Check for updates

SYSTEMATIC REVIEW

Influences on User Trust in Healthcare Artificial Intelligence: A Systematic Review [version 1; peer review: awaiting peer review]

Eva Jermutus[1], Dylan Kneale[2], James Thomas[2], Susan Michie[3]

[1]Social Science Research Unit, UCL Social Research Institute, University College London, London, London, UK
[2]EPPI-Centre, UCL Social Research Institute, University College London, London, London, UK
[3]Centre for Behaviour Change, Department of Clinical, Educational and Health Psychology, University College London, London, London, UK

v1 **First published:** 18 Feb 2022, **7**:65
https://doi.org/10.12688/wellcomeopenres.17550.1

Latest published: 18 Feb 2022, **7**:65
https://doi.org/10.12688/wellcomeopenres.17550.1

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Abstract

Background: Artificial Intelligence (AI) is becoming increasingly prominent in domains such as healthcare. It is argued to be transformative through altering the way in which healthcare data is used. The realisation and success of AI depend heavily on people's trust in its applications. Yet, influences on trust in healthcare AI (HAI) applications so far have been underexplored. The objective of this study was to identify aspects related to users, AI applications and the wider context influencing trust in HAI.
Methods: We performed a systematic review to map out influences on user trust in HAI. To identify relevant studies, we searched seven electronic databases in November 2019 (ACM digital library, IEEE Explore, NHS Evidence, ProQuest Dissertations & Thesis Global, PsycINFO, PubMed, Web of Science Core Collection). Searches were restricted to publications available in English and German. To be included studies had to be empirical; focus on an AI application (excluding robotics) in a health-related setting; and evaluate applications with regards to users.
Results: Three studies, one mixed-method and two qualitative studies in English were included. Influences on trust fell into three broad categories: human-related (knowledge, expectation, mental model, self-efficacy, type of user, age, gender), AI-related (data privacy and safety, operational safety, transparency, design, customizability, trialability, explainability, understandability, power-control-balance, benevolence) and context-related (AI company, media, users' social network). The factors resulted in an updated logic model illustrating the relationship between these aspects.
Conclusion: Trust in HAI depends on a variety of factors, both external and internal to AI applications. This study contributes to our understanding of what influences trust in HAI by highlighting key

influences, as well as pointing to gaps and issues in existing research on trust and AI. In so doing, it offers a starting point for further investigation of trust environments as well as trustworthy AI applications.

Keywords
trust, artificial intelligence, human-AI interaction, health care, healthcare AI, systematic review, MMAT, logic model

Corresponding author: Eva Jermutus (eva.jermutus.18@ucl.ac.uk)

Author roles: Jermutus E: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kneale D**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Thomas J**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Michie S**: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the Wellcome Trust [201524; an award to the Human Behaviour-Change Project]. The substantive work of the review was part of an ESRC-funded PhD studentship [2166011; awarded to the Industrial Challenge Project 'Advancing and Applying the Science of Behaviour Change through Machine-Live Learning']. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Jermutus E, Kneale D, Thomas J and Michie S. **Influences on User Trust in Healthcare Artificial Intelligence: A Systematic Review [version 1; peer review: awaiting peer review]** Wellcome Open Research 2022, **7**:65 https://doi.org/10.12688/wellcomeopenres.17550.1

First published: 18 Feb 2022, **7**:65 https://doi.org/10.12688/wellcomeopenres.17550.1

Introduction

Applications of Artificial Intelligence (AI) are increasingly considered for integration into healthcare systems to increase efficiency and address issues such as staff shortages. The success of this integration will, in parts, depend on people's trust in the applications of AI. However, there is little evidence about what influences trust in healthcare Artificial Intelligence (HAI). Understanding what influences trust in AI in healthcare settings is an important step to making informed decisions around its integration and implementation. The aim of this systematic review is to summarise published, empirical data and to provide an overview of factors related to humans and AI applications that influence users' trust in healthcare AI. This will provide insight into gaps in our knowledge and the research questions that need to be addressed in future research as well as issues in our current conceptualization of trust in AI.

AI broadly refers to a *"set of advanced technologies that enable machines to carry out highly complex tasks effectively […] tasks that would require intelligence if a person were to perform them"*[1]. The term AI, however, is used inconsistently and can describe a range of different applications. As a result, there is no universally accepted definition and attempts at defining AI result in a somewhat imprecise and fuzzy definition that is challenging to operationalise, as is arguably the case with the definition above. The problem is exacerbated due to definitions evolving as the technology itself evolves. As a result, what once was considered AI may not be considered AI today. While we will employ the term AI throughout this report, we acknowledge that it encompasses a range of fields, applications and techniques.

It is worth noting that the review focuses on *explicit* AI by which we mean applications that explicitly state they contain an AI component and where it is possible to discern that they do (e.g., by using respective terminology). The decision to focus on explicit AI may be contested with respect to some applications. For instance, there is a long-standing tradition around Medical Imaging Analysis and Decision Support Systems (DSS) in healthcare. While DSSs can be AI, they are not necessarily so and often DSS are simply referred to as 'computerised' or 'automated'. Without additional information, it is unclear if these systems are AI. Including such examples in the review would result in conflating AI with other technologies which would cloud our understanding of features specific to AI. The question of trust in AI also means that the review requires studies that are clearly AI so as to not conflate trust of AI with trust of technology more broadly. Put differently, the present review focuses on AI that is communicated as such in the respective publications.

AI is argued to have the potential to improve healthcare by transforming the way in which we use data, treat patients and develop diagnostic tools[2]. Accordingly, it is often perceived as part of a solution to tackling healthcare issues such as increasing costs and staff shortages. The promises of AI have resulted in large investments for its development. For example, the UK Government recently announced it planned to spend £250m on integrating AI into health services[3]. Similarly, worldwide spending on AI is expected to rise to $232 by 2025[4]. While investments are steadily growing, adoption rates remain low[5]. The lack of adoption partly stems from applications of AI still being in development and use cases still being identified and explored as well as a lack of evaluation studies. Engagement with end-users is necessary in order to develop fit-for-purpose systems, although defining the 'end' user can be complicated by users of HAI not always synonymous with beneficiaries of HAI. For instance, HAI systems like the one developed in the Human Behaviour-Change Project (HBCP)[6] may be used by a practitioner to support decision-making, although the additional support from the HBCP system is expected to benefit a larger group. For instance, practitioners can query the HBCP system to get a better understanding of which behaviour change interventions work best for a particular group of people and use this to inform their decision on how to help. Hence, a practitioner would be the user but the people benefitting from the use of the system would be a larger group of people that did not actually use the system. Furthermore, the optimal point at which users should switch to new technology is not clear and will depend upon many factors such as their context, attitude towards risk taking, baseline trust in or knowledge of the new technology. At the same time, there are structural challenges such as the lack of a suitable data infrastructure as well as ethical challenges that influence the adoption of AI[2,7–9]. A central concern is the issue of trust which has previously been suggested as key to realising a technology's potential[10]. Trust in technology, however, can have negative implications if we trust too much. For instance, trusting too may result in errors in electronic prescribing[11]. Similarly, trusting too little may result in underutilizing a technology, thereby missing out on opportunities a technology offers.

A myriad of disciplines and researchers have analysed the formation and antecedents of trust in automation or technology in general[10,12–14]. There is also a lot of research on the effectiveness of specific AI applications to a particular medical problem. Similar to previous healthcare technologies, however, realising AI's potential is not just about establishing the effectiveness of its applications but also about resolving issues such as trust[15]. While the number of publications on trust in AI is growing at a fast pace[16], most work remains theoretical. For instance, recent work has introduced an incremental model of trust[17] and discussed how trust in AI can be formalized[18]. While theoretical work is important, it is no sufficient evidence base to inform practice.

Given that there are plans to integrate AI in healthcare systems, a better understanding of the underlying mechanisms of how users of AI decide whether or not to trust an AI application and how they judge an AI's trustworthiness is required to inform decisions about the implementation of AI. This review was developed to support and inform decisions about the implementation of AI systems such as that developed in the HBCP[6].

What is Trust?

Trust is an elusive concept and its definition and operationalization vary within and across disciplines and contexts, resulting in a somewhat fragmented understanding of what trust is 19. Trust is oftentimes described as a function of the trustor (e.g. user), trustee (e.g. machine) and situation[20–22]. Trust becomes relevant when uncertainty and risk are involved as discussed in 23. The decision to trust someone or something depends, at least in part, on the trustee's *trustworthiness,* i.e., its attribute of being reliable and predictable. It reflects an evaluation of the trustee's attributes[24]. Technology, however, has neither volition nor moral agency. Trust in technology therefore is based on beliefs about the characteristics of a technology rather than will or motives as it has none[24].

Yet, technology and technological corporations are often conflated. Focusing on the technology, the present review adopts a working definition where trust in AI is defined as an individual's attitude towards an AI application about its ability to perform a particular action important to the individual. The attitude is comprised of a set of beliefs about the AI's capabilities and characteristics.

Previous research on automated systems suggests that trustworthiness is fostered by characteristics such as perceived competence, responsibility and dependability[21], as well as certain design features such as communication style or level of control[20]. The studies reviewed in the trust in automation literature overwhelmingly focus on contexts such as monitoring tasks, autonomous driving or flight simulation tasks rather than healthcare. This prompts the question how transferable the findings are to the healthcare context. A recent scoping review provides an initial overview of personal, institutional and technological enablers and impediments of trust in digital health[25]. Digital health, however, encompasses a wide range of technologies and thus conflates AI with other technologies. This is an issue given the peculiarities of AI techniques. For example, it is possible for models which are understandable to be based upon intuition-defying statistical relationships; this is known as non-intuitiveness[26]. The quality of being nonintuitive presents issues such as people using the AI being unable to make sense of relationships between variables. As a result, assessing whether the basis for a decision is sound, is difficult, which creates further issues. For instance, a doctor may find it difficult to decide whether to rely on or trust an AI, yet, has to explain or justify that decision to other parties without having the insight required to make a strong case. AI's ability to learn exacerbates such issues as the same input does not necessarily result in same output, creating an additional layer of complexity.

AI is also often singled out in the media and communicated as a superior technology with great opportunities and threats[27]. It therefore seems essential to investigate AI and its unique characteristics separately in order to understand people's, especially users', perceptions and understanding of AI and, ultimately, their trust in AI. Existing reviews have focused on subcomponents of AI such as trust in robotics in as well as different types of trust in AI[28–30]. While these reviews and more recent ones[31] considered trust in AI across domains, this study is the first review to specifically focus on trust in HAI.

Understanding which characteristics internal and external to an AI system convey trustworthiness not only aids the development of new AI applications but also allows us to better understand and meet people's expectations of such applications. The present systematic review seeks to contribute to this endeavour by analysing influences on user trust in HAI.

Direct vs Indirect Trust Judgements– The Importance of Learned Trust

The public's perception of and trust in HAI are important matters to examine as they can influence not only the uptake but also the regulation of AI[32]. Similarly, understanding patients' trust in a decision that their physician reached with the assistance of an AI application and their trust in the AI itself is of great importance. However, both scenarios pose a problem: there is *no direct interaction* between the patient and the AI application. The patient could draw on publicly available information and the physician, but not the AI itself to reach a trusting decision. As a result, it is difficult for researchers to disentangle concerns regarding the specific AI application from broader concerns such as data privacy if the patient mistrusts the application. Similarly, a measure of public trust would include individuals that have not actually interacted with an AI application. The public's perception may also reflect the image of AI as a single or homogenous phenomenon (i.e., general view of AI) when in fact it is very context and application specific.

The lack of direct interaction has important implications for the way a trust(worthiness) judgement (i.e., an answer to the question 'Do I trust this AI to do x?') is reached as illustrated in the logic model in Figure 1.

Figure 1 distinguishes between two scenarios: The indirect scenario (solid lines) represents individuals and groups that do not interact with AI but may be affected by its existence or use (e.g., patients). In absence of an interaction with an AI application, an individual does not have the option to use experience with the application to inform his/her decision. Accordingly, the individual has to rely on factors external to the AI (e.g., media narratives or their own dispositions) to reach a trust judgement. Conversely, the direct scenario (solid and dotted lines) represents cases where individuals directly interact with an AI application (e.g., users of an AI application). Interacting with the application allows the user to experience the AI's features, enabling the user to utilize information specific to the AI on top of the other influences to reach a trust judgement. For both paths, the trust(worthiness) judgement will be made with a certain level of confidence based on the input of the different influences[33].

The model is based on the finding that trust is a function of person, technology and environment[20,34]. Accordingly, there are three trusting inputs: human, AI and contextual influences. The separation of trust judgement and trusting behaviour is based on the fact that previous research distinguishes between

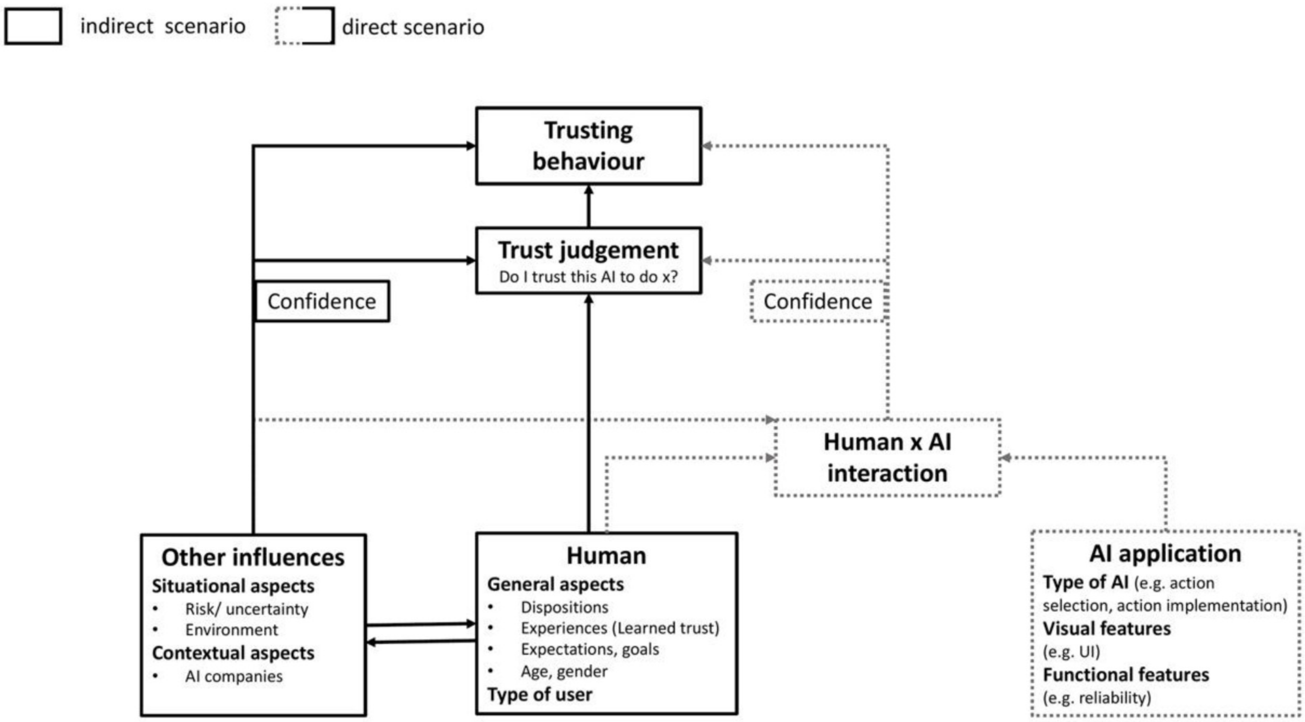

Figure 1. Logic model depicting the two different scenarios of AI usage (indirect, direct) used to determine the scope of the review. The indirect scenario (solid lines) represents cases where individuals do not interact with an AI application. The dotted lines represent the additional paths that occur when individuals directly interact with the AI application. I.e., the direct scenario is the full logic model whereas the indirect scenario includes only the solid lines.

trust as a behaviour and trust as an attitude[35]. Finally, the differentiation between direct and indirect trust scenarios draws onto the concept of learned trust (LT) as defined and used in Hoff and Bashir (2015)[20]. The authors reviewed factors influencing trust in automation and grouped the evidence into three broad dimensions: dispositional trust, situational trust and learned trust. Situational trust (ST) is highly context- and interaction-specific and variability within ST stems from the external environment as well as internal characteristics of the operator[20]. External aspects include factors such as type or complexity of the system whereas internal aspects entail factors such as self-confidence or subject matter expertise. Dispositional trust (DT) defines an "enduring tendency to trust in automation" and thus is considered as trait-like and relatively stable over time[20]. Finally, LT is based on experiences relevant to the specific automated system. It develops from interacting with the system and informs the user's evaluation of the system (e.g., a specific AI application).

The model accounts for both, direct and indirect scenarios as the future of healthcare likely means patients and other stakeholders will be involved in both. However, given that LT emerges from an interaction between a human and an AI application, a better understanding of user trust requires a consideration of LT (Human x AI interaction in Figure 1) along with the other two dimensions (i.e., human and other influences). This is only possible through considering the direct scenario in Figure 1 as LT is largely absent in the indirect scenario. Accordingly, the present review focuses on the direct scenario to better understand influences on user trust in HAI.

The model was used throughout the review by guiding the inclusion criteria, developing the initial coding tool as well as building the starting point for the synthesis of included papers.

Review questions

- What characteristics, that is functional and/ or aesthetic features of a HAI, influence user trust in the AI?
- What factors external to the AI application (i.e., human and context) influence user trust in the AI?

Methods

We used a three-part search strategy to identify studies meeting the inclusion criteria: (1) A search of electronic databases for published work and grey literature, using a comprehensive search strategy for user trust in HAI. (2) We then searched the reference lists of primary studies included in the review and the reference list of relevant, previously published reviews such as 25. (3) We contacted authors of included papers to identify further relevant literature. Results are reported in line with the PRISMA reporting standards for systematic reviews. For the completed checklist, see *Extended data*[36].

Data sources

Seven electronic databases were searched between 17th to 19th November 2019 with a pre-determined search strategy (see *Extended data* for search strategy[36]), these included: ACM digital library, IEEE Xplore, NHS Evidence, Ovid ProQuest Dissertations & Thesis Global, Ovid PsycINFO, PubMed, Web of Science Core Collection. The databases were chosen in consultation with a research librarian to ensure that literature from different disciplines investigating the topic under review were represented (e.g., computer science, psychology). The searches were restricted to publications available in English and German, with no publication date restriction. There were no limits on study participants in terms of age, gender, ethnicity or profession. There also was no limit on study setting and studies of all levels of healthcare settings (primary, secondary and tertiary) were considered.

Search strategy

The search strategy varied between databases as the strategy was adapted to the specificities of the different database search interfaces. For instance, databases such as PubMed are focused on literature from the healthcare environment and as such did not necessarily require the healthcare component in the search terms whereas others (e.g., Web of Science) did.

An information specialist at the EPPI-Centre was consulted to develop the search strategy. The search strategy was based on formulating the below keywords around the main themes of the review:

1) Trust

Since trust is an ill-defined concept in the literature[37], the search strategy did not only use the word trust but also use related terms such as trustworthiness, credibility, distrust, mistrust and confidence which are often used synonymously to trust.

2) AI

The theme of AI was disaggregated into terms related to AI such as machine learning, intelligent agent, expert systems etc. to capture the multitude of terms used to describe AI applications. Preliminary searches revealed that more specific AI terminology such as natural language processing and artificial neural networks result in literature more focused on the implementation of a specific algorithm rather than an evaluation of an application involving a user and thus were not used to construct the AI term to reduce the noise in the search. However, the terms were not explicitly excluded to allow for potential overlap between specific and more general AI terminology.

3) Healthcare

As mentioned above, some databases did not require a healthcare search term. For those that did, healthcare was described as healthcare, health care or indicated through words such as medic* or clinic*.

Eligibility criteria of included studies

Each study was required to me*et al*l of the following criteria (see Figure 2):

(1) Be set in a healthcare decision-making context; (2) Focus on an AI application and communicate the presence of AI explicitly (e.g., through appropriate language); (3) Be an empirical study investigating the relationship between trust and another variable with trust being the/an outcome variable. (4) The AI application is such that it gathers information, analyses information or provides a recommendation rather than implementing an action (i.e., human remains ultimate authority). (5) Focus on an AI application that is *not* robotics.[1] (6) Evaluate the AI application with regards to users (e.g., user perceptions, experiences of AI) rather than reporting the implementation and performance-based evaluation of the AI.

Failure to meet any of the eligibility criteria resulted in exclusion from the review. Any disagreement between the two review authors (EJ, DK) over the eligibility of a particular study was resolved through discussion with a third review author (JT). The number of excluded studies and the reasons for exclusion were recorded at each stage (see Figure 3).

Data extraction

The following information was documented for each included article: authors' names, year of publication, country of origin, type of publication, type of user, type of AI application, trust concept, factors pertaining to trust in the AI as well as information to conduct a risk of bias assessment (e.g., information on study method, results, conclusions etc.). All records were screened and processed in EPPI-Reviewer 4[38].

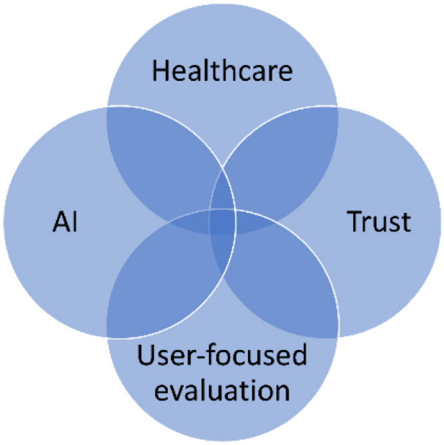

Figure 2. Venn diagram depicting the inclusion criteria. Only studies meeting all four criteria (i.e., intersection of the four aspects) were included.

[1] Robotics was excluded as a) not all robotics is AI and b) the physical presence of robots with AI adds a dimension to trust that lays outside the scope of the current review.

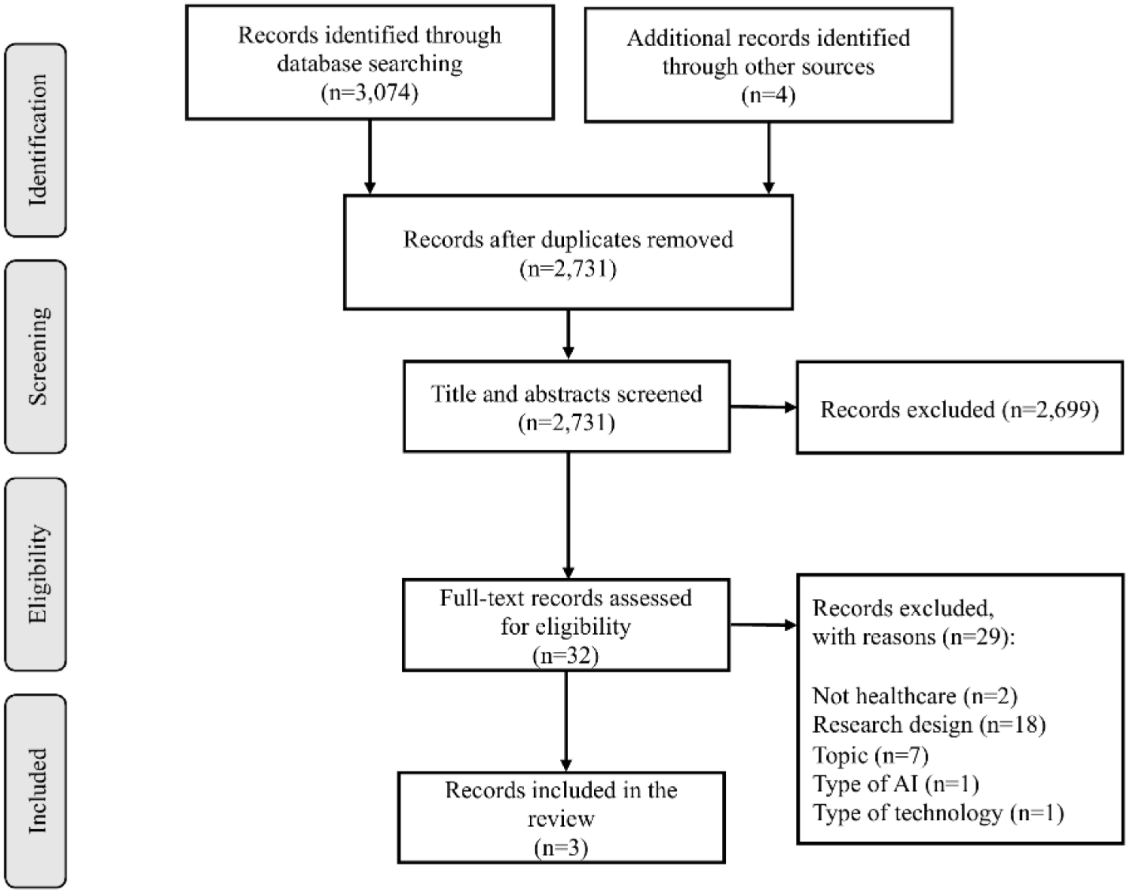
Identification
Screening
Eligibility
Included
Records identified through database searching (n=3,074)
Additional records identified through other sources (n=4)
Records after duplicates removed (n=2,731)
Title and abstracts screened (n=2,731)
Records excluded (n=2,699)
Full-text records assessed for eligibility (n=32)
Records excluded, with reasons (n=29):
Not healthcare (n=2)
Research design (n=18)
Topic (n=7)
Type of AI (n=1)
Type of technology (n=1)
Records included in the review (n=3)

Figure 3. ***P*RISMA flow diagram of study selection process**.

Data synthesis

The study set out to synthesize included studies using framework synthesis driven by a logic model (for review protocol see 39). However, given the nature and number of studies, this approach was deemed no longer appropriate. Narrative synthesis was chosen as an alternative approach allowing to consider and discuss the heterogeneity of included studies. Data analysis followed the steps outlined in Popay *et al.'s* (2006) guidance on conducting narrative synthesis in systematic reviews[40]. The preliminary synthesis involved creating tables for an initial comparison of studies as well as a textual summary of each study including its trust concept, decision context, type of AI, type of user and influences on trust. This was followed by a translation of primary concepts to reflect conceptual overlap. The next step explored the relationships within and between studies by the means of conceptual triangulation. We developed a conceptual model based on the initial logic model which led us to abandon some aspects whereas others became more explicit allowing us to look at different types of relationships between constructs (e.g., influences vs moderators/ mediators).

Results

Study selection

The first author (EJ) reviewed all titles and abstracts resulting from the search. A second reviewer (DK) reviewed titles and abstracts of 10% (181 papers) of the sample once duplicates had been removed. To assess the level of agreement between reviewers, Cohen's kappa was computed as measure of inter-rater-reliability for titles and abstract. The score was .961 signifying a strong agreement between the reviewers[41].

Figure 3 shows a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the study selection process. The search returned 3,078 articles. After removing duplicates, 2,731 articles remained for screening. Of these, 32 were retained for more detailed review. The majority of the papers (18) were excluded after reading the full text as their research design did not match the required criterion, i.e., they did not constitute a user evaluation; were not empirical; did not have trust as one or the outcome variable or only measured the level of trust as a background variable without investigating influences on trust. Three articles m*et al*l inclusion criteria and were included. The authors of included studies were contacted to identify further (grey) literature. All authors responded. Sent papers were assessed for relevance, resulting in the inclusion of one additional record. The relevant paper reported the same study as an already included one but did so in more detail, resulting in four included records reporting on three studies.

Assessment of risk of bias

No tool was decided upon before the review as the criteria and research questions allowed for a range of study types to be included. Upon familiarizing with studies, the Mixed Methods Appraisal Tool (MMAT) was chosen. The MMAT is a recently developed tool that allows researchers to appraise studies of different types with one tool rather than applying several ones. In line with guidance on the MMAT, no scores were calculated. Rather, a detailed report for each included study was written (see *Extended data*[36]). Each study raised some concerns (see completed MMAT table, Table 1). For instance, Cai *et al*., (2019)[42] suffered from insufficient reporting on the quantitative and qualitative components and their integration. In contrast, Fritz (2015)[43] gave a detailed description of their methodology though the way participants were interviewed

Table 1. Scoring of included papers on the MMAT. Please note that the scoring table was adjusted from 36 by a) excluding items that were not applicable (i.e. categories 2 and 4) and b) replacing the response columns at the end by titles of included papers in order to provide the reader with an overview of the quality assessment of all studies.

	Methodological quality criteria	**Cai *et al.* (2019)**	**Fritz (2015)**	**Hengstler *et al.* (2016)**
Screening questions	Are there clear research questions?	Yes	Yes	Yes
	Do the collected data allow to address the research questions?	Yes	Yes	Yes
Qualitative	Is the qualitative approach appropriate to answer the research question?	Yes	Yes	Yes
	Are the qualitative data collection methods adequate to address the research question?	Can’t tell	No	Can’t tell
	Are the findings adequately derived from the data?	Can’t tell	Yes	Yes
	Is the interpretation of results sufficiently substantiated by data?	Yes	Yes	Yes
	Is there coherence between qualitative data sources, collection, analysis and interpretation?	Can’t tell	Yes	Yes with reservations
Quantitative	Are the participants representative of the target population?	Can’t tell		
	Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes		
	Are there complete outcome data?	Can’t tell		
	Are the confounders accounted for in the design and analysis?	Can’t tell		
	During the study period, is the intervention administered (or exposure occurred) as intended?	Can’t tell		
Mixed methods	Is there an adequate rationale for using a mixed methods?	No		
	Are the different components of the study effectively integrated to answer the research question?	No		
	Are the outputs of the integration of qualitative and quantitative components adequately interpreted?	Can’t tell/ Not applicable		
	Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?	Can’t tell		
	Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?	No		

and included in the results was deemed inappropriate. A similar concern arose when assessing Hengstler *et al*., (2016)[44] as there seemed to be a bias towards recruiting people higher up in the company hierarchy, which appeared problematic as there was only one person per case study. Nonetheless, no study was excluded based on the quality assessment.

Overview of included studies

Given the small number of included records, we first provide a description of each study in terms of its type of AI, type of user, as well as the decision context (if present) and their trust concept to illustrate key differences between the papers. The summaries are followed by the synthesized trust influences and an updated logic model.

Cai et al. (2019)

In Cai *et al*., (2019)[42] pathologists were presented with a prototype AI application called SMILY which uses deep neural networks to identify visually similar medical images. The tool helped to inform the following decision context: "When making differential diagnosis, pathologists need to generate hypotheses, compare and contrast evidence for those hypotheses, and then determine which diagnosis is most likely." Pathologists first make a hypothesis and generate a set of alternative hypotheses to rule out. They then consider the hypotheses in light of the information they have (e.g., biopsy) to determine which one is more likely. When they are unsure, they often look at similar images. The AI includes a range of refinement tools to guide the algorithm's retrieval process (e.g., refine by region/example/concept), i.e., offers different options to refine the search for similar images. The authors draw onto dimensions of Mayer *et al*.'s (1995)[45] conceptualization of trust: benevolence and capability. Capability (also ability) is "the group of skills, competencies, and characteristics that enable a party to have influence within some specific domain" whereas benevolence refers to "the extent to which a trustee is believed to want to do good to the trustor".

Hengstler et al. (2016)

Interviewees were not users, but representatives of companies (CEO, Executive Consultant, Group Manager and Director of Solutions and Sales Support) releasing applied AI who were asked about strategies to foster trust in their respective AI. Accordingly, the decision contexts of the four included AI applications as well as the applications itself vary. Since Hengstler *et al*., (2016)[44] focused on the cross-case analysis of strategies fostering trust, we used the collective evidence of the case studies rather than talking about each individual case study and the respective trust influences. It should be noted however, that different AI applications may have different trust influences. The authors take the approach that trust in an AI application requires considering not only trust in the technology but also in the company releasing it and the company's respective communication strategies. As for trust in the technology, Hengstler *et al*., (2016)[44] draw on Lee and Moray's (1992)[12] grouping of factors influencing trust in automation: performance (information describing what the AI does), process (information describing how the AI operates and refers to its understandability) and purpose (why the AI was developed). It should be noted that not all applications were health related, i.e., five of the nine use cases were from the transportation industry.

Fritz (2015)

Participants were adults aged 65 and above who expressed an interest in learning about and being interviewed on the topic of smart homes. The smart home "combines artificial intelligence software with sensor monitoring for the purpose of maintaining safety and health. This smart home learns the residents' motion patterns and can take an action on behalf of the resident living in the home.". Participants did not interact or live in a smart home but rather learned about this particular smart home via textual information. The aim of the study was to gain a better understanding of older adults' "knowledge, perceptions and description of smart home monitoring as these relate to self-identified culturally based expectations" and "to understand the influence of socially constructed predictors and barriers to adoption of smart home monitoring". Trust was not pre-defined as a focal variable but emerged as a theme relevant to older adults' understanding and perception of smart homes. However, the author does not give a definition of what exactly they mean by 'trust'.

Influences on trust in HAI: an updated logic model

The results of the synthesis are summarized in the updated logic model in Figure 4. The new model started with the original model introduced at the beginning of this review and was refined based on the analysis of included studies. The new model no longer explicitly distinguishes between indirect and direct scenarios, given that only one of the included studies involved an interaction, whereas the other two were a hypothetical interaction and a non-user perspective about the user's needs.

In the following, we provide a brief explanation of the model before discussing the different inputs and their respective attributes in more detail.

Explanation of the new logic model

Similar to the initial theoretical model, influences on trust broadly fall into three categories: user, AI application and broader contextual aspects (inputs in Figure 4). Unlike the initial logic model, contextual aspects are not considered a group of inputs but represent different and distinct inputs such as media or social group. Each input is characterised by attributes. For instance, a user is characterised by its age, gender and knowledge of AI, whereas an AI application is characterized by attributes such as its design features or level of transparency. Please note that some attributes were more supported by the included studies than others.

Together, AI and user influence the human x AI interaction. The experience of the interaction influences the level of trust the user places in the AI (trusting judgement) which in turn influences the trusting behaviour (e.g., reliance). These two outputs are further influenced by broader contextual factors – some of which may influence each other. The new model also

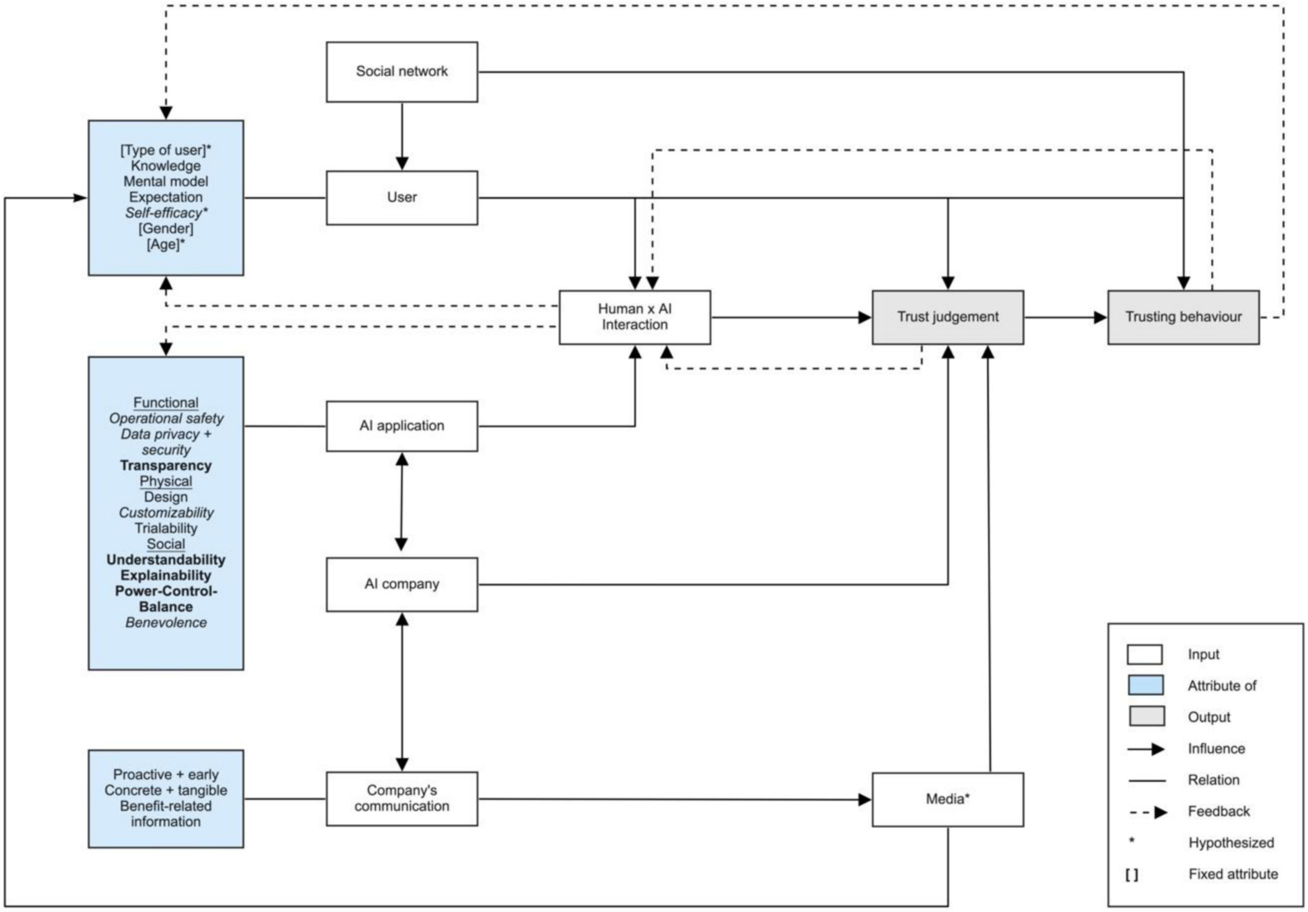

Figure 4. Updated logic model of influences on trust in HAI. Transparent boxes indicate inputs. Inputs are high level factors influencing the trustworthiness judgement. Blue boxes indicate attributes and characterise the respective input. The relation between input and attributes are marked by a black line without arrowhead. Solid black arrows indicate the influence of an input. Feedback loops of outputs (grey boxes) are marked with dotted arrows. Stars indicate hypothesised attributes. Attributes in bold were supported by all included studies; those in italics by two studies and the remaining by one study. Attributes in square brackets (i.e. []) indicate fixed user attributes, i.e., those unaffected by a feedback loop.

shows some likely feedback loops. Feedback loops to the user only affect flexible (e.g., knowledge) but not fixed attributes such as gender.

Overview of inputs and their respective attributes

User attributes

User attributes are factors that influence the trustworthiness judgement at the human level.

Firstly, *Knowledge* (also *AI literacy*) was identified as a user attribute influencing trust in AI. Fritz (2015)[43] observed “Both knowledge and trust levels were low across multiple participants and there were no participants where knowledge and trust were both high or where knowledge was low and trust high.”.

Moreover, the users’ mental model affected trust in an AI application. Mental model refers to the user’s idea of how AI works. If the model did not match the experience of the AI, users distrusted and/ or questioned it. For instance, pathologists in Cai *et al*., (2019) doubted the AI when “their mental model of similarity did not match that of SMILY [the AI]”[42]. Similarly, users’ expectations influenced their trust in the AI.

One study also indicated that age and gender may moderate a users’ trust level. In particular, older participants in Fritz (2015)[43] generally lacked trust in the technology and female participants more frequently reported privacy-related concerns than male participants. While older age may be linked to less familiarity or exposure to and knowledge of AI, the studies gave no indication as to why gender may play a role. It should be noted that Fritz was the only study to obtain input from the public and this was home-based AI technology. *Self-efficacy*, an individual’s belief in their own abilities, emerged as a further factor that may moderate an individual’s trust in AI. Based on Fritz (2015)[43] this appeared to be linked to the level of knowledge people perceived themselves to have. Overall, the included studies indicate that the *type of user* (e.g., personal vs professional) may influence trust in an AI application as it may be linked to different motivations, mental models or levels of AI literacy. We therefore included this aspect as hypothesized characteristic of the user.

Attributes of the AI application

Attributes of AI applications can influence the human x AI interaction and the respective trusting judgement. These attributes may be a physical, functional or social feature of the AI.

Functional features

Functional features refer to functions of the AI. Firstly, the *operational safety* of the AI matters. That is, it should work as intended and meet certain standards[44] that minimize the risk of AI going awry – a concern raised in Fritz (2015)[43]. Moreover, data security and privacy were important for people to trust the AI. In Fritz (2015)[43], this aspect referred to questions about data storage, access and the concern of stolen data. In contrast, Hengstler *et al*., (2016)[44] identified it as a key ingredient of fostering trust, given the sensitive nature of data in healthcare. Transparency also influenced trust in AI. If the application was not transparent about why it made a particular decision, users could not understand the reasoning going into a decision, making them more likely to not trust as stated in Cai *et al*., (2019)[42].

Physical features

We grouped design and customizability as physical aspects of the AI. design influenced trust by communicating the purpose of the AI, i.e., why does it exist, as well as influencing the ease of use. Customizable features influenced trust by allowing users to tailor the application to their needs, thus giving them control (see below).

Social features

Social features describe attributes of the AI that are interwoven with the user. They may also be considered processes occurring in the human-AI interaction.

All studies mentioned the need for understanding as a central influence on trust: If users did not understand the AI application, their trust decreased. Understandability was needed at various levels: in order to trust, users needed to understand the AI's reasoning regarding a decision[42]; understand how to use it (ease of use) as well as understand the purpose of the AI[44]. Company representatives in Hengstler *et al*., (2016)[44] further emphasized the need for understanding the technology to actively foster trust and avoid misunderstandings and misrepresentations. Trialability, i.e., letting people test and interact with an application before the final product, was presented as a strategy to foster knowledge and understandability. Across studies, understandability was closely linked to the need for explanation (explainability in Figure 4). In fact, explanation appeared to be the counterpart to understandability. For instance, Hengstler *et al*., (2016)[44] suggests that if users receive an explanation as to why an AI was developed (i.e. what problem it is solving), it is easier to understand its purpose and application context.

Power and control were two other aspects that appeared in all three studies, resulting in the construction of the Power-Control-Balance (PCB) attribute. PCB refers to the level of autonomy a human has when using an AI application. Included studies indicated a need for balance between giving away power and the user still being (and feeling) in control. In Hengstler *et al*., (2016)[44], company representatives emphasized that users need to know that the AI is under control of a human in order to foster trust. However, Cai *et al*., (2019)[42] suggests that too much control could result in "fear of over-influencing the AI". However, power may not only refer to being in or giving up control but may also be linked to a user's level of knowledge. Specifically, Fritz (2015)[43] suggests that a lack of knowledge of AI can make people feel powerless (i.e. not knowledgeable enough) to make a decision without outside support.

Customizability, i.e., setting the level or type of control in relation to users' needs or preferences, was another attribute that influenced trust. For instance, when users interacted with an AI-driven tool that could be adjusted to user's needs versus one that could not, customizability resulted in higher trust ratings[42]. It also represented a strategy to achieve a PCB.

Finally, studies indicated that benevolence of the AI, defined by Mayer *et al*. (1995) as the "extent to which a trustee [AI] is believed to want to do good to the trustor [human]"[45], plays a role. For instance, in Cai *et al*., (2019)[42], participants trusted the AI more that they rated more benevolent.

The tool that was considered more benevolent was rated as more useful, supportive and requiring less effort. Applying the notion of benevolence to AI, however, is debatable and may in fact point towards the importance and influence of an AI's innovating company (see below).

Contextual inputs

This group of inputs refers to factors that influence the trustworthiness judgement but are not an attribute of the AI or the user.

Innovating company

Fritz (2015)[43] and Hengstler *et al*., (2016)[44] reported the need to consider the innovating company of the AI. The company is inherently linked to the technology and influences trust in the AI application for instance via its reputation. However, the company may also influence trust in AI via the notion of benevolence. For instance, in Cai *et al*., (2019)[42] the AI tool was rated as highly benevolent when reducing workload and adjusting to user preferences. By designing the tool in this way, the company may have covertly expressed its benevolence to users.

Company's communication

Linked to the company is the company's communication which acted as a strategy to practice explainability and achieve understandability. Hengstler *et al*., (2016)[44] indicated that in order to be effective, communication needs to be tangible, concrete and contextualize the AI's benefit.

Media

Media was another aspect mentioned in Hengstler *et al*., (2016)[44] that could help foster trust in AI. For one, media is a channel that a company can use to communicate its technology. At the same time, the mass media are consumed by (potential) users and as such may influence them by shaping their expectations or mental models of AI as well as having an educational role as stated by participants in Hengstler *et al*., (2016)[44]. It may also act as a frame by which we learn about the reputation and benevolence of a company.

Social network

Social network describes the notion that an individual is influenced by input from others. Fritz (2015)[43] indicated that this can take the form of people asking for advice because they do not feel knowledgeable enough to make an independent decision. Similarly, people may get recommendations from others that influence their expectations or mental model of AI. However, social network may also refer to the importance of norms whereby people consider what is socially acceptable.

Discussion

Principal findings

This review highlights influences on user trust in HAI. The results show that these influences broadly fall into one of three categories: user-related, AI-related and wider contextual factors. Overall, the influences appear to be a subset of factors that Hoff and Bashir (2015)[20] found to constitute trust in automation. While Hoff and Bashir (2015)[20] divided the influences into different types of trust (dispositional, situational and learned), we took a different approach by grouping the trust influences not according to trust type but according to input (user, AI, context) in line with the initial logic model.

A prevalent theme throughout this review was the need for understanding aspects or actions of the AI. Technical communities currently focus on 'explainable AI' (XAI), i.e., how the AI is configured. Providing an explanation can prove challenging. For example, the HBCP[6] is using deep learning to make predictions about behaviour change. While such models can make a prediction, they cannot currently provide a rationale as to *why* they have made a particular prediction. Our review supports the notion that understandability, i.e., how a human can understand the AI, may be more important than explainable AI. Similar patterns have been observed in the more general automation literature[46]. Explaining is necessary but insufficient if people cannot understand it. Thus, we need to know what kind and level of explanation different users require. This involves gaining a deeper understanding of users' needs, understandings and mental models of AI. The need for a better understanding of the user should be accompanied by education of users. Currently, people's understanding of AI is broad but not deep and they lack an understanding of data and privacy protection implications[47]. Given the suggested influence of knowledge on trust, we need to educate people on the capabilities and limitations of AI (among other aspects) to increase AI literacy[48].

Literacy may also help the power-control balance (PCB) in the human-AI interaction. According to the cognitive view on trust introduced in Castelfranchi and Falcone (2000)[49], delegating control to a trustee, i.e. the AI, is the origin of perceived risk in the human-AI interaction. Risk in turn brings us back to the heart of trust: trust matters when risk is involved. As pointed out in the results, power may be associated with how knowledgeable users feel. Increasing their knowledge of AI may help them evaluate the AI application and its associated risks more accurately, and in extension, help them feel more in control and empowered.

Explainability and understandability are inevitably linked to a certain level of transparency. All included studies considered transparency an important aspect influencing trust though the way they did so differed. Transparency for Hengstler *at al*., (2016)[44] mainly meant being transparent about the development process of the AI. In contrast, Cai *et al*., (2019)[42] emphasized the transparency of AI's reasoning whereas for participants in Fritz (2015)[43] transparency was about how personal data is used. This highlights that we cannot simply say "AI needs to be transparent". Rather, we need to be specific about what we mean by transparency and to which aspect(s) of an AI application it refers. The same applies to other attributes of an AI application. We need to be clear about which part(s) of a specific application need to be explained, understood or customizable.

Similar to previous research on trust in automation such as Lee and See (2004)[9], our review also found that the wider context of the human-AI interaction matters (e.g. cultural and social influences). In particular, participants rather than the authors in Hengstler *et al*., (2016)[44] indicated that the media is an important influence to consider. The media act as a communication channel for the company while simultaneously acting as a source of information for (potential/ future) users. Previous research revealed that AI issues are highly politicized in the media[50] and that the way a technology is covered in media can affect perceptions of it[51]. In so doing, media can influence our expectations and understanding of AI as well as our trust in its applications. The question is less about if the media play a role, but rather what role they play. How does media coverage of AI influence our expectations, mental models and our trust in HAI? Future research can address this question by investigating the language in media coverage of AI.

Heterogeneity of trust concepts

While the focus of the review was on influences on trust in HAI, we made another observation that is worth discussing in order to advance our understanding and research agenda of AI in healthcare: the use of different trust concepts. Fritz (2015)[43] did not specify what she meant when using the term trust though the context suggests that it is the general idea of relying on technology which the participants (older people) viewed as loss of autonomy. While Hengstler *et al*., (2016)[44] consider trust in AI through the lens of Lee and Moray's (1992)[12] three-dimensional construct of performance, process and purpose, their general trust definition stems from Mayer *et al*., (1995)[45]. Cai *et al*., (2019)[42] also adapted Mayer *et al*.'s (1995)[45] trust dimensions of capability and benevolence. Mayer *et al*. (1995)[45] developed their trust concept in context of trust in organizations and emphasized that their definition of trust is applicable to "relationships with other identifiable party who is perceived to act and react with volition towards the trustor". However, as McKnight *et al*. (2011)[24] pointed out: technology lacks volition. Thus, the question arises as to how appropriate the concept of 'benevolence' and, in extension, Mayer *et al*.'s (1995)[45] trust definition is in the context of AI.

AI's lack of volition may explain why company and application cannot be separated when studying trust. While AI does not have moral agency or volition, the related company does and may therefore act as a proxy. That is, people may think about motives of the company when encountering an application. This might, in parts, explain why participants in 43 voiced concerns about big companies: Company and application are not seen as separate entities. A recent analysis of public perception of AI in medical care supports this notion: distrust of AI companies was one of the main reasons for negative attitudes towards AI on social media[52]. The European Commission's High-level Expert Group on AI also argued that trusting AI involves trusting the technology itself as well as the designers and organizations developing, deploying and using the AI[53]. Yet, companies also have the power to manufacture a benevolent image of the technology itself. For instance by giving it a name such as SMILY. How well the notion of benevolence fits into the world of AI is beyond the scope of this review and for future work. However, the matter illustrates that we need to carefully choose our trust concepts and consider its implications to our research and understanding of trust in AI.

Limitations

A drawback to this review is the small number of included studies. While the included papers provided rich data, the small number of papers limits the generalizability of findings as well as the extent to which concepts could be synthesized. Due to the heterogeneous nature of the papers, there were only a few themes that were supported by all three papers. The papers also only represent some disciplines investigating the topic. Moreover, there are inherent issues in the terminology used for the concepts of trust and AI[2]. Both concepts are characterized by diffuse and inconsistent use of terminology which became apparent when developing the search strategy. The issue is exacerbated by the terminological differences between disciplines. While efforts were made to include various disciplines and their respective terminology, it is likely that the review missed relevant studies. Therefore, the review should be considered as an exemplary rather than exhaustive review on the matter. Finally, we acknowledge that studies on trust in AI have been published since the search phase of this review concluded. However, to our knowledge, these studies do not meet the concepts or criteria of the present review because they are either is not empirical such as 54 or not focused on healthcare [e.g. 55].

Aside from general limitations, we also acknowledge the limits of the synthesis. As outlined in the results section, the included studies applied varying trust concepts. We also did not exclude any studies on the basis of their quality assessment. Yet, all papers had shortcomings in the quality of reporting and/ or methodological issues. Furthermore, while providing interesting data on the commercial perspective on what influences trust in AI, Hengstler *et al*., (2016)[44] show-cased trust strategies rather than critiquing or challenging them. At times, it was also difficult to separate design from human factors forcing us to make a choice about assigning attributes. For instance, we chose to frame two aspects as understandability and explainability. By choosing the ending -ability we frame the two aspects as an attribute of the AI (i.e., we want AI to be explainable). Alternatively, we could have grouped them into user attributes by framing them as need to understand or a need for explanation. While both ways of framing the two constructs bear potential, the authors felt that framing it on side of the AI facilitated their integration into the logic model.

Conclusion

Applications of HAI are becoming more prevalent and have the potential to transform healthcare. To realize this potential, applications need to be used appropriately which in turn relies on trust in these applications. Influences on trust in HAI have so far remained underexplored. In this systematic review, we sought to fill this gap by analysing which AI-related, human-related and contextual factors influence trust in HAI. Overall, two qualitative and one mixed-methods study were included in this review. Trust influences clustered into user-related, AI-related and contextual factors.

The included studies illustrate how different applications of AI in the health domain can be, and, in extension, show why it is not appropriate to talk about AI as if it were one homogenous concept. The findings also indicate that we need to gain a better understanding of the interaction between human and AI in order to foster appropriate trust in the applications. The foundation for this lies in an appreciation of users' needs, understandings, and mental models of AI. This needs to be accompanied by education on AI. Our findings also indicate the importance of broader societal influences that appear to be less accounted for in empirical work. Therefore, future research is tasked with studying the influence of external influences such as the media and AI companies to create a holistic understanding of the trust environment. Finally, the review highlighted issues more general to the study of trust and AI in healthcare such as the need for clear definitions and operationalizations of

[2] For a discussion on issues around a theory of trust in HAI see: F. Gille, A. Jobin, and M. Ienca, "What we talk about when we talk about trust: Theory of trust for AI in healthcare," *Intell. Med.*, vol. 1–2, no. June, p. 100001, Nov. 2020, doi: 10.1016/j.ibmed.2020.100001.

key terms including trust and transparency. Accordingly, this review is a first step to gain a better understanding of which characteristics of an AI system convey trustworthiness by highlighting influences on user trust in HAI, as well as broader issues that need to be addressed.

Data availability

Underlying data

All data underlying the results are available as part of the article and no additional source data are required.

Extended data

Open Science Framework: Influences on User Trust in Healthcare Artificial Intelligence (HAI) - A Systematic Review https://doi.org/10.17605/OSF.IO/GP8RM[36]

This project contains the following extended data:

- Example of search strategy.docx (Search string for Web of Science Core Collection)
- Quality-reports.docx (Short version of quality reports as part of the MMAT process for each included record)
- PRISMA-checklist.doc (Completed PRISMA reporting checklist for systematic reviews)
- PRISMA_flow_diagram.docx (Completed PRISMA flow diagram)

References

1. Harwich E, Laycock K: **Thinking on its own: AI in the NHS.** Reform. January, 2018.
Reference Source

2. Rajkomar A, Dean J, Kohane I: **Machine Learning in Medicine.** *N Engl J Med.* 2019; **380**: 1347–58.
Publisher Full Text

3. Gallagher J: **NHS to set up national artificial intelligence lab.** BBC News. Aug. 08, 2019.
Reference Source

4. Oshri I, Ryan D, Plugge A: **Ready, Set, Failed? Avoiding setbacks in the intelligent automation race**. 2018.
Reference Source

5. Bughin J, Hazan E, Ramaswamy S, *et al.*: **ARTIFICIAL INTELLIGENCE THE NEXT DIGITAL FRONTIER?** McKinsey Global Institue, 2017; 1–75.
Reference Source

6. Michie S, Thomas J, Mac Aonghusa P, *et al.*: **The Human Behaviour-Change Project: An artificial intelligence system to answer questions about changing behaviour [version 1; peer review: not peer reviewed].** *Wellcome Open Res.* 2020; **5**: 122.
PubMed Abstract | Publisher Full Text | Free Full Text

7. AHSN and Network: **Accelerating Artificial Intelligence in health and care: results from a state of the nation survey**. London, 2018; Accessed: May 12, 2019.
Reference Source

8. Fenech M, Strukelj N, Buston O: **Ethical, Social, and Political Challenges of Artificial Intelligence in Health.** *Futur Advocacy Wellcome Trust.* 2018; 56.
Reference Source

9. Whittlestone J, Nyrup R, Alexandrova A, *et al.*: **Ethical and societal implications of algorithms, data, and artificial intelligence: a roadmap for research**. 2019.
Reference Source

10. Lee JD, See KA: **Trust in Automation: Designing for Appropriate Reliance.** *Hum Factors.* 2004; **46**(1): 50–80.
PubMed Abstract | Publisher Full Text

11. Lyell D, Magrabi F, Raban MZ, *et al.*: **Automation bias in electronic prescribing.** *BMC Med Inform Decis Mak.* 2017; **17**(1): 28.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Lee J, Moray N: **Trust, control strategies and allocation of function in human-machine systems.** *Ergonomics.* 1992; **35**(10): 1243–1270.
PubMed Abstract | Publisher Full Text

13. Pop VL, Shrewsbury A, Durso FT: **Individual Differences in the Calibration of Trust in Automation**. *Hum Factors.* 2015; **57**(4): 545–56.
PubMed Abstract | Publisher Full Text

14. Yang XJ, Wickens CD, Hölttä-Otto K: **How users adjust trust in automation: Contrast effect and hindsight bias.** *Hum Fac Erg Soc P.* 2016; **60**(1): 196–200.
Publisher Full Text

15. Powell J: **Trust me I'm a chatbot: How artificial intelligence in health care fails the turing test.** *J Med Internet Res.* 2019; **21**(10): e16222.
PubMed Abstract | Publisher Full Text | Free Full Text

16. Shi ZR, Wang C, Fang F: **Artificial Intelligence for Social Good: A Survey.** arXiv. 2020; Accessed: Feb. 03, 2021.
Reference Source

17. Ferrario A, Loi M, Viganò E: **In AI We Trust Incrementally: a Multi-layer Model of Trust to Analyze Human-Artificial Intelligence Interactions.** *Philos Technol.* 2020; **33**(3): 523–539.
Publisher Full Text

18. Jacovi A, Marasović A, Miller T, *et al.*: **Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI**. 2020.
Publisher Full Text

19. Chopra K, Wallace WA: **Trust in electronic environments.** *Proceedings of the 36th Annual Hawaii International Conference on System Sciences, HICSS 2003.* 2003.
Publisher Full Text

20. Hoff KA, Bashir M: **Trust in automation: Integrating empirical evidence on factors that influence trust.** *Hum Factors.* 2015; **57**(3): 407–434.
PubMed Abstract | Publisher Full Text

21. Merritt SM, Ilgen DR: **Not all trust is created equal: Dispositional and history-based trust in human-automation interactions.** *Hum Factors.* 2008; **50**(2): 194–210.
PubMed Abstract | Publisher Full Text

22. Siau K, Wang W: **Building trust in artificial intelligence, machine learning, and robotics.** *Cut Bus Technol J.* 2018; **31**(2): 47–53.
Reference Source

23. Kini A, Choobineh J: **Trust in Electronic Commerce: Definition and Theoretical Considerations.** *IEEE.* 1998.
Publisher Full Text

24. Mcknight DH, Carter M, Thatcher JB, *et al.*: **Trust in a specific technology: An investigation of its components and measures.** *ACM Trans Manag Inf Syst.* 2011; **2**(2): 1–25.
Publisher Full Text

25. Adjekum A, Blasimme A, Vayena E: **Elements of trust in digital health systems: Scoping review.** *J Med Internet Res.* 2018. **20**(12): e11254.
PubMed Abstract | Publisher Full Text | Free Full Text

26. Selbst AD, Barocas S: **The intuitive appeal of explainable machines.** *Fordham Law Rev.* 2018; **87**(3): 1085–1139.
Publisher Full Text

27. Ciupa M: **Is AI in Jeopardy? The Need to Under Promise and Over Deliver - The Case for Really Useful Machine Learning**. 2017; 59–70.
Publisher Full Text

28. Hancock PA, Billings DR, Schaefer KE, *et al.*: **A meta-analysis of factors affecting trust in human-robot interaction.** *Hum Factors.* 2011; **53**(5): 517–527.
PubMed Abstract | Publisher Full Text

29. Song Y, Luximon Y: **Trust in AI agent: A systematic review of facial anthropomorphic trustworthiness for social robot design.** *Sensors (Basel).* 2020; **20**(18): 5087.
PubMed Abstract | Publisher Full Text | Free Full Text

30. Glikson E, Woolley AW: **Human trust in artificial intelligence: Review of**

empirical research. *Acad Manag Ann.* 2020; **14**(2): 627–660.
Publisher Full Text

31. Kaplan AD, Kessler TT, Brill JC, *et al.*: **Trust in Artificial Intelligence: Meta-Analytic Findings.** *Hum Factors.* 2021; 187208211013988.
PubMed Abstract | Publisher Full Text

32. Cave S, Coughlan K, Dihal K: **'Scary robots' examining public responses to AI.** In *AIES 2019 - Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society.* 2019; 331–337.
Publisher Full Text

33. Felici M: **How to Trust: A Model for Trust Decision Making.** *Int J Adapt Resilient Auton Syst.* 2012; **3**(3): 20–34.
Publisher Full Text

34. Schaefer KE, Chen JY, Szalma JL, *et al.*: **A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems.** *Hum Factors.* 2016; **58**(3): 377–400.
PubMed Abstract | Publisher Full Text

35. Meyer J, Lee JD: **Trust, Reliance, and Compliance.** Oxford University Press, 2013.
Publisher Full Text

36. Jermutus E: **Influences on User Trust in Healthcare Artificial Intelligence (HAI) - A Systematic Review**. 2021.
http://www.doi.org/10.17605/OSF.IO/GP8RM

37. Montague ENH, Kleiner BM, Winchester WW: **Empirically understanding trust in medical technology.** *Int J Ind Ergon.* 2009; **39**(4): 628–634.
Publisher Full Text

38. Thomas J, Graziosi S, Brunton J, *et al.*: **EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis**. 2020.
Reference Source

39. Jermutus E: **Influences on User Trust in Artificial Intelligence in Healthcare: A Systematic Review Protocol.** EPPI Centre, 2020; Accessed: Feb. 13, 2021.
Reference Source

40. Popay J, *et al.*: **Narrative Synthesis in Systematic Reviews: A Product from the ESRC Methods Programme.** *ESRC Methods Program.* 2006; (2006): 93.

41. McHugh ML: **Interrater reliability : the kappa statistic.** *Biochem Med (Zagreb).* 2012; **22**(3): 276–282.
PubMed Abstract | Free Full Text

42. Cai CJ, Reif E, Hegde N, *et al.*: **Human-centered tools for coping with imperfect algorithms during medical decision-making.** *Conf Hum Factors Comput Syst - Proc.* 2019; (**45**): 1–14.
Publisher Full Text

43. Fritz RL: **THE INFLUENCE OF CULTURE ON OLDER ADULTS ' ADOPTION OF SMART HOME MONITORING.** A QUALITATIVE DESCRIPTIVE STUDY BY ROSCHELLE LYNNETTE FRITZ A dissertation submitted in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY College o. 2015.
Publisher Full Text

44. Hengstler M, Enkel E, Duelli S: **Applied artificial intelligence and trust-The case of autonomous vehicles and medical assistance devices.** *Technol Forecast Soc Change.* 2016; **105**: 105–120.
Publisher Full Text

45. Mayer RC, Davis JH, David Schoorman F: **An Integrative Model of Organizational Trust**. 1995.
Reference Source

46. Balfe N, Sharples S, Wilson JR: **Understanding Is Key: An Analysis of Factors Pertaining to Trust in a Real-World Automation System.** *Hum Factors.* 2018; **60**(4): 477–495.
PubMed Abstract | Publisher Full Text | Free Full Text

47. Bristows: **Artificial Intelligence_ Public Perception Attitude and Trust**. 2018.
Reference Source

48. Long D, Magerko B: **What is AI Literacy? Competencies and Design Considerations.** *Conf Hum Factors Comput Syst - Proc.* 2020; 1–16.
Publisher Full Text

49. Castelfranchi C, Falcone R: **Trust is much more than subjective probability: mental components and sources of trust.** *Proc Hawaii Int Conf Syst Sci.* 2000; **64**: 132.
Publisher Full Text

50. Brennen AJS, Howard PN, Nielsen RK: **An Industry-Led Debate: How UK Media Cover Artificial Intelligence**. 2018.
Reference Source

51. Nisbet MC, Scheufele DA, Shanahan J, *et al.*: **Knowledge, reservations, or promise? A media effects model for public perceptions of science and technology.** *Communication Research.* 2002; **29**(5): 584–608.
Publisher Full Text

52. Gao S, He L, Chen Y, *et al.*: **Public Perception of Artificial Intelligence in Medical Care: Content Analysis of Social Media.** *J Med Internet Res.* 2020; **22**(7): e16649.
PubMed Abstract | Publisher Full Text | Free Full Text

53. Hleg A: **Ethics Guidelines For Trustworthy AI**. 2018.
Reference Source

54. Asan O, Bayrak AE, Choudhury A: **Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians.** *J Med Internet Res.* 2020; **22**(6): e15154.
PubMed Abstract | Publisher Full Text | Free Full Text

55. Ashoori M, Weisz JD: **In AI We Trust? Factors That Influence Trustworthiness of AI-infused Decision-Making Processes.** *arXiv.* 2019.
Reference Source