Machine learning and human-machine trust in healthcare: A systematic survey

Review Article

Author(s):
Lin, Han; Han, Jiatong; Wu, Pingping; Wang, Jiangyan; Tu, Juan; Tang, Hao; Zhu, Liuning

Publication date:
2024-04

Permanent link:
https://doi.org/10.3929/ethz-b-000635466

Originally published in:
CAAI Transactions on Intelligence Technology 9(2), https://doi.org/10.1049/cit2.12268

Received: 3 April 2023 | Revised: 1 June 2023 | Accepted: 15 June 2023

CAAI Transactions on Intelligence Technology

DOI: 10.1049/cit2.12268

The Institution of Engineering and Technology WILEY

REVIEW

Machine learning and human-machine trust in healthcare: A systematic survey

Han Lin[1] | Jiatong Han[2] | Pingping Wu[1] | Jiangyan Wang[3] | Juan Tu[4] | Hao Tang[5] | Liuning Zhu[6]

[1]Jiangsu Key Laboratory of Public Project Audit, School of Engineering Audit, Nanjing Audit University, Nanjing, China

[2]School of Information Engineering, Nanjing Audit University, Nanjing, China

[3]School of Statistics and Data Science, Nanjing Audit University, Nanjing, China

[4]Key Laboratory of Modern Acoustics (MOE), School of Physics, Nanjing University, Nanjing, China

[5]Department of Information Technology and Electrical Engineering, ETH Zurich, Zurich, Switzerland

[6]The First Affiliated Hospital with Nanjing Medical University, Jiangsu Province Hospital, Nanjing, China

Correspondence

Pingping Wu and Liuning Zhu.
Email: wupingping@nau.edu.cn and drzln2013@163.com

Funding information

Qinglan Project of Jiangsu Province of China, Grant/Award Number: BK20180820; National Natural Science Foundation of China, Grant/Award Numbers: 12271255, 61701243, 71771125, 72271126, 12227808; Major Projects of Natural Sciences of University in Jiangsu Province of China, Grant/Award Numbers: 21KJA630001, 22KJA630001; Postgraduate Research and Practice Innovation Program of Jiangsu Province, Grant/Award Number: KYCX23_2343

Abstract

As human-machine interaction (HMI) in healthcare continues to evolve, the issue of trust in HMI in healthcare has been raised and explored. It is critical for the development and safety of healthcare that humans have proper trust in medical machines. Intelligent machines that have applied machine learning (ML) technologies continue to penetrate deeper into the medical environment, which also places higher demands on intelligent healthcare. In order to make machines play a role in HMI in healthcare more effectively and make human-machine cooperation more harmonious, the authors need to build good human-machine trust (HMT) in healthcare. This article provides a systematic overview of the prominent research on ML and HMT in healthcare. In addition, this study explores and analyses ML and three important factors that influence HMT in healthcare, and then proposes a HMT model in healthcare. Finally, general trends are summarised and issues to consider addressing in future research on HMT in healthcare are identified.

KEYWORDS

human-machine interaction, machine learning, trust

1 | INTRODUCTION

Machine learning (ML), one of the core technologies of artificial intelligence (AI), is developing by leaps and bounds [1–3]. With the development of technologies, more and more intelligent machines are entering human life and embedded in all aspects of people's production and life. Machines incorporating ML technology are gradually being used in military, healthcare [4–6], education, transport [7] and other fields. The development of technology has also changed the relationship between humans and machines [8]. Most previous studies are based on distrust of AI and unilaterally demand that AI must conform to human ethical standards from all perspectives; however, the trustworthiness of AI and ML is not only determined by its trustworthy algorithms, but also related to both sides of human-machine interaction (HMI). The introduction of trustworthy machine learning can build a bridge between HMI and intelligent machines [9]. The application of ML methods in healthcare is still imperfect so far. ML techniques can have an impact on human-machine trust (HMT) from the machine perspective in HMI. Therefore, it is important to explore ML and HMT in healthcare.

In healthcare, do patients trust doctors [10]? Do patients trust doctors or medical machines more [11]? As the use of ML in healthcare deepens [12], the question of whether medical intelligent machines can be trusted has sparked controversy. Trust plays an important role in HMI in healthcare. For doctors and patients, exposure to medical machines is inevitable [13]. In human-machine collaboration, the maturity of the machine's technical capability is the first premise, then the second premise is the trust relationship between humans and machines [14]. The degree of trust between humans and machines largely affects the effectiveness of human-machine collaboration. Establishing a good trust mechanism [15] can make rational use of the high performance of machines while making use of human intelligence. How to enhance trust between humans and medical machines to make HMI in healthcare more harmonious and comfortable to meet human needs deserves extensive research. In HMI in healthcare, the proper use of the machine depends on both the human and the machine [16]. It is necessary to properly consider the factors that affect HMT and combine trustworthy machine learning to establish a good HMT relationship to make HMI more effective [17].

However, so far, there is a lack of research reviews that systematically and accurately sort out HMT and ML in healthcare. This paper can help researchers understand the important role of trustworthy machine learning in HMI and how to build good HMT, which is important for optimising HMI in healthcare. Based on an extensive literature reading, this paper provides a systematic and in-depth summary of ML and HMT in healthcare, and outlines the development of ML and HMT, leading to a synthetic HMT model in healthcare.

The main contributions of this paper are: (1) exploring the characteristics of ML and HMT in healthcare with reference to previous studies; (2) presenting the problem of HMT in healthcare, exploring ML and the three important factors (machine, human, and environment factors) influencing HMT in healthcare, and discussing and proposing our HMT model in healthcare; (3) exploring promising future research directions for HMT in healthcare. The rest of the paper is organised as follows: Section 2 describes and analyses the concept and development of ML and HMT in detail; Section 3 specifically analyses that trustworthy machine learning in HMI in healthcare can facilitate HMT and gives a trust model; Section 4 concludes and points out the direction of future work.

2 | RELATED WORKS

2.1 | Machine learning

As a rapidly developing technology, ML is widely used in medical intelligent machines. In 1956, the term AI was defined by John McCarthy [18] as the science and engineering of building intelligent machines. This marked the formal birth of the emerging discipline of AI. AI is a branch of computer science that attempts to understand the essence of intelligence and produce a new intelligent machine capable of responding in a manner similar to human intelligence. ML is the main branch of AI technologies, and deep learning (DL) is one of the subfields of ML [19]. The relationship between AI, ML, and DL is shown in Figure 1, where DL is a subset of ML, and ML is a subset of AI.

The goal of AI is to make algorithms simulate 'intelligence'. ML is the core of AI and is a multidisciplinary discipline. ML is the study of how machines can simulate human learning behaviours [20], so that they can learn to acquire new knowledge and skills, and optimise their own machine performance based on what they have learned.

ML algorithms build a mathematical model of sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to perform the task. The general process of ML is to train the input training data, and use the trained machine model to operate on the unknown test data according to its own algorithm and strategy to get the running results. ML has been widely used to solve various complex challenges in various fields such as healthcare, finance and industry. In the healthcare field, ML technologies are used in disease diagnosis [21] and surgical robotics [22] etc. Of course, ML also faces difficulties and challenges. One important issue is the opacity of ML techniques [23], which makes it difficult to explain the inner workings of intelligent machines, and which drives the development of explainable machine learning. The core idea of explainable machine learning is to make a model that considers both prediction accuracy and interpretability [24], and try to find the best balance between the two. It needs to consider not only the accuracy of the model, but also to give the reason for getting that result, and thus to achieve the properties of safety, transparency, and fairness of the model. In short, explainability is the transformation of machine learning from a black-box model to a white-box model.

2.2 | Human-machine trust

Human-machine trust refers to the relationship that occurs between a human and an intelligent machine and is established

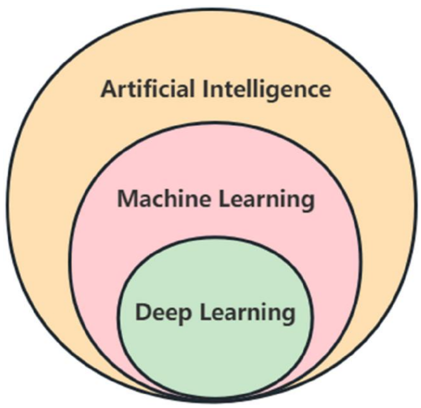

FIGURE 1 The relationship between artificial intelligence (AI), machine learning (ML) and deep learning (DL).

by the criterion of mutual trust between the two parties. With the continuous development and improvement of various advanced technologies, machines incorporating ML and other technologies are widely integrated into cell phones, hospitals, and homes [25]. HMI [26] and HMT have become hot topics of research today.

Advances in technology have driven the development of HMT research. Researchers first focused on the trust relationship between humans. Whereas in HMI, the interaction participants changed from human to human to human to machine, and the research focus changed from interpersonal trust to HMT. The corresponding history of HMT development is shown in Figure 2. As demonstrated by previous studies, there are various influencing factors of trust, which need to be used rationally to improve HMT and promote human-machine cooperation.

The concept of trust is abstract, complex, and has different analyses and definitions in different fields. In 1958, The American psychologist Deutsch introduced the study of trust to the field of psychology with his famous Prisoner's Dilemma experiment. Subsequently, numerous researchers have defined and analysed the concept of trust. For illustration, Rotter defines trust as the general expectation that an individual or group can rely on the verbal statements or written expressions of other individuals or groups in 1967 [27]. Although definitions of trust vary, the consensus view among researchers is that trust is the basis for relationships involving transactions or exchanges. Mayer, Davis, and Schoorman [28] considered characteristics of the trustor, the trustee, and the role of risk, and proposed a definition of trust and a model of its antecedents and outcomes. Mayer, et al. [28] define trust as the willingness of one party to accept the vulnerability associated with the potential risks or unfavorable actions of another party.

In this paper, we employ the classical definition given by Rousseau, Sitkin, Burt, and Camerer [29]. Trust is a state of mind that includes a willingness to take losses based on a positive expectation of another person's intentions or actions. The stronger this willingness is, the greater the trust in the other party. Trust is an important component of interpersonal and organisational behaviour that influences people's decisions about the behaviour of others in personal and organisational settings, and trusting others also means taking the risk of possible harm from the other person's behaviour [30].

Although HMT is different from interpersonal trust, trust also plays an important role in HMI. Muir [31] extended Barber's definition of interpersonal trust and the model of trust from Rempel, Holmes, and Zanna [32] to human-machine relationships, and developed a comprehensive framework for studying automated trust [33]. Trust in human-machine cooperation does not occur directly between humans, but during the interaction between humans and machines. The concept of HMT, which has been accepted by many researchers, was proposed by Lee and See [34]. Based on their definition of trust from an attitudinal perspective, combined with the healthcare scenario we focus on. Trust is the attitude of an individual (e.g. patient) who believes that an agent (e.g. AI therapy robot) can help him/her achieve a certain goal (e.g. disease detection) in an uncertain or vulnerable situation. On the other hand, Siau and Wang [35] considered trust in HMI as human trust in AI systems or trust in AI algorithm developers. Hoffman [36] argued that trust in human-machine cooperation is a change process. An inappropriate level of trust may have negative consequences [34, 37, 38]. Parasuraman and Riley [38] discussed multiple mis-calibrations: misuse, disuse, and abuse of automation techniques. Trust is not always properly calibrated.

In HMT, the trustor is humans and the trustee is machines. It is difficult to generate mutual emotional interaction between humans and machines, so trust is usually generated by instantaneous judgement of humans on machines. The building and sustaining of trust depend on many factors [39]. Jian, Bisantz, and Drury [40] conducted a three-stage experiment that included a word elicitation study, a questionnaire study, and a paired comparison study. They explored three types of trust, including human–human trust, HMT, and trust in general. The results of the experiment identified 12 potential factors of trust between human and automated machines. The 12 factors were used to develop a proposed scale to measure HMT. Trust affects human dependence on machines in HMI [34]. Researchers have found that in HMI, trust is influenced by three main types of factors: human, machine (technology), and environment. Hancock, et al. [41] developed a three-factor model of HMT based on the HMT model by reviewing relevant data and quantitatively analysing the effects of human, machine, and environment characteristics [42]. Schaefer, Chen, Szalma, and Hancock [43] used meta-analysis to assess trust in machines. Human-related factors and automation-related factors provide the moderating effect [43]. The machine performance and attributes were the largest contributors to the development of trust in HMI [41]. In 2011, researchers proposed a MT model of healthcare trust, which abstracts and simplifies the object of study in a healthcare system into a doctor agent, a patient agent, and a central trust agent that perform interactive behaviours with each other [44]. The MT model can integrate direct interactive trust and indirect trust relationships to establish a dynamic equilibrium. Bahtiyar and Çağlayan [45] proposed a model to assess the security trust of e-health services from the perspective of entities. The eHealth trust model is a patient privacy research framework. The eHealth trust model suggests that a patient's eHealth trust is informed by their perceptions, experiences, and environment [46]. The eHealth Trust is the primary determinant in a patient's behavioural reaction to health information exchange. Shareef, et al. [47] proposed a trust-disposition model for understanding adoption of an autonomous homecare system.

3 | MACHINE LEARNING AND HUMAN-MACHINE TRUST IN HEALTHCARE

Machine learning techniques are widely used in the healthcare field, but also face difficulties and challenges. With the widespread application of machine learning in intelligent healthcare field, the issue of HMT is more and more worth exploring.

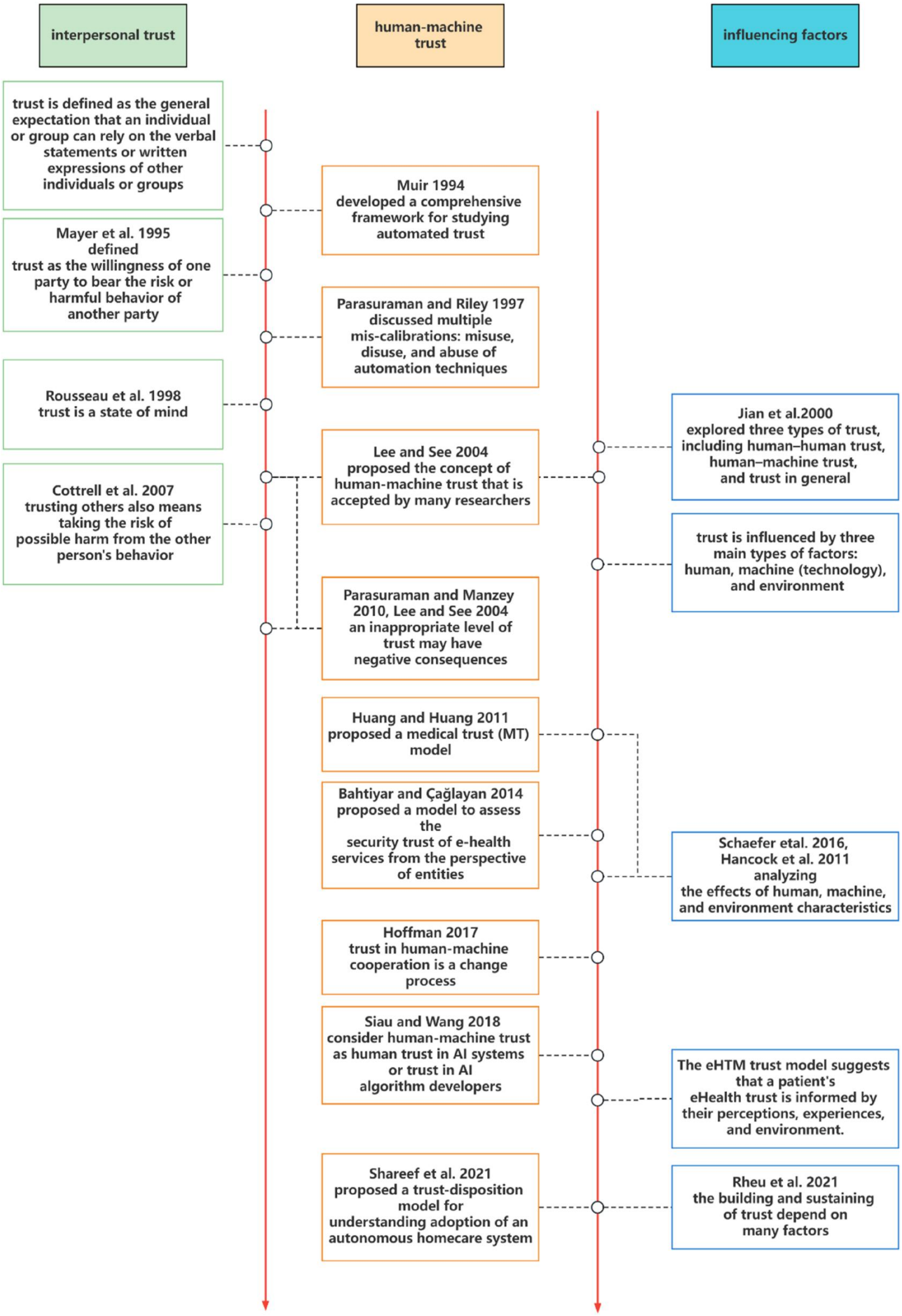

FIGURE 2 Development history of human-machine trust (HMT).

3.1 | Trust issue

Currently, medical information sharing has become a trend in medical informatics. In healthcare, a serious trust security issue stems from data breaches. Once a data breach occurs, users' privacy is not protected and users lose trust in the machine. In 2015, the U.S. health insurance company Anthem was hacked to steal the personal information of more than 80 million customers and employees. In the same year, UCLA's healthcare system was hacked and about 500,000 medical data were compromised. In 2017, a system failure at a U.S. healthcare company resulted in a data breach of approximately 47.5 GB involving sensitive private information of approximately 150,000 patients. The security of healthcare data is a matter of personal privacy and healthcare organisations. Without data security and privacy protection, it is impossible to gain the trust of users (doctors and patients).

In the healthcare field, another serious trust security issue is misdiagnosis. Although ML can help ensure a more accurate diagnosis and reduce the rate of misdiagnosis to some extent, it still has the potential for misdiagnosis. As it is known, the famous IBM Watson Health had sometimes given wrong recommendations for cancer treatment, prescribing drugs that may cause bleeding for cancer patients who are already bleeding heavily [48]. Misdiagnosis can occur if there are problems with medical diagnosis. The consequences of misdiagnosis in the medical process can be very serious. In the healthcare field, incorrect diagnostic decisions from machine learning models can threaten the lives of patients. Fortunately, Memorial Sloan-Kettering Cancer Centre gave the explanation that the aforementioned cancer patient was only fictitiously created to train Waston in tumour diagnosis and was not a real patient, and no one was harmed as a result.

In addition, the unexplainability of medical diagnostic machines that incorporate machine learning poses a significant challenge to trust. On the one hand, it reduces the trustworthiness of the models and makes it difficult to build trust between users and machines. On the other hand, an uninterpretable model is extremely limited in practical deployment in many domains because it does not provide more reliable information to users. When machine learning models are unexplainable and it is impossible to determine how the models make decisions, doctors are afraid to easily use the results provided by ML for diagnosis, and patients do not trust the diagnosis [49].

So how to build a good HMT relationship? Multiple perspectives need to be considered in HMI, including human, machine, and environmental perspectives. In the next section, taking into account various factors of HMT, a synthetic HMT model is proposed to make HMI in healthcare more effective.

3.2 | Human-machine trust model

Trust as a multidimensional and dynamic construct has been extensively studied in its models or frameworks in the psychological field. In interaction, the cognition and behaviour of human decision makers are influenced by multiple dynamic and uncertain factors [50, 51]. Conducting evaluations can help understand collaborative decision-making processes [52] and drive machine design and human-machine collaborative design forward in healthcare.

HMT in healthcare involves the collaboration and integration of humans and intelligent machines to improve patient care, clinical decision making, and medical outcomes. Various approaches and solutions have been explored to rationalise the use of HMT in healthcare. A clinical decision support system integrates ML algorithms and clinical knowledge to provide real-time decision support to healthcare professionals. It can help diagnose diseases, select appropriate treatment options, and predict patient outcomes. These systems combine patient data, medical literature, and expert guidelines to provide evidence-based recommendations that reduce errors and improve the efficiency of clinical decision making. HMT-based telehealth platforms can facilitate remote collaboration between healthcare professionals and patients. These platforms utilise video conferencing, wearables and remote monitoring tools to enable virtual consultations and remote patient monitoring. Robotic systems are used in surgical procedures under the guidance of human surgeons to improve precision, dexterity, and minimally invasive techniques. Surgeons collaborate with robotic systems to perform complex procedures with improved precision and reduced risk. Intelligent assistive technologies, such as exoskeletons or robotic prostheses, can help physically disabled or injured patients regain mobility and independence. These technologies all require trust as a prerequisite, working in coordination with individuals to amplify their capabilities and provide support. HMT-based health monitoring systems use machine learning algorithms to analyse patient data and detect early signs of deterioration or disease progression. By continuously monitoring vital signs, biomarkers and other patient data, these systems can generate predictive models to anticipate adverse events, enabling healthcare professionals to proactively intervene and improve patient outcomes. These approaches highlight the potential of HMT to improve healthcare delivery, patient outcomes, and decision-making by healthcare professionals. It is critical to ensure that these HMT systems are carefully designed, validated, and integrated into existing healthcare workflows, taking into account human factors, ethical considerations, and regulatory requirements to ensure their safe and effective use in healthcare.

Trust plays an important role in HMI in healthcare. Based on the analysis discussed above, we explain and analyse the influencing factors of HMT in healthcare from four perspectives: machine (technology), human, interaction, and environment, and propose a synthetic HMT model in healthcare as shown in Figure 3. The model is an interaction trust mechanism model based on both machine and human interaction subjects in healthcare, and integrating interaction environment and interaction behaviours. This model can be applied to healthcare to improve the trust between humans and machines during healthcare interactions, thus contributing to a better development of the healthcare field. From the machine perspective, trust is closely related to machine technology. We fully consider machine anthropomorphism and trustworthy machine learning to incorporate into our trust model [53]. From the human perspective, we divide them into two categories: healthcare professionals, patients and attendants. We include their age, professional level and experience in the model separately to analyse their effects on trust. Among the interaction environment factors, we mainly consider the impact on trust from the perspective of legal and ethical. Interaction behaviour, meanwhile, occurs during HMI, which affects HMT and is a dynamic process. Different from general HMI, the main participants are

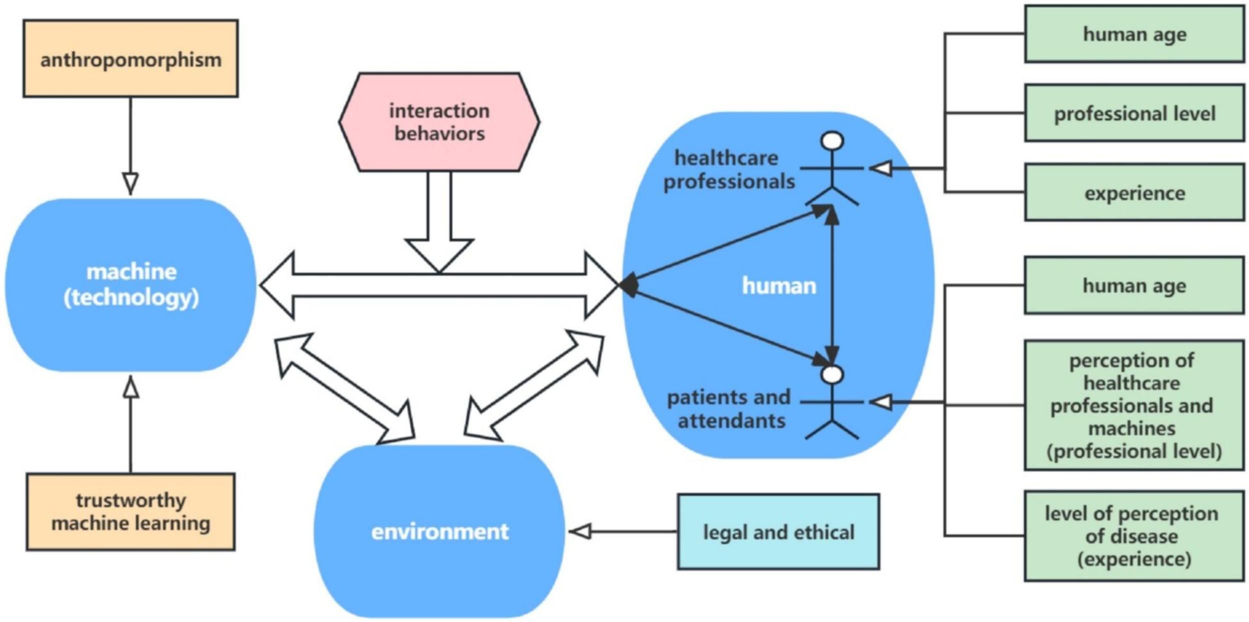

FIGURE 3 Human- machine trust model.

the healthcare machines and the people using the healthcare machines in HMI in healthcare. Nowadays, intelligent machines are widely used in various medical scenarios [54], and we are committed to propose the HMT model to promote trust in HMI and improve medical interaction relationships. By fostering the trust of healthcare professionals and patients to increase their acceptance of intelligent machines, we aim to establish a collaborative human-machine healthcare system with mutual trust and reliability [55], high quality and efficiency in the process of healthcare intelligence. The HMT model in healthcare proposed in this paper has the following implications and effects. First, it provides a reference for the design and development of healthcare machines. It can provide help for researchers to select trustworthy machine learning techniques and provide a theoretical basis for developing trustworthy healthcare machines. Second, it solves the problem of healthcare professionals' and patients' trust in machines in healthcare HMI based on specific trust influencing factors. Only if the healthcare professionals trust the healthcare machines used, the application can be better. And it can promote patients to trust the machine. Finally, it can provide suggestions for building a good external healthcare environment, which can be used to establish trust norms in healthcare to protect the power of healthcare organisations and patients, and thus improve the HMT in healthcare. Below are the specific trust influencing factors in our model.

3.2.1 | Machine (Technology)

From the machine (technology) perspective, trustworthy machine learning models should be built. The trustworthiness of machine learning techniques can be enhanced from six perspectives, including privacy, transparency, explainability, reliability, accuracy, and autonomy as shown in Figure 4.

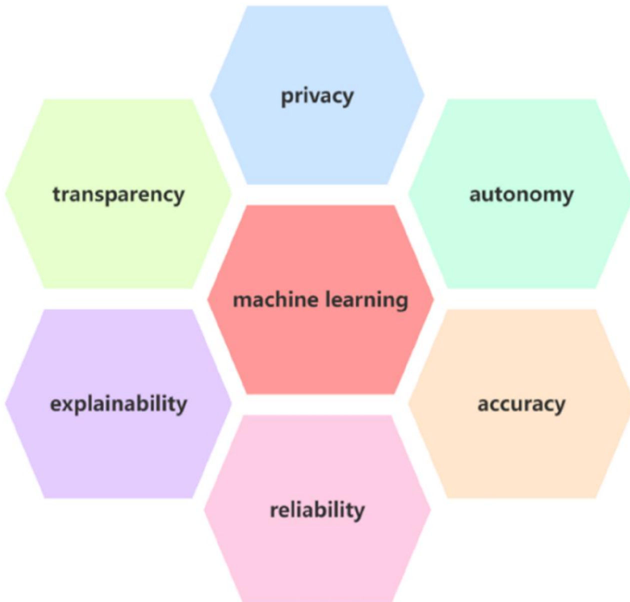

FIGURE 4 Trustworthy machine learning (ML).

Privacy. Do a good job of protecting user data privacy.

Machine learning techniques need to be backed by good trust technology, which in turn ensures privacy. And the commonly applied trust technology is blockchain technology. Blockchain technology can be used for privacy protection [56]. It can break the original more closed medical environment and facilitate the collection and use of medical data and information. And it can protect information security to a certain extent. Blockchain can provide a trusted data. The greatest quality of blockchain is decentralisation, creating trust from mistrust. Blockchain can be used to solve the problem of data privacy and security [57]. First, blockchain is built on asymmetric encryption, hash functions, secure multi-party computing, smart contracts and other technologies, and has a deep technical foundation. Second, blockchain can solve the problem of online data privacy and security [58]. Two main features of blockchain are the immutability of data and decentralisation. This also makes the information collected and stored by blockchain more authentic and reliable, which can help solve the problem of people distrusting each other. Decentralisation can eliminate centres of trust and create chained networks of trust. Smart contracts on the blockchain provide a trusted way of computing [59]. Smart contracts are written to the blockchain, and a series of operations such as storing, reading, and executing data are guaranteed by the characteristics of the blockchain, and the execution process is transparent and traceable. A successfully constructed smart contract waits for the fulfilment of conditions and then automatically executes the contents of the contract. This can solve the problem of trust between users, between users and medical institutions, and between users and machines.

Transparency and explainability. Improving transparency and explainability.

First, ML provides advice or helps make decisions, then the people making the decisions need to understand why the machine is giving them that advice and choose whether to take it. For example, when a doctor uses ML to diagnose a disease, he or she must be able to understand why the medical diagnostic machine is making such a recommendation and decide whether or not to adopt it. Second, for patients affected by ML techniques, they need to be able to understand the decisions made by the ML model in order to choose whether to accept the treatment plans proposed by ML. Third, for developers, understanding the black-box models of ML can lead to better research, improved methods and models, and increased system capabilities. Last but not the least, for the general public, it is essential to popularise the relevant scientific knowledge to them, so that they can gradually accept it.

Transparency is the degree to which machine behaviour can be understood and predicted [60]. Improving transparency in human-machine cooperation is important [61]. Studies have found that bilateral transparency helps foster trust [62]. Fischer, Weigelin, and Bodenhagen [63] investigated the effect of transparency on HMT in blood pressure measurement scenarios. The results showed that improved transparency increased patient's trust.

There is a need to make the 'black box' transparent and to enhance the explainability of machine behaviour. How algorithms derive the conclusion 'whether the patient has cancer' from 'the pixel values of a medical image' is a process that is often difficult to explain to doctors and patients [64].

Explainable artificial intelligence (XAI) attempts to provide explanations for black-box models in ML that we humans can understand. It aims to enable more transparent and interpretable machine learning to help users better understand machine behaviour as well as make predictions and decisions [65], improving the reliability and trustworthiness of machine learning [66]. In recent years academics have conducted extensive and in-depth research and proposed explainable machine learning models that can improve the explainability and transparency of machine learning models and enable a trust relationship between the user and the decision mode [67]. In addition, some models give appropriate explanations along with decisions to gain users' trust and understanding [62]. The Google team has developed an interpretable medical diagnostic machine for cardiovascular disease [68]. It can generate heat maps that show the pixels in medical images that have a significant impact on the diagnostic results. Zhang, Xie, Xing, McGough, and Yang [69] proposed a semantically and visually interpretable medical image diagnostic network. Researchers proposed a post hoc interpretable system for Doctor XAI. Kang Zhang developed a medical diagnostic machine for retinal diseases and pneumonia [70]. The machine is interpretable, and when it gives a diagnosis, it shows people the areas activated by the neural network during the decision-making process, showing the basis for its decision.

For some traditional machine learning models such as linear regression, logistic regression, decision trees etc., the models are relatively simple. It can be understood the internal structure of the model and the internal parameters of the model in the process of using. Machine model results are more explainable.

However, for some complex machine learning models such as neural networks, their internal structures are very complex, and it is difficult for us to observe the changes of data information by layer-by-layer neural networks or neuron-by-neuron. The machine model results are less explainable.

The two types of models are explained in different ways. As shown in Figure 5, to improve the explainability of the models, we can use the following two approaches: ex post facto explanation for complex models [71, 72], and for some traditional models, explanation from the internal nature of the model [73–75]. Intrinsically explainable models, such as decision tree quantitative explanation [76] and decision tree regularisation [77]. The complexity of the model structure can be reduced, such as for random forests, by reducing the depth of the tree model [78], sacrificing the accuracy of the model for the explainability. In addition, the model can be visualised. For example, for neural network models, the original accuracy of the model can be maintained, and after the model is trained, post-hoc assisted attribution resolution methods and visualisation tools are used to obtain the explainability of the model. Visualisation tools are CNNVis [79] and Lucid [80] which is a neural network visualisation library built on Deep Dream etc. Visualisation techniques can help to address the transparency interpretability factor. The space of ML model may be high level spatial, which humans cannot imagine and understand. High-dimensional abstract information can be transformed into readable and interpretable information through visualisation techniques.

Our discussion and analysis can be summarised in Table 1. As shown in Table 1, some popular explainable machine learning and their explanation types with specific methods are

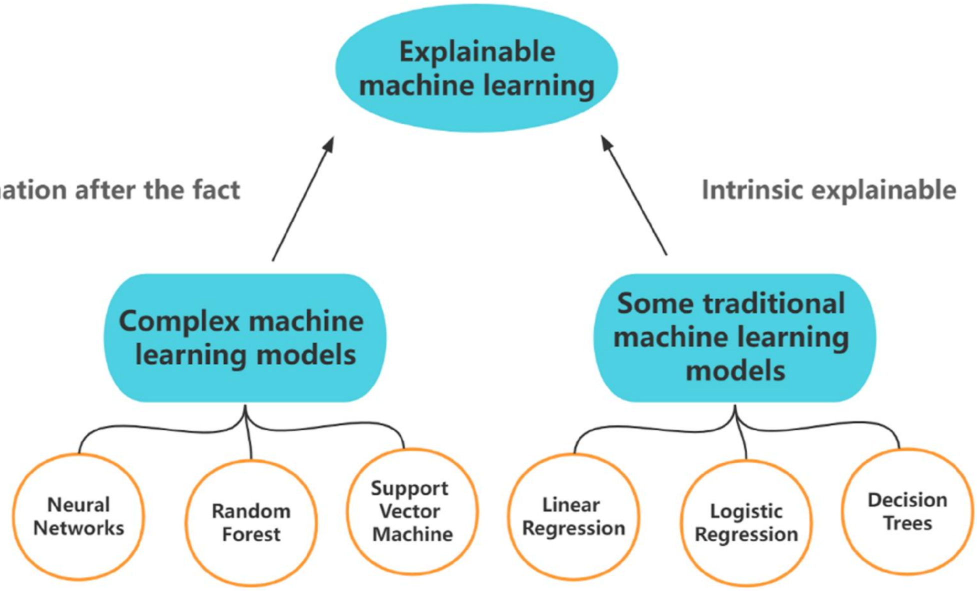

FIGURE 5 Explainable machine learning (ML).

TABLE 1 Some analysis of explainable Machine learning (ML).

Machine learning	Paper	Type of explanation	Method
Neural networks	Bach, et al. [72]; Liu, et al. [79]; Olah, et al. [80]	Explanation after the fact	Using visualisation tools such as CNNVis, Lucid etc. to obtain the explainability of the model after it has been trained
Random forest	Zhao, et al. [78]	Explanation after the fact	Reducing the complexity of the model structure
Support vector machine	W. Zhang, et al. [71]	Explanation after the fact	Providing an explanation of the structural risk of the model
Linear regression	Caywood, et al. [75]	Intrinsic explainable	Viewing parameter or feature statistics within the interpretable model
Logistic regression	Itani, et al. [73]	Intrinsic explainable	Using feature weight estimation to account for the model
Decision trees	Q. Zhang, et al. [71]; Ming, et al. [74]; Moraffah, et al. [77]	Intrinsic explainable	Reflecting the decision process and data characteristics directly, quantitative interpretation of decision trees, decision tree regularisation

listed. While these techniques and approaches enhance the explainability of DL models, it is important to note that complete explainability and transparency are still challenging to achieve. DL models are highly complex, and their decision-making is influenced by numerous interconnected parameters. DL techniques, particularly neural networks, are often considered as black boxes because they can be difficult to interpret and understand. The high complexity and non-linearity of DL models make it challenging to directly interpret their decisions and reasoning processes. The level of explainability achieved also depends on the specific architecture, data, and problem domain.

While ML models are generally considered less explainable compared to traditional techniques, there has been ongoing research and efforts to make ML models more explainable and transparent. Ongoing research and the development of XAI methods are making progress in improving their explainability. XAI is a field of study that aims to develop techniques and methods to provide insights into the decision-making process of complex ML models. And improving explainability is important for humans to understand intelligent machines and achieve better medical human-machine cooperation during healthcare interactions [81]. In this paper, we summarise some current research on methods to enhance explainability and make ML as trustworthy as possible. In addition, explainability enhancement is performed by studying the explainability of the model themselves and post hoc explanations. Moreover, improving the explainability and transparency of DL models will help debug models, guide future data collection, provide truly reliable information for feature construction and human decision making, and ultimately build trust between humans and machines.

Accuracy and reliability. Improved accuracy and reliability can lead to good machine performance.

Good performance is crucial to the trustworthiness of medical machines. The algorithm performance should be improved to enhance the accuracy of medical virtual machines such as Dr. Watson. Researchers at Stanford University had developed an algorithm that could perform melanoma diagnosis better than dermatologists, but with an accuracy rate of less than 75%. Although it can provide effective suggestions to doctors, there is a possibility of misdiagnosis. Improvements to the algorithm are needed before it can be used. Otherwise, the accuracy rate is too low to gain the trust of both doctors and patients. Although users do not necessarily trust the diagnostic results just because it is highly precise, high precision is a guarantee for the credibility of a virtual machine such as Dr. Watson. The Da Vinci surgical robot has been developed and iterated and updated. The fourth and fifth generation of the robots are now introduced to China and can be put into clinical use. Compared with the previous version, the latest Da Vinci surgical robot has a clearer vision, more stable operation, and more flexible operation. Its mechanical arm is thinner and longer and has a greater range of motion, allowing it to rotate 360°. It can help doctors complete difficult and risky surgeries. Continued technological and algorithmic improvements in physical machines can make HMI more effective and thus gain the trust of users (doctors and patients).

While DL techniques can be effective in processing large amounts of data and extracting patterns, there are considerations to be made when applying them to individualised predictions or monitoring. At the same time, we need to consider the impact of data type on reliability. Since different individuals have different physical conditions, whether the approach of using DL based on large data for medical diagnosis such as disease prediction is reliable and whether its diagnosis results are trustworthy.

DL models are highly dependent on the quality and representativeness of the training data. If the training data predominantly represents a certain demographic or specific subgroups of individuals, the model's performance may be biased towards those groups. It is crucial to ensure that the training data adequately represents the diversity of individuals to avoid biased predictions for specific populations. Individuals exhibit variations in their physiology, genetics, lifestyle, and other factors that influence their health. DL models trained on large and diverse datasets may capture general trends, but they may not be able to capture individual variations accurately. The generalisation of the model's predictions to specific individuals should be done cautiously, taking into account individual characteristics and context.

While employing big data-based DL methods for individualised predictions or monitoring presents challenges [82], it is possible to address these concerns through careful consideration of data representativeness, personalised models, transfer learning, interpretability, and feedback mechanisms. To address the challenge of individual variability, personalised DL models can be developed. These models can be trained on data specific to an individual or a smaller cohort that represents their characteristics more accurately. Personalised models can leverage a combination of big data and individual-specific data to provide more reliable predictions and monitoring. Transfer learning is a technique where a pre-trained DL model is used as a starting point and fine-tuned on a smaller dataset specific to an individual or a subgroup [83]. By leveraging the knowledge learned from big data, this approach can be used to adapt the model to an individual's unique physiology and improve its reliability for individual predictions. To build trust and ensure the reliability of predictions, it is crucial to incorporate interpretability techniques into DL models. By providing explanations or highlighting the factors contributing to a prediction, individuals can better understand the model's reasoning and assess its reliability. Additionally, incorporating feedback mechanisms allows individuals to provide input and correct any inaccuracies, enhancing the reliability of the monitoring or prediction system. These approaches can help improve the reliability and trustworthiness of ML-based systems when applied to individual physiology.

Medical machines not only need to be trained with large amounts of data, but also need to be personalised for different medical individuals to assist in treatment [83]. That is, during the training process, researchers need to train the characteristics for different patients, based on their individual medical data. ML techniques require not only large amounts of data, but also small data about an individual. Generally, for data that do not require consideration of individual differences, we use machine learning methods based on large amounts of data [82], while for targeted diagnosis we use a combination of individual data and large amounts of data for prediction and diagnosis. The outcome of the same disease may vary greatly depending on factors such as the patient's age, physical condition, and the period of disease development in which the patient is receiving treatment. Targeted treatment approaches are needed for different patients. Disease prediction can be performed based on information about multiple patient characteristics and used to guide personalised medicine. For example, skin cancer detection is performed by inputting skin image data of patients using a trained model [84]. With electronic medical data, the risk of an individual developing various diseases is assessed using noise-reducing autoencoders [85]. The long-term records of patients' personal medical information can provide enough personalised training for the machines to improve their diagnostic capabilities, provide safe and reliable technology, and increase trust in medical machines.

Big data-based machine learning learns large amounts of data and can provide more robust and reliable predictions because they have been trained on different examples. Small data-based machine learning focuses on using specific data from individuals or smaller groups. This approach recognises the uniqueness of physiological characteristics and individual differences. By incorporating personalised or individual-specific data, it is possible to create models that are better suited to the specific needs of individuals and take into account their unique physiology. In practice, the choice between big data-based or small data-based machine learning depends on the specific application, the available data, and the desired level of personalisation. A combination of the two approaches, such as transfer learning or personalised models trained on a combination of big data and individual-specific data, can strike a balance between general trends and personalised predictions, providing reliable and customised insights for individual physiological monitoring or prediction.

In addition, the stable performance of the physical machines will also increase user's trust [86, 87]. Machines that consistently perform well are more likely to be trusted than those that perform poor [14, 37, 88, 89]. Therefore, there is a need to improve technical accuracy and reliability to make the machine perform well and thus gain trust in HMI in healthcare.

Healthcare organisations are focussing research on big data analytics, disease diagnosis, risk prediction and quality of care monitoring. In healthcare, ML plays an essential role. And the results launched by ML models are related to the quality of care and have an impact on patient safety. Therefore, for healthcare organisations, there is a need to establish a common performance metric to evaluate the performance of different ML techniques and their application effectiveness for model selection. In order to choose trustworthy machine learning, we should consider a combination of accuracy, reliability, and explainability [90]. ML methods should not be chosen solely for their high accuracy. Hence, a compromise option that takes into account multiple key indicators is required. The more complex the model, the less generalisable the model is and the less explainable it is. We need to use models with different levels of complexity depending on the specific medical application. In addition, data can also affect the reliability and accuracy of the model. Therefore, in the selection process, ML methods should be selected for different types of medical data. Moreover, the intelligent machines should be tested and optimised during the medical use, which in turn will better improve the HMT.

Some recommendations for healthcare organisations to guide their evaluation and selection process are summarised below:

(1) Data privacy and security perspective: Evaluate the data governance practices for ML technology. Ensure that it complies with relevant data protection regulations and follows best practices for data security, privacy and confidentiality. Evaluate the effectiveness of technologies used to anonymise and protect sensitive patient data during the training and inference phases. Ensure that appropriate measures are in place to mitigate privacy risks. (2) Accuracy and reliability perspective: Evaluate the performance of ML algorithms by validating them on independent datasets and comparing them to the state-of-the-art methods available. Look for evidence of high accuracy, robustness, and generalisability across different populations or environments. (3) Explainability perspective: Prioritise ML models that provide transparency and explainability. Models with clear explanations of their decision-making processes can help build trust and promote human understanding of the underlying factors that contribute to predictions. (4) Increase clinical validation: Work with healthcare professionals and domain experts to evaluate the clinical validity of ML technology. Assess whether the predictions or results provided by the technology are consistent with established medical knowledge and guidelines.

Autonomy. As autonomy increases, HMT decreases.

Autonomy can be described simply as the ability of a machine to perform tasks independently. The U.S. Department of Defence has identified four levels of autonomy, including human operated, human delegated, human supervised, and fully autonomous [91]. There are two tasks for humans and machines: developing protocols and selecting actions. The division is based upon whether the task agency is in the humans or in the machines [92]. Murray, Rhymer, and Sirmon [93], Raisch and Krakowski [94] explored the level of machine automation of the HMI process.

Levels of automation may complicate the human trust in machines [95]. Higher levels of automation are not better. One study found that higher levels of automation machines may lead to human mistrust. The higher the level of automation of the machine, the harder it is for the user to understand the machine, which may lead to a decrease in trust [96]. Users tend to trust better-controlled machines [97]. In addition, in some cases, adaptive automation can be an effective solution to the problem of trade-offs between different levels of machine automation [98]. In healthcare, in general, as the level of automation increases, the trust of doctors decreases and the trust of patients decreases. However, perceptions of the level of automation can change depending on factors such as the age and experience of the doctors and patients, which in turn can affect the level of trust.

Anthropomorphism. Add anthropomorphic features to the machine.

Machines should be appropriately anthropomorphised, not anthropomorphised the higher the better.

Adding anthropomorphic features can improve the trust of machines. There are many studies on the role of anthropomorphism in HMT. Waytz, Heafner, and Epley [99] found that when self-driving cars have anthropomorphic features [100], such as name and gender, humans are more confident that the vehicle is up to the task of autonomous driving [101]. Thus, the vehicles can be better put into use. Similarly, anthropomorphic features can be applied to HMI in healthcare to improve patients' trust in machines. In the field of machine behaviour, face trustworthiness can be used to improve communication and interaction between medical machines (such as medical guide robots) and humans [102–104].

Ramachandran et al. modified the personality design of the robot to make it more useful in a healthcare setting, with animated eyes, a voice with a local accent, and polite contextual phrases to mimic the behaviour of a nurse when interacting with a patient. The results of the experiment showed that if the robot communicated with the user in a polite and friendly manner, this would increase the user's trust in it. When the robot behaved in a stricter manner, the user's trust in the robot decreased. And psychological anthropomorphic robots that exhibit empathy have higher participant acceptance. Different facial expressions and design elements of the robot can affect HMT. The Uncanny Valley is a hypothesis about how humans feel about robots and non-human objects. When the similarity between robots and humans reaches a certain level, human reactions to robots suddenly become extremely negative and repulsive, and when the similarity between robots and humans continues to rise to the level of similarity between ordinary people, human emotional reactions to them will return to positive. In human-robot interaction, machines that are human-like in appearance are perceived to be friendlier and more reliable, while non-human machines are perceived to be colder. Broadbent, et al. [105] argue that appearance can influence the interaction between machines and humans, and that a robot with a more human face display is more trusted. Researchers found that machines with a distinctly non-human appearance were more popular than machines that were extremely human-like [106]. The Uncanny Valley Effect is also an important reference for designing the degree of machine anthropomorphism.

As shown in Table 2, typical machine influencing factors and their targets in HMI in healthcare are listed. In addition, HMT in healthcare can be improved by taking different methods for these influencing factors. By studying the relevance between influencing factors and HMT in healthcare, better machine design development and use can be carried out to improve trust in healthcare and human-machine cooperation. Machines with anthropomorphic designs can better interact with healthcare professionals and patients. For example, Pepper is a social humanoid robot, 120 cm tall, with an anthropomorphic design and body language [107]. It can synthesise its surroundings, actively response and interact with healthcare professionals and patients. Anthropomorphism can make patients trust the machine more, which in turn allows the healthcare robot to fully play its role as an aid to treatment and companionship.

3.2.2 | Human

As shown in Table 3, typical influencing factors in HMT in healthcare are listed. HMT is a two-sided relationship, and in spite of machine influencing factors, we need to fully consider human influencing factors and incorporate such features into the HMT model in healthcare. Human age affects the user's trust in the machines during such interactions [115]. Children have grown up with intelligent machines, whereas adults did not. Adults have more confidence in rejecting the machine's advice. Children have not yet developed this kind of confidence. Studies have shown that older people trust automation more than younger people [109]. The specific effects of age on trust may vary with the situation [108]. In addition, professional level and experience can affect healthcare trust.

For healthcare professionals, increasing the healthcare professionals' professional level usually contributes to their trust in machines [112]. The higher the level of profession, the less likely healthcare professionals are to rely on machines [110, 111]. The experience of healthcare professionals can influence

TABLE 2 Some analysis of machine influencing factors in human-machine trust (HMT) in healthcare.

Paper	Influencing factors	Target	Method	Relevance
Sharma, et al. [56]; de Moraes Rossetto, et al. [57]; Mathis, et al. [58]; Al Omar, et al. [59]	Privacy	Protecting data and user privacy and security	Using technologies such as blockchain to provide a trusted data and a trusted way of computing	Increasing
Chien, et al. [60]; Ishowo-Oloko, et al. [61]; Edmonds, et al. [62]; Fischer, et al. [63]	Transparency	Turning a black-box model into a white-box	Enhancing the transparency of the machine	Increasing
Edmonds, et al. [62]; Nazar, et al. [65]; Glikson and Woolley [66]; Eshete [67]; Poplin, et al. [68]; Zhang, et al. [69]; Adadi and Berrada [70]	Explainability	Turning a black-box model into a white-box	Improving machine interpretability and predictability	Increasing
Yin, et al. [14]; Parasuraman and Manzey [37]; Aggarwal, et al. [88]; Buchlak, et al. [89]	Accuracy	Making the machine have good performance	Improving machine model accuracy	Increasing
Mathis, et al. [58]; Kraus, et al. [87]	Reliability	Making the machine have good performance	Improving reliability	Increasing
Argall [91]; Wen and Imamizu [92]; Murray, et al. [93]; Raisch and Krakowski [94]; Moray, et al. [95]; Klugman, et al. [96]; Verberne, et al. [97]; de Visser and Parasuraman [98]	Autonomy	Designing the level of machine automation	Appropriate level of automation	Reducing (As machines become more automated, human trust decreases.)
Waytz, et al. [99]; Niu, et al. [100]; Song and Luximon [104]; Broadbent, et al. [105]; Lim, et al. [106]	Anthropomorphism	Adding anthropomorphic elements to machines	Moderate anthropomorphism	Fluctuating (As the degree of anthropomorphism increases, trust increases and then decreases.)

TABLE 3 Some analysis of human influencing factors in human-machine trust (HMT) in healthcare.

Paper	Influencing factors		Relevance
Steinke, et al. [108]; Ho, et al. [109]	Human age		Relatively speaking, older people and children trust intelligent machines more.
Yamani, et al. [110]; Sanchez, et al. [111]; Rajaonah, et al. [112]	Professional level	Perception of healthcare professionals and machines	As professional level increases, humans will have increased trust in machines, but will not rely on them.
		Professional level of healthcare professionals	
Rajaonah, et al. [112]; Balfe, et al. [113]; Merritt and Ilgen [114]	Experience	Level of perception of disease	Experience can improve human trust in machines.
		Experience of healthcare professionals	

their acceptance of medical machines. Experience can facilitate human understanding of machines [113], and it has a direct impact on the development of HMT [112, 114]. The age of healthcare professionals can also affect their trust in machines. These factors can be used to maintain a good level of trust and build a good HMT relationship.

For patients and attendants, their age, professional level and experience can also affect healthcare trust. The level of profession here is reflected in their perception of the healthcare professionals and the medical machines. The experience is mainly reflected in their level of perception of the disease. The level of patients and attendants information about the disease affects trust. One study found that improving patients' disease perceptions by providing information tailored to their needs can help patients have a more consistent understanding of their disease and may lead to a better health-related quality of life [116, 117]. In addition, a high level of profession and experience of the doctors can lead to a good sense of patient perception of the doctor and the machine. Together with good machine performance, this can improve patients' perceptions and feelings about the healthcare system, which in turn can improve trust. And in HMI in healthcare, maintaining mutual trust between healthcare professionals and patients is also the basis of good HMT.

3.2.3 | Interaction behaviours

Between humans, gaze and other forms of interaction behaviours can affect trust. Does a similar effect occur during HMI? Experiments have found that robot gaze has an effect on trust. Establishing eye contact with a robot has a positive effect on the perceived sociality of the robot and the quality of HMI [118]. Kompatsiari, Ciardo, Tikhanoff, Metta, and Wykowska [119] conducted a study on how users evaluate human-like machines based on established eye contact. Participants felt a higher level of interaction with the machine when the machine established eye contact. Therefore, establishing eye contact should be considered when designing robot behaviour for HMI in healthcare. In addition, the healthcare robot Pepper can be used for care in preventive gymnastics exercises for the elderly [120]. Pepper can communicate with healthcare professionals and patients through expressions, movements, sounds, and feedback, and can even dance and joke. Pepper is believed to have the ability to move, act, and communicate like a human. The healthcare robot Paro can stimulate verbal and behavioural interactions in stroke patients [121]. One study found that Paro could alleviate patients' depression and help them interact and communicate better [122]. A healthcare social robot typically has four primary senses, including vision, hearing, balance, and touch. Lifelike behaviours can be created from these to interact with health care professionals and patients. Therefore, increasing multi-sensory interaction behaviours should be considered when designing machine for HMI in healthcare. Fratczak, Goh, Kinnell, Justham, and Soltoggio [123] investigated whether the use of robot control strategies has a positive effect on human post-accident behaviour. In the designed scene, a robot first made a sudden and unexpected action, and then it may apologise for their behaviour. The results suggested that this act of apology can improve HMT after a machine malfunction. Some researchers further found that a robot that can recognise errors and communicate its intent to correct the situation is considered more capable than a robot that simply apologises for the error. However, the latter was considered more popular and people were more willing to use such robots [124]. Researchers studied 326 people's perceptions of a mobile guidance robot that employed synthetic social behaviours to elicit trust in its use after an error [124]. This could be applied to a healthcare guidance robot in healthcare HMI to allow the robot to show self-awareness and ownership of its errors to mitigate the effects of errors, increase affinity and trust of healthcare professionals and patients for the robot, and make the robot's interactive behaviour appear more genuine.

3.2.4 | Environment

A good external environment can facilitate trust building. With the constraints of relevant laws, policies and ethics, the safety of users (doctors and patients) in using machines can be guaranteed and the trust of users in using them can be enhanced [125, 126].

Regulatory Compliance and Standards: Ensure that machine learning technology complies with relevant healthcare regulations. Assess whether the technology adheres to recognised standards or has received relevant certifications specific to healthcare. The following are the relevant laws and regulations of some typical countries and regions. (1) In 2017, the Development Plan of New Generation Artificial Intelligence proposed by the Chinese government emphasised the need to strengthen the innovative application of AI in the healthcare field and accelerate the development of intelligent medical treatment, which provides a platform and opportunity for further development of medical robots. It also listed the hybrid intelligence of human-robot cooperation as one of the bottlenecks that urgently need to be broken. In 2021, the Professional Committee on the Governance of China's Next Generation Artificial Intelligence released the Code of Ethics for Next Generation Artificial Intelligence, which focuses attention on algorithmic bias and other technical governance issues. With the policy support, it can make people have more trust in such technologies. (2) Regulatory agencies such as the FDA have approved its use or the corresponding national laws and regulations can improve human trust in the intelligent machines [127]. (3) In 2018, The EU General Data Protection Regulation (GDPR) came into force, mandating AI algorithms to be interpretable. AI can better gain the trust of humans if it has the ability to explain its decisions. The European Commission's Senior Expert Group on Artificial Intelligence has defined trusted AI. Trustworthy AI should meet three necessary conditions: Machines with AI should comply with all applicable laws and regulations, adhere to ethical principles and values, and be safe and secure.

Human doctors are guided by a set of legal and ethical principles, as well as industry norms and professional ethics. However, when it comes to intelligent machines, the question arises: can they truly comply with medical ethics? Some individuals argue that ML is merely a tool and utilising it may lead to potentially unethical outcomes. The involvement of humans in the ML process is limited, and the generated results may lack a comprehensive explanation. This unpredictability raises concerns about unexpected outcomes.

Moreover, intelligent machines lack the inherent qualities of human doctors, such as empathy and compassion. As a result, they may make decisions that differ from those made by human practitioners [128]. It is crucial to consider the responsibility associated with employing machines for diagnostic purposes and whether they can effectively shoulder this responsibility.

Therefore, due to existing ethical constraints and legal norms, the complete reliance on machines for making medical judgements is currently impractical. It remains necessary for a human doctor to be involved in the decision-making process. This collaboration ensures that the medical judgements are signed off by a qualified professional who can be held accountable.

Consequently, for the widespread application of ML technology in healthcare, it is essential to further enhance social and moral constraints, as well as refine legal norms. These improvements will establish a framework that addresses the ethical considerations and provides a reliable and responsible integration of ML technology in medical practice.

Ethical considerations: (1) Bias assessment: Investigate whether ML technology has been assessed for potential bias in its predictions, particularly regarding sensitive attributes such as race or gender. Ensure that the technology does not perpetuate or amplify existing biases in healthcare. (2) Fairness and impartiality: Assess the fairness and impartiality of the technology, considering how it affects different population groups and whether it contributes to health disparities. Assess whether measures are in place to address and mitigate bias and ensure equitable outcomes. (3) Human oversight and accountability: Consider the role of human oversight in the deployment of ML technology. Determine how the technology can be used to enhance the decision-making of healthcare professionals, rather than replace their expertise and ethical judgement.

Collaboration and Vendor evaluation: (1) Collaborate with experts: Collaborate with reputable research institutions, healthcare organisations, or ML experts in the healthcare field. Their input and assessment can provide additional insight into the reliability and ethical implications of the technology. (2) Vendor evaluation: Assess the track record, reputation and expertise of the vendor or organisation providing the ML technology. Consider factors such as their experience in healthcare, previous successful deployments, and commitment to ethical practices.

For healthcare organisations, it is important to have a thorough understanding of the clinical context, ethical considerations, and potential impact on patient outcomes when evaluating and selecting ML technologies. Engaging a multidisciplinary team that includes healthcare professionals, data scientists, and ethicists can help facilitate a thorough evaluation process and ensure the selection of trustworthy and reliable ML technologies.

4 | CONCLUSION AND FUTURE WORK

Trust plays an important role in the process of HMI in healthcare, and the establishment of a good trust relationship in HMI in healthcare requires the joint efforts of the fields of computer science, psychology, and medicine. Establishing effective trust in HMI in healthcare is of great significance for both academic research and real-world applications. As described in this paper, with the development of technology, trust is no longer only between human and human, but also exists between human and machine. With the AI technologies such as ML being widely used in medical machines, it is becoming increasingly important to better apply the trust influencing factors of HMI in healthcare and use these to improve machine development and design.

AI is currently in an early stage, what will happen to HMT when it reaches an advanced stage in the future? Will people trust machines better? Or even everyone will have to have a intelligent robot dog as a pet. Although much progress has been made in the application of intelligent machines in the healthcare field, HMT is still a key issue that needs to be researched, there is still a long way to go for its specific application. In order to achieve breakthroughs, the following challenges must be considered and overcome. The following issues and challenges are worth considering:

(1) Enhance communications in the fields of healthcare and computer science, psychology etc. HMI in healthcare is not the same as other scenarios. It is necessary to collect and extract information from medical institutions such as hospitals. Therefore, future research should increase communication and cooperation with medical and other related professionals. Focusing on intelligent healthcare and exploring HMT in it is a long-term and important subject.
(2) Improve accuracy and explainability of ML in healthcare. Although DL methods improve accuracy, they cannot be explained, and DL models vary. Therefore, it is necessary to find a balance between the two directly and develop relevant visualisation techniques to transform high-dimensional information into interpretable information. Deep learning machine models are often not available at the same time, and most studies focus on one perspective. It is a necessary study to balance the relationship between both improving accuracy and enhancing explainability to get better DL machine models.
(3) Create a unified human-machine trust index system. Since the concept of trust is inclined to abstraction, there is a lack of unified quantitative indicators. So far, there is a lack of unified evaluation indexes for HMT models in healthcare. Trust is influenced by multiple factors, and it is an inevitable trend for future research to combine multiple trust influencing factors and establish an index system based on these factors to improve machine models accurately and effectively.
(4) Explore decision-making relationship between human and machine in HMI in healthcare. More attention needs to be paid to the influence of human perspective on HMT and further research on the decision-making relationship between human and machine in HMI in healthcare. How to ensure that the machine does not affect human autonomy, maintain human-centeredness, and promote human-machine cooperation deserves deeper investigation.

ACKNOWLEDGEMENTS
This work is supported by the National Natural Science Foundation of China (no. 72271126, 12271255, 61701243, 71771125, 12227808) and the Major Project of Natural Science Foundation of Jiangsu Education Department (no. 22KJA630001, 21KJA630001), Qinglan Project of Jiangsu Province, Natural Science Foundation of Jiangsu Province (BK20180820), and Postgraduate Research and Practice Innovation Program of Jiangsu Province (no. KYCX23_2343).

CONFLICT OF INTEREST STATEMENT
The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT
The raw data supporting the results of this research will be made accessible without restriction to any competent researcher by the authors.

ORCID
Han Lin https://orcid.org/0000-0001-5136-5059
Pingping Wu https://orcid.org/0000-0001-7822-5208

REFERENCES
1. Rajpurkar, P., et al.: AI in health and medicine. Nat. Med. 28(1), 31–38 (2022). https://doi.org/10.1038/s41591-021-01614-0
2. Dwivedi, Y.K., et al.: Artificial Intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int. J. Inf. Manag. 57, 101994 (2021). https://doi.org/10.1016/j.ijinfomgt.2019.08.002
3. Narin, A., Kaya, C., Pamuk, Z.: Automatic detection of coronavirus disease (Covid-19) using x-ray images and deep convolutional neural networks. Pattern Anal. Appl. 24(3), 1207–1220 (2021). https://doi.org/10.1007/s10044-021-00984-y
4. Coronato, A., et al.: Reinforcement learning for intelligent healthcare applications: a survey. Artif. Intell. Med. 109, 101964 (2020). https://doi.org/10.1016/j.artmed.2020.101964
5. Vaishya, R., et al.: Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes Metabol. Syndr.: Clin. Res. Rev. 14(4), 337–339 (2020). https://doi.org/10.1016/j.dsx.2020.04.012
6. Giger, M.L.: Machine learning in medical imaging. J. Am. Coll. Radiol. 15(3), 512–520 (2018). https://doi.org/10.1016/j.jacr.2017.12.028
7. Jahangiri, A., Rakha, H.A.: Applying machine learning techniques to transportation mode recognition using mobile phone sensor data. IEEE Trans. Intell. Transport. Syst. 16(5), 2406–2417 (2015). https://doi.org/10.1109/tits.2015.2405759
8. Gebru, B., et al.: A review on human–machine trust evaluation: human-centric and machine-centric perspectives. IEEE Tran. Hum. Mach. Syst. 52(5), 952–962 (2022). https://doi.org/10.1109/thms.2022.3144956
9. Bryndin, E.: Development of artificial intelligence by ensembles of virtual agents with mobile interaction. Autom. Control Intell. Syst. 8, 1–8 (2020). https://doi.org/10.11648/j.acis.20200801.11
10. Vallelonga, F., Elia, F.: Doctor–patient relationship at the time of COVID-19: travel notes. Intensive Care Med. 46(9), 1802 (2020). https://doi.org/10.1007/s00134-020-06152-w
11. Promberger, M., Baron, J.: Do patients trust computers? J. Behav. Decis. Making 19(5), 455–468 (2006). https://doi.org/10.1002/bdm.542
12. Naylor, C.D.: On the prospects for a (deep) learning health care system. JAMA 320(11), 1099–1100 (2018). https://doi.org/10.1001/jama.2018.11103
13. Huo, W., et al.: Interacting with medical artificial intelligence: integrating self-responsibility attribution, human–computer trust, and personality. Comput. Hum. Behav. 132, 107253 (2022). https://doi.org/10.1016/j.chb.2022.107253
14. Yin, R., et al.: Wearable sensors-enabled human–machine interaction systems: from design to application. Adv. Funct. Mater. 31(11), 2008936 (2021). https://doi.org/10.1002/adfm.202008936
15. Andras, P., et al.: Trusting intelligent machines: deepening trust within socio-technical systems. IEEE Technol. Soc. Mag. 37(4), 76–83 (2018). https://doi.org/10.1109/mts.2018.2876107
16. Kora, P., et al.: Transfer learning techniques for medical image analysis: a review. Biocybern. Biomed. Eng. 42(1), 79–107 (2022). https://doi.org/10.1016/j.bbe.2021.11.004
17. Asan, O., Bayrak, A.E., Choudhury, A.: Artificial intelligence and human trust in healthcare: focus on clinicians. J. Med. Internet Res. 22(6), e15154 (2020). https://doi.org/10.2196/15154
18. Hamet, P., Tremblay, J.: Artificial intelligence in medicine. Metabolism 69, S36–S40 (2017). https://doi.org/10.1016/j.metabol.2017.01.011
19. Jiang, F., et al.: Artificial intelligence in healthcare: past, present and future. Stroke Vasc. Neurol. 2(4), 230–243 (2017). https://doi.org/10.1136/svn-2017-000101
20. Jordan, M.I., Mitchell, T.M.: Machine learning: trends, perspectives, and prospects. Science 349(6245), 255–260 (2015). https://doi.org/10.1126/science.aaa8415
21. Zhang, K., et al.: Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia

using computed tomography. Cell 181(6), 1423–1433 (2020). https://doi.org/10.1016/j.cell.2020.04.045
22. Andras, I., et al.: Artificial intelligence and robotics: a combination that is changing the operating room. World J. Urol. 38(10), 2359–2366 (2020). https://doi.org/10.1007/s00345-019-03037-6
23. Mak, K.-K., Pichika, M.R.: Artificial intelligence in drug development: present status and future prospects. Drug Discov. Today 24(3), 773–780 (2019). https://doi.org/10.1016/j.drudis.2018.11.014
24. Confalonieri, R., et al.: A historical perspective of explainable Artificial Intelligence. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 11(1), e1391 (2021). https://doi.org/10.1002/widm.1391
25. Tzachor, A., et al.: Responsible artificial intelligence in agriculture requires systemic understanding of risks and externalities. Nat. Mach. Intell. 4(2), 104–109 (2022). https://doi.org/10.1038/s42256-022-00440-4
26. Rahwan, I., et al.: Machine behaviour. Nature 568(7753), 477–486 (2019). https://doi.org/10.1038/s41586-019-1138-y
27. Siegrist, M.: Trust and risk perception: a critical review of the literature. Risk Anal. 41(3), 480–490 (2021). https://doi.org/10.1111/risa.13325
28. Mayer, R.C., Davis, J.H., Schoorman, F.D.: An integrative model of organizational trust. Acad. Manag. Rev. 20(3), 709–734 (1995). https://doi.org/10.5465/amr.1995.9508080335
29. Rousseau, D.M., et al.: Not so different after all: a cross-discipline view of trust. Acad. Manag. Rev. 23(3), 393–404 (1998). https://doi.org/10.5465/amr.1998.926617
30. Cottrell, C.A., Neuberg, S.L., Li, N.P.: What do people desire in others? A sociofunctional perspective on the importance of different valued characteristics. J. Pers. Soc. Psychol. 92(2), 208–231 (2007). https://doi.org/10.1037/0022-3514.92.2.208
31. Muir, B.M.: Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics 37(11), 1905–1922 (1994). https://doi.org/10.1080/00140139408964957
32. Rempel, J.K., Holmes, J.G., Zanna, M.P.: Trust in close relationships. J. Pers. Soc. Psychol. 49(1), 95–112 (1985). https://doi.org/10.1037/0022-3514.49.1.95
33. Schilke, O., Reimann, M., Cook, K.S.: Trust in social relations. Annu. Rev. Sociol. 47(1), 239–259 (2021). https://doi.org/10.1146/annurev-soc-082120-082850
34. Lee, J.D., See, K.A.: Trust in automation: designing for appropriate reliance. Hum. Factors 46(1), 50–80 (2004). https://doi.org/10.1518/hfes.46.1.50_30392
35. Siau, K., Wang, W.: Building trust in artificial intelligence, machine learning, and robotics. Cut. Bus. Technol. J. 31, 47–53 (2018)
36. Hoffman, R.R.: A taxonomy of emergent trusting in the human–machine relationship. Cognit. Systems Eng., 137–164 (2017). https://doi.org/10.1201/9781315572529-8
37. Parasuraman, R., Manzey, D.H.: Complacency and bias in human use of automation: an attentional integration. Hum. Factors 52(3), 381–410 (2010). https://doi.org/10.1177/0018720810376055
38. Parasuraman, R., Riley, V.: Humans and automation: use, misuse, disuse, abuse. Hum. Factors 39(2), 230–253 (1997). https://doi.org/10.1518/001872097778543886
39. Rheu, M., et al.: Systematic review: trust-building factors and implications for conversational agent design. Int. J. Hum. Comput. Interact. 37(1), 81–96 (2021). https://doi.org/10.1080/10447318.2020.1807710
40. Jian, J.-Y., Bisantz, A.M., Drury, C.G.: Foundations for an empirically determined scale of trust in automated systems. Int. J. Cognit. Ergon. 4(1), 53–71 (2000). https://doi.org/10.1207/s15327566ijce0401_04
41. Hancock, P.A., et al.: A meta-analysis of factors affecting trust in human-robot interaction. Hum. Factors 53(5), 517–527 (2011). https://doi.org/10.1177/0018720811417254
42. Hoff, K.A., Bashir, M.: Trust in automation: integrating empirical evidence on factors that influence trust. Hum. Factors 57(3), 407–434 (2015). https://doi.org/10.1177/0018720814547570
43. Schaefer, K.E., et al.: A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. Hum. Factors 58(3), 377–400 (2016). https://doi.org/10.1177/0018720816634228
44. Huang, Q., Huang, S.: Simulating multi-Agent based medical trust model. Comput. Appli. Software 28, 129–130 (2011)
45. Bahtiyar, Ş., Çağlayan, M.U.: Trust assessment of security for e-health systems. Electron. Commer. Res. Appl. 13(3), 164–177 (2014). https://doi.org/10.1016/j.elerap.2013.10.003
46. Carter-Langford, A., Wiljer, D.: The eHealth trust model: a patient privacy research framework. Stud. Health Technol. Inf. 257, 382–387 (2019)
47. Shareef, M.A., et al.: A new health care system enabled by machine intelligence: elderly people's trust or losing self control. Technol. Forecast. Soc. Change 162, 120334 (2021). https://doi.org/10.1016/j.techfore.2020.120334
48. Strickland, E.: IBM Watson, heal thyself: how IBM overpromised and underdelivered on AI health care. IEEE Spectr. 56(4), 24–31 (2019). https://doi.org/10.1109/mspec.2019.8678513
49. Printz, C.: Artificial intelligence platform for oncology could assist in treatment decisions. Cancer 123(6), 905 (2017). https://doi.org/10.1002/cncr.30655
50. Ordóñez, L.D., Benson, L., III, Pittarello, A.: Time-pressure Perception and Decision Making, vol. 2, pp. 517–542. The Wiley Blackwell Handbook of Judgment and Decision Making (2015)
51. Cokely, E.T., Kelley, C.M.: Cognitive abilities and superior decision making under risk: a protocol analysis and process model evaluation. Judgment Decision Making 4(1), 20–33 (2009). https://doi.org/10.1017/s193029750000067x
52. Damacharla, P., et al.: Common metrics to benchmark human-machine teams (HMT): a review. IEEE Access 6, 38637–38655 (2018). https://doi.org/10.1109/access.2018.2853560
53. Elia, R., et al.: A real-world data collection framework for a fused dataset creation for joint human and remotely operated vehicle monitoring and anomalous command detection. CAAI Trans. Intell. Technol. 7(3), 432–445 (2022). https://doi.org/10.1049/cit2.12119
54. Yu, J., et al.: Humanoid control of lower limb exoskeleton robot based on human gait data with sliding mode neural network. CAAI Trans. Intell. Technol. 7(4), 606–616 (2022). https://doi.org/10.1049/cit2.12127
55. Xiong, W., et al.: Challenges of human—machine collaboration in risky decision-making. Front. Eng. Manag. 9(1), 89–103 (2022). https://doi.org/10.1007/s42524-021-0182-0
56. Sharma, A., Kaur, S., Singh, M.: A comprehensive review on blockchain and Internet of Things in healthcare. Trans. Emerg. Telecommun. Technol. 32(10), e4333 (2021). https://doi.org/10.1002/ett.4333
57. de Moraes Rossetto, A.G., Sega, C., Leithardt, V.R.Q.: An architecture for managing data privacy in healthcare with blockchain. Sensors 22(21), 8292 (2022). https://doi.org/10.3390/s22218292
58. Mathis, F., Vaniea, K., Khamis, M.: Prototyping useable privacy and security systems: insights from experts. Int. J. Hum. Comput. Interact. 38(5), 468–490 (2022). https://doi.org/10.1080/10447318.2021.1949134
59. Al Omar, A., et al.: A transparent and privacy-preserving healthcare platform with novel smart contract for smart cities. IEEE Access 9, 90738–90749 (2021). https://doi.org/10.1109/access.2021.3089601
60. Chien, S.-Y., et al.: Influence of culture, transparency, trust, and degree of automation on automation use. IEEE Trans. Hum. Mach. Syst. 50(3), 205–214 (2019). https://doi.org/10.1109/thms.2019.2931755
61. Ishowo-Oloko, F., et al.: Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation. Nat. Mach. Intell. 1(11), 517–521 (2019). https://doi.org/10.1038/s42256-019-0113-5
62. Edmonds, M., et al.: A tale of two explanations: enhancing human trust by explaining robot behavior. Sci. Robot. 4(37), eaay4663 (2019). https://doi.org/10.1126/scirobotics.aay4663
63. Fischer, K., Weigelin, H.M., Bodenhagen, L.: Increasing trust in human–robot medical interactions: effects of transparency and adaptability. Paladyn. J. Behav. Rob. 9(1), 95–109 (2018). https://doi.org/10.1515/pjbr-2018-0007

64. He, J., Zhang, P.: An investigation on path from algorithm trust to man-machine trust. Stud. Dialectics Nat. 36, 81–85 (2020) (in Chinese)
65. Nazar, M., et al.: A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques. IEEE Access 9, 153316–153348 (2021). https://doi.org/10.1109/access.2021.3127881
66. Glikson, E., Woolley, A.W.: Human trust in artificial intelligence: review of empirical research. Acad. Manag. Ann. 14(2), 627–660 (2020). https://doi.org/10.5465/annals.2018.0057
67. Eshete, B.: Making machine learning trustworthy. Science 373(6556), 743–744 (2021). https://doi.org/10.1126/science.abi5052
68. Poplin, R., et al.: Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. Nat. Biomed. Eng. 2(3), 158–164 (2018). https://doi.org/10.1038/s41551-018-0195-0
69. Zhang, Z., et al.: Mdnet: a semantically and visually interpretable medical image diagnosis network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6428–6436 (2017)
70. Adadi, A., Berrada, M.: Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6, 52138–52160 (2018). https://doi.org/10.1109/access.2018.2870052
71. Zhang, W., Wang, L., Wang, Z.: Interpretable machine learning study of the many-body localization transition in disordered quantum Ising spin chains. Phys. Rev. B 99(5), 054208 (2019). https://doi.org/10.1103/physrevb.99.054208
72. Bach, S., et al.: On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. PLoS One 10(7), e0130140 (2015). https://doi.org/10.1371/journal.pone.0130140
73. Itani, S., Lecron, F., Fortemps, P.: Specifics of medical data mining for diagnosis aid: a survey. Expert Syst. Appl. 118, 300–314 (2019). https://doi.org/10.1016/j.eswa.2018.09.056
74. Ming, Y., Qu, H., Bertini, E.: Rulematrix: visualizing and understanding classifiers with rules. IEEE Trans. Visual. Comput. Graph. 25(1), 342–352 (2018). https://doi.org/10.1109/tvcg.2018.2864812
75. Caywood, M.S., et al.: Gaussian process regression for predictive but interpretable machine learning models: an example of predicting mental workload across tasks. Front. Hum. Neurosci. 10, 647 (2017). https://doi.org/10.3389/fnhum.2016.00647
76. Zhang, Q., et al.: Interpreting cnns via decision trees. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6261–6270 (2019)
77. Moraffah, R., et al.: Causal interpretability for machine learning-problems, methods and evaluation. ACM SIGKDD Explor. Newsl. 22(1), 18–33 (2020). https://doi.org/10.1145/3400051.3400058
78. Zhao, X., et al.: iforest: interpreting random forests via visual analytics. IEEE Trans. Visual. Comput. Graph. 25(1), 407–416 (2018). https://doi.org/10.1109/tvcg.2018.2864475
79. Liu, M., et al.: Towards better analysis of deep convolutional neural networks. IEEE Trans. Visual. Comput. Graph. 23(1), 91–100 (2016). https://doi.org/10.1109/tvcg.2016.2598831
80. Olah, C., et al.: The building blocks of interpretability. Distill 3, e10 (2018). https://doi.org/10.23915/distill.00010
81. Zhang, J., et al.: A spatial attentive and temporal dilated (SATD) GCN for skeleton-based action recognition. CAAI Trans. Intell. Technol. 7(1), 46–55 (2022). https://doi.org/10.1049/cit2.12012
82. Yan, Z., et al.: Multi-instance deep learning: discover discriminative local anatomies for bodypart recognition. IEEE Trans. Med. Imag. 35(5), 1332–1343 (2016). https://doi.org/10.1109/tmi.2016.2524985
83. Li, L., Li, T.: Traceability model based on improved witness mechanism. CAAI Trans. Intell. Technol. 7(3), 331–339 (2022). https://doi.org/10.1049/cit2.12124
84. Esteva, A., et al.: Correction: corrigendum: Dermatologist-level classification of skin cancer with deep neural networks. Nature 546(7660), 686 (2017). https://doi.org/10.1038/nature22985
85. Miotto, R., et al.: Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci. Rep. 6, 1–10 (2016). https://doi.org/10.1038/srep26094
86. Dong, W., Fang, W.: Trust in automation: research review and future perspectives. Acta Autom. Sin. 47, 1183–1200 (2021) (in Chinese)
87. Kraus, J.M., et al.: Two routes to trust calibration: effects of reliability and brand information on trust in automation. Int. J. Mobile Hum. Comput. Interact. 11(3), 1–17 (2019). https://doi.org/10.4018/ijmhci.2019070101
88. Aggarwal, R., et al.: Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. NPJ Digit. Med. 4(1), 65 (2021). https://doi.org/10.1038/s41746-021-00438-z
89. Buchlak, Q.D., et al.: Machine learning applications to clinical decision support in neurosurgery: an artificial intelligence augmented systematic review. Neurosurg. Rev. 43(5), 1235–1253 (2020). https://doi.org/10.1007/s10143-019-01163-8
90. Huong, A., et al.: A hierarchical optimisation framework for pigmented lesion diagnosis. CAAI Trans Intell. Technol. 7(1), 34–45 (2022). https://doi.org/10.1049/cit2.12073
91. Argall, B.D.: Autonomy in rehabilitation robotics: an intersection. Annu. Rev. Control, Robotics, Autonom. Syst. 1, 441–463 (2018). https://doi.org/10.1146/annurev-control-061417-041727
92. Wen, W., Imamizu, H.: The sense of agency in perception, behaviour and human–machine interactions. Nat. Rev. Psychol. 1(4), 211–222 (2022). https://doi.org/10.1038/s44159-022-00030-6
93. Murray, A., Rhymer, J., Sirmon, D.G.: Humans and technology: forms of conjoined agency in organizations. Acad. Manag. Rev. 46(3), 552–571 (2021). https://doi.org/10.5465/amr.2019.0186
94. Raisch, S., Krakowski, S.: Artificial intelligence and management: the automation–augmentation paradox. Acad. Manag. Rev. 46(1), 192–210 (2021). https://doi.org/10.5465/amr.2018.0072
95. Moray, N., Inagaki, T., Itoh, M.: Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. J. Exp. Psychol. Appl. 6(1), 44–58 (2000). https://doi.org/10.1037/1076-898x.6.1.44
96. Klugman, C.M., et al.: The ethics of smart pills and self-acting devices: autonomy, truth-telling, and trust at the dawn of digital medicine. Am. J. Bioeth. 18(9), 38–47 (2018). https://doi.org/10.1080/15265161.2018.1498933
97. Verberne, F.M., Ham, J., Midden, C.J.: Trust in smart systems: sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. Hum. Factors 54(5), 799–810 (2012). https://doi.org/10.1177/0018720812443825
98. de Visser, E., Parasuraman, R.: Adaptive aiding of human-robot teaming: effects of imperfect automation on performance, trust, and workload. J. Cogn. Eng. Decis. Mak. 5(2), 209–231 (2011). https://doi.org/10.1177/1555343411410160
99. Waytz, A., Heafner, J., Epley, N.: The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. J. Exp. Soc. Psychol. 52, 113–117 (2014). https://doi.org/10.1016/j.jesp.2014.01.005
100. Niu, D., Terken, J., Eggen, B.: Anthropomorphizing information to enhance trust in autonomous vehicles. Hum. Factors Ergon. Manuf. Serv. Ind. 28(6), 352–359 (2018). https://doi.org/10.1002/hfm.20745
101. Gao, Z., et al.: Trust in automated vehicles. Adv. Psychol. Sci. 29(12), 2172–2183 (2021). (in Chinese). https://doi.org/10.3724/sp.j.1042.2021.02172
102. Qi, Y., et al.: Regulation of facial trustworthiness evaluation: the proposal and empirical verification of the experience transfer hypothesis. Adv. Psychol. Sci. 30(4), 715–722 (2022). (in Chinese). https://doi.org/10.3724/sp.j.1042.2022.00715
103. Song, Y.: Building a 'deeper' trust: mapping the facial anthropomorphic trustworthiness in social robot design through multidisciplinary approaches. Des. J. 23(4), 639–649 (2020). https://doi.org/10.1080/14606925.2020.1766871
104. Song, Y., Luximon, Y.: Trust in AI agent: a systematic review of facial anthropomorphic trustworthiness for social robot design. Sensors 20(18), 5087 (2020). https://doi.org/10.3390/s20185087
105. Broadbent, E., et al.: Robots with display screens: a robot with a more humanlike face display is perceived to have more mind and a better personality. PLoS One 8, e72589 (2013). https://doi.org/10.1371/journal.pone.0072589

106. Lim, V., Rooksby, M., Cross, E.S.: Social robots on a global stage: establishing a role for culture during human–robot interaction. Int. J. Soc. Robot. 13(6), 1307–1333 (2021). https://doi.org/10.1007/s12369-020-00710-4
107. Kyrarini, M., et al.: A survey of robots in healthcare. Technologies 9(1), 8 (2021). https://doi.org/10.3390/technologies9010008
108. Steinke, F., Fritsch, T., Silbermann, L.: Trust in ambient assisted living (AAL)-a systematic review of trust in automation and assistance systems. Int. J. Adv. Life Sci. 4, 77–88 (2012)
109. Ho, G., Wheatley, D., Scialfa, C.T.: Age differences in trust and reliance of a medication management system. Interact. Comput. 17(6), 690–710 (2005). https://doi.org/10.1016/j.intcom.2005.09.007
110. Yamani, Y., Long, S.K., Itoh, M.: Human–automation trust to technologies for naïve users amidst and following the COVID-19 pandemic. Hum. Factors 62(7), 1087–1094 (2020). https://doi.org/10.1177/0018720820948981
111. Sanchez, J., et al.: Understanding reliance on automation: effects of error type, error distribution, age and experience. Theor. Issues Ergon. Sci. 15(2), 134–160 (2014). https://doi.org/10.1080/1463922x.2011.611269
112. Rajaonah, B., Anceaux, F., Vienne, F.: Trust and the use of adaptive cruise control: a study of a cut-in situation. Cognit. Technol. Work 8(2), 146–155 (2006). https://doi.org/10.1007/s10111-006-0030-3
113. Balfe, N., Sharples, S., Wilson, J.R.: Understanding is key: an analysis of factors pertaining to trust in a real-world automation system. Hum. Factors 60(4), 477–495 (2018). https://doi.org/10.1177/0018720818761256
114. Merritt, S.M., Ilgen, D.R.: Not all trust is created equal: dispositional and history-based trust in human-automation interactions. Hum. Factors 50(2), 194–210 (2008). https://doi.org/10.1518/001872008x288574
115. Matthews, G., et al.: Evolution and revolution: personality research for the coming world of robots, artificial intelligence, and autonomous systems. Pers. Indiv. Differ. 169, 109969 (2021). https://doi.org/10.1016/j.paid.2020.109969
116. Husson, O., et al.: Illness perceptions in cancer survivors: what is the role of information provision? Psycho Oncol. 22(3), 490–498 (2013). https://doi.org/10.1002/pon.3042
117. Botti, S., Orfali, K., Iyengar, S.S.: Tragic choices: autonomy and emotional responses to medical decisions. J. Consum. Res. 36(3), 337–352 (2009). https://doi.org/10.1086/598969
118. Belkaid, M., et al.: Mutual gaze with a robot affects human neural activity and delays decision-making processes. Sci. Robot. 6(58), eabc5044 (2021). https://doi.org/10.1126/scirobotics.abc5044
119. Kompatsiari, K., et al.: It's in the eyes: the engaging role of eye contact in HRI. Int. J. Soc. Robot. 13(3), 525–535 (2021). https://doi.org/10.1007/s12369-019-00565-4
120. Tanioka, R., et al.: Potential legal issues and care implications during care-prevention gymnastic exercises for the elderly using Pepper in long term health care facilities. Intell. Control Autom. 9(03), 85–93 (2018). https://doi.org/10.4236/ica.2018.93007
121. Karner, S., et al.: Effects of a robot intervention on visuospatial hemineglect in postacute stroke patients: a randomized controlled trial. Clin. Rehabil. 33(12), 1940–1948 (2019). https://doi.org/10.1177/0269215519865993
122. Chen, S.-C., et al.: A social robot intervention on depression, loneliness, and quality of life for Taiwanese older adults in long-term care. Int. Psychogeriatr. 32(8), 981–991 (2020). https://doi.org/10.1017/s1041610220000459
123. Fratczak, P., et al.: Robot apology as a post-accident trust-recovery control strategy in industrial human-robot interaction. Int. J. Ind. Ergon. 82, 103078 (2021). https://doi.org/10.1016/j.ergon.2020.103078
124. Cameron, D., et al.: The effect of social-cognitive recovery strategies on likability, capability and trust in social robots. Comput. Hum. Behav. 114, 106561 (2021). https://doi.org/10.1016/j.chb.2020.106561
125. Surden, H.: Artificial Intelligence and Law: An Overview, vol. 35, pp. 19–22. Georgia State University Law Review (2019)
126. Leenes, R., Lucivero, F.: Laws on robots, laws by robots, laws in robots: regulating robot behaviour by design. Law, Innov. Technol. 6(2), 193–220 (2014). https://doi.org/10.5235/17579961.6.2.193
127. Khor, W.S., et al.: Augmented and virtual reality in surgery—the digital surgical environment: applications, limitations and legal pitfalls. Ann. Trans. Med. 4(23), 454 (2016). https://doi.org/10.21037/atm.2016.12.23
128. Nsugbe, E., Al-Timemy, A.H.: Shoulder girdle recognition using electrophysiological and low frequency anatomical contraction signals for prosthesis control. CAAI Trans Intell. Technol. 7(1), 81–94 (2022). https://doi.org/10.1049/cit2.12058

How to cite this article: Lin, H., et al.: Machine learning and human-machine trust in healthcare: a systematic survey. CAAI Trans. Intell. Technol. 9(2), 286–302 (2024). https://doi.org/10.1049/cit2.12268