Review Article

Factors influencing trust in medical artificial intelligence for healthcare professionals: a narrative review

Victoria Tucci[1]^, Joan Saary[2,3], Thomas E. Doyle[4,5,6]

[1]Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada; [2]Division of Occupational Medicine, Department of Medicine, University of Toronto, Ontario, Canada; [3]Canadian Forces Environmental Medicine Establishment, Toronto, Ontario, Canada; [4]Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada; [5]School of Biomedical Engineering, McMaster University, Hamilton, Ontario, Canada; [6]Vector Institute of Artificial Intelligence, Toronto, Ontario, Canada

Contributions: (I) Conception and design: All authors; (II) Administrative support: All authors; (III) Provision of study materials or patients: All authors; (IV) Collection and assembly of data: V Tucci; (V) Data analysis and interpretation: V Tucci; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Thomas E. Doyle, PhD. Engineering Technology Building ETB-106, McMaster University, 1280 Main Street West, Hamilton, ON L8S 0A3, Canada. Email: doylet@mcmaster.ca.

Objective: We performed a comprehensive review of the literature to better understand the trust dynamics between medical artificial intelligence (AI) and healthcare expert end-users. We explored the factors that influence trust in these technologies and how they compare to established concepts of trust in the engineering discipline. By identifying the qualitatively and quantitatively assessed factors that influence trust in medical AI, we gain insight into understanding how autonomous systems can be optimized during the development phase to improve decision-making support and clinician-machine teaming. This facilitates an enhanced understanding of the qualities that healthcare professional users seek in AI to consider it trustworthy. We also highlight key considerations for promoting on-going improvement of trust in autonomous medical systems to support the adoption of medical technologies into practice.

Background: Artificially intelligent technology is revolutionizing healthcare. However, lack of trust in the output of such complex decision support systems introduces challenges and barriers to adoption and implementation into clinical practice.

Methods: We searched databases including, Ovid MEDLINE, Ovid EMBASE, Clarivate Web of Science, and Google Scholar, as well as gray literature, for publications from 2000 to July 15, 2021, that reported features of AI-based diagnostic and clinical decision support systems that contribute to enhanced end-user trust. Papers discussing implications and applications of medical AI in clinical practice were also recorded. Results were based on the quantity of papers that discussed each trust concept, either quantitatively or qualitatively, using frequency of concept commentary as a proxy for importance of a respective concept.

Conclusions: Explainability, transparency, interpretability, usability, and education are among the key identified factors thought to influence a healthcare professionals' trust in medical AI and enhance clinician-machine teaming in critical decision-making healthcare environments. We also identified the need to better evaluate and incorporate other critical factors to promote trust by consulting medical professionals when developing AI systems for clinical decision-making and diagnostic support.

Keywords: Artificial intelligence (AI); trust; healthcare; medical technology adoption; decision support system

Received: 25 August 2021; Accepted: 25 November 2021; Published: 30 March 2022.

doi: 10.21037/jmai-21-25

View this article at: https://dx.doi.org/10.21037/jmai-21-25

^ ORCID: 0000-0002-2344-2560.

Introduction

Rapid development of healthcare technologies continues to transform medical practice (1). The implementation of artificial intelligence (AI) in clinical settings can augment clinical decision-making and provide diagnostic support by translating uncertainty and complexity in patient data into actionable suggestions (2). Nevertheless, the successful integration of AI-based technologies as non-human, yet collaborative members of a healthcare team, is largely dependent upon other team users' trust in these systems. Trust is a concept that generally refers to one's confidence in the dependability and reliability in someone or something (3). We refer to AI as a computer process that algorithmically makes optimal decisions based on criteria utilizing machine learning-based models (2). Although trust is fundamental to influencing acceptance of AI into critical decision-making environments, it is often a multidimensional barrier that contributes to hesitancy and skepticism in AI adoption by healthcare providers (4).

Historically, the medical community has sometimes demonstrated resistance to the integration of technology into practice. For instance, the adoption of electronic health records was met with initial resistance in many locales. Identified barriers included cost, technical concerns, security and privacy, productivity loss and workflow challenges, among others (5). Therefore, we could expect that similar factors may contribute to the issues related to lack of trust in medical AI, which is also considered a practice-changing technology.

It is imperative to elucidate the factors that impact trust in the output of AI-based clinical decision and diagnostic support systems; enhancing end-user trust is crucial to facilitating successful human-machine teaming in critical situations, especially when patient care may be impacted. Ahuja *et al.* [2019] summarize numerous studies that have assessed the importance of optimizing medical AI systems to enhance teaming and interactions with clinician users (6). Similarly, Jacovi *et al.* [2021] outline several concepts established in the discipline of engineering that are well understood to contribute to increased end-user trust in AI systems, including, but not limited to, interpretability, explainability, robustness, transparency, accountability, fairness, and predictability (7). However, in contrast to the engineering literature, there appears to be a gap in the medical literature regarding exploration of the specific factors that contribute to enhanced trust in medical AI amongst healthcare providers.

We, therefore, wished to address this gap by performing a review of the literature to better understand the trust dynamics between healthcare professional end-users and medical AI systems. Further, we sought to explore the key factors and challenges that influence end-user trust in the output of decision support technologies in clinical practice and compare the identified factors to established concepts of trust in the domain of engineering. We recognize the challenges in AI research regarding inconsistency and lack of universally accepted definitions of key trust concepts. Since these are terms that are relevant for understanding the concept of trust, we accepted that terminology may be applied inconsistently throughout the literature. The aim of this paper is to delineate the qualitatively and quantitatively assessed factors that influence trust in medical AI to better understand how autonomous systems can be optimized to improve both decision-making support and clinician-machine teaming. A quantitative summary of the discourse in the healthcare community regarding factors that influence trustworthiness in medical AI will provide AI researchers and developers relevant input to better direct their work and make the outcomes more clinically relevant. We specifically focus on healthcare professionals as the primary end-users of medical AI systems and highlight challenges related to trust that should be considered during the development of AI systems for clinical use.

A literature review is appropriate at this time as there currently does not exist a comprehensive consolidation of available literature on this topic. As such, this review is an important primary step to synthesize the current literature and consolidate what is already known about trust in medical AI amongst healthcare professionals. Since this has not been previously performed, it will enable identification of knowledge gaps and contribute to further understanding by summarizing relevant evidence. By first collating information in the form of a literature review, this paper describes the breadth of available research and provides a foundational contribution to an eventual evidence-based conceptual understanding. We aim to facilitate an understanding of the landscape in which this information applies to make valuable contributions to medical AI. By elucidating the discourse in the medical community regarding key factors related to trust in AI, this paper also provides the foundation for further research into the adoption of medical AI technologies, as well as highlights key considerations for promoting trust in autonomous medical systems and enhancing their capabilities to support healthcare professionals. We present the following article in accordance with the Narrative Review reporting

checklist (available at https://jmai.amegroups.com/article/view/10.21037/jmai-21-25/rc).

Methods

Search strategy and selection criteria

A comprehensive review of the literature was conducted to identify studies describing factors that contribute to trustworthy medical AI according to healthcare providers. To evaluate this phenomenon, we reviewed articles that reported a quantitative or qualitative delineation of the factors that impact trustworthiness in medical AI for the focus population (past or currently practicing certified healthcare professionals). Both peer-reviewed and pre-print articles published or available between 2000 to July 15, 2021, were eligible for inclusion. Studies of interest included those that reported features of AI-based diagnostic and clinical decision support systems that contribute to enhanced end-user trust. Papers discussing implications and applications of medical AI in clinical practice were also recorded.

Relevant articles were identified by searching databases including Ovid MEDLINE, Ovid EMBASE, Clarivate Web of Science, and Google Scholar. Grey literature sources identified through the Google search engine, medRxiv pre-print server, and medical organization statements and repositories, were manually searched for additional relevant information.

Search terms included "machine intelligence", "computer intelligence", "cognitive computing", "robot*", "expert system*", "intelligent system*", "autonomous agent*", "autonomous medical advisory system*", "artificial intelligence", "artificial* intelligen*", "decision support", combined with "trust*", "perspective*", "opinion*", "thought*", "comfort*", "perception*", "barrier*", "adopt*", "physician*", "doctor*", "clinician*", "practitioner*", "healthcare professional*", "health care professional*". The search terms were used in combination with MeSH terms ('artificial intelligence', 'trust', 'physicians') to extend the comprehensiveness of the literature searches. Additional references were identified from the citations of relevant articles. Only papers published in English were reviewed. We did not exclude articles based on study design. The final reference list was generated based on relevance to the scope of this review.

Definitions of key terms

We had suspicion that because our team is interdisciplinary, topics in one field, like medicine, may be presented or defined differently than concepts in another discipline like engineering, despite appearing to have the same meaning. Therefore, we also undertook a preliminary comparison of the definitions of several common AI trust topics between these two disciplines (medicine *vs.* engineering). We did not undertake this process for every concept; going forward, it is important for researchers in both fields to be cognizant of potential discrepancies in the meaning of similar terminologies. For example, according to healthcare professionals, explainability tends to be conceptualized as providing information about the mechanisms regarding how decisions are generated (2). Engineering adds the specific ability to reveal the relevant inputs and signals in the AI reasoning process (7).

In medicine, transparency generally entails understanding why certain decisions are made and which factors drive outputted recommendations in order for users to assess the logic of the model and understand the applicability of the data to their patients (8). This aligns reasonably well with the definition of this concept in engineering, where it is viewed as a dimension of explainability (9). In both healthcare and engineering, interpretability is understood as the ability of a user to understand the connection between the features extracted by an AI program and its output (10).

Study information extraction, recording, and concept mapping

Data extracted from studies under consideration for inclusion were: concepts of trust in medical AI, study type (quantitative or qualitative), category of the concepts considered (i.e., a concept of trust or an implication/application of AI in clinical settings), and concepts that were specific to certain medical specialties. It was noted that select papers discussed factors that are generally understood as being relevant to trust enhancement in AI yet were not examined in the specific and direct context of trust in the respective article. These studies were flagged, yet still considered for inclusion because the discussed factors were established concepts known to impact trust in AI technology. As such, they contribute to an understanding of the features that healthcare professionals seek in trustworthy AI.

Recorded articles that were considered for inclusion were stratified by the type of information provided in terms of quantitative data (i.e., primary research-based studies that included participant cohorts and papers in which the purpose of the research was analytically examining

medical AI trust concepts), or qualitative information (i.e., studies that did not involve participants, including, but not limited to, review papers and perspective articles). There is a substantial quantity of grey literature that discusses these topics from an opinion-based standpoint, however, we were particularly interested in examining the literature that attempted to quantify or systematically examine these concepts qualitatively. We acknowledge that there is a significant amount of opinionated conversation that is neither quantifiable nor systematically examined. The total number of papers that discussed each recorded concept were hand-counted and a single reviewer recorded data that was then verified by two independent second reviewers.

When completing full-article reviews, we scanned for explicit terms relating to AI trust concepts. The 'Methods' section of included papers was reviewed and mention of key themes that are considered contributory to enhanced trust in medical AI (i.e., transparency, explainability, robustness, etc.), whether implicit or explicit, were recorded. In papers mentioning multiple themes, each was individually recorded. Implicit/alternative definitions, testimonials from medical professionals, and examples were mapped to explicit and commonly understood/established concepts that relate to trust in AI via a method of thematic synthesis. The implicit definitions were thus identified as, and recorded according to, the concept they were mapped to. Only implicit AI trust concepts (i.e., those that were alluded to in articles via direct participant quotes or indirect definitions, etc.) were mapped to an explicit definition so they could be accounted for in our quantitative analysis. These concepts were mapped by extracting the alternative/implicit definition from the respective article and matching it to the most appropriate, known AI trust concept to which it best aligns; the interested reader can find these details in Table S1 online. Result quantification was based upon prevalence i.e., the quantity of papers that discussed each trust concept. This paper uses frequency of concept commentary as a surrogate for importance of a respective topic. This approach demonstrates highly investigated items; if researchers and funding are directed to select areas at the expense of others, it can be assumed that there is a perceived correlation between frequency of investigation and importance of a topic. However, we acknowledge that there may exist other topics judged by other means to be more important yet discussed less frequently, and thus not fully reflected in the results.

Analysis

The initial literature searches generated a total of 194 potential articles. After eliminating duplicated articles from database and search engine results, as well as seven non-relevant studies with publication dates beyond the inclusion range, 147 relevant articles remained and were recorded to be considered for inclusion. The abstracts of these articles were then reviewed in detail to identify those that related to factors that contribute to enhanced trust in medical AI. Studies that were not related to healthcare professionals and medical AI, as well as those that discussed surgical robotics rather than AI clinical decision-making and/or diagnostic support, resulted in further elimination of 29 papers. As a result, 111 full-text articles were comprehensively assessed for eligibility. Those that did not discuss factors that contribute to medical professionals' trust in AI, but rather explored AI trust concepts at a high-level and discussed general perceptions, resulted in the exclusion of 26 papers. Further, three studies that discussed patients' trust in AI, as opposed to healthcare professionals' trust in medical AI, were also eliminated. Lastly, papers that did not delineate the contributing factors to trust in AI in medical settings resulted in the further exclusion of five papers. The remaining 77 articles that qualitatively or quantitatively assessed applications/implications of medical AI, and/or factors that contribute to trust in medical AI, were evaluated in detail to identify and record this information for future reference. However, since the focus of this paper is on concepts of trust in medical AI, 20 articles that exclusively discussed applications/implications of such technology were eliminated. There were 13 articles that discussed both applications/implications as well as concepts of trust in AI, and were thus included. In total, there were 57 remaining articles that were retained for inclusion in the final summary and analysis (see *Figure 1*).

Results

We aimed to highlight the medical AI trust concepts/factors that are most frequently discussed amongst healthcare professionals by using frequency of discussed topics as a proxy for importance. The concepts were categorized according to whether they were qualitatively and/or quantitatively evaluated in the respective articles in which they were investigated. The data is graphically displayed in

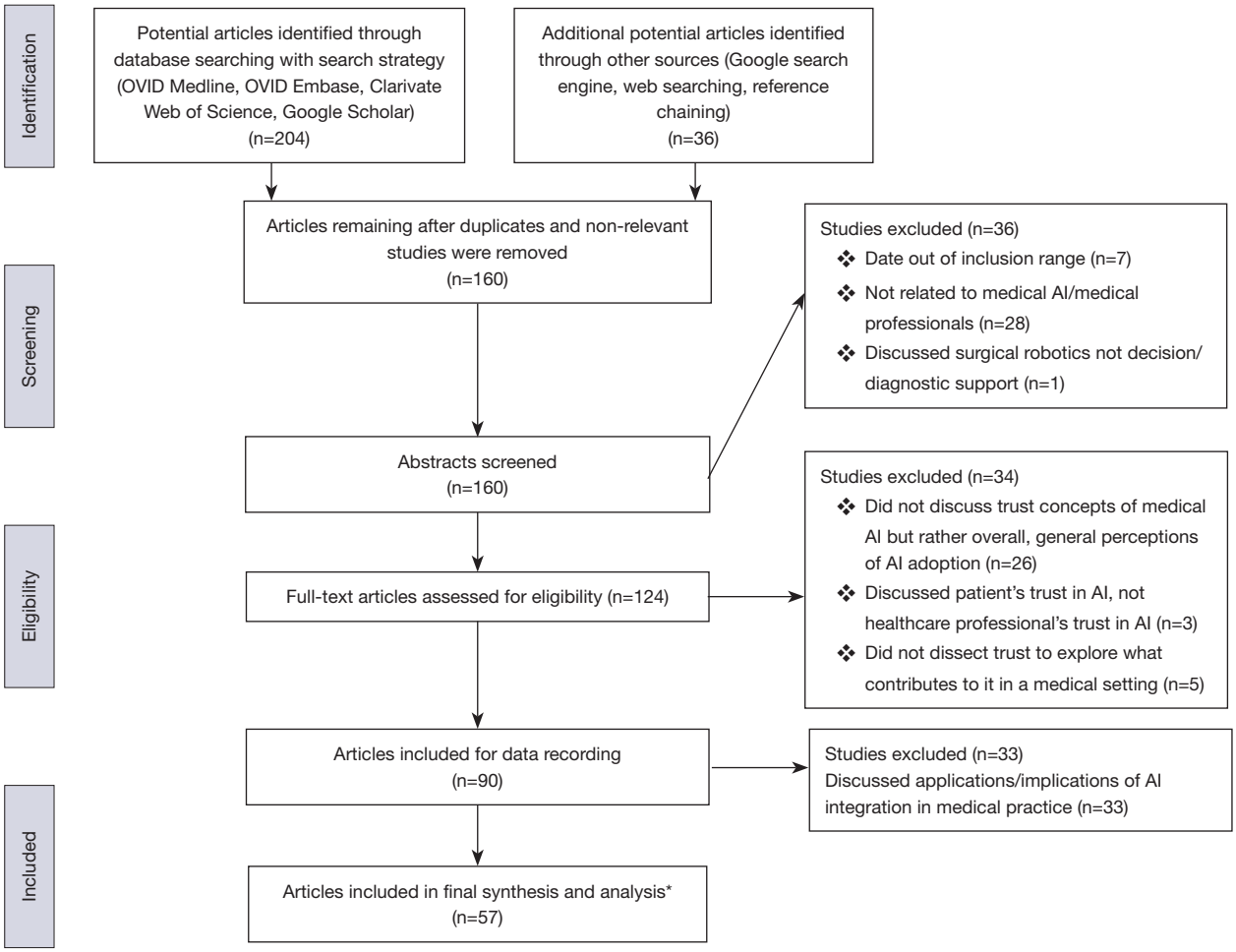

Figure 1 PRISMA flow diagram for study screening and inclusion/exclusion procedure. *, select papers included in the final synthesis and analysis discussed both applications/implications of medical AI, as well as factors that contribute to trust in medical AI, and were thus included in this total. AI, artificial intelligence.

two distinct presentation formats to highlight comparisons between the number and type of AI trust concepts that were quantified compared to those that were presented qualitatively in the literature (see *Figures 2,3*).

A quantitative presentation of the number of papers that discuss AI trust-related concepts identifies the areas of current focus of research endeavor. Identifying the key factors and how commonly each are found thereby highlights the most common as well as under-represented (or possibly overlooked) concepts. Overall, there were a total of 42 factors identified that contribute to enhanced end-user trust in medical AI systems. Of these 42 factors, 10 were solely analyzed quantitatively, 16 were solely analyzed qualitatively, and 16 were examined both qualitatively and quantitatively. Note that all percentages are based on the total number of articles identified.

There were nine trust concepts that were consistently identified through both qualitative and quantitative methodologies, however, were more frequently analyzed qualitatively, including:

(I) Complexity (5.3% of total articles identified) (2,11);
(II) Accuracy (5.3%) (12,13);
(III) Continuous updating of evidence base (7.0%) (12,14,15);
(IV) Fairness (8.8%) (2,4,16,17);
(V) Reliability (10.5%) (8,13,17-19);
(VI) Education (10.5%) (20-24);
(VII) Interpretability (14.0%) (4,8,19,21,24,25);
(VIII) Transparency (28.1%) (2,4,8,11,15-18,21,23,26-30);
(IX) Explainability (45.6%) (8,12,16,17,19,20,23-28,31-41).

The sixteen trust factors that were only analyzed qualitatively included:

(I) Data representativeness (1.8%) (17);
(II) Standardized performance reporting label inclusion (1.8%) (12);
(III) Fidelity (1.8%) (25);
(IV) Ethicality (1.8%) (25);
(V) Lawfulness (1.8%) (25);
(VI) Data discoverability and accessibility (1.8%) (15);
(VII) Compliance (1.8%) (15);
(VIII) Knowledge representation (1.8%) (29);
(IX) Computational reliability (1.8%) (22);
(X) Relevance/insight (1.8%) (11);
(XI) Consistency (3.5%) (42,43);
(XII) Causability (3.5%) (21,27);
(XIII) Predictability (5.3%) (8,13,25);
(XIV) Dependability/competence (5.3%) (8,19,28);
(XV) Validation (7.0%) (4,8,17,25);
(XVI) Robustness (7.0%) (2,16,17,25);

The additional 10 factors that were *only* analyzed through quantitative methods included:

(I) Availability (1.8%) (44);
(II) Effort expectancy (1.8%) (45);
(III) Endorsement by other general practitioners (GPs (1.8%) (46);
(IV) AI agreement with physician suspicions (1.8%) (47);
(V) Information security (3.5%) (48,49);
(VI) Performance expectancy (3.5%) (45,50);
(VII) Sensitivity to patient context (3.5%) (46,51);
(VIII) Alignment with clinical workflow (3.5%) (46,52);
(IX) Perceived usefulness (5.3%) (44,52,53);
(X) GP involvement in tool design and dissemination (5.3%) (42,43,46).

Overall, explainability was discussed consistently across 23 included articles, and was the factor examined the most often. This suggests it is one of the most important concepts for trust in medical AI. In articles that included both qualitative and quantitative examination of concepts, the trust factors that were more often quantitatively analyzed tended to focus on provenance (46,54), usability (42-44), and privacy (52,55,56) (see *Figures 2,3*). This reveals an apparent gap in the literature; these concepts may benefit from further qualitative analysis amongst medical professional populations to provide AI researchers more insight into aspects to consider when developing AI technology to facilitate clinical adoption.

Among the articles that solely explored AI trust concepts quantitatively, education and usability were discussed most often. Other top contributory factors to enhanced end-user trust in medical AI identified quantitatively were explainability, privacy, GP involvement in tool design and dissemination, and perceived usefulness (see *Figures 2,3*).

Considering the entirety of the available literature, and despite more articles focusing on qualitative analysis of this concept, explainability appears to be the most important factor to enhancing trust in medical AI systems according to healthcare professionals.

Discussion

Substantial investments in AI research and regulation suggest that this technology could become an essential clinical decision-making support tool in the near future (57). Existing literature in the domain of machine learning discusses methods that have been applied to successfully train data-driven mathematical models to support healthcare decision-making processes (2,58,59). This is a common basis for the engineering of medical AI, which has future applications in clinical decision-making and complex domains, like precision medicine, to optimize patient health outcomes (60). Similarly, the CONSORT-AI reporting guideline extension provides direction for research investigators to help promote transparency and completeness in reporting clinical trials for medical AI interventions (61). As such, it may enhance trust in the AI technology upon clinical integration and support clinical decision-making processes by facilitating critical appraisal and evidence synthesis for interventions involving medical AI. The use of this reporting guideline is important to consider for clinical trials involving AI to promote end-user trust by ensuring a comprehensive evaluation of the technology before deployment and integration into clinical environments.

However, while there is substantial discussion regarding the technical and engineering aspects of such intelligent systems, healthcare professionals often remain hesitant to adopt and integrate AI into their practice (62). As such, it is necessary to better understand the factors that influence the trust relationship between medical experts and AI systems. This will not only augment decision-making and diagnostics, but also facilitate AI adoption into healthcare settings.

The unique aspect of our study is the focus on a synthesis of factors that are considered contributory to the enhancement of trust in the output of medical AI systems from the perspective of healthcare experts. We identified a disparity in the volume of literature that qualitatively versus

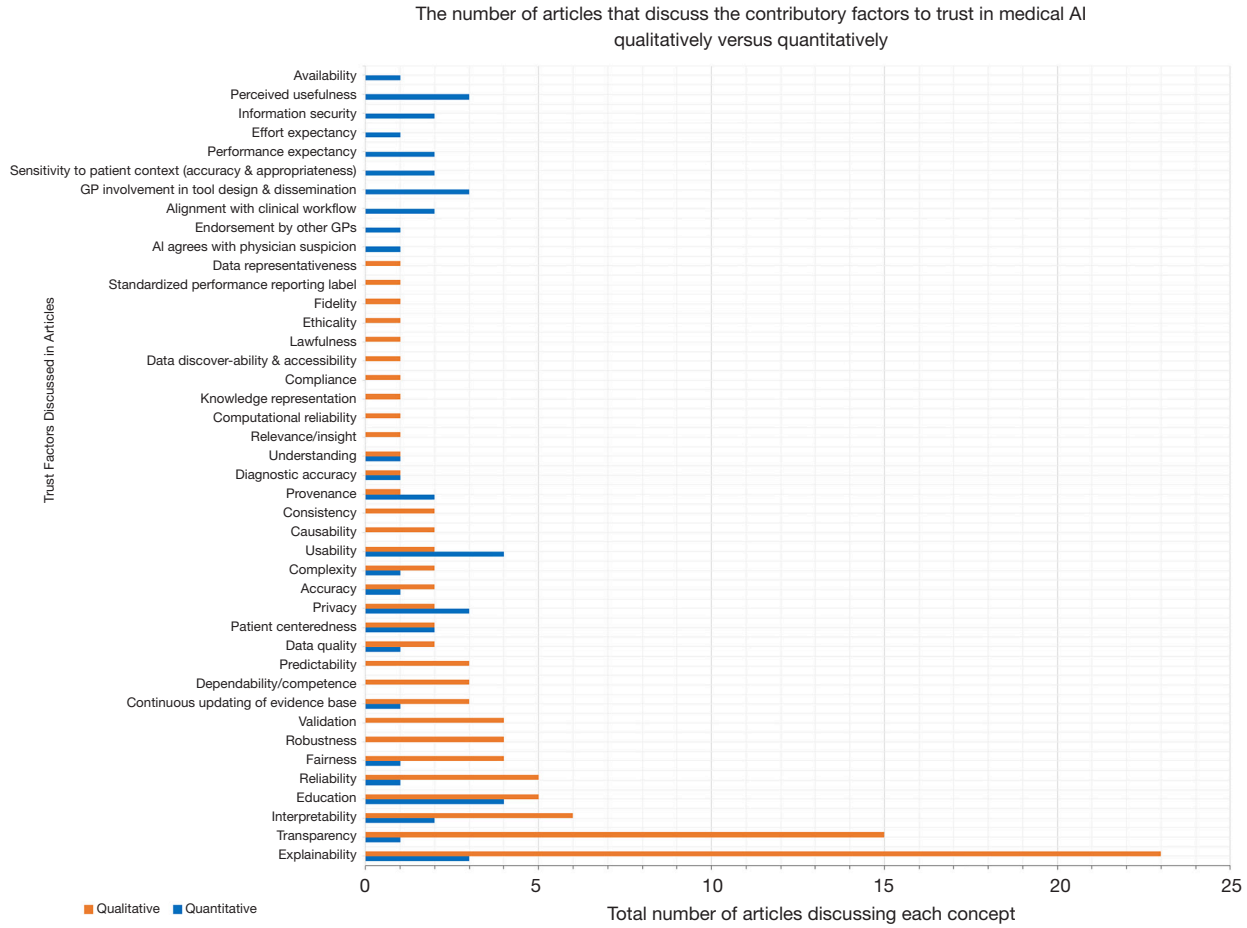

Figure 2 Bar graph depiction of the results comparing the number of articles that discuss medical AI trust concepts qualitatively versus quantitatively. AI, artificial intelligence.

quantitatively discusses each AI trust factor. This highlights a gap in the analytic assessment of AI trust factors and identifies trust topics that still require quantification amongst the medical community.

Explainability is discussed the most frequently overall across the qualitative and quantitative literature, and it is thus considered to be the most important factor influencing levels of trust in medical AI. Our findings tend to be consistent with literature in both medicine and engineering that have a large focus on supporting clinicians in making informed judgments using clinically meaningful explanations and developing technical mechanisms to engineer explainable AI (40,63). Since frequency was used as a surrogate for presumed importance, we acknowledge that there may exist other concepts related to trust in medical AI that are more important to healthcare professionals yet discussed less frequently, and thus may not be fully reflected in this paper. As well, explainable AI models, like decision trees, come at the expense of algorithm sophistication and are limited by big data (2). As such, achieving balance between algorithm complexity and explainability is necessary to enhancing trust in medical AI.

We note that there is heterogeneity between the medical AI trust factors that are commonly quantified as compared to qualitatively assessed. For instance, education and usability were important concepts that were frequently quantified in the included literature. Gaining a foundational background in AI is thought to be conducive to an increased trust and confidence in the system. By understanding the functionality of AI, healthcare trainees would be able to gradually develop a relationship with such systems (20,24). For instance, in a recent questionnaire by Pinto Dos Santos *et al.*, 71% of the medical student study population agreed that there was a need for AI to be included in the medical

Figure 3 Radar graph representation of the results comparing the number of articles that discuss medical AI trust concepts qualitatively versus quantitatively. Note that factors are presented counterclockwise and the vertical scale has been compressed beyond ten. †, number of articles qualitatively discussing explainability =23; ‡, number of articles qualitatively discussing transparency =15. AI, artificial intelligence.

training curriculum (64). Further, usability is a component of user experience and is dependent upon the efficacy of AI features in accommodating for, and satisfying, user needs to enhance ease of use (65). This is particularly important in clinical decision-making environments where healthcare provider end-users rely on ease of use to make critical judgements regarding patient care. As such, poor usability, i.e., complex AI user interfaces, can hinder the development of trust in the AI system (12).

Transparency and interpretability were also common factors particularly among qualitative assessments. These are also topics of focus in the engineering literature regarding the technical development of trustworthy AI programs (7). Transparency enables clinician users to make informed decisions when contemplating a recommendation outputted by a medical AI system. It also supports trust enhancement as transparent systems display their reasoning processes. In this way, healthcare professionals can still apply their own decision-making processes to develop differential diagnoses and complement the AI's conclusions because they can understand the methodical process employed by the system.

It follows that fairness is also an important factor to the user-AI trust relationship in healthcare (4,16). AI algorithms are inherently susceptible to discrimination by assigning weight to certain factors over others. This introduces the risk of exacerbating data biases and disproportionately affecting members of protected groups, especially when under-representative training datasets are used to develop the model (2,66). As such, implementing explainable and transparent AI systems in medical settings can aid physicians in detecting potential biases reflected by algorithmic flaws. Interpretability was also found to be commonly considered impactful to the AI trust relationship because it allows medical professionals to understand the AI's reasoning process. This prevents clinicians from feeling constrained by an AI's decision (67).

Although not quantified, robustness is the ability of a computer system to cope with errors in input datasets

and characterizes how effective the algorithm is with new inputs (2). It is also an important factor that is commonly discussed in the engineering discipline as a critical component of trustworthy AI programs. Even with small changes to the initial dataset, poor robustness can cause significant alterations to the output of an AI model (2).

Advances in AI capabilities will expand the role of this decision-support technology from automation of repetitive and defined tasks, towards guiding decision-making in critical environments, which is typically performed exclusively by medical professionals. As such, healthcare providers may increasingly rely on AI. Increased reliance on this technology requires a foundational trust relationship to be established in order to execute effective decision-making; this is also referred to as 'calibrated trust' (68). As such, our findings are relevant for those working to develop and optimize medical AI software and hardware to facilitate adoption and implementation in healthcare settings. These results are also relevant to inform healthcare professionals about which AI trust factors are quantified in terms of importance amongst the medical community.

An end-user's perception of the competence of an AI system not only impacts their level of trust in the technology, but also has a significant influence on how much users rely on AI, and thus impacts the effectiveness of healthcare decisions (69). However, Asan *et al.* [2020] explain that the level of trust in medical AI may not necessarily be positively correlated with clinical or patient outcomes (2). They introduce the concept of 'optimal trust' and note that trust maximization does not necessarily result in optimal decision-making via human-AI collaboration, since the user accepts the outcomes generated by the AI system without critical judgment. This can be particularly dangerous in clinical settings where patient life is at risk. Optimal trust entails maintaining a certain level of mutual skepticism between users and AI systems regarding clinical decisions. Since both are susceptible to error, the development of AI should incorporate mechanisms to sustain an optimal trust level (2,69).

Our study provides insight into the breadth of factors that may contribute to achieving an optimal trust relationship between human and machine in medical settings and exposes components that may have been previously overlooked and require further consideration. Our findings are consistent with the notion that incorporation of explainability, transparency, fairness, and robustness into the development of AI systems contributes to achieving a level of optimal trust. According to our results, these factors are frequently assessed qualitatively in the healthcare literature as topics that are considered important to trusting an AI system for medical professionals. It follows that quantification of these factors would provide better insight into the demand to incorporate them in medical AI development.

Achieving optimal trust in AI likely entails consideration of a range of factors that are important for healthcare professionals. However, *Figures 2,3* clearly depict that to date, the focus has been skewed to only a few commonly discussed factors, and that numerous other factors may require further consideration during the AI technology engineering and development phase, prior to deployment into clinical settings. This is important because the consequence of failing to consider this breadth of factors is insufficient trust in the AI system, which itself constitutes a barrier to adoption and integration of AI in medicine.

We also acknowledge that the discrepancies in linguistics within, and between, professional domains are a limitation of the field and a barrier to a fully comprehensive search strategy. Although the search strategy was developed to capture relevant articles discussing factors influencing trust in medical AI, it also returned a significant quantity of articles that were not actually discussing trust concepts related to medical AI, but rather focused on direct implications and applications of this technology in healthcare settings.

Although the focus of this paper was a summative assessment of the qualitatively and quantitatively measured factors impacting trust in medical AI, the implications of *integrating* this technology may also influence trust in AI in healthcare settings. Implementing AI-based decision-support systems inevitably disrupts the physician's practice model, as they are required to adopt a new thinking process and mechanism for performing a differential diagnosis that is teamed with an intelligent technology. Willingness to adopt AI, therefore, likely relates in part to the impact of AI on the medical practice, which in turn may affect trust in AI. For instance, physicians have been trained with the Hippocratic Oath, thus introducing a machine that can interact, as well as interfere with, the patient-physician relationship may increase hesitancy in trusting and adopting AI. A recent qualitative survey by Laï *et al.* [2020] found that while healthcare professionals recognize the promise of AI, their priority remains providing optimal care for their patients (62). Physicians generally avoid relinquishing entrusted patient care to a machine if it is not adequately trustworthy. So, integration of medical AI into clinical

settings also has deeper philosophical implications.

Lastly, we noted greater expansion of AI applications in certain distinct medical specialties, including dentistry and ophthalmology (70,71). This confirms that AI adoption in some areas of healthcare may be broader than others, and they tended to be disciplines in which sophisticated technical instrumentation is already commonly utilized.

Limitations of this study

Although the users of AI technology can be diverse, the focus of this paper is limited to the healthcare discipline. We acknowledge that trust relationships with AI systems could significantly differ for other relevant stakeholders such as patients, and insurance providers. As such, the incorporation of more perspectives from unique stakeholders in the medical community may offer a more extensive perspective of the factors contributing to trust in medical AI.

We acknowledge that there are challenges in the realm of AI research regarding the inconsistency and lack of universally accepted definitions of key terms, including transparency, explainability, and interpretability. Given that these are concepts anticipated to be relevant for understanding the concept of trust, we were obligated to accept that terminology may be applied inconsistently in the literature.

We also recognize that although we use frequency of discussed topics as a surrogate for significance, it only reflects a degree of perceived importance, as it is also possible that the current research focus is misdirected, or that concepts deemed important in one discipline would not necessarily translate to those deemed relevant in another discipline using the same medical AI. We acknowledge that this also does not identify the unknown unknowns regarding factors that contribute to trust in medical AI for healthcare providers. Further, we acknowledge that there may be bias in the categorization of implicit AI trust concepts, as mapping these to an explicit concept was partly based on the authors' professional judgment; however, established definitions were consulted to increase objectivity when deciding the explicit concepts upon which the implicit ones would be mapped.

The timeframe of this study was limited to articles published beyond the year 2000 until July 2021. As AI is rapidly developing and the integration of medical AI in clinical practice is becoming more pertinent, we recommend on-going monitoring of this literature, as well as review of other domains that may discuss AI trust concepts that were not identified in this paper.

Conclusions

In order to facilitate adoption of AI technology into medical practice settings, significant trust must be developed between the AI system and the health expert end-user. Overall, explainability, transparency, interpretability, usability, and education are among the key identified factors currently thought to influence this trust relationship and enhance clinician-machine teaming in critical decision-making environments in healthcare. There is a need for a common and consistent nomenclature between primary fields, like engineering and medicine, for cross-disciplinary applications, like AI. We also identify the need to better evaluate and incorporate other important factors to promote trust enhancement and consult the perspectives of medical professionals when developing AI systems for clinical decision-making and diagnostic support. To build upon this consolidation and broad understanding of the literature regarding the conceptualization of trust in medical AI, future directions may include a systematic review approach to further quantify relevant evidence narrower in scope.

Acknowledgments

We would like to thank the Biomedic.AI Lab at McMaster University for their on-going support and feedback throughout the process of this study. We would also like to thank the Department of National Defense Canada for their funding of the Biomedic.AI Lab activities through the Innovation for Defence Excellence and Security (IDEaS) Program.

Funding: This work was supported by the Canadian Department of National Defence's Innovation for Defence Excellence and Security (IDEaS) Program, in which we were awarded a research contract to study Barriers to Adoption of Autonomy, specifically for medical advisory systems.

Footnote

Reporting Checklist: The authors have completed the Narrative Review reporting checklist. Available at https://jmai.amegroups.com/article/view/10.21037/jmai-21-25/rc

Peer Review File: Available at https://jmai.amegroups.com/article/view/10.21037/jmai-21-25/prf

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://jmai.amegroups.com/article/view/10.21037/jmai-21-25/coif). TED declares this work was supported by the Canadian Department of National Defence's Innovation for Defence Excellence and Security (IDEaS) Program for the study of Barriers to the Adoption of Autonomy, specifically autonomous medical advisory systems and that a US Patent Application is in process for "Method for Enabling Trust in Collaborative Research". The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Thimbleby H. Technology and the Future of Healthcare. J Public Health Res 2013; 2:e28.
2. Asan O, Bayrak AE, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. J Med Internet Res 2020;22:e15154.
3. Wilkins CH. Effective Engagement Requires Trust and Being Trustworthy. Med Care 2018;56 Suppl 10 Suppl 1:S6-8.
4. Quinn TP, Senadeera M, Jacobs S, et al. Trust and medical AI: the challenges we face and the expertise needed to overcome them. J Am Med Inform Assoc 2021;28:890-4.
5. Kruse CS, Kristof C, Jones B, et al. Barriers to Electronic Health Record Adoption: a Systematic Literature Review. J Med Syst 2016;40:252.
6. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ 2019;7:e7702.
7. Jacovi A, Marasović A, Miller T, et al. Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. New York, NY, USA: Association for Computing Machinery, 2021:624-35.
8. Gilbank P, Johnson-Cover K, Truong T. Designing for Physician Trust: Toward a Machine Learning Decision Aid for Radiation Toxicity Risk. Ergonomics in Design: The Quarterly of Human Factors Applications 2020;28:27-35.
9. Felzmann H, Fosch-Villaronga E, Lutz C, et al. Towards Transparency by Design for Artificial Intelligence. Sci Eng Ethics 2020;26:3333-61.
10. Reyes M, Meier R, Pereira S, et al. On the Interpretability of Artificial Intelligence in Radiology: Challenges and Opportunities. Radiol Artif Intell 2020;2:e190043.
11. Shortliffe EH, Sepúlveda MJ. Clinical Decision Support in the Era of Artificial Intelligence. JAMA 2018;320:2199-200.
12. Jones C, Thornton J, Wyatt JC. Enhancing trust in clinical decision support systems: a framework for developers. BMJ Health Care Inform 2021;28:e100247.
13. Hatherley JJ. Limits of trust in medical AI. J Med Ethics 2020;46:478-81.
14. Walsh K. The perspective of physicians on the value of online clinical decision support. Proc (Bayl Univ Med Cent) 2019;32:58-60.
15. Richardson JE, Middleton B, Platt JE, et al. Building and maintaining trust in clinical decision support: Recommendations from the Patient-Centered CDS Learning Network. Learn Health Syst 2019;4:e10208.
16. Pelaccia T, Forestier G, Wemmert C. Deconstructing the diagnostic reasoning of human versus artificial intelligence. CMAJ 2019;191:E1332-5.
17. Cutillo CM, Sharma KR, Foschini L, et al. Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. NPJ Digit Med 2020;3:47.
18. Wald B. Making AI more "explainable" in health-care settings may lead to more mistakes: U of T researcher [Internet]. Department of Medicine 2020 [cited 2021 Aug 17]. Available online: https://deptmedicine.utoronto.ca/news/making-ai-more-explainable-health-care-settings-may-lead-more-mistakes-u-t-researcher
19. Starke G, van den Brule R, Elger BS, et al. Intentional machines: A defence of trust in medical artificial intelligence. Bioethics 2021. doi: 10.1111/bioe.12891.
20. Kiser KJ, Fuller CD, Reed VK. Artificial intelligence in radiation oncology treatment planning: a brief overview. J Med Artif Intell 2019;2:9.
21. He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. Nat Med

2019;25:30-6.
22. Durán JM, Jongsma KR. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. J Med Ethics 2021. [Epub ahead of print].
23. Reddy S, Allan S, Coghlan S, et al. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020;27:491-7.
24. Kitamura FC, Marques O. Trustworthiness of Artificial Intelligence Models in Radiology and the Role of Explainability. J Am Coll Radiol 2021;18:1160-2.
25. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform 2021;113:103655.
26. McIntosh C, Conroy L, Tjong MC, et al. Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer. Nat Med 2021;27:999-1005.
27. Holzinger A, Biemann C, Pattichis CS, et al. What do we need to build explainable AI systems for the medical domain? arXiv:171209923 [cs, stat] [Internet]. 2017 Dec 28 [cited 2021 Aug 17]. Available online: http://arxiv.org/abs/1712.09923
28. Nundy S, Montgomery T, Wachter RM. Promoting Trust Between Patients and Physicians in the Era of Artificial Intelligence. JAMA 2019;322:497-8.
29. Alexander GL. Issues of trust and ethics in computerized clinical decision support systems. Nurs Adm Q 2006;30:21-9.
30. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLoS Med 2018;15:e1002689.
31. Gaube S, Suresh H, Raue M, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. NPJ Digit Med 2021;4:31.
32. Holzinger A, Langs G, Denk H, et al. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev Data Min Knowl Discov 2019;9:e1312.
33. Challen R, Denny J, Pitt M, et al. Artificial intelligence, bias and clinical safety. BMJ Qual Saf 2019;28:231-7.
34. Chen T, Keravnou-Papailiou E, Antoniou G. Medical analytics for healthcare intelligence - Recent advances and future directions. Artif Intell Med 2021;112:102009.
35. Grote T. Trustworthy medical AI systems need to know when they don't know. Journal of Medical Ethics 2021;47:337-8.
36. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019;17:195.
37. Feldman R, Aldana E, Stein K. Artificial Intelligence in the Health care Space: How We Can Trust What We Cannot Know. Stanford Law & Policy Review 2019;30:399.
38. Meskó B, Görög M. A short guide for medical professionals in the era of artificial intelligence. NPJ Digit Med 2020;3:126.
39. Stoel B. Use of artificial intelligence in imaging in rheumatology - current status and future perspectives. RMD Open 2020;6:e001063.
40. Amann J, Blasimme A, Vayena E, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020;20:310.
41. Lee SS. Philosophical evaluation of the conceptualisation of trust in the NHS' Code of Conduct for artificial intelligence-driven technology. J Med Ethics 2021. [Epub ahead of print]. doi: 10.1136/medethics-2020-106905
42. Petkus H, Hoogewerf J, Wyatt JC. What do senior physicians think about AI and clinical decision support systems: Quantitative and qualitative analysis of data from specialty societies. Clin Med (Lond) 2020;20:324-8.
43. Liberati EG, Ruggiero F, Galuppo L, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci 2017;12:113.
44. Petitgand C, Motulsky A, Denis JL, et al. Investigating the Barriers to Physician Adoption of an Artificial Intelligence-Based Decision Support System in Emergency Care: An Interpretative Qualitative Study. Stud Health Technol Inform 2020;270:1001-5.
45. Fan W, Liu J, Zhu S, et al. Investigating the impacting factors for the healthcare professionals to adopt artificial intelligence-based medical diagnosis support system (AIMDSS). Ann Oper Res 2020;294:567-92.
46. Ford E, Edelman N, Somers L, et al. Barriers and facilitators to the adoption of electronic clinical decision support systems: a qualitative interview study with UK general practitioners. BMC Med Inform Decis Mak 2021;21:193.
47. Dreiseitl S, Binder M. Do physicians value decision support? A look at the effect of decision support systems on physician opinion. Artif Intell Med 2005;33:25-30.
48. Dünnebeil S, Sunyaev A, Blohm I, et al. Determinants of physicians' technology acceptance for e-health in ambulatory care. Int J Med Inform 2012;81:746-60.
49. Antwi WK, Akudjedu TN, Botwe BO. Artificial intelligence in medical imaging practice in Africa: a

qualitative content analysis study of radiographers' perspectives. Insights Imaging 2021;12:80.
50. Heselmans A, Aertgeerts B, Donceel P, et al. Family physicians' perceptions and use of electronic clinical decision support during the first year of implementation. J Med Syst 2012;36:3677-84.
51. Harada Y, Katsukura S, Kawamura R, et al. Effects of a Differential Diagnosis List of Artificial Intelligence on Differential Diagnoses by Physicians: An Exploratory Analysis of Data from a Randomized Controlled Study. Int J Environ Res Public Health 2021;18:5562.
52. Paranjape K, Schinkel M, Hammer RD, et al. The Value of Artificial Intelligence in Laboratory Medicine. Am J Clin Pathol 2021;155:823-31.
53. Kortteisto T, Komulainen J, Mäkelä M, et al. Clinical decision support must be useful, functional is not enough: a qualitative study of computer-based clinical decision support in primary care. BMC Health Serv Res 2012;12:349.
54. Wiens J, Saria S, Sendak M, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019;25:1337-40.
55. Sassis L, Kefala-Karli P, Sassi M, et al. Exploring Medical Students' and Faculty's Perception on Artificial Intelligence and Robotics. A Questionnaire Survey. Journal of Artificial Intelligence for Medical Sciences 2021;2:76-84.
56. Castagno S, Khalifa M. Perceptions of Artificial Intelligence Among Healthcare Staff: A Qualitative Survey Study. Front Artif Intell 2020;3:578983.
57. IEEE-USA Board of Directors. Artificial Intelligence Research, Development and Regulation [Internet]. IEEE-USA; 2017. Available online: https://ieeeusa.org/wp-content/uploads/2017/10/AI0217.pdf
58. Miotto R, Wang F, Wang S, et al. Deep learning for healthcare: review, opportunities and challenges. Brief Bioinform 2018;19:1236-46.
59. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019;25:24-9.
60. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019;6:94-8.
61. Liu X, Cruz Rivera S, Moher D, et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. Nat Med 2020;26:1364-74.
62. Laï MC, Brian M, Mamzer MF. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020;18:14.
63. Antoniadi AM, Du Y, Guendouz Y, et al. Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. Appl Sci 2021;11:5088.
64. Pinto Dos Santos D, Giese D, Brodehl S, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol 2019;29:1640-6.
65. Yang F, Shi S, Zhu J, et al. Analysis of 92 deceased patients with COVID-19. J Med Virol 2020;92:2511-5.
66. Gianfrancesco MA, Tamang S, Yazdany J, et al. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. JAMA Intern Med 2018;178:1544-7.
67. Sethi T, Kalia A, Sharma A, et al. Chapter 1 - Interpretable artificial intelligence: Closing the adoption gap in healthcare. In: Barh D. editor. Artificial Intelligence in Precision Health. Academic Press, 2020:3-29.
68. Hoffman RR, Johnson M, Bradshaw JM, et al. Trust in Automation. IEEE Intelligent Systems 2013 Jan;28(1):84–8.
69. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46:50-80.
70. Chen YW, Stanley K, Att W. Artificial intelligence in dentistry: current applications and future perspectives. Quintessence Int 2020;51:248-57.
71. Gunasekeran DV, Wong TY. Artificial Intelligence in Ophthalmology in 2020: A Technology on the Cusp for Translation and Implementation. Asia Pac J Ophthalmol (Phila) 2020;9:61-6.

doi: 10.21037/jmai-21-25
Cite this article as: Tucci V, Saary J, Doyle TE. Factors influencing trust in medical artificial intelligence for healthcare professionals: a narrative review. J Med Artif Intell 2022;5:4.

References

72. Singh A, Sengupta S, Lakshminarayanan V. Explainable Deep Learning Models in Medical Image Analysis. J Imaging 2020;6:52.
73. van derWaa J, Schoonderwoerd T, vanDiggelen J, et al. Interpretable confidence measures for decision support systems. International Journal of Human-Computer Studies 2020;144:102493.
74. Komninos A. An Introduction to Usability [Internet]. The Interaction Design Foundation. 2020 [cited 2021 Aug 17]. Available online: https://www.interaction-design.org/literature/article/an-introduction-to-usability

Table S1 Concept mapping of alternative/implicit concepts, testimonials, or examples

Alternative definition/implicit concepts, testimonials, or examples as explained in the individual article	Explicit concept this is mapped to
Saliency mapping (20)	Explainability (72)
Display and calculate measures of confidence in prediction accuracy (31,33)	Explainability (73)
“Tools that converted inputted data into an easy-to-follow management plan (e.g., the ‘Sick Child Template’) were perceived as useful” (46)	Usability (74)
“It’s got a kind of column of green things, a column of orange things and a column of red things. Then there’s a really clear next page about what you should do if they’re kind of, if they’ve got lots of greens, you know, what the process would be if they’ve got lots of reds in terms of, you know, one side and I think that’s really helpful just because it kind of combines the data you’re putting in with actually a useful plan.” (clear communication and guidance) (46)	Interpretability (10)
“Transparency: understanding factors driving the prediction to assess the logic behind the model and understanding whether the data were applicable to their patients” (8)	Transparency (9)
“The advice given by the CDSS can be difficult to interpret” (42)	Interpretability (10)
“Some CDSS act like black boxes with no insight into their conclusions” (42)	Explainability (7)
"Concerns that CDSS output may not be worded clearly” (42)	Usability (74)
“Giving—where possible—some account of the mechanism for how decisions are arrived at; the quality, size and source of any datasets relied on; and assurance that standard guidelines for training the algorithm were followed (as well as monitoring appropriate learning diagnostics) will probably assuage some clinicians’ concerns” (12)	Explainability (7)
“There are developments towards opening the ‘black box’ by providing so-called class-discriminating attention maps, which may give at least an indication of where the network had focused on, in order to come to a certain classification” (39)	Explainability (7)

https://dx.doi.org/10.21037/jmai-21-25