Future Healthcare Journal 11 (2024) 100172

Contents lists available at ScienceDirect

Future Healthcare Journal

journal homepage: https://www.sciencedirect.com/journal/future-healthcare-journal

Quality of interaction between clinicians and artificial intelligence systems. A systematic review

Argyrios Perivolaris[a,b,*], Chris Adams-McGavin[c], Yasmine Madan[d], Teruko Kishibe[e], Tony Antoniou[f,g,h], Muhammad Mamdani[b,i,j], James J. Jung[a,b,c]

[a] *Institute of Medical Sciences, University of Toronto, Canada*
[b] *St. Michaels Hospital, Unity Health Toronto, Canada*
[c] *Department of Surgery, Temetry Faculty of Medicine, University of Toronto, Canada*
[d] *Department of Health Sciences, McMaster University, Canada*
[e] *MISt, Library Services, Unity Health Toronto, Canada*
[f] *Department of Family and Community Medicine, St. Michael's Hospital, Canada*
[g] *Li Ka Shing Knowledge Institute, St. Michael's Hospital, Canada*
[h] *Department of Family and Community Medicine, University of Toronto, Canada*
[i] *Leslie Dan Faculty of Pharmacy, Temerty Faculty of Medicine, University of Toronto, Canada*
[j] *Dalla Lana School of Public Health, University of Toronto, Canada*

ARTICLE INFO

Keywords:
Artificial intelligence
Clinicians
Interaction
Systematic review
Interviews

ABSTRACT

Introduction: Artificial intelligence (AI) has the potential to improve healthcare quality when thoughtfully integrated into clinical practice. Current evaluations of AI solutions tend to focus solely on model performance. There is a critical knowledge gap in the assessment of AI–clinician interactions. We systematically reviewed existing literature to identify interaction traits that can be used to assess the quality of AI–clinician interactions.
Methods: We performed a systematic review of published studies to June 2022 that reported elements of interactions that impacted the relationship between clinicians and AI-enabled clinical decision support systems. Due to study heterogeneity, we conducted a narrative synthesis of the different interaction traits identified from this review. Two study authors categorised the AI–clinician interaction traits based on their shared constructs independently. After the independent categorisation, both authors engaged in a discussion to finalise the categories.
Results: From 34 included studies, we identified 210 interaction traits. The most common interaction traits included usefulness, ease of use, trust, satisfaction, willingness to use and usability. After removing duplicate or redundant traits, 90 unique interaction traits were identified. Unique interaction traits were then classified into seven categories: usability and user experience, system performance, clinician trust and acceptance, impact on patient care, communication, ethical and professional concerns, and clinician engagement and workflow.
Discussion: We identified seven categories of interaction traits between clinicians and AI systems. The proposed categories may serve as a foundation for a framework assessing the quality of AI–clinician interactions.

Introduction

The field of medicine has witnessed a remarkable transformation with the rapid rise of artificial intelligence (AI). AI has demonstrated tremendous potential in healthcare, from aiding in medical diagnosis to predicting disease outbreaks.[1] Specifically, several studies have found that AI image recognition systems perform at the same level or higher when compared to human clinicians.[2] AI may also help manage patient data and medical records, reducing the potential for human error and streamlining the healthcare process.[3]

Despite these developments, AI integration into healthcare settings has remained challenging. Reasons for the lack of integration remain poorly characterised in the literature. When considering end users, concerns about system reliability and accuracy caused by non-transparent and inappropriate training data are a reason for the reluctance to integrate AI into practice.[4] Concerns about the potential to compromise patient privacy or autonomy may also limit uptake.[5] The need for clinician training and education to integrate AI into daily practice is another potential barrier to integration.[6] However, in contrast to studies

This article reflects the opinions of the author(s) and should not be taken to represent the policy of the Royal College of Physicians unless specifically stated.

* Corresponding author at: 36 Queen St E, M5B 1W8 Toronto, Ontario, Canada.
E-mail address: argyrios.perivolaris@mail.utoronto.ca (A. Perivolaris).

https://doi.org/10.1016/j.fhj.2024.100172
Received 28 February 2024; Received in revised form 15 July 2024; Accepted 4 August 2024
2514-6645/

describing system development and performance, relatively few studies have explored clinician experiences with AI.[7,8] One important consequence of this lack of research is an absence of a standardised approach for ascertaining the quality of interaction between clinicians and AI.

Quality of interaction is a perception associated with a service during an encounter with said service. Therefore, in this review, the quality of interaction construct pertains to how clinicians perceive the experience before, during and after engaging with an AI system. Quality of interaction is a critical component to consider because the perception of quality will likely be a key determinant of AI integration into practice.[9] In addition, identifying the specific features or traits that comprise the quality of interaction between clinicians and AI systems is needed to understand the elements warranting consideration when AI systems are developed, implemented, and evaluated. Yet, there have been no systematic reviews identifying these traits and synthesising them to create a framework that assesses the AI–clinician quality of interaction construct.

Our specific research question was 'What are the components that characterise the quality of AI–clinician interactions?'. To address this question, we conducted a systemic review of studies that reported clinician experiences and perceptions after an intervention with an AI-enabled clinical decision support system. Our objectives were to (1) characterise and review studies exploring clinician experiences and perceptions of AI and (2) summarise the interaction traits as discrete categories that will form the foundation of a standardised approach to evaluating the quality of interaction between AI-enabled clinical decision support systems and clinicians. The results will serve as item generation that will be reduced and refined in future work when developing a comprehensive framework.

Methods

Study design and search strategy

We conducted a systematic review according to the Cochrane Collaboration Handbook and reported the findings following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statements.[10,11]

We used a combination of Medical Subject Headings (MeSH) and text words relating to AI, informed by an established AI search filter, with terms relating to clinicians and perceptions (supplemental text S1 for search strategies).[12] In addition, the references of the included studies were searched manually for additional eligibility.

We included original research studies that reported on interaction traits with AI systems in clinical practice settings, including but not limited to hospitals, for-profit private care facilities, and telehealth care settings. We included studies presenting measures of clinician-reported experiences and quality of interaction with AI-enabled clinical decision support systems, defined as information technology systems that learn and support clinicians in decision-making. We excluded studies that pertained to the use of surgical robots because their use is not for decision aid. We also excluded AI technology that relies on simple rule-based or if-then-based strategies as we focused on AI systems that utilise more complex algorithms. We included studies of any healthcare workers, except those that solely studied students or learners. We included randomised trials, cohort, case–control, cross-sectional, case-series, qualitative and mixed methods studies. We excluded abstracts, dissertation/thesis work, unpublished reports or data, reviews, protocols, opinions and letters to editors. We also excluded animal-only studies, case reports, comments, editorials, letters, and studies published in languages other than English.

Data sources

We searched the Medline (OVID), PsychINFO (OVID), Embase (OVID, CINAHL (EBSCO) and Scopus (ELSEVIER) databases from inception to June 2022 with an experienced information specialist (TK), to identify published studies that reported on clinician experience with AI systems and quality of interaction between AI systems and the clinicians. In addition, the references of the included studies were searched manually for additional eligibility.

Study selection

Two reviewers (AP and YM) independently screened titles and abstracts to identify studies for full-text review using Covidence software.[13] The same two reviewers independently applied the inclusion and exclusion criteria in the full-text review to select studies for data extraction.

Data extraction

A data extraction sheet (supplemental appendix S2) was developed using the data extraction tab in Covidence and was pilot-tested for feasibility and acceptability. We extracted study characteristics, including year of publication, country, study design, clinician population and types, method used to evaluate interaction quality (surveys, interviews, questionnaires, etc.), description of the AI tool, and specific interaction trait types between clinicians and AI systems that reflected the quality of interaction. An interaction trait is an element of interaction that impacts the relationship between the AI system and the clinician. The two reviewers completed the data extraction independently. Study selection and data extraction disagreements were resolved through discussions or by a third reviewer (JJ) if a consensus was not reached. No attempts were made to contact the authors of the included studies for supplementary information. A definition of all the study objectives can be found in Table 1.

We extracted terms that described the quality of interaction traits between clinicians and AI systems verbatim from the included studies. These extracted interaction traits were then paraphrased to remove redundancy and duplicate concepts. A statistical meta-analysis was not possible due to the heterogeneity of the data and thus, we reported a narrative synthesis that summarises and defines the interaction traits. All unique interaction traits were listed in a spreadsheet. In the initial step of categorisation, we sorted interaction traits based on their similarity. Next, two authors (AP and JJ) independently grouped individual traits into categories reflecting similar ideas or meanings about the quality of interaction between clinicians and AI systems and provided each category with a name that reflected the theme of the summarised ideas. Categorisation and title changes continued until each interaction trait could be placed in one discrete category without qualifying for another. Authors were blinded to each others' categorisations to maintain independence. Both authors then engaged in a virtual consensus meeting to evaluate and discuss all agreements and any discrepancies in their independent categorisations. Authors were allowed to provide justifications for their categorisations. Discrepancies were resolved and consensus was achieved by the two authors independent of a third party. Once the two authors agreed with the final categories that represented the interaction traits and the quality of interaction construct, iterative refinement of the consensus results would take place by the process of obtaining feedback and insights from other study authors (TA, MM, YM). Each author independently sedanalysed the categorisation results and provided feedback to the two original evaluators who would make the necessary improvements to the categories. Additional rounds of independent analysis continued until all three authors had no further feedback.

Quality and risk of bias assessment

The Critical Appraisal Skills Programme (CASP) tool was used to evaluate the quality of qualitative studies and qualitative components of mixed-methods studies.[14] We only considered the qualitative questions

Table 1
Definitions of the study objective terms.

Study objective	Definition
Quality of interaction	A perception associated with a service during an encounter with said service.
Interaction trait	An element of interaction that impacts the relationship between the AI system and the clinician.
Clinical decision support system	Information technology systems that learn and support clinicians in decision-making.
Complex AI	An AI system that is trained on a set of data to learn the underlying patterns to allow it to excel in pattern creation.

asked and responses of clinicians relevant for interaction trait collection. Any possible quantification of responses did not contribute to the categorisation analysis. All questions were rated as 'yes', 'no', or 'cannot determine'. The CASP tool does not report a summative score. Instead, an overall assessment of the studies as 'not valuable', 'semi-valuable', 'valuable', or 'very valuable' was reported. These overall assessments were based on a judgemental approach where the reviewers evaluated how valuable the research was for providing relevant and reasonable interaction traits for the AI–clinician quality of interaction construct. Two reviewers (AP and YM) independently performed the quality assessments for the studies and differences were resolved by reaching consensus when needed.

Semi-structured interviews

We further conducted in-depth semi-structured interviews with clinicians and AI experts in order to confirm the contents of the systematic review findings and to obtain and categorise additional interaction traits not found in the systematic review. We interviewed six participants who were currently or formerly employed with Unity Health and who had at least one interaction with an AI system at the hospital in their role as clinicians or data scientists. Potential participants were approached via recruitment emails by either the study coordinator (AP) or the principal investigator (JJ). An interview guide was developed to elicit participant perspectives and experiences regarding their interactions with AI systems and items for the AI–clinician quality of interaction construct that were not identified in the systematic review. The interview guide was pilot-tested with three participants of varying background knowledge for feasibility and comprehensiveness. These included a medical student with prior knowledge of AI, a nurse with no experience in AI, and an AI content expert. After feedback integration, the interview guide was finalised. All interviews were conducted using video conference calls, were between 30 and 45 min in duration, and audio recorded. Informed consent was obtained from each participant.

Following transcription, we sedanalysed the data using a descriptive content analysis approach, with the categories generated from the systematic review serving as the initial guiding framework. Additional interaction traits that were raised during the interviews were included to capture new insights. The new interaction traits obtained from the interviews that were not found in the systematic review were listed in a spreadsheet and classified into one of the categories. All data were anonymised, and participant identities were safeguarded throughout the study to ensure confidentiality. All interview participants provided informed consent for their participation and the publishing of anonymised data. This study was reviewed and approved by the Unity Health Toronto Research Ethics Board (REB# 23–006).

Results

Our search identified 17,660 articles. After excluding 1,927 duplicates, a total of 14,864 articles were excluded as they did not meet the inclusion criteria. Of the remaining 869 full texts, we excluded 834 for various reasons, including wrong measurement outcomes, no AI intervention, and incorrect population. One final study was excluded from the narrative synthesis after it was deemed low quality based on the quality assessment evaluation.[46] Therefore, the final sample comprised 34 studies (Fig. 1).

Quality assessment results

All studies had qualitative analyses and thus, were assessed using the CASP checklist (Table 2). From the 35 studies, 29 received an overall rating of valuable or very valuable in quality assessment. Reasons for studies receiving 'semi-valuable' or 'not valuable' designation included inappropriate utilisation of qualitative methods for measuring non-subjective outcomes, sample recruitment strategies, and lack of consideration for bias. Quality assessment was not part of the systematic review inclusion criteria, but rather was undertaken to provide an overview of the quality of the literature identified as being eligible for inclusion. Therefore, studies that received a 'not valuable' designation could be included in the data analysis. The studies were denoted as 'not valuable' in their methodology for providing meaningful insight into the tool in question; however, they provided unique viewpoints to the AI–clinician quality of interaction that were not obtained from the other included studies.

Study characteristics

The publication years range from 2006 to 2022, and the participant samples comprised 39 different clinician types (Table 3). Study designs included qualitative research, randomised controlled trials, and mixed-methods studies, with the latter representing the majority (n = 23; 67.6%) of studies in this review. Evaluation methods included questionnaires, interviews, surveys, scenario observations and evaluations, usability scales, focus group discussions, and case studies. There were 33 different AI systems included in the 34 studies. A brief description of each AI system is included in Table 3.

Characteristics of interaction traits from the systematic review

We identified a total of 210 quality of interaction traits from the 34 included studies. After removing duplicates and redundant traits, there were 90 unique interaction traits (Table 3). For example, 'ease of learning', 'learnability, and 'impact on clinician learning' were reduced to 'ease of learning'. The most frequently studied interaction trait was usefulness which appeared in 32.4% of the studies. Other interaction traits reported with high frequency were ease of use (29.4%), trust (23.5%), satisfaction (23.5%), willingness to use (23.5%), and usability (14.7%). Interaction traits were grouped based on a judgemental approach into seven categories (Tables 4 and 5). The final seven categories for evaluating the quality of interaction between clinicians and AI were usability and user experience, system performance, clinician trust and acceptance, impact on patient care, clinician engagement and workflow, communication and collaboration, and ethical and professional concerns. The seven categories are listed and defined in Table 5. The following paragraphs summarise and define seven AI–clinician interaction categories and their associated interaction traits.

Usability and user experience

Eight studies assessed ease of use of the AI system.[19,21,23,25,36,37,41,44] Ease of use is defined as how easily users can utilise an AI system on their own. Four studies measured clinician-reported responses on the learnability of the system.[21,25,27,36] Two of those studies further specified by asking about ease of learning for the system.[21,27] Clinicians were asked about their confidence in

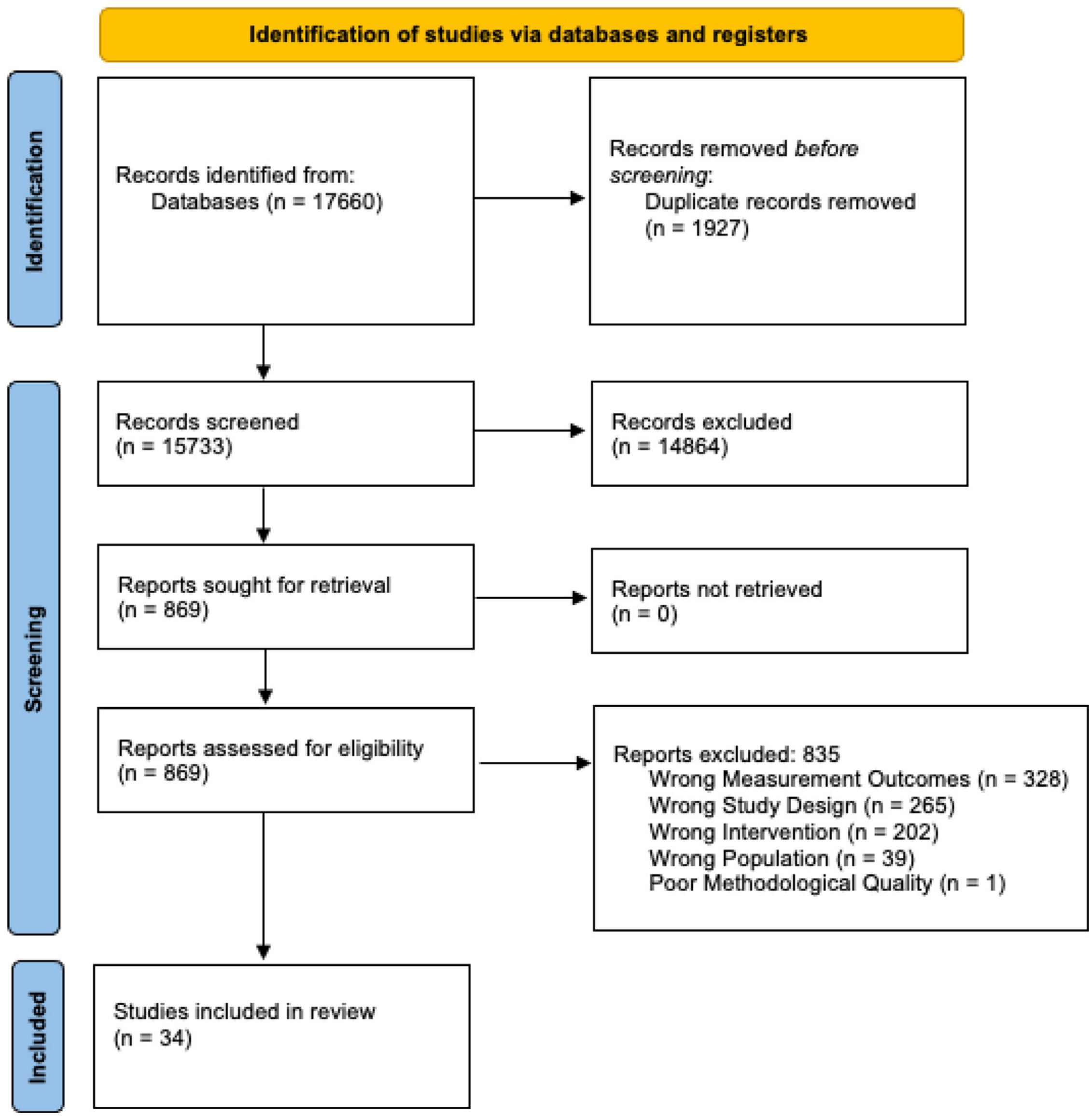

Fig. 1. Preferred reporting items for systematic reviews and meta-analyses flow diagram.

their ability to use the AI system in three studies.[21,26,36] Similar to clinician confidence, one study measured the clinician's comfort when using the system.[49] One study measured how the AI system impacted the clinician's physical and mental demands.[22] Two studies asked clinicians how they felt the information was organised visually and its comprehensiveness.[27,35] One study measured whether clinicians perceived the AI system's visual appearance as appealing.[19] Five studies measured clinicians' overall satisfaction with the AI system's performance.[15,27,34,40,46]

System performance

Seven studies reported interaction traits relevant to clinicians' perceptions of the technological capabilities of the AI system. Two studies asked clinicians about their beliefs on the accuracy of the AI system's outputs.[32,36] Two studies evaluated clinicians' perceptions of the speed of the AI system.[36,37] Two studies measured clinicians' perceptions of the overall quality and computational efficiency of the system.[19,20] Two studies measured clinicians' perceptions of the performance of AI.[17,25]

Clinician trust and acceptance

There were 18 studies that evaluated clinicians' trust and acceptance of AI systems. Six studies assessed clinicians' beliefs about trust in the system accuracy and development,[22,25,31,45] security,[16] and patient trust.[20] Two studies measured clinicians' beliefs about the potential that AI could have in healthcare in the future.[15,43] Six studies measured the clinicians' opinions of the AI system's output, including their acceptance or disagreement with the system's output.[19,22,26,28,29,40] Two studies evaluated the clinicians' beliefs about the reliability of the AI system based on the outputs, training data used, or transparency of system development.[31,44] One study measured clinicians' scepticism about the system, which shared many parallels with system reliability.[18] Three studies evaluated the clinician's beliefs on inherent risks with AI including risks to the workplace, patient care and patient information privacy.[25,33,47]

Impact on patient care

Some of the studies were not only concerned with the clinician's opinions on how AI impacts their roles and responsibilities, but also with their opinions about patient care. Two studies assessed how AI integration would affect the patient–physician interaction and if this would hinder AI deployment in their clinical practice.[20,42] Two studies measured clinicians' concerns about the sensitive nature of patient information being used in AI systems.[26,47] One study asked clinicians how they thought the AI system would impact the quality of care for

Table 2
The reported answers of the Critical Appraisal Skills Programme for each study included in the analysis.

Study ID	1. Was there a clear statement of the aims of the research?	2. Is a qualitative methodology appropriate?	3. Was the research design appropriate to address the aims of the research?	4. Was the recruitment strategy appropriate to the aims of the research?	5. Was the data collected in a way that addressed the research issue?	6. Has the relationship between researcher and participants been adequately considered?	7. Have ethical issues been taken into consideration?	8. Was the data analysis sufficiently rigorous?	9. Is there a clear statement of findings?	10. How valuable is the research?
Abdulaal 2021[15]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Valuable
Aldughayfiq 2022[16]	Yes	Yes	Yes	Yes	Yes	No	Can't tell	Yes	Yes	Valuable
Allen 2021[17]	Yes	Yes	Yes	Yes	Yes	Can't tell	No	Yes	Yes	Very valuable
Ankolekar 2022[18]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Very valuable
Bajorek 2012[19]	Yes	Yes	Yes	Can't tell	Yes	Yes	Yes	Yes	Yes	Very valuable
Benrimoh 2021[20]	Yes	Yes	Yes	Yes	Yes	Can't tell	Yes	Yes	Yes	Valuable
Calisto 2021[21]	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Valuable
Calisto 2022[22]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Very valuable
Carlile 2020[23]	Yes	Yes	Yes	Can't tell	Yes	Yes	Yes	Yes	Yes	Very valuable
Cheikh 2022[24]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Valuable
Choudhury 2022[25]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Very valuable
Creed 2022[26]	Yes	Yes	Yes	Yes	Yes	Can't tell	Yes	Yes	Yes	Very valuable
Dontchos 2021[27]	Yes	Yes	Yes	Can't tell	Yes	Yes	Yes	Yes	Yes	Valuable
Dunsmuir 2008[28]	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Valuable
Garrett Fernandes 2021[29]	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Semi-valuable
Ginestra 2019[30]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Very valuable
Goel 2022[31]	Yes	Yes	Yes	Yes	Can't tell	Yes	Can't tell	Yes	Can't Tell	Not Valuable
Hirsch 2018[32]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Very valuable
Hogue 2021[33]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Valuable
Im 2006[34]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Very valuable
Jaber 2022[35]	Yes	Yes	Yes	No	Yes	No	Can't tell	Yes	Yes	Very Valuable
Jones 2021[36]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Semi-Valuable
Juluru 2021[37]	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Very Valuable
Kim 2022[38]	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Not Valuable
Kumar 2020[39]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Very valuable
Künzel 2022[40]	Yes	Yes	Yes	No	Yes	No	No	Yes	Can't Tell	Valuable
Moret-Tatay 2022[41]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Semi Valuable
Romero-Brufau 2020[42]	Yes	Yes	Yes	Yes	Yes	Can't tell	Yes	Yes	Yes	Very valuable
Scheder-Bieschin 2022[43]	Yes	Yes	Yes	Can't tell	Yes	No	Yes	Yes	Yes	Very Valuable
Scheetz 2021[44]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Very Valuable
Tanguay-Sela 2022[45]	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Very Valuable
Wong 2021[46]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Can't Tell	Yes	Very valuable
Zhai 2021[47]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Valuable
Zhang (2021)[48]	No	No	No	No	No	No	No	No	No	Not Valuable
Zhang 2022[49]	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Very Valuable

patients.[30] One study reported on how AI affects levels of patient engagement in their care.[49]

Clinician engagement and workplace

Nine studies measured the perceived usefulness of the AI system.[18,19,22,32,35,39,43,45,47]] Terminology utilised in the literature to describe usefulness included usefulness, relevance and benefit. Another interaction trait similar to usefulness measured in three studies was helpfulness.[20,30,45] Helpfulness was deemed distinct from usefulness as this trait pertains to an evaluation of how the AI system assists the clinician as opposed to just solving a problem.

Seven studies reported measurements of the clinician's willingness to use or recommend the system to colleagues.[18,21,36,41,43,46,47] One paper further addressed this by measuring resistance bias, which included factors such as fear, anger or lack of awareness when using the AI system.[47] Five studies evaluated how the AI system impacted task completion time, including both time saving and increases in time.[16,18,37,39,43]

Communication and collaboration

Four studies evaluated the impact that AI has on clinician–clinician communication and collaboration. One study measured an AI ePrescription tool's impact on pharmacist communication with prescribers.[16] One study evaluated how AI systems would impact supervisory processes and interpersonal communication between psychotherapists in their practice.[26] Two studies evaluated AI's impact on team collaboration and team care coordination.[30,42]

Ethical and professional concerns

Eight studies reported measurements on interaction traits reflecting how AI will impact clinical occupational roles and responsibilities. Three studies measured the interaction trait of efficient integration techniques to ascertain a higher quality of interaction between AI and clinicians.[21,37,46] Clinicians in two studies reported concerns about the uncertainty of how AI may change their clinical practice.[45,49] Other interaction traits that pertain to clinician workplace and occupational changes were measured in one study each including liability concerns,[33] the need for additional training and supervision,[32] a complication of their job,[42] and concerns about job security.[42]

Interview results

A total of six participants were interviewed. All participants were male and between the ages of 27 and 50. One of the participants was a computer scientist / AI content expert and the other five participants were physicians (two surgeons and one each of general internal medicine internist, family physician and hospitalist) with experience in AI. The clinicians' average years of practice was 2.4 years. The partici-

Table 3
Study characteristics and quality of interaction trait frequencies.

Study ID	Year	Geography	Study design	Population	Number of clinicians	Method of evaluating interaction quality	Description of AI System	Evaluated Interaction Traits
Abdulaal et al.	2021	United Kingdom	Mixed-methods study	Physicians; senior house officers; registrars; consultants; primary care GPs	31	Semi structured end user interviews	Artificial neural network that produces patient-specific mortality predictions for COVID-19 and graphical user interface to facilitate the use of the system at the bedside	- User satisfaction - Ease of use - Likelihood of providing surprising predictions - Potential for system performance - Impact on clinical management
Aldughayfiq et al.	2022	Canada, United States, United Kingdom	Qualitative research	Prescribers, pharmacists	284	Web-based questionnaire	ePrescription system that uses machine learning to safely prescribe medication based on patient medication history and health conditions	- Trust in security - Willingness to use - Impact on timesaving - Impact on misinterpretation - Impact on communication with prescribers
Allen et al.	2021	United States	Qualitative research	Radiologists	489	Electronic surveys	Any AI-based system used by the radiologist for breast, thoracic and neurological imaging	- Impact on image interpretation - Assessment of AI performance - Clinicians' willingness to pilot test AI system before adoption
Ankolekar et al.	2022	Netherlands	Mixed-methods study	Radiologists; nurses; pulmonologists	9	Qualitative interviews	AI-enabled clinical decision support system to generate personalised lung cancer treatment decisions	- Impact on value added to treatment decisions - Impact on time saving - Applicability of shared decision-making using the system - Usability of the system - Usefulness of the system - System value for clinicians - System value for patients - Scepticism about the system - Trust
Bajorek et al.	2012	Australia	Qualitative research	Cardiology clinicians; geriatrics clinicians; neurology clinicians, haematology clinicians	27	Structured questionnaire	Computerised risk management system based on developed algorithms to aid decision making regarding antithrombotic therapy in older patients	- Quality of content output - Overall appearance of the system - Content organisation - Quality of system screen layout - Quality of typography - Ease of use - Usefulness - Clinician agreement with system output
Benrimoh et al.	2021	Canada	Qualitative research	Family medicine clinicians; psychiatrists	20	Self-report questionnaires, scenario observations, and interviews	AI-enabled clinical decision support system for the treatment of major depression based on the 2016 Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines for depression treatment	- Helpfulness for patient's understanding - Impact on patient–physician interactions - Impact on patient trust - Trust in system - Clinical usefulness - Impact on quality of information output - Impact on time saving
Calisto et al.	2021	Portugal	Mixed-methods study	Radiologist; medical general interns; surgeons; immunotherapist; oncologists	45	Semi structured interviews	Neural network and deep learning method to support automatic and reliable medical diagnosis workflow and classification of breast images	- Willingness to use - Ease of use - Ease of learning - Degree of need for learning before using system - Need for technical support - Confidence in using the system - Perception of system integration to workflow - Level of consistency in the system's output

(*continued on next page*)

Table 3 (*continued*)

Study ID	Year	Geography	Study design	Population	Number of clinicians	Method of evaluating interaction quality	Description of AI System	Evaluated Interaction Traits
Calisto et al.	2022	Portugal	Mixed-methods study	General clinicians	45	Interviews and observations	The BreastScreening framework that utilises AI-based techniques such as deep learning to offer radiologists an autonomous second reader opinion during the breast cancer diagnosis	- Trust in the system - Acceptability of the system's output - Usability of the system - Understandability of the system - Usefulness of the system - Learnability - User satisfaction - Impact on user's mental and physical demands - Willingness to use - Future potential for use
Carlile et al.	2020	United States	Mixed methods study	Emergency physicians	202	Surveys	Novel deep learning AI algorithm designed to enhance identification of consolidation on chest radiographs	- Ease of use - Impact on medical decision-making
Creed et al.	2022	United States	Mixed-methods study	Psychotherapists	30	Focus group discussions	AI and performance-based feedback fidelity measurement for motivational interviews that uses speech signals from recordings to generate a clinician performance score	- Systems potential utility for supervision of psychotherapists motivational interview performance - How the AI system impacts typical interpersonal interactions and supervisory meetings - System's potential to help with training and education of clinicians - System's potential to promote professional growth - Concerns on how recording-based systems manage non-verbal content such as body language - Impact on rapport between clinicians and patients - Concerns about potential risk related to cultural differences - Organisation's ability to meet technological requirements necessary to host system - Impact on patient privacy - System's impact on clinician's confidence in their practice and techniques - Acceptability - Appropriateness for use - Feasibility
Cheikh et al.	2022	France	Mixed-methods study	Radiologists	79	Survey	AI-based algorithm system to help with diagnostic performance for pulmonary embolism	- Usability of the system
Choudhury et al.	2022	United States	Qualitative research	Physicians; nurses	119	Validated online survey	AI-enabled clinical decision support system that provides standardisation of red blood cell transfusion without compromising organ function	- Usability - Learnability - Ease of use - Trust in the system - Perceived risk of the system - System performance - Impact on efficiency of task performance - Impact on effectiveness
Dontchos et al.	2021	United States	Mixed-methods study	Radiologists	13	Screening mammogram review	Breast Imaging Reporting and Data System (BI- RADS) that utilises deep learning for breast imaging practices	- Acceptance of the system

(*continued on next page*)

Table 3 (*continued*)

Study ID	Year	Geography	Study design	Population	Number of clinicians	Method of evaluating interaction quality	Description of AI System	Evaluated Interaction Traits
Dunsmuir et al.	2008	Canada	Mixed-methods study	Anaesthesiologists	10	Usability questionnaire	AI system that enables clinicians to create knowledge rules without the need of a knowledge engineer or programmer	- Ease of learning - User satisfaction - System information's on-screen organisation - Impact on productivity
Garrett Fernandes et al.	2021	Netherlands	Mixed-methods study	Radiation oncologists	3	Radiologist evaluation	Neural network automatic cardiac contouring algorithm for radiotherapy planning computed tomography images by employing a 3D deep learning model	- Acceptability
Ginestra et al.	2019	United States	Mixed-methods study	Bedside clinicians; nurses	287	Web based questionnaire	Machine learning algorithm to predict severe sepsis or septic shock	- Impact of the system output to lead to additional clinical information - Agreement with the system output - Understandability of the system output - Impact on patient management - Helpfulness of the system - Impact on quality of care - Impact on team communication - Impact on level of patient monitoring - Impact on resource utilisation - Interpretation of system outputs
Goel et al.	2022	India, Australia	Mixed-methods study	General clinicians	30	Radiologist evaluation	Deep learning model that predicts COVID-19 from chest computerised tomography images	- Reliability of the system outputs - Perceived understanding of the system - Trust in the system
Hirsch et al.	2018	United States	Mixed-methods study	Counsellors	21	Interviews	AI and performance-based feedback fidelity measurement for motivational interviews	- Usefulness of the system - Perception of the system layout - Comprehensiveness of the system and results - Need for user training and supervision - Accuracy of the systems results - Opinions of objectivity for the system's outputs - Impact on workplace concerns
Hogue et al.	2021	Canada	Mixed-methods study	Pharmacists	25	Focus group discussions and surveys	Machine learning system to help with identifying atypical medication orders	- Usefulness of the system - Satisfaction - Willingness to use - Perception of effective AI integration - Impact on care - Impact on human staffing needs - Impact on clinician responsibilities for adverse event risks - Impact on the user's professional role and recognition

(*continued on next page*)

Table 3 (*continued*)

Study ID	Year	Geography	Study design	Population	Number of clinicians	Method of evaluating interaction quality	Description of AI System	Evaluated Interaction Traits
Im et al.	2006	United States	Qualitative research	Nurses	122	Questionnaire	Intelligent computer decision assessment support system for dealing effectively with sex and ethnic differences in cancer pain experience	- Perception of system layout and design - Perception of system capabilities - User's reactions to terminology and system information - Impact on clinician learning - Satisfaction with system
Jaber et al.	2022	Lebanon	Mixed-methods study	Psychiatrists	3	Qualitative survey	Explainable AI used in stress prediction based on physiological measurements	- Usefulness of the system - Acceptance of interpretation of the system outputs - Organidation of system output - User's perception on other applications of the system
Jones et al.	2021	Australia	Mixed-methods study	Radiologists	11	Survey	AI algorithm system that utilised machine learning to help detect imaging features on chest X-rays	- Impact on task efficiency - Impact on task accuracy - Impact on user's attitude towards AI in general - System's output is inconsistent - User satisfaction - Willingness to use the system - Ease of use - Need for technical support - Learnability - Confidence using the system
Juluru et al.	2021	United States	Mixed-methods study	Radiologists	14	Survey	AI algorithm for evaluating lymphoscintigraphy examinations	- Ease of use - Impact on task efficiency - Impact on reducing errors - Perception on the format and consistency of the system output - Perception of system integration to clinical workflow
Kim et al.	2022	South Korea	Mixed-methods study	Radiologists; physicians	23	System usability scale	AI deep learning algorithm-based decision support system for chest radiography	- Usability
Kumar et al.	2020	United States	Randomised controlled trial	Physicians	43	Case based studies	Machine learning-based electronic order recommendation system	- Usefulness - Impact on ease of task completion - Impact on user productivity - Impact on efficient use of time - Impact on user job performance
Künzel et al.	2022	Germany	Qualitative research	Radiation oncologists	5	Likert scale questionnaire	Deep learning-based annotation software and AI automatic particle swarm optimisation planning system for contouring	- Agreement of the system output - Agreement on system's automatic generated treatment plan - Satisfaction using the system
Moret-Tatay et al.	2022	Spain	Qualitative research	Medical practitioners; nurses; psychologists; occupational therapists; speech therapists	30	Survey	AI algorithm based virtual assistant for screening cognitive impairment	- Utility of the system - Perception of user experience using the system - Ease of use - Willingness to use

(*continued on next page*)

Table 3 (*continued*)

Study ID	Year	Geography	Study design	Population	Number of clinicians	Method of evaluating interaction quality	Description of AI System	Evaluated Interaction Traits
Romero-Brufau et al.	2020	United States	Qualitative research	Physicians; nurses; clinical assistants; other users	81	Survey	Various AI-enabled clinical decision support systems	- System effectiveness - Impact on management of patient's condition- Impact on care coordination - Impact on patient complications - Beliefs about job security - Perception about AI's ability to understand clinician's job - Perception of familiarity with AI - Perception around excitement about AI
Scheder-Bieschin et al.	2022	Germany	Mixed-methods study	Physicians; nurses	88	Likert scale questionnaire	AI system with adaptive Bayesian reasoning-based techniques to gather relevant symptoms and history for handover to clinicians	- Usefulness - Perception of the system's potential - Impact on rapport with patient - Impact on provision of medically helpful information - Impact on time saving - Clinicians would recommend the system to other clinicians
Scheetz et al.	2021	Australia	Mixed-methods study	Nurses; endocrinologists; ophthalmologists; optometrists; Aboriginal health workers	8	Satisfaction questionnaire	An offline automated AI-assisted model to screen for diabetic retinopathy and age-related macular degeneration	- Ease of use - Ease of interpretability of system output - Need for training to integrate the system - Efficiency of the system - Reliability of the system output - Clinician's trust in the system - User confidence in communicating the system output with patients
Tanguay-Sela et al.	2022	Canada	Mixed-methods study	Psychiatrists; primary care physicians	20	Self-report questionnaires; scenario observations; and interviews	AI-enabled clinical decision support system for the treatment of major depression based on the 2016 Canadian Network for Mood and Anxiety Treatments (CANMAT) guidelines for depression treatment	- Usefulness of the system - Helpfulness of the system - Perceived 'reasonableness' of the system - Trust in the system - User's comfort level with the system - Communicability and interpretability of the system - Impact on treatment decision and clinical practice
Wong et al.	2021	Canada	Mixed-methods study	Radiologists; radiation therapists; dosimetrists; radiation oncologists	203	Post-contouring surveys	Deep learning-based auto-segmented contour models for organs at risk and clinical target volumes	- User satisfaction

(*continued on next page*)

Table 3 (*continued*)

Study ID	Year	Geography	Study design	Population	Number of clinicians	Method of evaluating interaction quality	Description of AI System	Evaluated Interaction Traits
Zhai et al.	2021	China	Mixed-methods study	Radiologists; medical students	307	Questionnaire	AI-assisted contouring system that automates the primary tumour volume and normal tissue for radiation oncologists	- Performance expectancy of system - Usefulness of the system for tasks - Impact on task efficiency - Impact on user productivity - Impact on outcomes of clinician's work - Impact on clinician effort expectancy - Perception of system being clear and understandable - Ease of learning - Ease of use - Colleagues' influence on willingness to use the system - Perception of resources necessary to use the system - Perception of knowledge necessary to use the system - Willingness to use - Perceived risk using the system - Concern for malfunction and performance failure - Perception that more time is needed to fix errors caused by system - Impact on psychological distress on clinicians - Privacy concerns - Behavioural resistance to using the system - Job security concerns - Clinician's intention to use or recommend system - Clinician's ability to override system outputs
Zhang et al.	2022	United States	Mixed-methods study	Physicians, pharmacists	46	Online surveys	AI-enabled clinical decision support software that provides personalised treatment recommendations based on society guidelines for clinicians who treat patients with diabetes	- Impact on patient outcomes - Impact on patient engagement - Impact on physicians' clinical knowledge - Comfort using the system - Impact on change in practice - Concerns about system integration to workflow - Concerns about technical glitches - Concerns that the system uses outdated knowledge sources - Concerns about lack of integration to electronic health records

pant demographic information is listed in Table 6. The interviews provided many of the same interaction traits found in the systematic review, but also generated 15 additional interaction traits. These 15 traits have been listed in Table 7, while also being included in Table 4 and the relevant categorisation results. The interaction traits differed from the ones obtained from the literature by providing perceptions relevant to real-world clinical experiences pertinent to AI in healthcare as a concept in general and not limited to a certain AI tool. No new interaction categories were generated by the interviews.

Discussion

In this systematic review and interview study, we evaluated 34 studies and six semi-structured interviews that reported on the quality of interaction between clinicians and AI-enabled clinical decision support systems to uncover common interaction traits for the AI–clinician quality of interaction constructs. Interaction traits were summarised and arranged into seven discrete categories that represent the AI–clinician quality of interaction construct. These seven categories were usability and user experience, system performance, clinician trust and acceptance, impact on patient care, communication, ethical and professional concerns, and clinician engagement and workflow. Our findings provide a novel taxonomy of interaction traits that can be refined to a comprehensive framework for assessing the quality of interactions between clinicians and AI systems.

There continues to be an increase in research evaluating the role of AI in healthcare, with an emphasis on developing tools that can improve patient outcomes. AI research has frequently demonstrated acceptable

Table 4
Judgmental categorization of the interaction traits from the included studies.

Interaction item	Traits from systematic review and qualitative study
Usability and user experience	- User satisfaction - Ease of use - Quality of content output - Overall appearance of the system - Content organisation - Quality of system screen layout - Quality of typography - Ease of learning/learnability - Confidence in using the system - Usability - Understandability of the system - Impact on user's mental and physical demands - Need for user training and supervision - Impact on clinician learning - User's perception on other applications of the system - Impact on ease of task completion - Perception of user experience using the system - User's comfort level with the system - Communicability and interpretability of the system - Clinician's ability to override system outputs - Perception of knowledge necessary to use the system - *System accessibility*[*]
System performance	- Likelihood of providing surprising predictions - Assessment of AI performance - Applicability of shared decision-making using the system - Impact on quality of information output - Level of consistency in the system's output - Accuracy of the systems results - Impact on task efficiency - Impact on reducing errors - Impact on timesaving - Performance expectancy of system - Concerns about technical glitches - Perception of system capabilities
Trust and acceptance	- Trust in security - Acceptability of the system's output - Agreement with system output - Concerns on how recording-based systems manage non-verbal content such as body language - Concerns about potential risk related to cultural differences - Impact on patient privacy - Perceived risk of the system - Reliability of the system outputs - Opinions of objectivity for the system's outputs - Impact on user's attitude towards AI in general - Perception of familiarity with AI - Perception around excitement about AI - Concerns that the system uses outdated knowledge sources - *Training data transparency*[*] - *Concerns about system discriminatory behaviour*[*]
System impact on patient care	- Impact on clinical management - System value for patients - Helpfulness for patient's understanding - Impact on patient–physician interactions - Impact on patient trust - Impact on medical decision-making - Impact on quality of care - Impact on level of patient monitoring - Impact on patient complications - Impact on patient outcomes - Impact on patient engagement - *Patient education on AI*[*]

(*continued on next page*)

Table 4 (*continued*)

Clinician engagement and workflow	- Willingness to use - Impact on image interpretation - Impact on value added to treatment decisions - Impact on time saving - Usefulness of the system - Perception of system integration to workflow - Future potential for use - System's potential to help with training and education of clinicians - System's potential to promote professional growth - Organisation's ability to meet technological requirements necessary to host system - Appropriateness for use - Feasibility - Impact on efficiency of task performance - Impact on productivity - Impact of the system output to lead to additional clinical information - Helpfulness of the system - Impact on resource utilisation - Impact on task accuracy - Impact on user job performance - Impact on provision of medically helpful information - Clinicians would recommend the system to other clinicians - Impact on clinician effort expectancy - Impact on change in practice - *Concerns about inappropriate triaging** - *Dependency on AI systems** - *AI must provide novel information that clinicians would not know** - *Financial requirements of the system**
Communication and collaboration	- How the AI system impacts typical interpersonal interactions and supervisory meetings - Impact on team communication - Impact on care coordination - *System's social impact** - *Changes to medical consultations** - *Impact on number of clinician–clinician interactions**
Ethical and professional concerns	- Impact on clinician responsibilities for adverse event risks - Impact on the user's professional role and recognition - Impact on human staffing needs - Beliefs about job security - Perception about AI's ability to understand clinician's job - Impact on psychological distress on clinicians - Behavioural resistance to using the system - *Lack of standards and guidelines** - *Liabilities for patient complications** - *AI causes additional work for clinicians** - *Analysing data that patients did not consent to**

* Interaction traits obtained from the interviews.

Table 5
Proposed framework for the AI-clinician quality of interaction construct.

Interaction item	Definition
Usability and user experience	This includes characteristics related to the ease of use, learnability and system complexity.
System performance	This includes characteristics related to the system's performance, such as accuracy of the system's results, responsiveness of the system and comprehensiveness of the results.
Trust and acceptance	This includes characteristics related to clinician's trustworthiness and acceptance of the system, such as trust in the system, perceived risk and perceived reliability.
System impact on patient care	This includes characteristics related the quality of care for patients, such as patient–physician interaction, patient engagement, patient monitoring and patient management.
Clinician engagement and workflow	This includes characteristics related to the engagement of AI systems in practice, such as willingness to use, effective integration and AI impact on training and supervision.
Communication and collaboration	This includes characteristics related to clinician communication and collaboration, such as improving communication among clinicians, improving team communication and increasing patient monitoring.
Ethical and professional concerns	This includes characteristics related to concerns clinicians may have about the adoption of any AI system, such as AI impact on practice, job security concerns and clinician responsibilities for AI-caused adverse events.

levels of accuracy and performance metrics for healthcare AI systems. A study by Zeltzer *et al*. sought to evaluate the diagnostic accuracy of AI-generated diagnoses. The study results demonstrated that sampled providers accepted over 80% of AI diagnoses in virtual care. This is one example of how AI's appropriate accuracy and responsiveness may improve healthcare practices.[50] Although various studies provide the accuracy measurements of the AI system, the gap in the literature for how AI developers address the preferences of clinicians when integrating a new tool needs to be addressed. The proposed seven categories may provide a representation of the user's experiences when using an AI system. AI and clinicians' effective collaboration relies on the clinician's technical and conceptual skills in both technology and healthcare.[51] This re-

Table 6
Demographic table for the interview participants.

		Total
N		6
	% Female	0 (0)
Age, N (%)		
	< 35	4 (66.6)
	> 35	2 (33.3)
	Range (mean)	27, 50 (33)
Field of work, N (%)		
	Surgeon	2 (33.3)
	General internal medicine	1 (16.7)
	Hospitalist doctor	1 (16.7)
	Family medicine	1 (16.7)
	AI expert	1 (16.7)
Years of clinical practice*, N (%)		
	< 5	4 (80.0)
	> 5	1 (20.0)
	Range (mean)	1, 8 (2.4)

*Category is only reporting on the clinician participants.

Table 7
Additional interaction traits obtained from the semi-structured interviews and their respective interviews.

Interview number	New interaction trait
Interview #1	- Training data transparency - Lack of standards and guidelines - System accessibility - System's social impact
Interview #2	- Concerns about inappropriate triaging - Concerns about system discriminatory behaviour - Patient education on AI - Liabilities for patient complications
Interview #3	- Changes to medical consultations - Impact on number of clinician–clinician interactions - Dependence on AI systems
Interview #4	- Financial requirements of the system
Interview #5	- AI causes additional work for clinicians - Analysing data that patients did not consent to - AI must provide novel information that clinicians would not know
Interview #6	- N/A

view provides the first synthesised knowledge of the foundational components to evaluating a clinician's first-person experience with an AI system.

Clinicians play a pivotal role in the implementation and acceptance of AI systems in healthcare. A low level of trust in AI is currently one of the most prominent reasons for the lack of integration into the healthcare system.[4] As healthcare providers are expected to be exposed to AI systems more frequently, improving the average trust level clinicians have for AI in general is required to improve AI integration and shared decision-making. AI systems and patients would both employ clinicians as intermediaries to provide insights into patient needs and optimal treatment plans.[52] However, the modification of the patient–physician interaction to include AI cannot be accomplished without the willingness to use and acceptance of AI by clinicians.[53] Developing a common language to evaluate the interplay between clinicians and AI can benefit healthcare providers, AI scientists, developers and researchers. Providing a structured tool empowers clinicians to evaluate AI systems and their utility, improves their confidence in decision-making processes, promotes a stronger sense of teamwork between clinicians and AI-generated outputs, and helps overcome potential biases inherent in AI outputs.[54]

AI systems require interactions with human clinicians to have an impact on healthcare. A well-defined framework for the AI–clinician quality of interaction would catalyse successful collaboration between clinicians and AI systems. Usability scales such as the Health IT usability evaluation scale have demonstrated that when clinicians use AI systems, there were improvements in communication, resource management, and time saving. This positive impact is likely more pronounced when AI systems and human clinicians share a common language and set of criteria to improve user experience and refine system performance.[55] We argue that a standardised framework that measures the AI–clinician quality of interaction will promote a similar positive impact and provide clinicians with the opportunity to be an active contributor to AI development, reinforcing their continued importance in healthcare processes.

The standardisation of language used to describe the AI–clinician quality of interaction may help communication between clinicians and AI developers. An example of a framework that provides a standardised language for a clinical entity is the Dindo–Clavien classification of postoperative complications.[56] Prior to the advent of a standardised classification, there was a lack of consensus on how to report surgical complications, which hindered the progress of surgical outcomes research and finding ways to improve safety. Once the Dindo–Clavien classification was well validated and accepted, it helped streamline the reporting and comparison of postoperative complications by the surgical communities.[56] By standardising the quality of interaction construct, healthcare providers can make more informed decisions about their AI usage by having a quantitative method of presenting their first-person experiences with an AI system.

This systematic review has some limitations. First, this review did not include studies published in languages other than English and studies on AI systems from non-indexed journals. Second, the quality assessment utilised the CASP checklist to evaluate only the qualitative aspects of all included studies regardless of study type, which may have provided limited study validity for non-qualitative studies. Third, although thematic saturation was achieved and strong convergence between the review and interview results was found, a small sample size with limited participant variability was utilised for the interview portion of this study. Finally, our study was formative, and the categories and their components were generated using a subjective process. Additional research with a larger sample of content experts and end-users is needed to further refine categories and their components.

Conclusion

From 34 studies and six semi-structured interviews with clinical and content experts, we identified 90 unique interaction traits representing the quality of interaction between AI systems and clinicians. From these interaction traits, we were able to define seven categories, which were: usability and user experience, system performance, clinician trust and acceptance, impact on patient care, communication, ethical and professional concerns, and clinician engagement and workflow. Further research is needed to use the taxonomy of interaction traits identified in this review to develop and validate a standardised tool that can be used to comprehensively evaluate the quality of interaction between clinicians and AI systems in healthcare settings.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Argyrios Perivolaris: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Chris Adams-McGavin:** Writing – review & editing, Methodology, Conceptualization. **Yasmine Madan:** Writing – review & editing, Methodology, Data curation. **Teruko Kishibe:** Writing – review & editing, Methodology, Conceptualization. **Tony Antoniou:** Writing – review & editing, Supervision, Methodology, Data curation. **Muhammad Mamdani:** Writing – review & editing, Supervision, Methodology, Data curation. **James J. Jung:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Data curation, Conceptualization.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.fhj.2024.100172.

References

1. Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *J Ambient Intell Humaniz Comput*. 2023;14(7):8459–8486.
2. Shen J, Zhang CJP, Jiang B, *et al.* Artificial intelligence versus clinicians in disease diagnosis: systematic review. *JMIR Med Inform*. 2019;7(3):e10010.
3. Lee S, Kim HS. Prospect of artificial intelligence based on electronic medical record. *J Lipid Atheroscler*. 2021;10(3):282–290.
4. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. *J Med Internet Res*. 2020;22(6):e15154.
5. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Med Ethics*. 2021;22(1):122.
6. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. *JMIR Med Educ*. 2019;5(2):e16048.
7. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25(1):44–56.
8. Dahlin E. Mind the gap! On the future of AI research. *Humanit Soc Sci Commun*. 2021;8(2021):71.
9. Pagliari M, Chambon V, Berberian B. What is new with Artificial Intelligence? Human-agent interactions through the lens of social agency. *Front Psychol*. 2022;13:954444.
10. Higgins JPT, Thomas J, Chandler J, et al. (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.3 (updated February 2022). Cochrane, 2022.
11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev*. 2021;10:89.
12. McGowan J. *Library Services: Artificial Intelligence and Healthcare*. Toronto: St Michael's Unity Health; 2022.
13. Covidence Systematic Review Software, Veritas Health Innovation, Melbourne, Australia. Available at www.covidence.org.
14. Critical Appraisal Skills Programme (2018). CASP qualitative checklist.
15. Abdulaal A, Patel A, Al-Hindawi A, *et al.* Clinical utility and functionality of an artificial intelligence-based app to predict mortality in COVID-19: mixed methods analysis. *JMIR Form Res*. 2021;5(7):e27992.
16. Aldughayfiq B, Sampalli S. Patients', pharmacists', and prescribers' attitude toward using blockchain and machine learning in a proposed ePrescription system: online survey. *JAMIA open*. 2022;5(1):ooab115.
17. Allen B, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 ACR data science institute artificial intelligence survey. *J Am Coll Radiol*. 2021;18(8):1153–1159.
18. 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 Expect*. 2022;25(4):1342–1351.
19. Bajorek BV, Masood N, Krass I. Development of a computerised antithrombotic risk assessment tool (CARAT) to optimise therapy in older persons with atrial fibrillation. *Australas J Ageing*. 2012;31(2):102–109.
20. Benrimoh D, Tanguay-Sela M, Perlman K, *et al.* Using a simulation centre to evaluate preliminary acceptability and impact of an artificial intelligence-powered clinical decision support system for depression treatment on the physician-patient interaction. *BJPsych Open*. 2021;7(1):e22.
21. Calisto FM, Santiago C, Nunes N, Nascimento JC. Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification. *International Journal of Human – Computer Studies*. 2021;150:102607.
22. Calisto FM, Santiago C, Nunes N, Nascimento JC. BreastScreening-AI: Evaluating medical intelligent agents for human-AI interactions. *Artif Intell Med*. 2022;127:102285.
23. Carlile M, Hurt B, Hsiao A, Hogarth M, Longhurst CA, Dameff C. Deployment of artificial intelligence for radiographic diagnosis of COVID-19 pneumonia in the emergency department. *J Am Coll Emerg Physicians Open*. 2020;1(6):1459–1464.
24. Cheikh AB, Gorincour G, Nivet H, *et al.* How artificial intelligence improves radiological interpretation in suspected pulmonary embolism. *Eur Radiol*. 2022;32(9):5831–5842.
25. Choudhury A, Asan O, Medow JE. Effect of risk, expectancy, and trust on clinicians' intent to use an artificial intelligence system – Blood Utilization Calculator. *Appl Ergon*. 2022;101:103708.
26. Creed TA, Kuo PB, Oziel R, *et al.* Knowledge and attitudes toward an artificial intelligence-based fidelity measurement in community cognitive behavioral therapy supervision. *Adm Policy Ment Health*. 2022;49(3):343–356.
27. Dontchos BN, Yala A, Barzilay R, Xiang J, Lehman CD. External validation of a deep learning model for predicting mammographic breast density in routine clinical practice. *Acad Radiol*. 2021;28(4):475–480.
28. Dunsmuir D, Daniels J, Brouse C, Ford S, Ansermino JM. A knowledge authoring tool for clinical decision support. *J Clin Monit Comput*. 2008;22(3):189–198.
29. Garrett Fernandes M, Bussink J, Stam B, *et al.* Deep learning model for automatic contouring of cardiovascular substructures on radiotherapy planning CT images: dosimetric validation and reader study based clinical acceptability testing. *Radiother Oncol*. 2021;165:52–59.
30. Ginestra JC, Giannini HM, Schweickert WD, *et al.* Clinician perception of a machine learning-based early warning system designed to predict severe sepsis and septic shock. *Crit Care Med*. 2019;47(11):1477–1484.
31. Goel K, Sindhgatta R, Kalra S, Goel R, Mutreja P. The effect of machine learning explanations on user trust for automated diagnosis of COVID-19. *Comput Biol Med*. 2022;146:105587.
32. Hirsch T, Soma C, Merced K, *et al.* It's hard to argue with a computer:" investigating psychotherapists' attitudes towards automated evaluation. *DIS (Des Interact Syst Conf)*; 2018:559–571.
33. Hogue SC, Chen F, Brassard G, *et al.* Pharmacists' perceptions of a machine learning model for the identification of atypical medication orders. *J Am Med Inform Assoc*. 2021;28(8):1712–1718.
34. Im EO, Chee W. Nurses' acceptance of the decision support computer program for cancer pain management. *Comput Inform Nurs*. 2006;24(2):95–104.
35. Jaber D, Hajj H, Maalouf F, El-Hajj W. Medically-oriented design for explainable AI for stress prediction from physiological measurements. *BMC Med Inform Decis Mak*. 2022;22(1):38.
36. Jones CM, Danaher L, Milne MR, *et al.* Assessment of the effect of a comprehensive chest radiograph deep learning model on radiologist reports and patient outcomes: a real-world observational study. *BMJ Open*. 2021;11(12):e052902.
37. Juluru K, Shih HH, Keshava Murthy KN, *et al.* Integrating Al algorithms into the clinical workflow. *Radiol Artif Intell*. 2021;3(6):e210013.
38. Kim EY, Kim YJ, Choi WJ, *et al.* Concordance rate of radiologists and a commercialised deep-learning solution for chest X-ray: real-world experience with a multicenter health screening cohort. *PLoS ONE*. 2022;17(2):e0264383.
39. Kumar A, Aikens RC, Hom J, *et al.* OrderRex clinical user testing: a randomized trial of recommender system decision support on simulated cases. *J Am Med Inform Assoc*. 2020;27(12):1850–1859.
40. Künzel LA, Nachbar M, Hagmüller M, *et al.* Clinical evaluation of autonomous, unsupervised planning integrated in MR-guided radiotherapy for prostate cancer. *Radiother Oncol*. 2022;168:229–233.
41. Moret-Tatay C, Radawski HM, Guariglia C. Health professionals' experience using an azure voice-bot to examine cognitive impairment (WAY2AGE). *Healthcare (Basel)*. 2022;10(5):783.
42. Romero-Brufau S, Wyatt KD, Boyum P, Mickelson M, Moore M, Cognetta-Rieke C. A lesson in implementation: a pre-post study of providers' experience with artificial intelligence-based clinical decision support. *Int J Med Inform*. 2020;137:104072.
43. Scheder-Bieschin J, Blümke B, de Buijzer E, *et al.* Improving emergency department patient-physician conversation through an artificial intelligence symptom-taking tool: mixed methods pilot observational study. *JMIR Form Res*. 2022;6(2):e28199.
44. Scheetz J, Koca D, McGuinness M, *et al.* Real-world artificial intelligence-based opportunistic screening for diabetic retinopathy in endocrinology and indigenous healthcare settings in Australia. *Sci Rep*. 2021;11(1):15808.
45. Tanguay-Sela M, Benrimoh D, Popescu C, *et al.* Evaluating the perceived utility of an artificial intelligence-powered clinical decision support system for depression treatment using a simulation center. *Psychiatry Res*. 2022;308:114336.
46. Wong J, Huang V, Wells D, *et al.* Implementation of deep learning-based auto-segmentation for radiotherapy planning structures: a workflow study at two cancer centers. *Radiat Oncol*. 2021;16(1):101.
47. Zhai H, Yang X, Xue J, *et al.* Radiation oncologists' perceptions of adopting an artificial intelligence-assisted contouring technology: model development and questionnaire study. *J Med Internet Res*. 2021;23(9):e27122.
48. Zhang H, Huang W. Joint deep-learning-enabled impact of holistic care on line coagulation in hemodialysis. *J Healthc Eng*. 2021;2021:3413692.
49. Zhang X, Svec M, Tracy R, Ozanich G. Clinical decision support systems with team-based care on type 2 diabetes improvement for Medicaid patients: a quality improvement project [published online ahead of print, 2021 Nov 18]. *Int J Med Inform*. 2021;158:104626.
50. Zeltzer D, Herzog L, Pickman Y, *et al.* Diagnostic accuracy of artificial intelligence in virtual primary care. *Mayo Clinic Proceedings: Digital Health*. 2023:480–489.
51. Zirar A, Ali SI, Islam N. Artificial Intelligence (AI) coexistence: emerging themes and research agenda. *Technovation*. 2023;124:102747.
52. Snyder CF, Wu AW, Miller RS, Jensen RE, Bantug ET, Wolff AC. The role of informatics in promoting patient-centered care. *Cancer J*. 2011;17(4):211–218.
53. Lorenzini G, Arbelaez Ossa L, Shaw DM, Elger BS. Artificial intelligence and the doctor-patient relationship expanding the paradigm of shared decision making. *Bioethics*. 2023;37(5):424–429.
54. Zou J, Schiebinger L. AI can be sexist and racist - it's time to make it fair. *Nature*. 2018;559(7714):324–326.
55. Yen PY, Sousa KH, Bakken S. Examining construct and predictive validity of the Health-IT Usability Evaluation Scale: confirmatory factor analysis and structural equation modeling results. *J Am Med Inform Assoc*. 2014;21(e2):e241–e248.
56. Dindo D, Demartines N, Clavien PA. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg*. 2004;240(2):205–213.