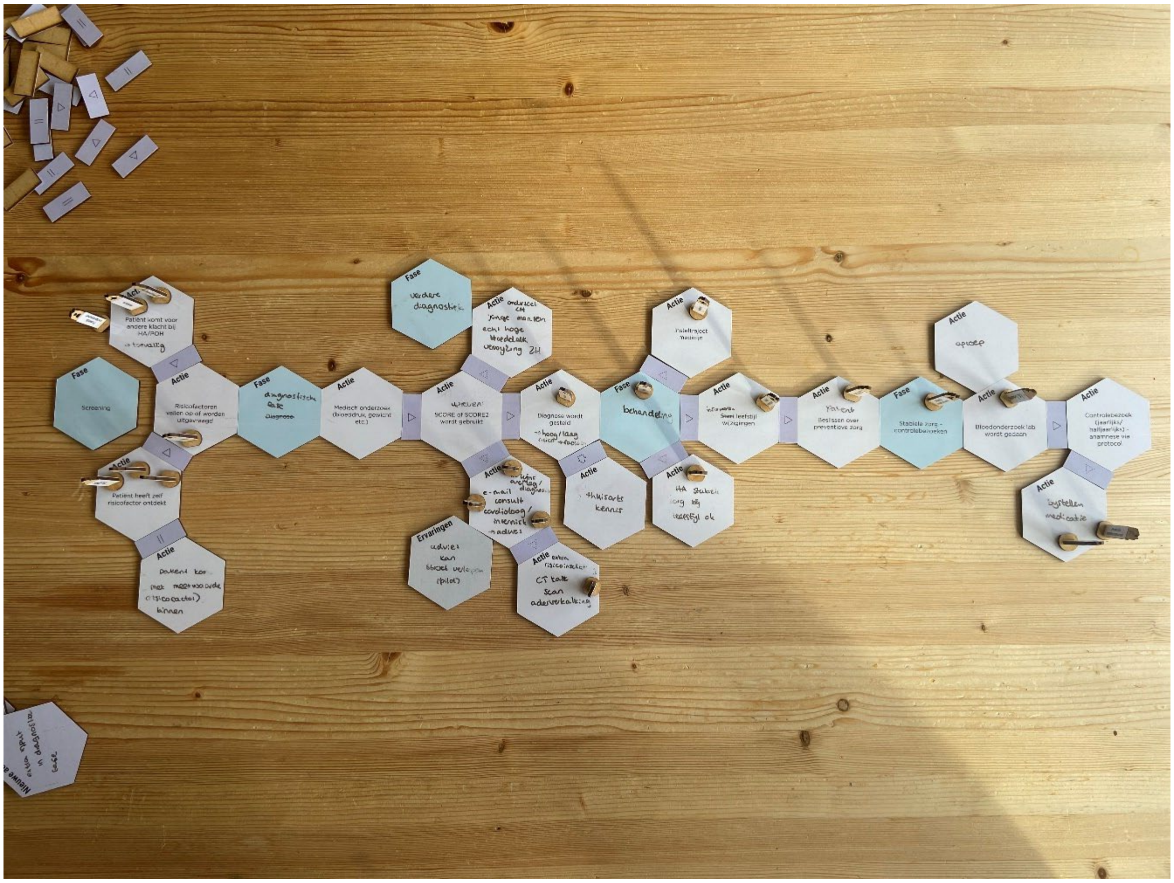


Figure A9.4 Participant 1 results phase 1

Figure A9.5 Participant 1 results phase 2

Figure A9.6 Participant 1 results phase 3

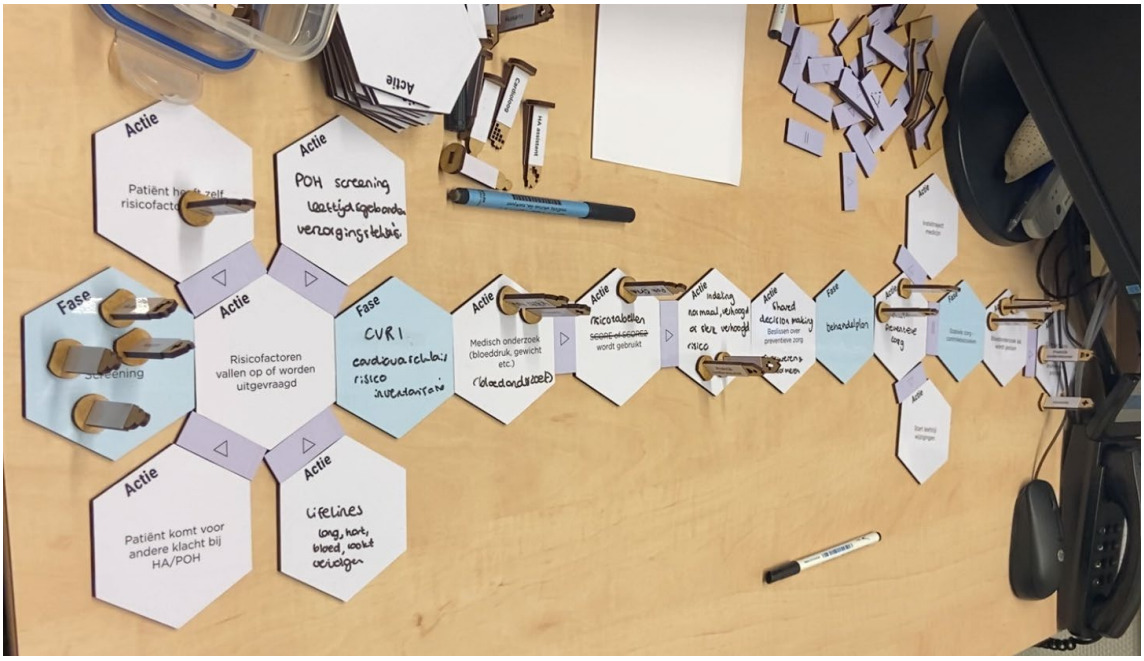


Figure A9.7 Participant 2 results phase 1

Figure A9.8 Participant 2 results phase 2

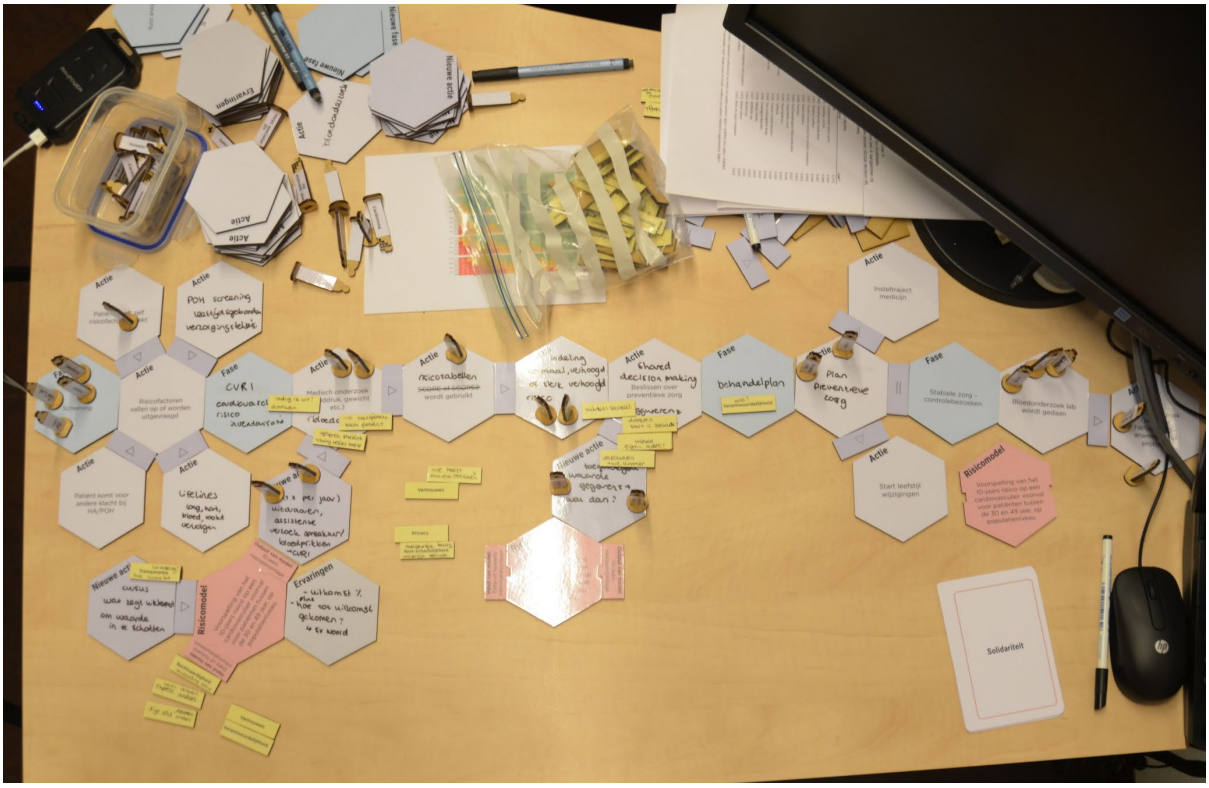


Figure A9.9 Participant 2 results phase 3

Figure A9.10 Participant 3 results phase 1

Figure A9.11 Participant 3 results phase 2

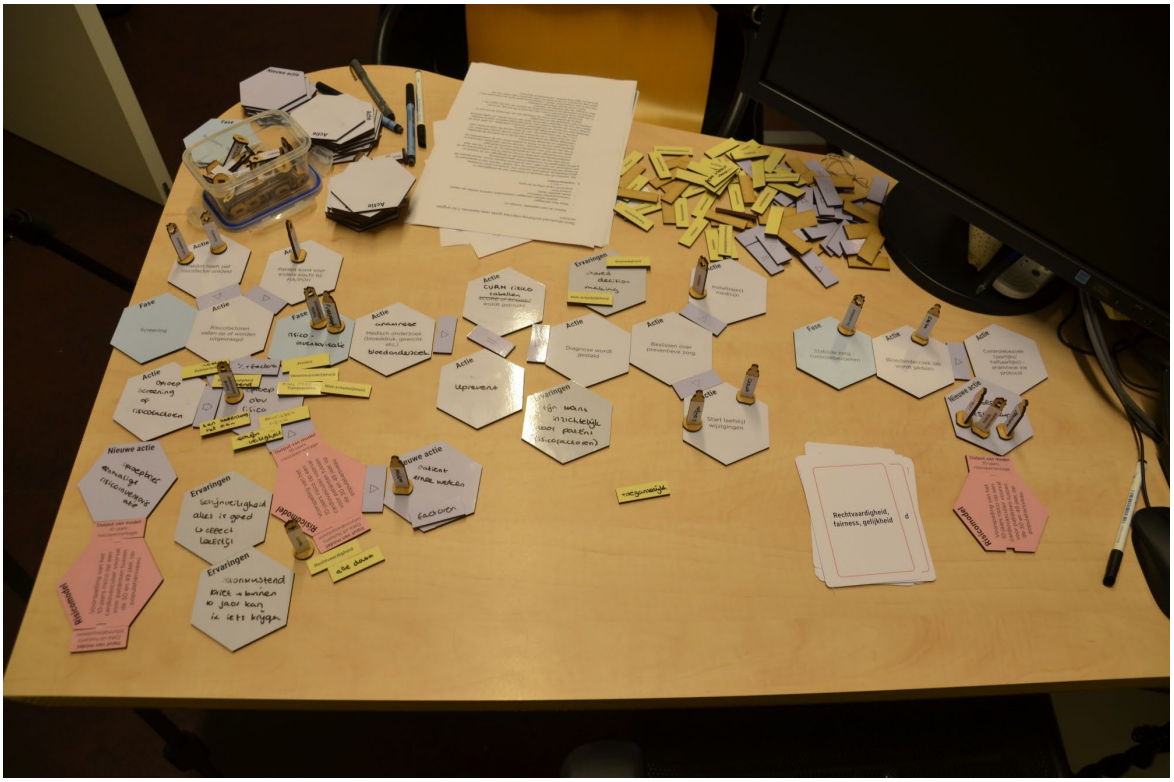


Figure A9.12 Participant 3 results phase 3

Figure A9.13 Participant 4 results phase 1

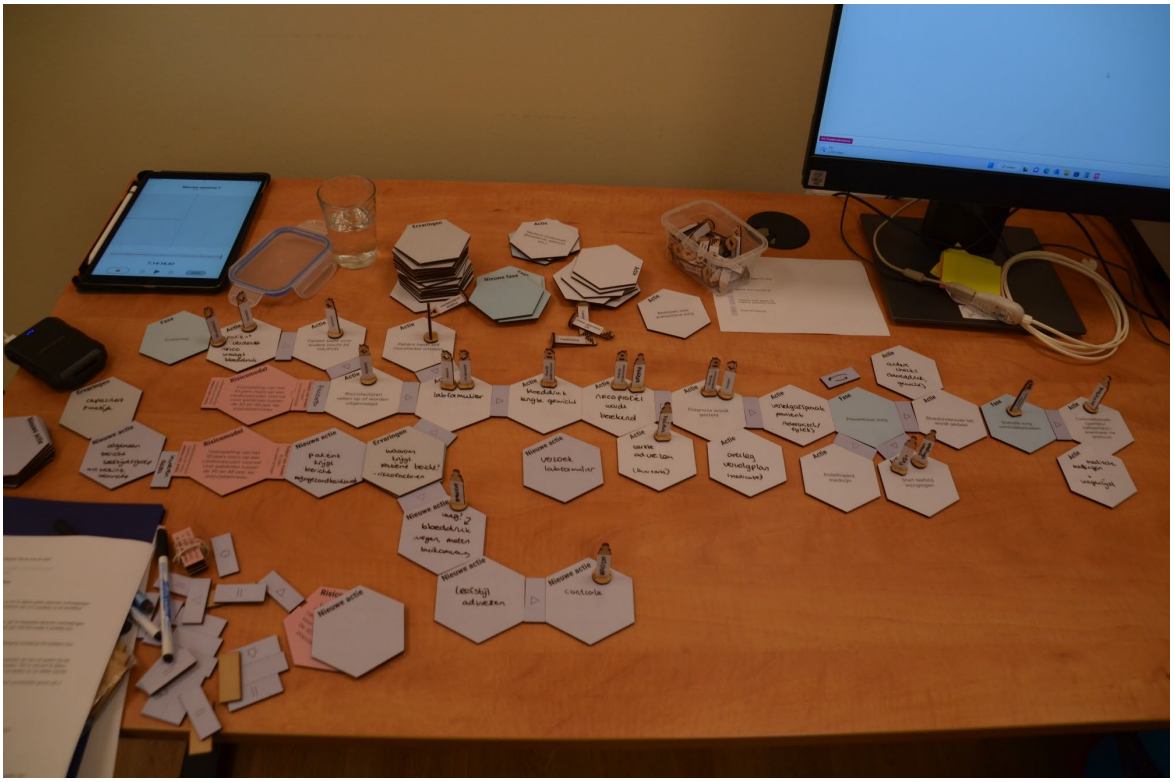


Figure A9.14 Participant 4 results phase 2

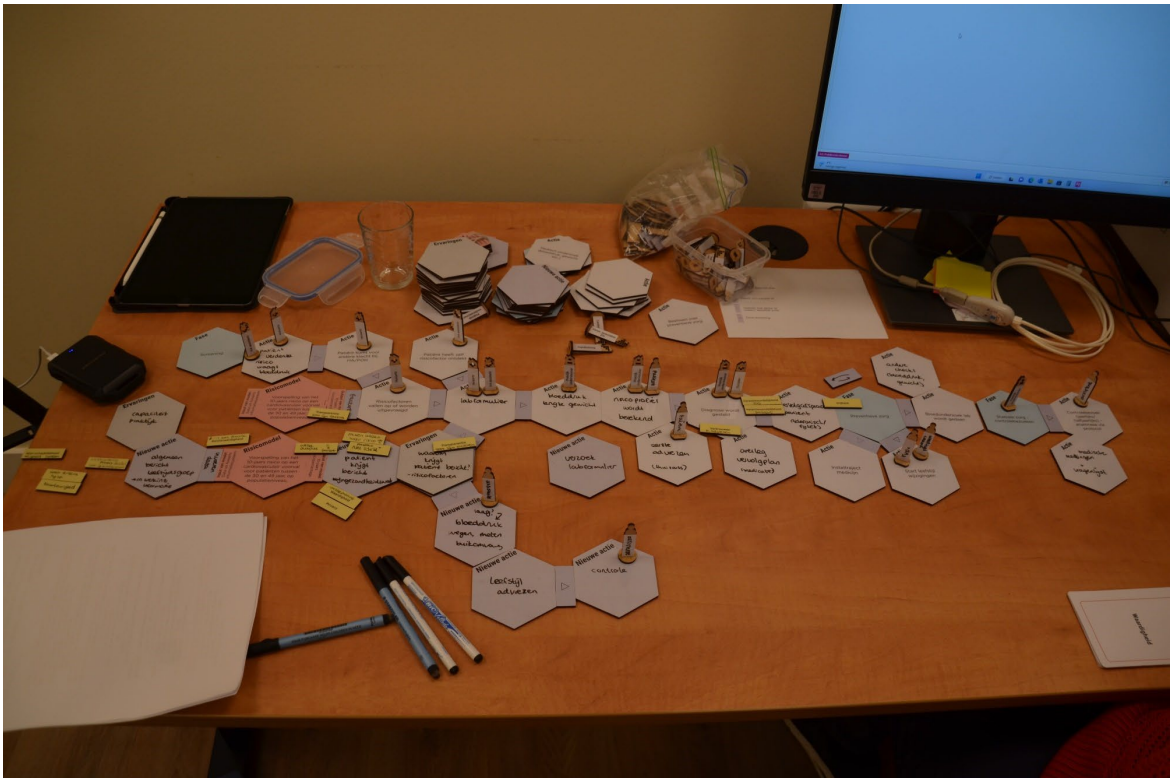


Figure A9.15 Participant 4 results phase 3

A10 THEMES AND CODES

Themes	Codes
Usage: mapping the workflow	- Questions/ uncertainties about/ checking basic journey - Change in workflow - Understanding specific actions - Discussing current tools - Listing actions without journey
Usage: stakeholders	- Stakeholder is mentioned - Stakeholder action or relation
Usage: mapping the ML model	Understanding the ML - Clinician (wrong) interpretation or (technical) assumption of ML system - Question/uncertainty by clinician on ML or extra explanation/ repetition explanation - Clinician compares ML with current systems Implementing the ML - Opinion on/ value of ML - Clinician implementation (general) - Mapping new actions - Clinician comments ability to map ML
Utility of mapping workflow and ML model (values resulting from mapping)	- Effect of mapping workflow so specifically
Utility: ethical values	- Clinician ethical concern: personal effect on clincian, eg own emotions/ way of working - Clinician ethical concern: effect on (interaction) patient - Clinician ethical concern: practical issue - Clinician ethical concern: technical, in and output model (/ interaction HA) - Other: ethical concerns in general - Value is prerequisite/ need for clinician - Comments on value implication on work/ model development - Designer interpretation of clinician explanations/ linking to values Jobin et al.
Usage: ethics/value cards	- Ethical values cards effect (e.g. guidance vs vague) - Ethical values overview (e.g. fitting terms, overlap) - Ethical values placement (designer placing, clinician placing) - Ethical values content (e.g. explanations)
Utility: expressing ethical values	- Expressing ethical values - Importance ethical values (evaluation)

Usage and utility: workshop overall	- Amount of information - Division stages (effect on each other) - Interactions with the tool - Workshop goal - Physicality - Relevant but not a theme/code

Table A10.1 Used themes (left cell) and codes (right cell) within the thematic analysis.

A11 VALUE TABLES

A11.1 Values mentioned

	P0	P1	P2	P3	P4
Transparency	2c,6c	1c	1b1,1c1, 4c, 8c	1c, 3c, 5c, 10c	3b, 6c, 8c
Trust	1c	1b, 5c	1b1, 3c		3c
Freedom and autonomy	1c	1a,1c,2b	4c		1c1, 4c
Beneficence		3c		7c	
Non-maleficence	3c	*2b,*1c	7c	3c	5c
Solidarity		2c		6c	
Justice, fairness, equity	5c		2c	4c	2c
Responsibility and accountability	1c,5c	1a, 1c, 2c	2b1, 5c	8c	4c
Privacy		1c, 2c	1c1, 6c	3c	1c, 7c
Dignity	4c	4c			7c

Table A11.1.1 Values mentioned per participant. (a,b,c = phase 1, 2, 3; b1= initial reaction model, b= during action mapping, c1= initial, c= with value cards. number = order of mentioning; *interpretation but not mentioned or researcher suggestion*).

A11.2 Values and participant interpretation

	P0	P1	P2	P3	P4
Transparency	**2c (explaining reason risk)** **2c (perhaps knowing factors)** ***6c (how much details on system?)***	***1c (GP understanding model)*** **1c (clear benefits) (honest about risk, why this patient group) (for patient what is risk based on, understand, what happens with it)**	1b1 (how does computer decide?) (if not transparent I tend to distrust it) (needed to explain it to patients) *(prediction reasoning to estimate value)*	**1c (understanding to justify for yourself and be honest to patient)** **3c (patient own decision, shared decision making) (privacy and non-maleficence (?))** **5c (how transparent in patient letter)**	*3b (knowing why person came out of model)* **6c (interpretability, why which values)** ***8c output (factors patient, general wording)***

			1c1 (understanding general and per risk) **4c (social skills per patient (less) transparent)** **8c (factors)**	**10c (consequences, responsible for people)**	
Trust	**1c (patient in practice)**	1b (data input HIS) *1b (substantiation input)* **5c (trustworthiness based on data selection, and filled in)**	2b1 (discrepancies me or computer- who is smarter?) **3c (word choice HIS for desired outcome - trust self bigger)** **3c (when does patient trust it?)**		**3c (patient trust in advice)**
Freedom and autonomy	**1c (patient refusing)**	*1a (patient refusing)* **1c (patient trust government)** *2c (does patient act on outcome?)*	**4c (GP acting on output or not - social skill) (having to explain oneself)**		1c1 (patient behavior) **4c (patient and clinician)**
Beneficence		**3c general (prerequisite for screening is that it benefits greater whole)**		**7c (general fitting word, not perse model) (group setting, lifestyle)**	
Non-maleficence	**3c (check-in with permission)**	*2b (ethical)* **1c (patient impact)**	**7c (basis of GP) (sensitive and specific - true positives)**	**3c (patient own decision, shared decision making) (privacy and transparency)**	**5c (handle data well)**
Solidarity	***(overlap dignity)***	**2c (does patient act on outcome?)**		**6c (not labeling)**	
Justice, fairness, equity	**5c (treat everyone - responsibility)**		***2c (re) (word choice HIS patient groups)***	**4c (which data collected) (labeling in sending letter) (transparency)**	**2c (model itself, developer)** **2c (ethical background inherent discrimination)**
Responsibility and accountability	**1c (duty of care), 5c (treat everyone-justice)**	1a (extent interference GP) **1c (patient, clear screening)** **2c (does patient act on outcome?)**	2b1 (who when wrong?) **5c (me or developer? --GP)**	**8c (integrity overlap transparency and privacy -patient knows choice) (clear on who is responsible care)** **(patient resp. for privacy as well)**	**4c (freedom patient behavior) (freedom clinician to give advice)**

Privacy		**1c (governmental, data) (not open to others) 2c (permission patient, understanding input data)**	1c1, **6c (basis of GP)**	**3c (patient own decision, shared decision making) (transparency and nonmaleficence)**	**1c (permission), 7c**
Dignity	**4c (shared decision making) (decent letter)**	**4c (let people make own choices)**			**7c (permission, privacy)**
	1b1 workload 2b1, 1c1 labeling (justice/dignity) 2c1 prioritizing (workload)	1b1 data (substantiation) 1b data (how SES from HIS) 1b permission (required from patient) (what happens with risk) 1c1 (is risk influenceable) (forcing risk on young group) 2b *accuracy* (good yield, health benefits and no harm) 1c1 factors 1c1 understand model & grounded	1b1 input data (depending on my notes) 1b1 input data (adapting writing style over time) 1b1 input data (interpersonal differences) 2b *accuracy* (different outcomes needed for value but difficult to trust) 3b workload (tension more screening) **1c workload (true positives) 7c sensitive and specific**	1b1 workload 2b1 how preventative to work? (disturbing/ false sense of security) 1c1 lifetime data 2c1 workload 3c1 accessibility (less administration) **9c data (how selection factors made)**	1b1 workload 2b1 permission 4b data (up to date HIS, filled in well) **9c *accuracy (*data input, performs well)**

Table A.11.2.1 Values mentioned per participant. (a,b,c = phase 1, 2, 3; b1= initial reaction model, b= during action mapping, c1= initial, c= with value cards. number = order of mentioning; *interpretation but not mentioned or researcher suggestion*).

A12 EXTENSIVE RESULTS

Results will be discussed adhering globally to the thematic structure of the toolkit's usage and utility. P1 and p2 were GP's, p0, p3 and p4 GPA's. Again, p0 was a pilot due to accuracy numbers being mentioned. Because of the small sample size and broad workshop coverage extending far beyond accuracy, their insights are included in results, always showing participant numbers. The created toolkit layouts are displayed per workshop phase and participant in appendix A9.

A12.1 Mapping the implementation context

As for usage of phase 1, in all tests, the clinician questioned or elaborated on the tiles of the toolkit that were laid out as basic journey on own initiative, the facilitator subsequently suggested summarizing changes in the physical workflow to which participants agreed or suggested otherwise. Pro-activeness differed: p0 and p2 shifted elements themselves, p1 and p3 suggested concrete changes yet were hesitant in interacting and p4 was tentative about any changes. Stakeholder relations and actions or workflow summaries often inspired each other. Furthermore, the physical journey was often pointed to, and reminded both participant and facilitator to return to previously mentioned actions, or aided participants in revising the journey's completeness (p0,p1,p4). The overarching 'Phase' tiles were confusing (all), 'Experience' tiles were mostly used for additional insights than their intended purpose and connectors were mostly for the facilitators understanding (p1,p2,p3,p4). In the evaluation, P1 was often mentioned as fun, giving new insights into existing work patterns (p1,p3,p4). P3: *"[mapping stakeholders] was also very clarifying [referring to the workflow mapping]. You are really examining the system, the care, who does what. I am thinking, we actually do a lot as GPA."*

As for the usage of phase 2, two participants immediately envisioned the model as population screening, as the developers intended. However, other participants placed it here later, as assumptions, questions and possible misunderstandings by the participant surfaced throughout initial reaction and action mapping, such as creating an action of selecting patient risk groups to be assessed by the model, leading to the facilitator explaining this is a model feature (p0), or discussing patient inclusion surfacing the misunderstanding that the patient delivers data through smartwatches (p2). Only p1 deviated in their envisioned mapping, suggesting governmental implementation as this could meet requirements of permission and accuracy. All participants added implementation during risk assessment or treatment. In implementation the participants often seemed hesitant, asking for validation or leaving the formulation of new actions to the facilitator. New actions were often formulated compared to existing manners of screening or risk prediction, for example mentioning the benefits of using uPrevent and suggesting similar use, or comparing it with how- and by who-screening was approached in the past. These comments often led to model requirements (such as uPrevent's visualization) or misunderstandings surfacing through the model being able to do more or less than past alternatives. All participants mentioned difficulties mapping the model or envisioning the future. P0 and p4 did mention satisfaction with the end overview, while p3 indicated ongoing uncertainty on her model understanding.

A12.2 Expressing ethical values

While intended to be separated between phase 2 and 3, the model placement, new actions, prerequisites and the values and their placement turned out to all influence each other. In phase 2 ethical issues already came up in the initial reaction to the model (p1,p2) and while making implementation or actions specific (appendix A9.5). E.g. while discussing the letter calling in patients, p0 reasoned: *"if you properly explain, 'research has showed that...'... That's difficult, it becomes ethical. Because people with lower SES... You can't put that in a letter."*. On the other hand, implementation was sometimes altered in phase 3. This interrelation was most strong for p1, whose model placement depended on requirements, and placement in turn influenced values.

As for cards usage, p0, p1 mostly used the term side of the cards to consider values. P2 did not use the explanations at all, p1 and p0 when terms were unclear, and p0 when reviewing discussed values. P3 and p4 used the explanations throughout, especially when asked which values are important to them concerning the model. The explanations did not always provide clarity: p1 *"This is too difficult. […] It is way too much information."*. Interaction-wise, participants mapped cards in the journey or grouped them while discussing, p4 placed them in a static row and p3 stacked discussed ones.

Both in phase 2 and 3, the mentioned values or model requirements did not always fit within Jobin et al.'s values. As a result the facilitator often used an empty tile, or proposed an interpretation from Jobin et al. The participants interpretation of this term often started a new value discussion. Interaction-wise, the designer placed value tiles and noted the participants' interpretations or meaning in a certain context on the small empty space on them (see appendix A9). Participants commented on Jobin et al.'s terms overlapping, such as p0 mentioning "solidarity" and "dignity" overlapping and p3 "integrity" and "transparency". P1 commented on 'trust' not grasping 'trustworthiness' (implying accuracy). As for placement, the facilitator often validated an interpretation from the participants' explanations or actively asked participants. Participants expressed difficulties placing due to terms being general (p0,p2,p3,p4), e.g. p4 *"Yeah… When do people start to, for example, trust you?"* and p2 in the end evaluation: *"These are things for the whole process actually, so I would want to place them everywhere."*.

Overall, participants mentioned the value cards providing guidance: *"If you have to think of them yourself I probably would not have mentioned them, but if you see them you think 'Ah! That one and that one."* , (p2), or specificity: *"It is funny, you make the aspects more concrete. […] What element does it influence?"* (p1). P2 also mentioned even more values could be added, such as "accuracy" and "humanity", or to let participants think of values themselves: *"This also limits people right? If you only provide a few."*(p2). P0 mentioned being glad not having to think of the values herself, but was also happy she did not have to use all due to not knowing where to place them. The latter reflected the opinion of p3 and p4 as well, struggling with value placement, as terms are a natural part of care or do not consciously feature.

A12.2.1 Expressed values

Appendix A11 shows an overview of explicitly mentioned values by facilitator or participant, both in phase 2 and 3. Initially mentioned values or prerequisites in phase 2 included its medical substantiation (p1), data use (p1,p2) and workload (p0,p2,p3,p4). P2 stood out in already mentioning trust, transparency and responsibility: values overlapping with Jobin et al. While mapping new actions, mentioned values again included data (p4) and trust caused by it (p1), do people need/want to know their risk (p1,p3) and patient letters mentioning HIS or SES (p0,p2,p3,p4). P1 (and p4) mentioned permission of patient, p1 and p2 how well the system works (interpreted as accuracy), influencing placement. P2 again discussed more in-depth values, such as the tension of the added value yet distrust in the model predicting differently than the clinician. It also came up what information is required of the system (p1,p2,p3,p4), with p1 and p2 also mentioning general understanding of the model.

As for phase 3, as appendix A.11 shows, when using the value cards not previously mentioned values surfaced, while still mentioned as important or prerequisites. The discussed values became concrete through discussing them and often naturally translated into model requirements or functions, e.g. p2: researcher: *"what would you want to know of the system?"* participant: *"I would really appreciate knowing based on what they reached the prediction, so you can estimate its value. […] 'You have used the term pulmonary embolism five times, so we think the risk it raised with this percentage' I think that is essential for me to know what it means."*. In unique cases, this was harder, when values were more general or for example related to the developer, e.g. p4 when asked to explain how justice reflected itself: *"preventing or dealing with presumptions or discrimination'... Yes okay that is more.. from you, about the model itself."*.

As to the values, transparency was mentioned in all tests. Participants interpreted it as understanding the model in general or the outcomes of the system (all) to trust it (p1,p2) or explain why it is used to patients (p2,p3), and what details patients and GPs would need of the system (all), perhaps leading to intended less transparent to patients (p2,p3,p4). Trust was only not mentioned by p3, and usually included patient trust in the practice (p0,p2,p4) or the clinicians trust in the system caused by the HIS data input (p1,p2) or the smartness of the system (p2). It stands out data input often returned in the tests but mostly features in the extra cells in the value table, so not often mentioned explicitly with a value term. Considerations included how data is medically substantiated (p1), issues concerning data translation from the HIS (p1,p2,p4), and the selection made (p1,p3). There was a general pattern in clinician autonomy to choose what advice to give (p2,p4), simultaneous with their responsibility to deliver care (p0,p1,p2,p4) yet on the other hand clinician responsibility to follow up on this, and freedom to not do so (all). Responsibility, freedom and trust were mentioned together as general healthcare values by p0,p1,p3. The tension conformed with different values in different tests (responsibility, justice, privacy, freedom, dignity), showing the overlap in value interpretations. In the same line, the importance of shared decision making was mentioned under dignity, privacy, non-maleficence and transparency (p0,p3) and patient permission was mentioned amongst privacy and dignity. The reported vagueness of 'beneficence' (p0,p2,p3,p4) shows in its underrepresentation in the table. Solidarity was often mentioned only to repeat other values, and dignity is also mentioned less. Conclusions on the responsibility on the model (p2) and the case of a wrong outcome (p1), was that the GP was responsible by taking the model in practice. An information session or course on the model was mentioned by p0 (at the end of phase 3) and p2 (during phase 2). It stood out that some values, often mentioned as prerequisites, returned multiple times throughout the phases.

Overall, it stands out that, apart from p3, most values were mentioned in all tests, although sometimes differing in interpretation. Some values, such as transparency, privacy and trust were mentioned earlier in the conversation than others.

While overlapping, both values and specifically mentioned prerequisites also differed between participants. For example , only p1 and p2 were concerned with the models 'value' or accuracy, p0 and p1 and patient permission was a requirement for p1 to place the model and p4 in selecting actions, while this did not come up or was dismissed when asked about in other tests.

A12.3 Evaluating workshop utility

In the end evaluation, participants commented on the ethics being important to fit with the clinical practice (p0, p1,p4) : p1: *"What are the ethical aspects and where does that fit in daily practice, why does a risk model need to be used, and by who? […] That seems very valuable to me."* medical knowledge being needed for development (p4) and the workshops purpose in determining whether there is a pattern in ethical values (p0,p2). Within the workshop, GP's were most positive on the ML implementation (p1) and ethics discussion (p1,p2) were the nicest workshop aspect, yet p2 did indicate new insights on the complexity of ethics and more hesitation towards ML. Other participants indicated the ethical values were the most difficult part of the workshop (p0,p3,p4). For p0 this was due to general difficulty with the topic, for p3 because they needed to remind themselves of the goal of the workshop due to the unclarity of the objective being general ethics or ML ethics and the conceptual phase of the model development. P4 mentioned difficulties because of the general nature of the values. P1 questioned the workshops main objective since a large part was not related to the ethics. Some participants wondered how workshop mapping (p1) and values (p4) could be translated into the model. The workshop phases were mentioned to surface new insights throughout (p0,p2) and to be needed to build onto each other: p2: *"To clarify, what are you talking about? I need this to be able to think of it, where should this come [ML]? You build onto it"*), but followed each other up logically (p3). Lastly, physicality was seen as conversation starter (p1,p3) and helped in thinking (p2) P2:

"It is nice to actually shift the elements, and to then reassess whether it is right". However, p2 did mention interactions by the facilitator made her think less actively about tile placement. The workshop was experienced as intensive but no information was redundant (p0,p2,p3).

A13 DISCUSSION AND FUTURE WORK

This section discusses the interpretation of results, adhering to the research question: *How can we enable clinicians to express their ethical values regarding a Machine Learning Clinical Decision Support System, using a toolkit that promotes mutual understanding of the implementation context?* Insights and suggestions for future work cover the probe itself and abstractions to the field of clinical HCAI. The section concludes with limitations of the study.

A13.1 Creating a mutual understanding of the ML implementation context

The research results show how the interactive mapping of the current context (phase 1) and envisioned implementation context (phase 2) created shared comprehension between designer and clinician respectively. For the facilitator, the interactive mapping of the current workflow and stakeholders created context of the clinicians' needs and expectations of the model. Inversely, the concrete envisioned implementation mapping, interchanged with questions or clarifications, showed the facilitator the participant's vision of the model and iteratively increased clinicians knowledge. Abstracting this insight to clinician involvement in ML, establishing a common ground between developer and clinician is recommended due to its potential in creating both a shared mental foundation of knowledge, as well as a physical foundation to discuss specifics of the ML.

A13.2 Starting a conversation on ethical values

While the discussion on ethical values was the main goal of phase 3 of the workshop, unexpectedly these organically already surfaced in phase 2. This partially resulted in repeating values, yet the separation in phases could also be argued to make the workshop manageable. Still, on design level, future manners of integrating the discussion on ML implementation with ethical values more naturally could be explored. Furthermore, the phase value cards proved effective as well: its utility being mentioned in evaluation comments, and the set of value cards both sparking conversation on values previously mentioned – showing relevance - as well as providing new insights, and often resulted in model prerequisites. However, although differing in interpretation, all users selected almost all values, p3 indicated seeing the importance of all and some participants seemed to feel obliged to utilize all in the way they sorted the cards (p3, p4). Therefore, it could be argued that providing a set of values can bias toolkit users, also a known critique on VSD [2]. Concluding, both making implementation concrete (phase 2), as well as providing an ethical framework (phase 3) elicits value discussion, yet a recommendation for future work is to further investigate the balance of unguided elicitation and (different selections of) outlined values in consulting clinical end users.

Content-wise, mentioned values overlapped, yet unexpectedly some major prerequisites differed between participants. Because of this, a suggestion for future end-user value elicitation includes exploring the VSD "value dams and flows" method to take along restrictions right away and adopt the model throughout evaluation with different [9]. Another suggestion would be to explore the workshop in group sessions, to encourage stakeholder discussion.

Furthermore, unexpectedly, the value card term side was used predominantly, rather than consulting the explanation and especially the examples. One reason could be an already clear association with the term, or too little clarity resulting from the explanation. Besides, perhaps the workshop simply did not provide enough space – cognitive or time-wise - to explore the values in-depth. This links to another insight: although phase 1 and phase 2 were needed to establish a common

ground, they did diverge the workshop focus from the values. A future suggestion is to conduct the workshop in multiple sessions, separating the envisioned ML implementation from a session treating the ethical values in-depth. Besides, an iteration reviewing the card content could make their exploration threshold lower.

A13.2.1 Expressed values

Overall, the results of mentioned values show a general pattern: initial reactions seem to focus on ensuring the medical security of the model, and action mapping values still remain a bit practical in model value, what knowledge is desired of the model and how this will be communicated to- and received by the patient. In phase 3, using the value cards, new prerequisites still surface and discussed more facets of the model. Especially the often mentioned transparency, trust and privacy stand out, a contrast with the expected result that the clinicians would gravitate towards more general human values rather than the technical implications. However, mechanisms in patient-clinician interaction and freedom versus responsibility, featuring in current care and therefore also important in the model, did often feature.

P0 was a pilot due to its different informed accuracy. Accuracy was only mentioned by p1 and p2 on own initiative, and by p4 after it came up in conversation. However, an end evaluation comment of p0 stating that the basis of the model seems trustworthy due to the mentioned accuracy numbers shows that indeed, some outcomes of the pilot are perhaps altered. This insight does hint towards clinicians tying an opinion to technical numbers. Therefore, a suggestion for future research is to investigate the effect of more model details on the discussion with the clinician (using the workshop).

A13.3 Physicality and modularity

As mentioned before, the physicality of the tool made all discussed matter – workflows, the use of the ML and values – very concrete in an time-efficient manner, proving a canvas for later workshop phases. Through this concreteness, it became one of the unexpected manners of eliciting ethical values. Furthermore, physicality and modularity were mentioned as conversation catalyst. To further benefit from this potential, the toolkit itself should be iterated on with concrete changes such as making the phase tiles more clear or include writing space on the value tiles. Furthermore, it was found the workshop offered rich insights that the toolkit did not always account for, leading to a suggestion of adding more ambiguous or empty tiles.

It did stand out that the designer interacted more with the tool than the participant. While time efficient, and the mutual summarizing and adapting of content feeling like a joint mapping, clinician non-interaction was also commented on to cause less active thought on placement (p2). To align more with participatory design, if the workshop would consist of more sessions, this would also create space for more interaction by the participant.

A13.4 Multi-stakeholder perspectives: involving the clinical end user

While only being conducted amongst a limited sample size, it appeared GPs were more comfortable in discussing ethics than GPAs, both in their ease of discussing values in their initial reaction to the ML model, as their expressed interest in the topic in the end evaluation. One of the resulting suggestions for future work is to explore even better adaption of knowledge to the audience, for example first exploring ethics on a more general level, or again, performing the workshop in multiple sittings or group setting to reduce cognitive load and explore whether this increases of decreases participant confidence.

While both GPs and GPAs saw the importance of clinician involvement in ML development, multiple users indicated not being interested in such sessions. This could perhaps depend on the workshop's duration, or the users' unease with discussing ML and ethics due to their novelty to the topic and the resulting mental strain. This leads to the insight that

because of the abstract nature of envisioning a future with ML, concreteness is needed to create understanding and a feeling of relevance while also naturally adapting knowledge to different users.

Beyond the level of clinical end-user HCAI, the created tool is designed for, but not limited to the healthcare setting. Its different phases add a modularity to the workshop as well. Therefore, future developers in different contexts can also explore the use of the toolkit with the end users or stakeholders of their model, both for enabling the conversation on ethical values, as well as adopting solely the first phases of implementation mapping.

A13.5 Limitations of the study

A limitation of the study is that the facilitator did not have prior experience with ML development and was not part of the LUMC model development team. Mitigation of this affecting protocol explanations was attempted through coach revision, developer consultation and pilot testing. However, researcher background comprised (technical) details on the LUMC model and sometimes answers to clinician questions, possibly influencing participants' understanding of model capabilities and limitations. An example is the faulty and eventually eliminated pilot accuracy. However, researcher novelty did add a first-person perspective to the research, fitting the participants introduction to ML. Besides, the conceptual LUMC model description was perhaps more realistic in the research focus on the initial development stage.

Related to the researchers novelty, yet caused by a time-wise scoping restriction, the workshop's translation into practice was not incorporated in this research. Workshop value did come forward in expert interviews, indicating positivity towards more concrete workflow mapping and the novelty of involving end-users comprehensively in ethical values. However, future work could examine both the developers' ability to facilitate the workshop, as well as its utility in the translation of results to the developer team and to the technical implications for the ML. Examining utility for developers would also provide future guidelines on how explicitly values should conform with value sets, and whether specifying exact placement of values, which participants experienced difficulties with, is needed.

The use of a set of general AI values was chosen for its comprehensiveness and simultaneous ambiguity, offering overview yet inspiration. After the research, a literature review on ethical values specific to the healthcare context has been published [32]. This overview greatly overlaps with Jobin er al.'s [2019], though discrepancies show. Besides, this overview does contain sustainability, showing that eliminating the term from this research was a misjudgment. Building onto the previous suggestion of exploring the costs and benefits of a (different) set of values, future work could explore using ethical healthcare AI values.

Lastly, a general research limitation is the low sample size, potentially making individual differences between participants more apparent.