

Understanding Clinicians' Informational Needs for AI-driven Clinical Decision Support Systems: Qualitative Interview Study

Simone Mingels, Hannah Piehl, Madeline Therrien, Ekaterina Akhmad, Anniek van Hienen, Johan van Soest, Laura Hochstenbach, Andre Dekker, Olga Damman, Rianne Fijten

Submitted to: JMIR Medical Education
on: October 03, 2025

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript.....	5
Supplementary Files.....	30
Multimedia Appendixes	31
Multimedia Appendix 1.....	31
Multimedia Appendix 2.....	31
Multimedia Appendix 3.....	31
Multimedia Appendix 4.....	31
Multimedia Appendix 5.....	31
CONSORT (or other) checklists.....	32
CONSORT (or other) checklist 0.....	32

Understanding Clinicians' Informational Needs for AI-driven Clinical Decision Support Systems: Qualitative Interview Study

Simone Mingels¹; Hannah Piehl¹; Madeline Therrien¹; Ekaterina Akhmad¹; Anniëk van Hienen¹; Johan van Soest^{1, 2}; Laura Hochstenbach³; Andre Dekker¹; Olga Damman⁴; Rianne Fijten¹

¹ Department of Radiation Oncology (Maastr), GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, The Netherlands Maastricht University Maastricht NL

² Brightlands Institute for Smart Society (BISS), Faculty of Science and Engineering, Maastricht University, Heerlen, The Netherlands Heerlen NL

³ Department of Health Services Research, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands Maastricht NL

⁴ Department of Public & Occupational Health and Amsterdam Public Health research institute, section Quality of Care, Amsterdam UMC location VUmc, Amsterdam, The Netherlands Amsterdam NL

Corresponding Author:

Simone Mingels

Department of Radiation Oncology (Maastr), GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, The Netherlands Maastricht University
Paul Henri Spaaklaan 1
Maastricht
NL

Abstract

Background: Advancements in Artificial Intelligence (AI) are transforming healthcare, particularly through AI-driven Clinical Decision Support Systems (AI-CDSS) that aid in predicting disease progression and personalizing treatment. Despite their potential, adoption remains limited due to clinician concerns about algorithm misuse, misinterpretation, and lack of transparency.

Objective: This qualitative study explores the informational needs and preferences of clinicians to better understand and appropriately use AI-CDSS in decision-making. In parallel this study explores AI experts' perspectives on what information should be communicated to enable safe and appropriate use of AI-CDSS.

Methods: A qualitative description design study was conducted using semi-structured interviews with 16 participants (8 clinicians and 8 AI experts). Discussions focused on experiences with AI, informational needs, and feedback on existing reporting standards including Model Cards (Mitchell et al., 2019), Model Facts (Sendak et al., 2020), and the TRIPOD-AI checklist (Collins et al., 2015, 2024). The transcripts were analyzed through codebook thematic analysis.

Results: Four key themes were identified: (i) Clinicians need clear information on training data, its origin, size, and inclusion/exclusion criteria, to judge model applicability; (ii) Performance metrics must go beyond AUC and be clinically relevant to support informed decisions; (iii) Limitations and warnings about inappropriate use should be specific and clearly communicated to prevent misuse; (iv) Information should be presented in layered, customizable formats within existing clinical software, avoiding unnecessary jargon and allowing optional deeper explanations. While each of the reviewed reporting standards offered strengths, none were considered sufficient alone. Participants recommended a combined and clinician-centered approach to information delivery. Alignment of reporting standards with clinical workflows and decision thresholds was thought to be crucial to bridge the current usability gap.

Conclusions: To improve AI-CDSS adoption in clinical practice, reporting standards must be designed for better clinician comprehension and usability. Enhancing transparency, particularly regarding training data and performance, can likely help clinicians assess AI-CDSS more effectively. Information should be delivered in an accessible, layered format, fitting clinical workflows. Co-creation with clinicians throughout AI-CDSS development was a cross-cutting theme, highlighting its importance in ensuring tools are not only technically sound but also practically usable. Future research should explore how to structurally report on performance and validation metrics for clinician understanding and assess the impact of information provision on AI-CDSS adoption.

(JMIR Preprints 03/10/2025:85228)

DOI: <https://doi.org/10.2196/preprints.85228>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <https://www.jmir.org/>

No. Please do not make my accepted manuscript PDF available to anyone. I understand that if I later pay to participate in <https://www.jmir.org/>

Original Manuscript

Understanding clinicians' informational needs for AI-driven clinical decision support systems: Qualitative interview study

Simone M.B. Mingels ^a, Hannah S. Piehl ^a, Madeline S. Therrien ^a, Ekaterina Akhmad ^a, Anniek R. van Hienen ^a, Johan P.A. van Soest ^{a,b}, Laura Hochstenbach ^c, Andre L.A.J. Dekker ^a, Olga C. Damman ^d, Rianne R.R. Fijten ^a

a) Department of Radiation Oncology (Maastr), GROW Research Institute for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, The Netherlands

b) Brightlands Institute for Smart Society (BISS), Faculty of Science and Engineering, Maastricht University, Heerlen, The Netherlands

c) Department of Health Services Research, Care and Public Health Research Institute (CAPHRI), Maastricht University, Maastricht, The Netherlands

d) Department of Public & Occupational Health and Amsterdam Public Health research institute, section Quality of Care, Amsterdam UMC location VUmc, Amsterdam, The Netherlands

Corresponding Author:

Simone M.B. Mingels

Clinical Data Science

GROW Research Institute for Oncology and Reproduction

Maastricht University

Paul Henri Spaaklaan 1

6229 EN Maastricht

The Netherlands

Email: simone.mingels@maastrichtuniversity.nl

Telephone: +31615486010

Abstract

Background: Advancements in Artificial Intelligence (AI) are transforming healthcare, particularly through AI-driven Clinical Decision Support Systems (AI-CDSS) that aid in predicting disease progression and personalizing treatment. Despite their potential, adoption remains limited due to clinician concerns about algorithm misuse, misinterpretation, and lack of transparency.

Objective: This qualitative study explores the informational needs and preferences of clinicians to better understand and appropriately use AI-CDSS in decision-making. In parallel this study explores AI experts' perspectives on what information should be communicated to enable safe and appropriate use of AI-CDSS.

Methods: A qualitative description design study was conducted using semi-structured interviews with 16 participants (8 clinicians and 8 AI experts). Discussions focused on experiences with AI, informational needs, and feedback on existing reporting standards including Model Cards (Mitchell et al., 2019), Model Facts (Sendak et al., 2020), and the TRIPOD-AI checklist (Collins et al., 2015, 2024). The transcripts were analyzed through codebook thematic analysis.

Results: Four key themes were identified: (i) Clinicians need clear information on training data, its origin, size, and inclusion/exclusion criteria, to judge model applicability; (ii) Performance metrics must go beyond AUC and be clinically relevant to support informed decisions; (iii) Limitations and warnings about inappropriate use should be specific and clearly communicated to prevent misuse; (iv) Information should be presented in layered, customizable formats within existing clinical software, avoiding unnecessary jargon and allowing optional deeper explanations. While each of the reviewed reporting standards offered strengths, none were considered sufficient alone. Participants recommended a combined and clinician-centered approach to information delivery. Alignment of reporting standards with clinical workflows and decision thresholds was thought to be crucial to bridge the current usability gap.

Conclusion: To improve AI-CDSS adoption in clinical practice, reporting standards must be designed for better clinician comprehension and usability. Enhancing transparency, particularly regarding training data and performance, can likely help clinicians assess AI-CDSS more effectively. Information should be delivered in an accessible, layered format, fitting clinical workflows. Co-creation with clinicians throughout AI-CDSS development was a cross-cutting theme, highlighting its importance in ensuring tools are not only technically sound but also practically usable. Future

research should explore how to structurally report on performance and validation metrics for clinician understanding and assess the impact of information provision on AI-CDSS adoption.

Keywords: Artificial Intelligence; Delivery of Health Care; Informational needs; AI-CDSS; Reporting Standard; Transparency; AI Implementation; Clinicians and AI; Co-creation



1. Introduction

Advancements in Artificial Intelligence (AI) are rapidly transforming the healthcare sector [1]. AI has the potential to aid clinicians and patients in comparing various treatment options by predicting future events based on individual patient characteristics to determine which treatment would benefit the patient the most [2]. Implementing these AI-driven Clinical Decision Support Systems (AI-CDSS) offers the potential for highly personalized medicine by predicting disease survival probabilities and potential treatment side effects [3, 4]. Despite its potential, achieving sustained adoption of AI-CDSS in routine clinical care frequently encounters challenges [5].

One of these challenges is clinicians' resistance to adopting AI-CDSS, which is exacerbated by their concerns about the high potential risks to patient safety and quality of care [6-7]. Amongst clinicians, there can be a lack of in-depth knowledge of how an algorithm is constructed and how it comes to its predictions and recommendations [8]. This can make clinicians unprepared to assess if the algorithm is usable within their specific clinical setting, which can cause (unintended) misuse of algorithms [9-11]. An example of such misuse can be found in the study by Zhao et al. [12], where the INFLUENCE 2.0 tool was applied to guide decisions about primary therapy. However, this tool was originally designed to estimate individual, time dependent risks of recurrence or metastasis in breast cancer patients who have already completed curative treatment [13]. Using the model outside its intended context led to inadequate treatment recommendations [14].

The lack of knowledge that can exist amongst clinicians is not surprising, since there is an overall inadequate adherence to reporting standards within the AI development field [15-17]. Even when developers adhere to these reporting standards, the reports are often too technical for clinical end-users to adequately understand the potential risks or the intended use of the models [18-19]. Descriptions of AI-CDSS, such as intended use and target population, can help clinicians better understand the meaning of the outcome produced by a model before acting on its recommendation, which can prevent mistakes made due to misinterpretation of the outcome [20].

To stimulate reporting of prediction models in published literature, several initiatives have been introduced [21]. The TRIPOD-AI statement (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis – Artificial Intelligence update) is a set of recommendations for the reporting of studies developing, validating or updating a prediction model, and is commonly used and/or required by journal publications [22-24]. In addition, other guidelines such as the CONSORT-AI [25], DECIDE-AI [26] and SPIRIT-AI [27] were created. However, the focus of these reporting standards is on transparent reporting within clinical trial reports or protocols, not for improving understanding of AI for clinicians [28-29]. Furthermore, these reporting standards

are developed to report on an AI model, and not the environment or software system in which it is operating [30].

In an attempt to further standardize reporting of AI model information, Google established the concept of “Model Cards”, which is comparable to providing a recipe and nutritional facts to a meal [31]. Through Model Cards, developers provide a one- or two-page record highlighting characteristics of training data, intended use cases, and performance [32]. This Model Card can boost conformation to reporting standards but does not contain actionable information and guidance for applying AI in clinical practice. Sendak et al. [33] reconstructed the idea of Model Cards using the concept of drug fact boxes. Drug fact boxes are used to understandably communicate benefits and risks associated with medications to patients [34]. These concepts combined resulted in “Model Facts” [33], an overview to “collate relevant, actionable information in a compact overview to ensure that clinicians know how, when, how not, and when not to use model output in their clinical decisions” [33].

Although the Model Facts concept is not thoroughly tested yet, the initiative has sparked a discussion on how to effectively communicate the potential risks of AI models to healthcare professionals [33]. This highlights a broader need to understand what information clinicians require in order to evaluate and appropriately use AI-driven recommendations in their clinical decision-making. Reflecting these developments, the study investigates the following questions:

What are the informational needs and preferences of clinicians for understanding and appropriately using AI-driven recommendations on medical decision making?

What information, according to AI experts, should be communicated to clinicians to enable their safe and appropriate use of AI-driven Clinical Decision Support Systems?

2. Methodology and ethics

2.1 Study design

This qualitative description design study [35] was conducted using semi-structured interviews to explore how to effectively communicate information about AI-driven Clinical Decision Support Systems in medical decision-making. Interviews were conducted among two groups: (i) clinicians, to understand their preferences and informational needs, and (ii) AI experts, to gain insights into what experts consider necessary for clinicians to safely and appropriately use AI tools. The inclusion of both groups was based on the assumption that informational needs go beyond the clinicians expressed preferences, but also expert-identified information is necessary for informed and safe use of AI in clinical settings. This article was written in accordance with the Consolidated criteria for Reporting Qualitative Research (COREQ) checklist (Multimedia Appendix 6) [36]. Before the start of the study, the Research Ethics form of the Faculty of Health, Medicine and Life sciences (FHML-REC, FHML/HDT/2024.022.) was evaluated and ethical approval was granted.

2.2 Recruitment of participants

Study participants were recruited by convenience and snowball sampling between July and November 2024. The potential participants were contacted by email accompanied by study information detailing the study's objectives, procedures, and confidentiality measures for the recording of the interviews. The interviewees interested in participating filled in a consent form and sent back an email to schedule a meeting, either online or face-to-face. Inclusion criteria required participants to be either (i) practicing clinicians in the Netherlands or (ii) AI experts, entailing those with experience in developing and/or implementing diagnostic or treatment recommendation AI models within the healthcare setting. Participants who had less than 2 years of work experience were excluded.

2.3 Data collection

The semi-structured interviews were conducted by a researcher (S.M.) using an interview guide (Multimedia Appendix 1 and Multimedia Appendix 2), which contained different questions for clinicians and AI experts, each consisting of four parts (Table 1). The guides were translated to English for those with no sufficient command of the Dutch language. In part one, the participants were asked demographic questions concerning their age, medical or technical specialization and work experience history. In the second part, clinicians were asked about their experiences with AI in the healthcare setting, and about instructions they received before using AI. AI experts received

questions about their experience in developing AI for clinical use, and which information they provided to the end-users. Part three contained questions about which information they would ideally either receive as clinical end-users or provide as AI developers.

In the last part, all participants received an example of a Model Card [32], a Model Facts [33] and TRIPOD-AI checklist [22, 24] (Multimedia Appendix 3), presented in that sequence. These three formats were selected because they originate from distinct contexts and serve different purposes. Model Cards are intended for communication among developers and therefore more technical in nature. In contrast, Model Facts are specifically designed for clinicians. The TRIPOD-AI checklist is a reporting standard on AI models in academic publications. To create these examples, a Bayesian Network was selected from a public repository, as it provided a balanced level of complexity, being more complex than a linear regression, but still more transparent than a deep learning model. One researcher (E.A.) completed each format based on the available model information. To preview a variation in communication styles, the same type of information was in one format presented in bullet points, while in another format it was written in narrative form, and yet another displayed this information in a table. After examining the templates, participants were asked to provide feedback on the amount, structure and comprehensiveness of the information listed in these examples.

Table 1: Overview of the interview guide including differences between clinicians and AI experts

Clinicians	AI experts
Demographic information	
Experience with AI in care settings	Experience developing AI for care settings
Current received information about AI models	Provided information concerning AI models
Information they would prefer to receive	Information they think they should provide
Examples of the Model Card, TRIPOD-AI and Model Facts were shown, and participants were requested to provide feedback concerning amount, structure and comprehensiveness of the information	

2.4 The Research Group

S.M. is a PhD candidate in Clinical Data Science (CDS) at Maastricht University, working in a department that is generally tech-positive may have shaped her perspectives on AI in healthcare. Over the course of this research project, S.M. engaged with different paradigms, initially approaching research from a post-positivist stance but increasingly shifting toward a constructivist orientation. This perspective informs the study's recognition that a single "ground truth" of information provision is not attainable, and that knowledge is context-dependent and co-constructed.

H.P. is a PhD student researching trust in AI within CDS. M.T. is a PhD student at Maastricht Clinic, researching risk communication and shared decision-making in breast cancer radiation

therapy, with a focus on patient decision aids. A.H. is a PhD student in shared decision making within radiation oncology (CDS and Maastru), with advanced training in qualitative methodology, and advised S.M. on qualitative methods. R.F. is an assistant professor in AI and clinical decision making within CDS and Maastru, and supervises the PhD candidates.

2.5 Data analysis

After data collection, recordings were transcribed verbatim and pseudonymized. Participants were provided access to both their audio recording and the corresponding transcript, which they could review at any time. The transcripts were analyzed according to codebook thematic analysis [37] using the Atlas.ti software (Version 25.0.1). This method helped identify common themes and topics which were mentioned repeatedly in interviews. The analysis process started with inductive coding by a researcher (S.M.), in which there were no predefined codes. Two other researchers (M.T. & H.P.) independently applied codes from the codebook constructed by the first researcher (S.M.) to a subset of the interviews (N=4). Consensus was reached through discussion concerning code applicability and meaning, and a few codes were collaboratively rephrased. These codes were later grouped into general themes by identifying clusters within assigned codes for clinicians (Multimedia Appendix 4) and for AI experts (Multimedia Appendix 5). A member check of the analysis was not conducted, in order to avoid influencing participants' perspectives in potential follow-up research.

3. Results

3.1 Participant characteristics

In total, 26 clinicians and 11 AI experts were invited to participate. Among invited clinicians, eight declined due to lack of time, while ten did not provide a reason. Three invited AI experts also did not provide a reason for non-participation. Between August and November 2024, 16 interviews were conducted with clinicians (N=8) and AI experts (N=8). Participant characteristics are presented in Table 2.

Table 2: Participant characteristics

Characteristics	Clinicians (N = 8)	AI Experts (N = 8)
Gender (M / F)	6 / 2	5 / 3
Median Age (Range)	44 years (29–59)	37 years (28–48)
Median Work Experience (Range)	14 years (2–22)	11 years (3–21)
Specialties / Roles	Internal Medicine (2) Oncology (2) Gastroenterology (2) Cardiology (1) Radiology (1)	Medical Imaging AI (3) Innovation & implementation management (3) Data infrastructure & FAIR principles (2)

Through conducting thematic analysis, four key themes were identified which clinicians and AI experts expressed as essential for effective communication and use of AI-CDSS. These themes include: (1) Understanding the target population, which refers to the need to understand the data on which the model was trained; (2) Clinically meaningful outcomes, focusing on how model reliability and clinical relevance are communicated; (3) Warnings, limits and safe use, highlighting the importance of clearly defining boundaries, limitations and misuse risks; (4) Accessible design, which addresses the structure and presentation of content. These themes and subthemes can be found in table 3 and are discussed in detail below.

Table 3: Overview of themes and subthemes

Theme	Subthemes
Understanding the target population	Origin of the data Class imbalance Inclusion & Exclusion Training data size
Clinically meaningful outcomes	Difficult AUC Uncertainty bands Performance graphs & Comparative visualizations
Warnings, limits and safe use	Underused, poorly highlighted and filled in too vaguely Active warnings
Accessible design	Available at all times Language barrier Customizable or layered information provision

3.2 Understanding the target population

A central theme expressed by most clinicians and experts was the ability to compare an individual patient during consultation with the training population for the AI model. This was fundamental for clinicians to evaluate whether the recommendations from the model would be reliable in a specific clinical context or not.

"What I think they should know is on which population it (AI-CDSS) is based. So for which patient are they allowed to use it and for which patient are they not. If they have a new patient in front of them, when does it (the patient) belong within the population of the model."

- AI Expert 2

Initially, participants described this need as knowledge about the origin of the data, more specifically why, where and how the data was collected. Clinicians stressed that this way, they could extrapolate whether this reflected local populations, which was especially important when the model was not locally validated.

"Where is the population from? This is clearly a Dutch population, I can extrapolate this to our own thing, but I would want to see the patient characteristics." - Clinician 4

Although the concern about the representativeness of the data and its relevance to local populations was present from the start, it became more concrete when participants reviewed the TRIPOD-AI example. Clinicians remarked that the section "class imbalance" was of great importance to make the comparison. However, the way it was written in the example felt overwhelming and too technical. Solutions for this were given, such as a table, or graph to show the class imbalance, but also a digital and interactive solution for training data comparisons. AI experts already stressed the importance of providing information about the class imbalance before being shown the TRIPOD-AI example.

(Showed class balance in TRIPOD-AI example) “Yes, basically yes, yes. Look, and if you see that they’re all N1 patients, then is it also still suitable for N2? If your patient happens to be N2, so all those kind of things. Yeah I think that’s really important to take into account.” - Clinician 5

Later in the TRIPOD-AI document, the inclusion and exclusion graph for the training data, which is often provided in papers, was shown. Clinicians stressed that without understanding who was systematically left out of the training data, whether by design or by oversight, they felt the context to interpret the applicability lacked.

“For example, it doesn’t apply to pregnancy or people who are overweight with a BMI over 30, I’m just saying something. Because otherwise, I think you lose sight of that when you’re working in practice.” - Clinician 7

Finally, both clinicians and AI experts noted that information about the training data size was missing in the Model Facts example. This was, however, crucial information according to clinicians, since a great size of the training data is associated with a better model performance.

“And actually, I’d want to know how many patients there were, like I’d want to know for reliability, how many patients were in the model so I can estimate how good it is.” - Clinician 4

3.3 Clinically meaningful outcomes

Another key theme was the need for clear evidence of the performance and reliability of the model. Both groups stressed that for an AI-CDSS to be integrated into clinical practice, its performance must be demonstrably robust and easily interpretable. However, the manner in which the performance should be presented, as well as when it was acceptable, was not agreed upon by participants.

Clinicians in particular emphasized the importance of understanding what the reported performance metrics mean in practice. They noted that values like AUC, while technically informative, do not always translate into clinical insight without additional context. Some participants mentioned a desire for performance measures to align more closely with clinical outcomes, like sensitivity and specificity and their relevant thresholds. Clinicians and AI experts suggested that these measures can also easily be translated into statistical language such as false positives or negatives, which was thought to be more familiar by clinicians.

“The AUC just gives an overall picture of whether it is a good predictive model. But if it’s 0.88 and the balance between sensitivity and specificity for your specific question isn’t optimal, then it still isn’t a good model.” - Clinician 3

Experts mentioned that besides the performance measures, there should also be a focus on providing an uncertainty band to clinicians. This should give insights into when the model is handling an edge case. AI experts noted that in current AI development, developers chase a high-performance score

such as the AUC but might overlook clinical utility or model robustness by doing so.

"We often try to get that AUC as high as possible, but that applies to the entire population, so maybe we should start saying, maybe we need to develop models that actually filter out the worst-off patients or the ones who do the best." - AI expert 2

Ultimately, while clinicians overall claimed to be familiar with performance metrics, they expressed that information on model validation often remains buried in dense technical language or tables. However, to be actionable in a clinical setting, they emphasized that information must be presented in a more structured and visually digestible manner, such as performance graphs and comparative visualizations. Besides presenting the performance metrics, clinicians expressed a desire to know when they could use an outcome to make a treatment decision, in other words, when it was 'good enough' to rely on.

"Yes, these are always somewhat abstract concepts if you don't work with them daily. And yes, each time I can reason it out – like, what it means – but I can't immediately say whether it's good or not, right?" - Clinician 5

3.4 Warnings, limits and safe use

An important theme for both clinicians and experts was the need for clear warnings and boundaries for AI-CDSS use. Explicit indicators or limitations were desired, especially when the system may be applied in situations that fall outside its intended use or scope.

Participants praised that limitations and inappropriate uses were included in the presented reporting standards. However, many AI experts emphasized that these sections were often underused, poorly highlighted, and filled in too vaguely. AI experts also mentioned that biases and limitations of the model are difficult to formulate and fill in with only their background as developers. They mentioned that defining these biases and formulating necessary warnings for misuse of the model should be done in interdisciplinary settings. They also emphasized that misuses of the model can be identified during prototype testing or practice sessions.

"So I would say that, yes, this requires critical evaluation. The problem, I think, is that if someone developing the model doesn't have much clinical background, it becomes very difficult for them to judge what is critical and what should be reported." - AI expert 3

Besides explicit mentioning of these limitations, participants also stressed that more active warnings should be included. AI experts mentioned that it would be safest to lock down the scope of the AI-CDSS within the software, such as limiting input variables or blocking predictions outside the scope. However, clinicians had varying opinions about this, with some arguing that they can decide on the model scope themselves. Another suggestion made by AI experts to warn clinicians about the correct

use of an AI-CDSS was to provide pop-ups when using the model out-of-scope. However, opinions amongst clinicians were divided, with concerns regarding pop-up tiredness and the pop-up not being read or respected. Similarly, they thought that disclaimers were likely not to be read or forgotten about during the clinical use of the model.

“There are people in their 60s who function at the level of someone over 70, or even over 80, so to speak. And there are people who are 72 where you think, well, you actually function better than your calendar age would suggest. So it’s all a matter of interpretation. And I think that this is something that should be left to a clinician.” - Clinician 8

3.5 Accessible design

Participants emphasized that how information is presented is just as critical as the content itself. Even when essential data about the system is available, it may be ignored or misunderstood in clinical settings if it is not accessible, clearly structured, or presented in a usable format.

First, the information needs to be available at all times while using the AI-CDSS, according to participants. They stressed that information should preferably be integrated within the software. However, it was also thought to be important to distribute information in different ways, such as presentations, workshops and emails. According to participants, clinicians should also familiarize themselves with the model, which can be done through e-learning and practice sessions, which is important for retention of their understanding of the model.

“But if you, for example, give a presentation and a doctor can actually use your model that same day, then that might also have a big impact on retention.” - Clinician 1

Second, participants stressed the importance of the information being conveyed in an understandable manner. AI experts often spoke about the “language barrier” they experience when communicating about AI with clinicians, and clinicians confirmed this gap when they remarked on developer jargon. Participants suggested communicating in statistical language, such as p-values or false positives, which is generally more in line with clinicians’ education and academic training. Others suggested shrinking this “language barrier” by providing general AI training for clinicians, for example from the hospital or during medical training. Another familiar way of providing information would be in the form of papers, which clinicians are accustomed to reading. The TRIPOD-AI example was well received by participants, as its resemblance to the layout of traditional academic papers appeared to create a sense of familiarity. However, some clinicians stressed that they then might as well read the entire paper, while AI experts showed significant information was still missing due to word count restraints.

“I think that, in any case, education about AI will simply be necessary for everyone in the near future, because it’s gradually being used more and more in medicine. And it’s important to understand how it works or how something like that even comes about.” - Clinician 3

Third, clinicians often labelled themselves as being intrigued by AI and intrinsically motivated to gather information about new technologies. However, many clinicians also stressed that they themselves or their coworkers would not have enough time to read up about a model. This shows a difference in needs between clinicians, which calls for customizable or layered information **provision**. An example of how this works out in practice would be through providing information blocks / pop-ups on demand or linking to (validation) papers and other useful resources.

“And references – well, of course people aren’t going to look at them very much either, so they quickly become less important. And people don’t have endless time.” - Clinician 6

Last, of all the examples, the Model Card example was perceived as the least desired method by the participants, while the first page of the Model Facts was the most appreciated. However, according to participants, Model Facts still missed essential information blocks which were provided in the TRIPOD-AI, such as class imbalance, in- and exclusion criteria and population size. Participants recommended that this information be integrated as much as possible within the software to use the AI-CDSS.

“Well, if you want to know some of the background, then the TRIPOD would be my preference. But if it’s purely about how to apply it in the clinic, then the Model Facts is fine – it’s a bit shorter and more focused on clinical applicability.” - Clinician 2

4. Discussion

This qualitative study explored the informational needs and preferences of clinicians for understanding and appropriately using AI-driven clinical decision support systems (AI-CDSS) through semi-structured interviews. Besides this study explored AI experts' perspectives on what information should be communicated to clinicians to enable safe and appropriate use of AI-CDSS. From the thematic analysis, four key domains were identified which the participants considered essential: (i) understanding the population the model was trained on, (ii) interpreting performance metrics in a clinically meaningful way, (iii) being clearly informed about limitations and boundaries of use, and (iv) having the information presented in an accessible and integrated format. Elements of the presented Model Facts, Model Card and TRIPOD-AI examples were seen as valuable in addressing these needs, though none were considered fully sufficient on their own.

The first aspect concerns the ability to assess whether the patient in front of the clinician matches the training population of the AI-CDSS. Information on the data origin, distribution and size was thus seen as crucial to avoid inappropriate extrapolation and potential harm. This aligns with guidelines made for AI in clinical research in which proper use of predictive models increases when end-users understand the AI models better [28-29]. Currently, many technical solutions aim to address dataset limitations, such as undersampling in the presence of class imbalances [38]. However, little research exists on how to equip end-users to recognize and act upon known data limitations in real-time use. The TRIPOD-AI example was praised by participants for including elements like class imbalance and in- and exclusion criteria. However, participants suggested more use of visual tools or interactive comparison modules, which reflects a desire for more transparency delivered in comprehensive formats.

Additionally, there was a need to understand model performance in ways that directly inform clinical decision-making. Although clinicians appreciated seeing performance metrics, like the one shown in the Model Facts table, they found that metrics such as the AUC lack context. This was also the case in the study performed by Frey et al. [39], in which the Model Facts and the table containing information on performance and validation was not understood. These concerns are in line with other existing literature that argues for optimizing performance reporting to the clinical context and translating technical metrics into practical terms [40]. Instead, participants preferred metrics linked to treatment relevant thresholds. Several AI experts also pointed out that communicating uncertainty is vital for identifying edge cases where model performance might be unreliable. Performance should therefore not only be reported on a global level, but in ways that match the decision points clinicians face.

Another key finding concerned the need for transparent warnings and clear boundaries for appropriate model use. While participants acknowledged that current standards such as TRIPOD-AI and Model Facts do include these elements, these sections are often too vaguely phrased, making them less actionable for clinicians. The importance of clear, structured warnings is well established in other areas of healthcare, for example in drug labels, which provide standardized information on indications, contraindications and potential risks [41]. Such an approach would move beyond abstract disclaimers, offering clinicians tangible guidance to support responsible decision-making. Besides, some participants proposed integrating restrictions directly into the software, like locking out-of-scope inputs or showing warnings when the model is applied beyond its intended population. However, these suggestions were met with mixed reactions, due to concerns about alert fatigue, which is an already frequently experienced mental strain in Clinical Decision Support tools [42-43].

Finally, participants focused on how the information is structured and delivered to the end-user. Participants advocated for layered information design, in which summary-level information is accessible at first glance, with additional detail available on demand [44]. This design principle is widely used in both healthcare settings, for example in patient decision-aids [45-46], and in other industries or webpages, like Wikipedia. Interactive data visualizations were also suggested for understanding how individual patients compare to the broader training dataset. This preference mirrors trends in other industries, such as finance and data analytics, where interactive visual tools enhance sense-making and decision quality [47]. Clinicians also emphasized the importance of integrating information in the AI-CDSS interface and supporting information dissemination through workshops or e-learning modules. While the Model Facts format was seen as more accessible and clinically relevant, it lacked essential components including training data information which were better covered in TRIPOD-AI.

In general, participants throughout this study consistently emphasized that AI-CDSS can only be made safe, effective, and clinically relevant through co-creation with clinicians at every stage of development and deployment. This would start with jointly defining the model's purpose, identifying which clinical decisions it should support and what constitutes a 'good enough' model in clinical practice. Such early alignment likely helps to determine whether the model's goals are both clinically valuable and technically feasible. Beyond initial development, it seems that clinicians should also be involved in designing the model to fit existing workflows, identifying potential limitations, and establishing appropriate safeguards. AI experts noted they cannot define these elements alone, because these require clinical insight. While the value of co-creation in AI development is well recognized in theory, it remains inconsistently practiced [48-49]. One approach formalizing co-

creation is design thinking, which is a user-centered, iterative process that emphasizes empathy, problem definition, rapid prototyping, and testing in real-world contexts [50]. Embedding such methods more systematically into AI development is essential to close the gap between technical promise and clinical utility [51].

The results of this study should be interpreted in light of several limitations. First, the use of convenience sampling may have introduced selection bias, as individuals who voluntarily participate in research are often more engaged with and positive about the topic. Second, while the study aimed to explore clinicians' informational needs regarding AI-CDSS, many clinicians seemed to lack in-depth understanding of AI, which may have limited their ability to fully articulate their needs. To address this, perspectives from AI experts were included to enrich the findings. Third, the interdisciplinary approach introduced heterogeneity, due to participants coming from a range of professional backgrounds or clinical specialties. While this breadth is a strength, allowing for a more comprehensive understanding of cross-disciplinary needs, it may also obscure discipline-specific differences. For example, certain formats or tools may be more relevant or familiar to specific specialties. Future research should explore how disciplinary context shapes informational needs and preferences, ideally through larger-scale studies using quantitative methods that allow for broader demographic and professional representation and more robust comparisons across subgroups.

In conclusion, when using AI in clinical practice, it is important that clinicians use these technologies in a safe and informed way. However, current information provision likely falls short in meeting their informational needs, contributing to the lack of large-scale adoption. AI-CDSS developers should clearly communicate both who the model is for, including characteristics of its training data, and how it performs in clinically meaningful terms. This must include explicit limitations and appropriate warnings, presented in a way that is understandable, visual, and ideally interactive. While existing formats such as Model Cards, Model Facts, and TRIPOD-AI offer valuable starting points, none fully meet these needs on their own. Future efforts should consider combining and adapting elements from these formats to better support clinical use. Effective integration of AI-CDSS requires collaboration with clinicians during development and targeted education during implementation. Addressing informational needs of clinicians is critical to ensure appropriate use and supports the safe and effective integration of AI-CDSS into clinical practice.

Acknowledgments

The authors would like to thank the clinicians and AI experts who participated in this study. Their time and insights resulted in thorough discussions, which were essential to advancing the understanding of informational needs related to AI-driven clinical decision support systems.

Authors' contributions

Conceptualization: SMBM, EA, JPAS, LH, OCD, RRRF

Data curation: SMBM

Formal Analysis: SMBM, HSP, MST

Investigation: SMBM, HSP, MST

Methodology: SMBM, EA, ARH, JPAS, LH, OCD, RRRF

Project administration: SMBM, LH, RRRF

Resources: SMBM, EA, RRRF

Supervision: LH, OCD, RRRF

Writing – original draft: SMBM

Writing – review & editing: SMBM, HSP, MST, EA, AHR, JPAS, LH, ALAJD, OCD, RRRF

Funding

This research was funded by the Dutch Research Council (NWO).

Conflict of Interest

The primary author declares that they have no known competing financial interests. The primary author is affiliated with a department where AI-based systems are developed. Some of the authors receive research grants related to the development and implementation of AI. In addition, some of the respondents interviewed were known to the primary author prior to the study.

J.S. and A.D. are shareholders and receiving salary of Medical Data Works B.V.; although company activities are unrelated to this manuscript. Outside of the submitted work, R.F.'s research group receives funding from public-private research consortia, including contributions from Janssen-Cilag and Takeda. These companies had no role in the present study.

Data availability

The datasets gathered and analyzed during the study are not publicly available due to the presence of potentially identifiable information within the interview transcripts. However, de-identified excerpts or relevant portions of the data may be made available from the corresponding author upon reasonable request and subject to institutional and ethical approvals.

Abbreviations

AI: Artificial Intelligence

AI-CDSS: Artificial Intelligence-driven Clinical Decision Support Systems

AIOS: Arts in Opleiding tot Specialist (Physician in Specialist Training)

AUC: Area Under the Curve

CDSS: Clinical Decision Support Systems

CONSORT-AI: CONSolidated Standards Of Reporting Trials – Artificial Intelligence

COREQ: COnsolidated criteria for REporting Qualitative research

DECIDE-AI: Developmental and Exploratory Clinical Investigation of DEcision support systems driven by Artificial Intelligence

FHML-REC: Faculty of Health, Medicine, and Life Sciences – Research Ethics Committee

SPIRIT-AI: Standard Protocol Items: Recommendations for Interventional Trials – Artificial Intelligence

TRIPOD-AI: Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis – Artificial Intelligence

Multimedia Appendices

Appendix 1: Interview guide clinicians

Appendix 2: Interview guide AI experts

Appendix 3: Examples of Model Card, TRIPOD-AI and Model Facts

Appendix 4: Clusters with themes Clinicians

Appendix 5: Clusters with themes AI Experts

Appendix 6: COREQ checklist

References

1. Aung YYM, Wong DCS, Ting DSW. The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare. *British Medical Bulletin*. 2021;139(1):4-15. doi:10.1093/bmb/ldab016
2. Ankolekar A, van der Heijden B, Dekker A, et al. Clinician perspectives on clinical decision support systems in lung cancer: Implications for shared decision-making. *Health Expectations*. 2022;25(4):1342-1351. doi:10.1111/hex.13457
3. Goirand M, Austin E, Clay-Williams R. Implementing Ethics in Healthcare AI-Based Applications: A Scoping Review. *Science and Engineering Ethics*. 2021;27(5). doi:10.1007/s11948-021-00336-3
4. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial Intelligence and the Implementation Challenge. *Journal of Medical Internet Research*. 2019;21(7). doi:10.2196/13659
5. Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P. Implementation Frameworks for Artificial Intelligence Translation into Health Care Practice: Scoping Review. *Journal of Medical Internet Research*. 2022;24(1). doi:10.2196/32215
6. Petersson L, Larsson I, Nygren JM, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. *BMC Health Services Research*. 2022;22(1). doi:10.1186/s12913-022-08215-8
7. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future Healthcare Journal*. 2019;6(2):94-98. doi:10.7861/futurehosp.6-2-94
8. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*. 2019;17(1). doi:10.1186/s12916-019-1426-2
9. Esmailzadeh P. Challenges and strategies for wide-scale artificial intelligence (AI) deployment in healthcare practices: A perspective for healthcare organizations. *Artificial Intelligence in Medicine*. 2024;151. doi:10.1016/j.artmed.2024.102861
10. Smith H, Downer J, Ives J. Clinicians and AI use: Where is the professional guidance? *Journal of Medical Ethics*. Published online 2023. doi:10.1136/jme-2022-108831
11. Jabbour S, Fouhey D, Shepard S, et al. Measuring the Impact of AI in the Diagnosis of

- Hospitalized Patients: A Randomized Clinical Vignette Survey Study. *JAMA*. 2023;330(23):2275-2284. doi:10.1001/jama.2023.22295
12. Zhao A, Larbi M, Miller K, O'Neill S, Jayasekera J. A scoping review of interactive and personalized web-based clinical tools to support treatment decision making in breast cancer. *Breast*. 2022;61:43-57. doi:10.1016/j.breast.2021.12.003
 13. Völkel V, Hueting TA, Draeger T, et al. Improved risk estimation of locoregional recurrence, secondary contralateral tumors and distant metastases in early breast cancer: the INFLUENCE 2.0 model. *Breast Cancer Research and Treatment*. 2021;189(3):817-826. doi:10.1007/s10549-021-06335-z
 14. van Maaren MC, Hueting TA, Völkel V, van Hezewijk M, Strobbe LJ, Siesling S. The use and misuse of risk prediction tools for clinical decision-making. *Breast*. 2023;69:428-430. doi:10.1016/j.breast.2023.01.006
 15. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *The Lancet*. 2019;393(10181):1577-1579. doi:10.1016/S0140-6736(19)30037-6
 16. Fehr J, Citro B, Malpani R, Lippert C, Madai VI. A trustworthy AI reality-check: the lack of transparency of artificial intelligence products in healthcare. *Frontiers in Digital Health*. 2024;6. doi:10.3389/fdgth.2024.1267290
 17. Van Calster B, Wynants L, Timmerman D, Steyerberg EW, Collins GS. Predictive analytics in health care: how can we know it works? *Journal of the American Medical Informatics Association*. 2019;26(12):1651-1654. doi:10.1093/jamia/ocz130
 18. Campbell DJ. The clinical utility curve: A proposal to improve the translation of information provided by prediction models to clinicians. *BMC Research Notes*. 2016;9(1). doi:10.1186/s13104-016-2028-0
 19. Meshaka R, Pinto Dos Santos D, Arthurs OJ, Sebire NJ, Shelmerdine SC. Artificial intelligence reporting guidelines: what the pediatric radiologist needs to know. *Pediatric Radiology*. 2022;52(11):2101-2110. doi:10.1007/s00247-021-05129-1
 20. Yao S, Shen P, Dai T, et al. Human understandable thyroid ultrasound imaging AI report system — A bridge between AI and clinicians. *iScience*. 2023;26(4). doi:10.1016/j.isci.2023.106530
 21. Boag W, Hasan A, Kim JY, et al. The algorithm journey map: a tangible approach to implementing AI solutions in healthcare. *npj Digital Medicine*. 2024;7(1). doi:10.1038/s41746-024-01061-4
 22. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Medicine*. 2015;13(1):1. doi:10.1186/s12916-014-0241-z
 23. Collins GS, Dhiman P, Andaur Navarro CL, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic

- and prognostic prediction model studies based on artificial intelligence. *BMJ Open*. 2021;11(7). doi:10.1136/bmjopen-2020-048008
24. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. Published online April 16, 2024:e078378. doi:10.1136/bmj-2023-078378
25. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nature Medicine*. 2020;26(9):1364-1374. doi:10.1038/s41591-020-1034-x
26. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nature Medicine*. 2022;28(5):924-933. doi:10.1038/s41591-022-01772-9
27. Cruz Rivera S, Liu X, Chan AW, et al. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nature Medicine*. 2020;26(9):1351-1363. doi:10.1038/s41591-020-1037-7
28. Topol EJ. Welcoming new guidelines for AI clinical research. *Nature Medicine*. 2020;26(9):1318-1320. doi:10.1038/s41591-020-1042-x
29. Volovici V, Syn NL, Ercole A, Zhao JJ, Liu N. Steps to avoid overuse and misuse of machine learning in clinical research. *Nature Medicine*. 2022;28(10):1996-1999. doi:10.1038/s41591-022-01961-6
30. Lavin A, Gilligan-Lee CM, Visnjic A, et al. Technology readiness levels for machine learning systems. *Nat Commun*. 2022;13(1):6039. doi:[10.1038/s41467-022-33128-9](https://doi.org/10.1038/s41467-022-33128-9)
31. Fang H, Miao H. Introducing the Model Card Toolkit for Easier Model Transparency Reporting. GoogleResearch. Published online July 2020.
32. Mitchell M, Wu S, Zaldivar A, et al. Model cards for model reporting. In: Association for Computing Machinery, Inc; 2019:220-229. doi:10.1145/3287560.3287596
33. Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. *npj Digital Medicine*. 2020;3(1). doi:10.1038/s41746-020-0253-3
34. Schwartz LM, Woloshin S, Welch HG. Using a Drug Facts Box to Communicate Drug Benefits and Harms Two Randomized Trials. *Annals of Internal Medicine*. 2009;150(8):516-527. doi:10.7326/0003-4819-150-8-200904210-00106
35. Bradshaw C, Atkinson S, Doody O. Employing a Qualitative Description Approach in Health Care Research. *Global Qualitative Nursing Research*. 2017;4. doi:10.1177/2333393617742282
36. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for*

- Quality in Health Care. 2007;19(6):349-357. doi:10.1093/intqhc/mzm042
37. Braun V, and Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*. 2021;18(3):328-352. doi:10.1080/14780887.2020.1769238
38. Salmi M, Atif D, Oliva D, Abraham A, Ventura S. Handling imbalanced medical datasets: review of a decade of research. *Artificial Intelligence Review*. 2024;57(10):273. doi:10.1007/s10462-024-10884-2
39. Frey N, Agha-Mir-Salim L, Hinz E, Poncette AS, Balzer F. Assessing Healthcare Stakeholder Understanding of Machine Learning Documentation. In: Andrikopoulou E, Gallos P, Arvanitis TN, et al., eds. *Studies in Health Technology and Informatics*. IOS Press; 2025. doi:10.3233/SHTI250336
40. Reyna MA, Nsoesie EO, Clifford GD. Rethinking Algorithm Performance Metrics for Artificial Intelligence in Diagnostic Medicine. *JAMA*. 2022;328(4):329. doi:10.1001/jama.2022.10561
41. Sullivan HW, O'Donoghue AC, Aikin KJ. Primary Care Physicians' Use of FDA-Approved Prescription Drug Labels. *The Journal of the American Board of Family Medicine*. 2014;27(5):694-698. doi:10.3122/jabfm.2014.05.140039
42. Kesselheim AS, Cresswell K, Phansalkar S, Bates DW, Sheikh A. Clinical Decision Support Systems Could Be Modified To Reduce 'Alert Fatigue' While Still Minimizing The Risk Of Litigation. *Health Affairs*. 2011;30(12):2310-2317. doi:10.1377/hlthaff.2010.1111
43. McGreevey JD, Mallozzi CP, Perkins RM, Shelov E, Schreiber R. Reducing Alert Burden in Electronic Health Records: State of the Art Recommendations from Four Health Systems. *Appl Clin Inform*. 2020;11(01):001-012. doi:10.1055/s-0039-3402715
44. Boxwala AA, Rocha BH, Maviglia S, et al. A multi-layered framework for disseminating knowledge for computer-based decision support. *Journal of the American Medical Informatics Association*. 2011;18(Supplement 1):i132-i139. doi:10.1136/amiajnl-2011-000334
45. Van Strien-Knippenberg IS, Boshuizen MCS, Determann D, De Boer JH, Damman OC. Cocreation with Dutch patients of decision-relevant information to support shared decision-making about adjuvant treatment in breast cancer care. *Health Expectations*. 2022;25(4):1664-1677. doi:10.1111/hex.13510
46. Roumen C, Hasannejadasl H, Swart R, et al. Breast cancer patients' most important quality of life themes for a radiotherapy decision aid. *The Breast*. 2022;65:8-14. doi:10.1016/j.breast.2022.06.002
47. Perdana A, Rob A, Rohde F. Does Visualization Matter? The Role of Interactive Data Visualization to Make Sense of Information. *AJIS*. 2018;22. doi:10.3127/ajis.v22i0.1681

48. Laka M, Carter D, Milazzo A, Merlin T. Challenges and opportunities in implementing clinical decision support systems (CDSS) at scale: Interviews with Australian policymakers. *Health Policy and Technology*. 2022;11(3):100652. doi:10.1016/j.hlpt.2022.100652
49. Swan EL, Peltier JW, Dahl AJ. Artificial intelligence in healthcare: the value co-creation process and influence of other digital health transformations. *JRIM*. 2024;18(1):109-126. doi:10.1108/JRIM-09-2022-0293
50. Davis J, Docherty ,Catherine Ann, and Dowling K. Design Thinking and Innovation: Synthesising Concepts of Knowledge Co-creation in Spaces of Professional Development. *The Design Journal*. 2016;19(1):117-139. doi:10.1080/14606925.2016.1109205
51. Sreenivasan A, Suresh M. Design thinking and artificial intelligence: A systematic literature review exploring synergies. *International Journal of Innovation Studies*. 2024;8(3):297-312. doi:10.1016/j.ijis.2024.05.001

Supplementary Files

Multimedia Appendixes

Interview guide clinicians.

URL: <http://asset.jmir.pub/assets/85490c22f1b57d08846d07b97d101661.docx>

Interview guide AI experts.

URL: <http://asset.jmir.pub/assets/d99de2d1f8567f30f1a13d0e6c0f8d7d.docx>

Examples of Model Card, TRIPOD-AI and Model Facts.

URL: <http://asset.jmir.pub/assets/6faad6dd23a7e1324a5e30c64362af52.docx>

Clusters with themes clinicians.

URL: <http://asset.jmir.pub/assets/8c94b0ec9a7dd44af3d524719cb68f7e.docx>

Clusters with themes AI experts.

URL: <http://asset.jmir.pub/assets/54d5a09542614ebf04867d5df73db152.docx>

CONSORT (or other) checklists

COREQ checklist.

URL: <http://asset.jmir.pub/assets/a027fa8aa787f74015f2eeab97880a7c.pdf>